Video Background/Foreground Separation Model Based on Non-Convex Rank Approximation RPCA and Superpixel Motion Detection

YONGLI WANG[ID], HONGCHAO WEI[ID], XIAOYUN DING, AND JULIANG TAO

College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China

Corresponding authors: Hongchao Wei (1146565311@qq.com) and Yongli Wang (wangyongli@sdkd.net.cn)

ABSTRACT Traditional robust principal component analysis (RPCA) is very prone to voids in the process of background/foreground separation of complex scene videos and easy to misjudge the dynamic background as a moving target, which makes the separation effect unideal. In order to address this problem, this article introduces the super-pixel segmentation technique into the RPCA model. First, the Linear Spectral Clustering algorithm (LSC) is used to mark the super-pixel segmentation of the video sequence and a super-pixel grouping matrix is obtained. Then a new video background/foreground separation model is proposed based on the non-convex rank approximation RPCA and super-pixel motion detection (SPMD) technique. The Otsu algorithm is used to obtain the motion mask matrix and the augmented lagrange alternating direction method is used to solve the improved RPCA model. The results of numerical experiment show that the method proposed in this article has a higher accuracy in the detection of moving objects in dynamic background.

INDEX TERMS Video background/foreground separation, RPCA, superpixel segmentation, linear spectral clustering algorithm, Otsu algorithm, motion mask.

I. INTRODUCTION

Video background/foreground separation [1]–[3] is a key pre-processing step in the surveillance system. It has a wide range of applications in the fields of intelligent traffic management, intelligent video surveillance and sports behavior analysis. It is also one of the most active research topics in the field of computer vision, image processing and pattern recognition.

At present, the robust principal component analysis (RPCA) model based on matrix low rank sparse decomposition by Candès *et al.* [4] has been widely used in moving target detection [5]–[8]. In fact, background sequences can be modeled by a low-rank subspace that can change gradually over time, while moving foreground objects constitute associated sparse outliers [9]. The RPCA model achieves the detection of moving targets by decomposing the video matrix into a low-rank background matrix and a sparse moving foreground target matrix. This kind of method can estimate the background and separate the foreground moving target at the same time, effectively overcome the target false detection caused by the periodic change of the background, and has good robustness to noise and illumination changes. Specially, when

the background in the video is relatively static, the RPCA model can usually achieve more accurate foreground object detection [10], [11].

However, in most cases, a video sequence is always captured with a complex background in which the foreground objects may blend into the background [12], such as wind blowing leaves, waves, swaying vegetation, fountains, changes in light, ripples on the water, flags flying in the wind and so on. Because the background is not completely static (that is, the background also contains dynamic components), the performance of the foreground detection method will be affected by the dynamic pixel components in the background. It is easy to misjudge the dynamic background as the foreground moving target, resulting in the problems of incomplete and empty edge of the foreground moving object detection.

In order to overcome the above problems, scholars introduced different regularization terms and space-time continuity constraints into the original RPCA model, and used non-convex functions instead of nuclear norms to approximate the rank function, which greatly improved the foreground object detection results [13]–[16]. Especially in recent years, with the widespread application of super-pixel technique in the field of image processing, some scholars have proposed to use super-pixel segmentation technique to group

pixels that are adjacent in appearance, space or time to better preserve the boundary features of the foreground and achieved better results [10], [17]–[21].

Considering the success of super-pixel segmentation technique in image processing and the advantages of non-convex-rank-approximation-based RPCA model in video foreground/background separation and moving target detection under dynamic background, this article proposed a new video pre-background separation model (LSCNC-RPCA) based on the non-convex rank approximation RPCA and super-pixel motion detection technique. The new model is solved by the augmented lagrange alternating direction method. The experimental results of eight dynamic videos in the CDnet2014 database show that the model proposed in this article can effectively improve the accuracy of moving target extraction and has better visual effects of foreground target extraction than existed methods.

The rest of this article is organized as follows. Section 2 is a brief introduction to the background and some related work. The proposed model is given in section 3. Section 4 gives the experimental results and the conclusion is given in section 5.

II. BACKGROUND AND RELATED WORKS
In this section, the background and some related work are given.

A. BACKGROUND
1) RPCA MODEL
The traditional RPCA can be described as the following optimization problem

$$\min_{L,S} rank(L) + \lambda \|S\|_0 \quad s.t.\, X = L + S, \qquad (1)$$

where rank(L) is the rank function of the matrix, $\| \cdot \|_0$ is the l_0 norm of the matrix (that is, the number of non-zero elements in the matrix), and λ is a regularization parameter. Because the rank function and l_0 norm are non-convex discrete functions, the solution of this model is NP-hard. Note that the rank of a matrix is equal to the number of its nonzero singular values, the nuclear function and l_1 norm of the matrix can be used to convexly approximate the rank function and l_0 norm of the matrix, respectively. Hence, the above RPCA model can be approximated as the following convex optimization problem

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \quad s.t.\, X = L + S, \qquad (2)$$

where $\|L\|_* = \sum_{i=1}^{r} \sigma_i$ is the nuclear norm of the matrix L; σ_i is the non-zero singular value of the matrix L; $\| \cdot \|_1$ is the l_1 norm of the matrix, defined as: $\|S\|_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} |s_{ij}|$.

Under the assumption that the background of the video sequence is static or quasi-static, and the detection target in the video moves fast, the RPCA model (2) has a better detection effect on the foreground target.

2) SUPER-PIXEL SEGMENTATION
Super-pixel segmentation is an image segmentation technique proposed by Ren Xiaofeng in 2013 [22]. It refers to an image area with certain visual significance composed of adjacent pixels with similar physical characteristics such as texture, color, and brightness, etc [23]. Super-pixels can capture the redundant information of the image, thereby greatly reduces the complexity of image subsequent processing, such as target recognition and image segmentation. Therefore, it has been widely used in computer vision applications such as image segmentation, target tracking and target recognition, and so on.

At present, the main super-pixel segmentation algorithms include linear spectral clustering (LSC) [24], normalized cuts [25], meanshift [26], turbo-pixel [27], and SLIC [28],etc. Among them, the linear spectral clustering algorithm (LSC) uses an image segmentation method based on normalized cut sets. It uses the mathematical equivalence between two seemingly different methods to explicitly map the pixel data into a high-dimensional feature space. Then, through effective local operation, the global image structure can be successfully preserved, which effectively solved the problem of global high complexity. Compared with other existing super-pixel segmentation method, LSC algorithm not only avoids correlation matrix decomposition, but also avoids the creation of a large nuclear matrix. It has linear computational complexity and high storage efficiency, and the resulting super-pixel by LSC algorithm can well adapt to changes in the structure and texture of natural images. It can also connect the operation based on local features with the objective function of global optimization, and has good shape compactness and high boundary adhesion.

B. RELATED WORKS
For the traditional RPCA models that use the nuclear norm to approximate the rank function, there are two main problems. First, the nuclear norm is defined as the sum of all the singular values of the matrix. When one or several singular values of the matrix are very large, it will lead to the problem of excessive rank estimation, which will affect the recovery results of the low rank matrix. Secondly, in the solution of RPCA models based on nuclear norm, the singular value decomposition (SVD) must be performed at each step, which is time and computation consuming, especially when the matrix size is large.

To address this problem, scholars proposed different improved models. For example, in order to overcome the problem of excessive rank estimation of nuclear norms, Gu *et al.* [29] and Hu *et al.* [30] used weighted nuclear norms and truncated nuclear norms respectively to replace the normal nuclear norm and improved the detection effect to a certain extent. However, the calculation amount is still large. In order to overcome the problem of large computational complexity for solving the nuclear norm, scholars proposed RPCA models based on low-rank matrix decomposition technique [31], [32], which decomposes the low-rank matrix

into two or three smaller matrix, reducing the size of the matrix that requires singular value decomposition and the amount of calculation as well. In addition, in order to completely avoid singular value decomposition in the model solution and overcome the overestimation problem aroused by the nuclear norm, scholars proposed to use different non-convex functions to approximate the rank function, such as the non-convex-γ-norm [14], non-convex Laplace norm [15], logarithmic determinant function [33], [34],etc. In particular, Yang and Zou [6] proposed a general optimization model based on non-convex rank approximation to achieve background separation before surveillance video. A large number of experimental results show that the RPCA model based on non-convex rank approximation can restore a cleaner background and the operation time is shorter.

However, for the video sequences including complex dynamic background, such as waves, snowflakes, swinging leaves and so on, the above non-convex rank approximation-based RPCA model may misjudge the dynamic background as the foreground moving target, resulting in the problems of incomplete and empty edge of the foreground moving object detection. In order to overcome this problem, scholars consider making use of the spatiotemporal key information contained in video and image, that is, to extend RPCA model by imposing some additional constraints. Based on this consideration, Ebadi and Izquierdo [17] introduced a new sparse criteria and group structure sparse constraints into the foreground part, and proposed a sparse RPCA model with a dynamic tree structure. Javed *et al.* [18] proposed a matrix decomposition model that combines the maximum norm regularization and structural sparsity constraints simultaneously. Besides, with the wide use of super pixel segmentation technique, scholars began to introduce the super pixel segmentation technique into the RPCA model to construct the spatial-temporal continuity between pixels. For example, on the basis of local and global invariant assumptions, Javed *et al.* [10] proposed a super-pixel-based space-time manifold structured sparse RPCA model by using two different manifold regularizations to characterize the sparse parts and applied the model to the field of moving target detection. Based on the characteristics of super-pixels, Silva *et al.* [35]proposed an on-line swing single-class integration technique for feature selection in foreground / background separation. Chen *et al.* [21] proposed a background subtraction algorithm based on hierarchical super-pixel segmentation, spanning tree and optical flow by using multiple Gaussian mixture models (GMS) to generate super-pixel segmentation trees. Fang *et al.* [36] proposed a background subtraction method for video analysis based on multi-scale random super-pixel by replacing the custom region segmentation region with the super-pixel segmentation region with similar features, etc. The experimental results show that the method of super-pixel segmentation can effectively solve the problem of the lack of edge information in the extraction of moving objects by traditional models and improve the effect of foreground extraction and background restoration.

However, most of the existed video foreground and background separation models based on super-pixel segmentation technique use SLIC super-pixel segmentation technique. As a local feature-based algorithm, the relationship between SLIC and the global image attributes is not very clear. Hence, it is hard to connect the local feature-based operation with the global optimization objective function, which will greatly affect its boundary dependence and shape compactness and lead to the result that part of dynamic background information may be included in the foreground when dealing with videos with complex dynamic background.

Compared with SLIC, the LSC algorithm not only produces super pixels with high boundary dependence and good shape compactness by bridging the local and global methods, but also captures global image attributes, which effectively solves the high complexity of the global question [24]. Therefore, in this article we will combine the LSC super-pixel segmentation technique with the non-convex rank approximation based RPCA model to establish an improved video background foreground separation model. The non-convex rank approximation function used here is the non-convex-norm proposed in [14], which is defined as follows:

$$\|L\|_\gamma = \sum_j \frac{(1 + \gamma)\sigma_i(L)}{\gamma + \sigma_i(L)}, \quad \gamma > 0. \tag{3}$$

III. THE PROPOSED MODEL AND ALGORITHM

The establishment of the video background and foreground separation model in this article mainly includes three steps. First, use the LSC algorithm to perform super-pixel segmentation on the video sequence, and obtain the super-pixel grouping matrix, which contains the hyper-pixel grouping information of all pixels in each frame of the video sequence. Then, use the improved RPCA model to perform low-rank sparse decomposition on the video to obtain a sparse matrix. In the third step, the motion mask matrix is obtained by the Otsu algorithm (Otsu method) according to the obtained sparse matrix, and then the motion mask and super-pixel grouping matrix is used to further accurately locate the motion area and achieve effective separation of complex dynamic backgrounds. This method is referred as LSCNC-RPCA. Figure 1 shows the flow chart of this method.

A. LSC SUPER PIXEL SEGMENTATION

In the data preprocessing stage of this article, the LSC algorithm is used to segment each frame of the video to form the same region and record the corresponding grouping matrix. The homogenous area is an irregular and similar pixel block with similar texture, color and brightness. It not only has locality and continuity, but also provides effective boundary information for subsequent processing.

The LSC algorithm is proposed based on the study of the relationship between the objective function of the normalized cut set and the weighted K-mean. It first assigns a 5-dimensional feature vector $C_i = (L_i, a_i, b_i, x_i, y_i)^T$ to each color pixel, where (L, a, b) is the converted CIELab color
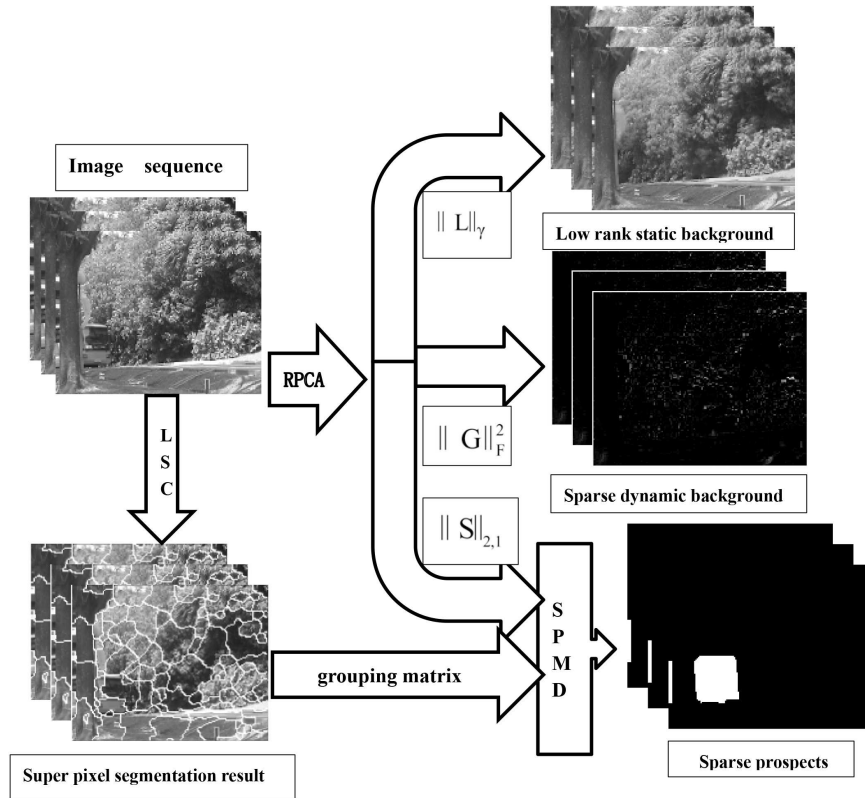
FIGURE 1. The flow chart of LSCNC-RPCA.

space and (x, y) is the spatial coordinates of the pixels. But the algorithm is not done in a 5-dimensional feature vector space. It uses the mathematical equivalence between two seemingly different methods to map data points into a high-dimensional feature space to improve linear separability.

Theorem 1 [24]: If the following equations (4) and (5) are satisfied at the same time, then the weighted K-means clustering and the minimized normalized cut set objective function are mathematically equivalent.

$$w(\mathbf{p})\phi(\mathbf{p}) \cdot w(\mathbf{q})\phi(\mathbf{q}) = \mathbf{W}(\mathbf{p}, \mathbf{q}), \forall \mathbf{p}, \mathbf{q} \in \mathbf{V}, \quad (4)$$

$$w(\mathbf{p}) = \sum_{\mathbf{q} \in \mathbf{V}} \mathbf{W}(\mathbf{p}, \mathbf{q}), \quad \forall \mathbf{p} \in \mathbf{V}, \quad (5)$$

where $w(\cdot)$ represents the weight value of the data point, $\phi(\cdot)$ is the function of mapping the data point to a high-dimensional feature space, and $W(\mathrm{p}, q)$ represents the similarity between the two data points.

In other words, the theorem can be expressed as: if the similarity of two points in the input space is equal to the weighted inner product between two corresponding vectors in the carefully constructed high-dimensional feature space, then the division result of the normalized cut set should be the same as the optimal clustering result of the weighted K-means clustering.

Algorithm 1 gives the steps to implement the LSC super-pixel segmentation algorithm, where V_x/V_y is approximately equal to the aspect ratio of the image, and $t > 0.5$ is a parameter that balances local compactness and global

---

Algorithm 1 LSC

Input: Video sequence, $t = 0.5$, super pixel number k;
Initialize: 1: Map each pixel $p = (L_p, a_p, b_p, x_p, y_p)^T$ to a vector $\phi(p)$ in the high-dimensional feature space;
    2: Uniformly sample k seed points on the image at a fixed horizontal interval V_x and vertical interval V_y;
    3: Initialize the weighted average m_k and search center c_k of the corresponding cluster through the seed point;
    4: Set label $L(p) = 0$ and distance $D(p) = \infty$ for each pixel p;
Iteration: 1: **for** pixels in the $tV_x * tV_y$ field of the search center p **do**
    2: Calculate the Euclidean distance D between $\phi(p)$ and m_k;
    3: **if** $D < d(p)$ **do**
    4:    $d(p) = D, L(p) = k$
    5: **end if**
    6: **end**
    7: Update weighted average and search center until convergence;
    8: merge small superpixels into its neighbors;
Output: Superpixel segmentation result.

---

optimality. In the clustering and merging stage, we have empirically merged small isolated pixels with a quarter of the expected super-pixel size with adjacent large pixels.

After the super-pixel segmentation of each frame of the video sequence is completed by LSC algorithm, the video sequence is labeled according to the super-pixel segmentation result. The pixels belonging to the i-th super-pixel in each frame are marked as $i(i = 1 \cdots k)$, and the grouping information of each video sequence frame is arranged as column vector to form a video grouping matrix $G \in R^{m \times n}$ (where m is the number of pixels in each video frame, n is the number of video frames). The grouping matrix can not only cover the entire video sequence, but also effectively locate the contour boundary of the moving target. Figure 2 shows the LSC super-pixel segmentation result of a frame image in the Watersurface sequence. From Figure 2, it can be found that the area segmented by the LSC algorithm has good shape compactness and high boundary adhesion. Moreover, the homogenous regions tend to have the same motion characteristics. Hence, the grouping matrix of the homogenous regions can be used to extract the sparse foreground.
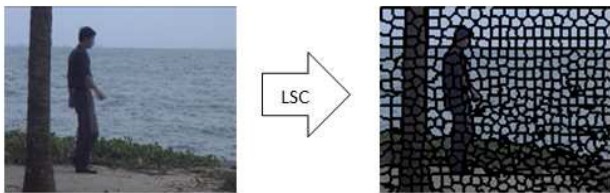

FIGURE 2. LSC super pixel segmentation effect diagram.

B. IMPROVED RPCA MODEL AND SOLUTION
1) PROPOSED MODEL
In order to better extract foreground objects in complex scenes with dynamic background, in this section, we introduce the noise analysis and decompose the visual frequency sequence D into three terms, i.e., $D = A + S + G$. Where A is the low rank static background, S is the sparse foreground, and G is the dynamic background. At the same time, in order to more accurately separate the sparse foreground from the dynamic background and prevent the moving objects from appearing in both S and G, we introduce an incoherent term to constrain S and G so as to improve the separability of the two.

The definition of incoherent terms is as follows:

$$\Theta(S, G) = \sum_{r=1}^{n} S_r^T G_r, \tag{6}$$

where S_r and G_r represent the rth column of S and G respectively.

In this article, we use the γ-norm to approximate the rank function of the matrix, and use the $l_{2,1}$ norm to describe the sparse prospects. Besides, note that the dynamically changing background is usually unstructured and non-sparse, we use the F-norm to describe the noise matrix. Then we can get the following improved RPCA model:

$$\min \left(\|L\|_\gamma + \lambda \|S\|_{2,1} + \beta \Theta(S, G) + \alpha \|G\|_F^2 \right)$$
$$\text{s.t. } D = L + S + G, \tag{7}$$

where $D, L, S, G \in R^{m \times n}$ are the original video data matrix, low rank background matrix, sparse foreground matrix and noise matrix, respectively. $\| \cdot \|_\gamma$ is defined by (3), $\gamma > 0$; $\| \cdot \|_{2,1}$ represents the $l_{2,1}$ norm, defined as $\|S\|_{2,1} = \sum_{j=1}^{n} \sqrt{\sum_{i=1}^{m} S_{i,j}^2}$; $\Theta(S, G)$ is the incoherent term; $\| \cdot \|_F$ is the Frobenius norm of the matrix; λ, α and β are penalty parameters.

2) SOLVING THE MODEL
In this section, we will use the alternating direction multiplier method (ADMM) [37] to solve the model (7). The augmented Lagrangian function of (7) is:

$$\mathcal{L}(L, S, Y, \mu)$$
$$= \|L\|_\gamma + \lambda \|S\|_{2,1} + \alpha \|G\|_F^2 + \beta \sum_{r=1}^{n} S_r^T G_r$$
$$+ \langle Y, D - L - S - G \rangle + \frac{\mu}{2} \|D - L - S - G\|_F^2, \tag{8}$$

where $Y \in R^{m \times n}$ is the Lagrange multiplier and μ is the penalty parameter.

Assume that the current number of iterations is k, then the variables and multipliers are updated as follows:
1) Fix S and G to update L^{k+1}:

$$L^{k+1} = \arg \min_L \|L\|_\gamma$$
$$+ < Y^k, D - L^k - S^k - G^k >$$
$$+ \frac{\mu^k}{2} \left\| D - L^k - S^k - G^k \right\|_F^2, \tag{9}$$

Equation (9) can be simplified as:

$$L^{k+1} = \arg \min_L \|L\|_\gamma$$
$$+ \frac{\mu^k}{2} \left\| L - \left(D - S^k - G^k + \frac{Y^t}{\mu^t} \right) \right\|_F^2. \tag{10}$$

In order to solve the problem (10), we need the following Theorem 2.

Theorem 2 [14]: Assume that $A = U \operatorname{diag}(\sigma_A) V^T$ is the singular value decomposition (SVD) of matrix $A \in R^{m \times n}$, where σ_G is the singular value of G. Let $F(Z) = f \circ \sigma(Z)$ be an unitary invariant function, and $\mu > 0$. Then the optimal solution for the problem

$$\min_Z F(Z) + \frac{\mu}{2} \|Z - A\|_F^2, \tag{11}$$

is $A^* = U diag(\sigma^*)V^T$, where

$$\sigma^* = \operatorname*{prox}_{f,\mu}(\sigma_A) = \arg \min_{\sigma \geq 0} f(\sigma) + \frac{\mu}{2} \|\sigma - \sigma_A\|_2^2. \tag{12}$$

In (10), let $F(L) = \|L\|_\gamma, A = D - S^k - G^k$, then according to Theorem 2 we can obtain

$$L^{k+1} = U diag(\sigma^*)V^T, \tag{13}$$

where σ^* is given by equation (12) and can be iterated by the following equation [14]:

$$\sigma^{k+1} = max(0, \sigma_A - \frac{\partial f(\sigma^k)}{\mu^k}). \qquad (14)$$

2) Fix A and G to update S^{k+1}:

$$S^{k+1} = \arg\min \lambda \|S\|_{2,1} + \beta \sum_{r=1}^{n} S_r^T G_r^k$$
$$+ < Y^k, D - L^{k+1} - S^k - G^k >$$
$$+ \frac{\mu^k}{2} \left\| D - L^{k+1} - S^k - G^k \right\|_F^2, \qquad (15)$$

Simplify the formula (15), we get:

$$S^{k+1} = \arg\min_{L} \frac{\lambda}{\mu^k} \|S\|_{2,1}$$
$$+ \frac{1}{2} \left\| S - \left(D - L^{k+1} - \left(\frac{\lambda}{\mu^k} + 1 \right) G^k + \frac{Y^k}{\mu^k} \right) \right\|_F^2. \qquad (16)$$

For (15), we use the following Theorem 3 proposed in [38] to iteratively update.

Theorem 3 [38]: For a given matrix $M \in R^{m \times n}$ and parameter $\tau > 0$, the following optimization problem

$$arg\min_{S} \tau \|S\|_{2,1} + \frac{1}{2} \|S - M\|_F^2, \qquad (17)$$

has a closed form of solution $S^* = (S_1^*, \cdots, S_n^*)$, where

$$S_j^* = max(0, \frac{\|M_j\|_2 - \tau}{\|M_j\|_2}) M_j, j = 1, \cdots, n, \qquad (18)$$

M_j is the jth column of matrix j-th column of matrix M.

In sub-problem (15), let $M = D - L^{k+1} - \left(\frac{\lambda}{\mu^k} + 1 \right) G^k + \frac{Y^k}{\mu^k}$, then we can obtain the solution of sub-problem (16) by (18).

3) Fix A and S to update G^{k+1}:

$$G^{k+1} = \arg\min_{L} \alpha \|F\|_F^2 + \beta \sum_{r=1}^{n} \left(S_r^{k+1} \right)^T G_r$$
$$+ \frac{\mu^k}{2} \left\| G - \left(D - L^{k+1} - S^{k+1} + \frac{Y^k}{\mu^k} \right) \right\|_F^2. \qquad (19)$$

Derive the objective function of equation (19) with respect to G, we can get the update formula of G^{k+1} as follows

$$G^{k+1} = \frac{\mu^k D - \mu^k L^{k+1} - \left(\mu^k + \beta \right) S^{k+1} + Y^k}{2\alpha + \mu^k}. \qquad (20)$$

4) Update parameters Y^{k+1}, μ^{k+1}:

$$Y^{k+1} = Y + \mu \left(D - L^{k+1} - S^{k+1} - G^{k+1} \right), \qquad (21)$$

$$\mu^{k+1} = max \left(\rho\mu^k, \mu_{\max} \right), \rho > 0. \qquad (22)$$

In conclusion, the algorithm framework for solving problem (7) is as follows:

Algorithm 2

Input: Original video sequence D, maximum number of iterations k_{max};

Initialize: $L = S = G = 0$, $Y = 0$, related parameter parameters $\lambda, \alpha, \beta, \mu_0, \varepsilon, \rho$, $k=0$,$\mu_{\max}$;

    while not converged **do**

        Update L^{k+1} by (13);

        Update S^{k+1} by (17);

        Update G^{k+1} by (19);

        Update Y^{k+1} by (20);

        Update μ^{k+1} by (21);

        Check the convergence condition: $k > k_{max}$ or

        $\frac{\|D-L^{k+1}-S^{k+1}-G^{k+1}\|_F^2}{\|D\|_F^2} \leq \varepsilon$;

    end while

Output: L,S,G.

C. SUPER-PIXEL MOTION DETECTION

In this section, the super-pixel motion detection is performed and the super-pixel segmentation results are combined with the sparse matrix generated by the RPCA model to analyze the correlation and obtain the motion mask of the image. Then the moving object is extracted from the relevant original video sequence.

Specifically speaking, for the sparse foreground matrix $S \in R^{m \times n}$(m is the number of pixels per frame of the video and n is the number of video frames) obtained by Algorithm 2, we use the Ostu method [39] to automatically calculate the motion mask $M \in R^{m \times n}$. The optimal decision threshold τ can be calculated, using the Ostu method, by minimizing the intra group variance, namely

$$\tau = \arg\min_{g} \left(\omega_0(g)\sigma_0^2(g) + \omega_1(g)\sigma_1^2(g) \right). \qquad (23)$$

For each column $S_j, j = 1 \cdots n$ of the sparse matrix S, $\omega_0(g)$ and $\omega_1(g)$ are the probability of moving pixels whose pixel value is smaller than the threshold τ while background pixels larger than the threshold τ,respectively, and $\omega_0(g) + \omega_1(g) = 1$, $\sigma_0^2(g)$ and $\sigma_1^2(g)$ are the corresponding intra-class variance. In this way, a motion mask based on the improved RPCA model can be obtained:

$$M_j(i, 1) = \begin{cases} 0, & S_j(i, 1) \geq \tau \\ 1 & S_j(i, 1) < \tau \end{cases}, \qquad (24)$$

where $M_j(i, 1)$ and $S_j(i, 1)$,$i = 1 \cdots m$ represent the values of column j of M and S, respectively, where $i = 1 \cdots m$.

Next, the motion mask M and the super-pixel segmentation grouping matrix G combined to obtain the final super-pixel motion detection motion mask N. Count the number of moving pixels in each homogeneous region in the same group. If $Q > $ ceil $\left(\frac{m}{\theta k} \right)$, the whole homogeneous region is classified as the moving pixels, where ceil$(\cdot)$ stands for the rounding-up function, m is the number of pixels per frame, k represents the number of super-pixels, and θ is the constraint parameter. Then the super-pixel motion detection motion mask N can be

expressed as follows:

$$\forall (x, y) \in SR :$$

$$N(x, y) = \begin{cases} 1, & Q\left\{(x, y') \in SR : M(x', y') = 1\right\} \\ & > \text{ceil}\left(\frac{m}{\theta k}\right) \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

Algorithm 3 presents the algorithm framework for super-pixel motion detection(SPMD).

Algorithm 3 SPMD

Input: Sparse foreground matrix S, hyperpixel segmentation and grouping matrix $G, m, k, \theta = 3$;
Initialize: 1:Formula (22) is used to calculate the optimal decision threshold τ;
  2:Equation (23) is used to calculate the sparse matrix motion mask M;
  3:**for** y=1: n;
  4: **for** j=1:g(y);
  5:    **for** x=1: m;
  6:    Statistic $Q(j)$ according to principle $(x, y) \in SR(j)$ and $M(x, y) = 1$;
  7:      **for** $Q(j) > q$
  8:      $N(x, y) = 1$;
  9:      **else**;
  10:      $N(x, y) = 0$;
  11:      **end** ;
  12:    **end**;
  13: **end**;
  14:**end**;
Output: N.

Figure 3 shows an example of super-pixel motion detection. For the traditional RPCA model, due to the uniformity of image texture or clustering errors, many holes are left in the foreground. By grouping adjacent pixels in the same homogenous region through the super-pixel grouping matrix, these holes can be correctly classified into motion plates, thereby reducing unexpected errors caused by static objects or object clustering.
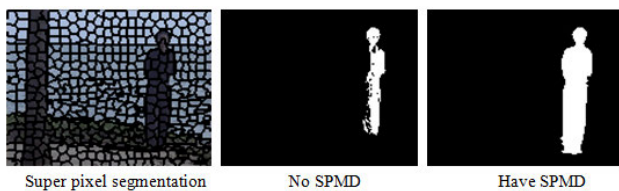

FIGURE 3. Example of superpixel motion detection.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we will verify the performance of the model proposed in this article through experiments, using a combination of quantitative and qualitative methods. The algorithm in this article was compared with GoDec [13], NC-RPCA [14], AccAltProj [40], KNN [41],

LRSD-TNNSR [42] and NCSC-RPCA [43]. We mainly selected eight video sequences in I2R dataset and CDnet2014 dataset for verification. All numerical experiments are based on the PC Inter Core i5-4210U 2.40Hz CPU, 4GB RAM environment, and are programmed using MATLAB R2019b.

A. PARAMETER SELECTION
In the model proposed in this article, we mainly use six parameters: $\lambda, \alpha, \beta, \mu_0, \varepsilon, \rho$. For the parameters λ, we adopt an adaptive value to different video sequences and set $\lambda = 5*10^{-3}/\sqrt{\min(\text{size}(D))}$. For the penalty parameter μ_0, we take $\mu_0 = 10^{-3}$. In order to better separate the dynamic background from the foreground, we take $\alpha = 5 \times 10^{-2}$ and $\beta = 10^{-3}$ to perform the relevant constraints. The parameter ρ is used to control the convergence rate, here we set $\rho = 1.1$.

In order to make different models comparable, all the methods here use the same stopping criterion, that is, when the number of iterations reaches $k_{max} = 300$ or the relative error is less than $\varepsilon = 10^{-6}$, the iteration stops.

B. NUMBER OF SUPERPIXELS
Fig.4 shows the FM values extracted from the moving trees, fountain, switchlight and override video sequences for different number of hyper-pixels. Among them, the moving objects in the moving trees and overpass video sequences are relatively large, while the moving objects in the fountain and switchlight video sequences are relatively small. It can be seen from the figure that the pixel block is not the more the better since the more the pixel blocks are, the less pixels are included in the block. On the other hand, the boundary of super-pixel block segmented by LSC can not completely coincide with the boundary of the moving object. Hence, when some of the background texture features are similar to the moving object, they will be divided into the same category, such that the moving object will be regarded as the background in the extraction process when there are too many background pixels in the super pixel block. In addition, for
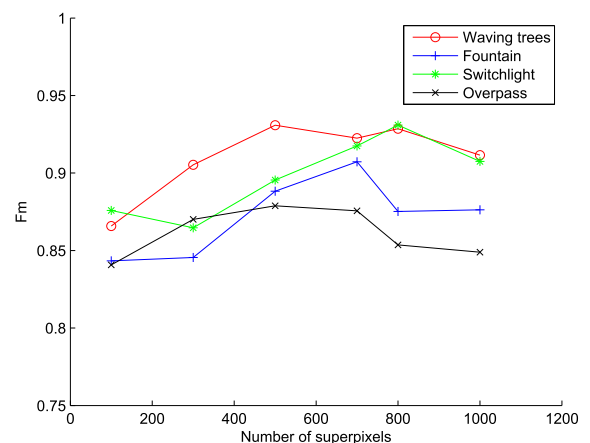

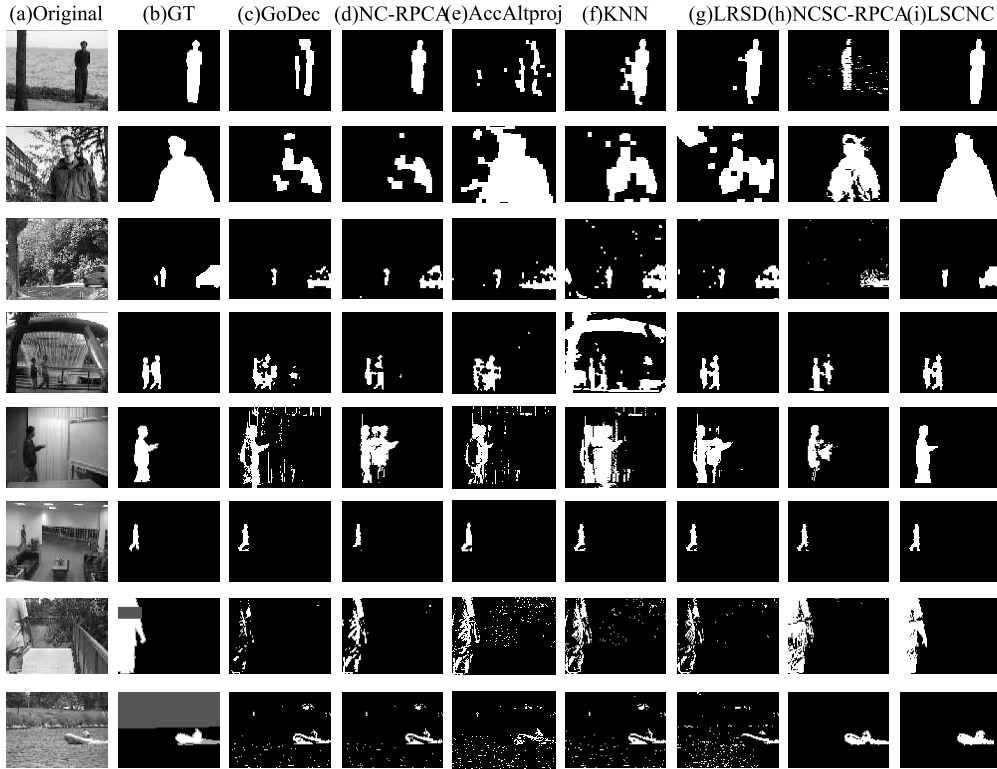**FIGURE 4.** Fm value in different number of superpixels.

FIGURE 5. The detection results of different algorithms in the data set.

video sequences with relatively small moving targets, the setting of super-pixel blocks should be as large as possible, because the smaller the moving targets are, the less obvious their texture will be.

Based on the above analysis, we divide the number of super pixels into two categories for the data set used in this article. The four types of moving objects, namely, water surface, waving trees, override and curtain, are relatively large, and the number of super pixels is set to 500, while Campous, fountain, boats and switchlight are relatively small, and the number of super pixels is set at 750.

C. QUANTITATIVE ANALYSIS INDICATORS

In order to evaluate the performance of each model more accurately and objectively in the quantitative comparative analysis, we refer to [44] and [45], and uses four indicators to evaluate the performance. To simplify the representation, the ground truth and estimated foreground are expressed as B_{GT} and B_E, respectively. The specific evaluation indexes are as follows:

1. Average gray error (AGE): The average gray error is the average of the absolute errors of B_{GT} and B_E, where B_{GT} and B_E are the values that have been converted to grayscale.

2. Percentage of error pixels (pEPs): Error pixels (EP) refers to the pixels whose difference between B_{GT} and B_E is non-zero (In the model results, white is used to represent the extracted target, and black is the background). PEPs is defined as the ratio of EPs to total pixels.

3. Peak signal-to-noise ratio (PSNR): PSNR is widely used to measure image quality. It is defined as

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX^2}{MSE}\right), \qquad (26)$$

where, the MAX value is 255, and MSE is the mean square error between B_{GT} and B_E.

4. F-measure: It is a comprehensive index based on comprehensive consideration of the performance of Precision (foreground segmentation precision) and Recall (previous spot accuracy), and is often used to evaluate the effect of dynamic video detection. It is defined as follows:

$$Precision = \frac{TP}{TP + FP}, \qquad (27)$$

$$Recall = \frac{TP}{TP + FN}, \qquad (28)$$

$$F - measure = 2\frac{Precision \times Recall}{Precision + Recall}. \qquad (29)$$

Among them, TP indicates the foreground pixel correctly marked as foreground, FP indicates that the background pixel is wrongly marked as foreground, and FN indicates that the foreground pixel is wrongly marked as background.

D. EXPERIMENTAL RESULT AND ANALYSIS

In this article, we mainly selects eight video sequences for verification, namely WaterSurface, Waving trees, Campous, Fountain, Overpass, Boats, Curtain sequence and SwitchLight. Among them, the first six video sequences are outdoor test sequences and the latter two are indoor video sequences.

TABLE 1. Comparison of objective results of foreground detection by different methods (Best: bold. Second best: underline).

Dataset	Index	GoDec	NC-RPCA	AccAltProj	KNN	LRSD-TNNSR	NCSC-RPCA	LSCNC-RPCA
watersurface	Fm	0.56794	0.87866	0.33679	0.81063	<u>0.90437</u>	0.46186	**0.97181**
	AGE	17.3049	5.1168	26.032	10.0572	<u>4.6011</u>	21.0645	**1.903**
	PSNR	29.9073	32.5531	29.0206	31.0857	<u>32.7838</u>	29.4803	**35.524**
	pEPs	0.067862	0.020066	0.10209	0.03944	<u>0.018043</u>	0.082606	**0.0051096**
Wavingtrees	Fm	0.54751	0.4444	0.80112	0.72874	0.59158	<u>0.80339</u>	**0.93086**
	AGE	51.7441	58.8323	40.1997	<u>36.522</u>	61.5775	28.9835	**11.2756**
	PSNR	25.5293	27.2505	28.0777	28.2861	27.1514	<u>28.7883</u>	**30.8399**
	pEPs	0.20312	0.23092	0.15789	0.14343	0.24168	<u>0.11386</u>	**0.044372**
Campous	Fm	0.67346	0.70514	0.6484	0.67705	<u>0.7584</u>	0.44224	**0.89352**
	AGE	7.8313	7.6277	10.2744	12.1067	<u>7.0306</u>	12.9753	**3.1217**
	PSNR	31.6289	31.6861	31.0393	30.683	<u>31.8632</u>	30.5325	**33.6262**
	pEPs	0.030711	0.029913	0.040292	0.047477	<u>0.027571</u>	0.050884	**0.012242**
Fountain	Fm	0.67282	0.81337	0.63026	0.18265	<u>0.898</u>	0.6087	**0.90725**
	AGE	10.0979	4.1668	10.383	82.6022	<u>2.4973</u>	9.1614	**2.3616**
	PSNR	31.0769	31.9991	31.0165	26.5131	<u>34.1107</u>	31.2883	**34.2321**
	pEPs	0.0396	0.01634	0.040717	0.32393	<u>0.0097935</u>	0.035927	**0.0092612**
Curtain	Fm	<u>0.70914</u>	0.54722	0.5252	0.4169	0.65895	0.50252	**0.9409**
	AGE	<u>15.5541</u>	27.2671	23.0189	47.9516	19.5987	20.128	**2.7009**
	PSNR	<u>30.1389</u>	28.9199	29.2877	27.6941	29.637	29.5791	**33.9405**
	pEPs	<u>0.060996</u>	0.10693	0.09027	0.18805	0.076858	0.078933	**0.010592**
Switchlignt	Fm	0.89394	0.787	0.89394	0.8869	<u>0.89939</u>	0.88535	**0.93093**
	AGE	0.95007	1.6016	0.95007	**0.10315**	0.89578	0.97722	<u>0.62433</u>
	PSNR	36.2093	35.0754	36.2093	36.0307	<u>36.3371</u>	36.1481	**37.121**
	pEPs	0.0037258	0.0062806	0.0037258	0.0040451	<u>0.0035129</u>	0.0038322	**0.0024484**
Overpass	Fm	0.16364	0.46456	0.50315	0.453	0.35788	<u>0.78169</u>	**0.87887**
	AGE	40.9107	32.4428	33.8218	35.5373	36.94188	<u>16.4294</u>	**9.5474**
	PSNR	28.2079	28.7578	28.6582	28.5403	28.484	<u>30.4614</u>	**32.0252**
	pEPs	0.18848	0.15496	0.16066	0.16707	0.17279	<u>0.088519</u>	**0.061816**
Boats	Fm	0.47286	0.45424	0.3079	0.45627	0.26683	**0.86633**	<u>0.84649</u>
	AGE	50.3092	50.2007	54.1759	54.9677	55.9833	**43.0975**	<u>43.898</u>
	PSNR	29.3489	29.3595	29.0029	29.286	28.8582	**30.1918**	<u>30.0802</u>
	pEPs	0.5198	0.51949	0.53465	0.52237	0.5412	**0.49383**	<u>0.49767</u>

All the sequences have light changes, and the light changes of SwitchLight sequence are more complicated than the Curtain sequence.

Figure 5 shows the detection results of GoDec, NC-RPCA, AccAltProj, KNN, LRSD-TNNSR, NCSC-RPCA and our proposed algorithm on the selected eight video sequences, where (a) represents the original video frame selected from 8 different dynamic background videos, (b) represents the true foreground of the corresponding video frame, (c) to (i) represent the foreground of the eight video sequences extracted by GoDec, NC-RPCA, AccAltProj, KNN, LRSD-TNNSR, NCSC-RPCA and our proposed algorithm.

The first line in Figure 5 is the WaterSurface dataset. In this dataset, the foreground is relatively simple, but the ripples on the sea surface will appear as background motion during detection, which increases the difficulty of pedestrian detection. It can be seen from the figure that although LRSD-TNNSR and KNN can suppress part of the disturbed background, some water surface ripples will still be detected

as the foreground, and the foreground area is not smooth enough. The calf part is missed from the NC-RPCA and NCSC-RPCA detection results. However, the proposed algorithm in this article can not only effectively extract the moving target, but also remove the influence of the water surface ripple disturbance, and the effect is significantly better than other methods.

Waving trees, Campous datasets and Overpass are all video sequences with a large number of undulating leaves. The background motion is relatively complex and can be easily extracted as foreground. The second, third and seventh lines show the detection results of these three sequences on these algorithms respectively. Irregular motion of shaking leaves in the background is the main obstacle to detecting moving objects. It can be seen from the figure that AccAltProj, KNN and LRSD-TNNSR cannot effectively eliminate the effects of shaking leaves in the background. Although GoDec, NC-RPCA and NCSC-RPCA can suppress the effects of shaking leaves, the detection area is very incomplete.

In contrast, our method can detect more complete moving targets. The NCSC-RPCA algorithm has a good effect on the extraction of Overpass data set extraction, but the void phenomenon is serious in the extracted target. For the algorithm in this article, since the target is close to the lens and the clothing texture is obvious, the LSC algorithm divides the target into several small hyper-pixel blocks. When more pixels are misdivided in the area of the small hyperpixel block, there will be a cavity in the extracted target. However, by contrast, our method can detect more complete moving targets.

The Fountain data set in the fourth row is a video sequence with a dynamic fountain in the background. It can be found that among the eight algorithms, LRSD-TNNSR and our algorithm have better moving target extraction effects. Besides, our algorithm extracts the most complete moving target area, and is least affected by the irregular movement of fountain.

The fifth and sixth lines are video sequences with illumination changes in the background, which are the Curtain and SwitchLight data sets. It can be found from the results that for the Curtain dataset, except for the algorithm in this article, there are obvious "phantom" phenomenon in the detection results of the other six algorithms. Meanwhile, the contour of the motion foreground detected by our algorithm is more complete and the background is cleaner. In general, for the SwitchLight dataset, our algorithm is hardly affected by illumination changes, and can effectively detect moving objects without extracting human shadow from the ground. From the comparison of the above experimental results, we can see that our algorithm can effectively eliminate the moving background and extract more complete foreground targets for video sequences with complex moving background.

The last line is the Boats data set, a data surface ripple that occupies a large portion of the image, which greatly affects the effect of foreground extraction. Although the denoising effect of the proposed algorithm is better than that of the previous five algorithms and the target extraction accuracy is similar to that of the CNSC-RPCA algorithm, due to the influence of water surface ripples, the LSC algorithm can't segment the moving object's boundary texture accurately, which makes the extracted moving object not smooth enough.

Table 1 shows the quantitative evaluation results of the proposed algorithm and other six methods on the test sequence. In this article, we use F-measure, Average Grayscale Error (AGE), Peak Signal-to-Noise Ratio (PSNR) and Error Pixel percentage scores (pEPs) for comparative analysis. From the definition, we know that the larger the values of F-measure and PSNR, the better the effect, and the smaller the values of AGE and pEPs, the better. From the comparison of the relevant evaluation index data, we can find that our algorithm has a great improvement in terms of F-measure, AGE, PSNR and pEPs, compared with the other six algorithms, in the eight data sets with complex dynamic background. Specially, for the data set with relatively large moving target area, the promotion effect is more significant. According to the

F-measure values of the eight selected datasets with different degrees of dynamic background interference, our algorithm has obvious advantages in target extraction when there is huge dynamic background disturbance in the video background, which shows that the model proposed in this article has strong robustness.

V. CONCLUSION

In order to effectively extract the moving target under complex dynamic background, this article proposes a video background separation model based on improved RPCA and super pixel motion detection. The new model combines the improved RPCA model and the super-pixel grouping matrix to realize the moving object detection, and effectively solves the problems of incomplete detection boundary and easy to appear holes in traditional foreground extraction. Experimental results show that when the moving target is large, the superiority of the algorithm is most obvious, while when the moving target is relatively small, the detection effect is slightly poor and prone to missing phenomenon. This will be the focus of our future research.

REFERENCES

[1] D. Zhang, O. Javed, and M. Shah, "Video object segmentation through spatially accurate and temporally dense extraction of primary object regions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 628–635.

[2] M. Hintermüller and T. Wu, *Robust Principal Component Pursuit Via Inexact Alternating Minimization on Matrix Manifolds*. Norwell, MA, USA: Kluwer, 2015.

[3] T. Bouwmans, F. Porikli, B. Hferlin, and A. Vacavant, *Background Modeling and Foreground Detection for Video Surveillance*. Boca Raton, FL, USA: CRC Press, 2014.

[4] E. Candès, X. Li, Y. Ma, and W. John, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, 2011.

[5] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, Mar. 2013.

[6] B. Yang and L. Zou, "Robust foreground detection using block-based RPCA," *Optik*, vol. 126, no. 23, pp. 4586–4590, Dec. 2015.

[7] Y. Chen, S. Wang, and Y. Zhou, "Tensor nuclear norm-based low-rank approximation with total variation regularization," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1364–1377, Dec. 2018.

[8] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, Jul. 2018.

[9] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Comput. Sci. Rev.*, vol. 23, pp. 1–71, Feb. 2017.

[10] S. Javed, A. Mahmood, S. Al-Maadeed, T. Bouwmans, and S. K. Jung, "Moving object detection in complex scene using spatiotemporal structured-sparse RPCA," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 1007–1022, Feb. 2019.

[11] X. Cao, L. Yang, and X. Guo, "Total variation regularized RPCA for irregularly moving object detection under dynamic background," *IEEE Trans. Cybern.*, vol. 46, no. 4, pp. 1014–1027, Apr. 2016.

[12] T. Bouwmans and B. Garcia-Garcia, "Background subtraction in real applications: Challenges, current models and future directions," Tech. Rep., 2019.

[13] T. Zhou and T. Dacheng, *Godec: Randomized Low-Rank & Sparse Matrix Decomposition in Noisy Case*. Berlin, Germany: Springer, 201, pp. 33–401.

[14] Z. Kang, C. Peng, and Q. Cheng, "Robust PCA via nonconvex rank approximation," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2015, pp. 211–220.

[15] Y. Chen, Y. Wang, M. Li, and G. He, "Augmented lagrangian alternating direction method for low-rank minimization via non-convex approximation," *Signal, Image Video Process.*, vol. 11, no. 7, pp. 1271–1278, Oct. 2017.

[16] Y. Chen, X. Xiao, and Y. Zhou, "Low-rank quaternion approximation for color image processing," *IEEE Trans. Image Process.*, vol. 29, pp. 1426–1439, 2020.

[17] S. Ebadi and E. Izquierdo, "Foreground segmentation via dynamic tree-structured sparse RPCA," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 314–329.

[18] S. Javed, S. H. Oh, A. Sobral, T. Bouwmans, and S. K. Jung, "Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 90–98.

[19] S. Javed, A. Mahmood, and T. Bouwmans, "Superpixels-based manifold structured sparse RPCA for moving object detection," in *Proc. Brit. Mach. Vis. Conf.*, 2017, pp. 4–7.

[20] S. Javed, S. Ki Jung, A. Mahmood, and T. Bouwmans, "Motion-aware graph regularized RPCA for background modeling of complex scenes," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 120–125.

[21] M. Chen, X. Wei, Q. Yang, Q. Li, G. Wang, and M. H. Yang, "Spatiotemporal GMM for background subtraction with superpixel hierarchy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1518–1525, Jun. 2018.

[22] X. Ren and M. Jitendra, "Learning a classification model for segmentation," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, p. 10.

[23] A. P. Moore, S. J. D. Prince, J. Warrell, U. Mohammed, and G. Jones, "Superpixel lattices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[24] J. Chen, Z. Li, and B. Huang, "Linear spectral clustering superpixel," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3317–3330, Jul. 2017.

[25] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[26] W. Tao, H. Jin, and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts," *IEEE Trans. Syst., Man, Cybern., B (Cybern.)*, vol. 37, no. 5, pp. 1382–1389, Oct. 2007.

[27] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "TurboPixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, Dec. 2009.

[28] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to State-of-the-Art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[29] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, Jan. 2017.

[30] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, Sep. 2013.

[31] R. Cabral, F. De la Torre, J. P. Costeira, and A. Bernardino, "Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2488–2495.

[32] G. Liu and S. Yan, "Active subspace: Toward scalable low-rank learning," *Neural Comput.*, vol. 24, no. 12, pp. 3371–3394, Dec. 2012.

[33] Y. Chen, Y. Wang, M. Li, and G. He, "Augmented lagrangian alternating direction method for low-rank minimization via non-convex approximation," Ph.D. dissertation, 2017.

[34] C. Peng, Z. Kang, H. Li, and Q. Cheng, "Subspace clustering using log-determinant rank approximation," Ph.D. dissertation, 2015.

[35] C. Silva, T. Bouwmans, and C. Frelicot, "Superpixel-based online wagging one-class ensemble for feature selection in foreground/background separation," *Pattern Recognit. Lett.*, vol. 100, no. 1, pp. 144–151, Dec. 2017.

[36] W. Fang, T. Zhang, C. Zhao, D. B. Soomro, R. Taj, and H. Hu, "Background subtraction based on random superpixels under multiple scales for video analytics," *IEEE Access*, vol. 6, pp. 33376–33386, 2018.

[37] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, *arXiv:1009.5055*. [Online]. Available: http://arxiv.org/abs/1009.5055

[38] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5366–5380, Sep. 2017.

[39] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[40] H. Q. Cai, J. F. Cai, and K. Wei, "Accelerated alternating projections for robust principal component analysis," *J. Mach. Learn. Res.*, vol. 20, no. 1, pp. 685–717, 2017.

[41] Z. Zhong, B. Zhang, G. Lu, Y. Zhao, and Y. Xu, "An adaptive background modeling method for foreground segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1109–1121, May 2017.

[42] Z. Xue, J. Dong, Y. Zhao, C. Liu, and R. Chellali, "Low-rank and sparse matrix decomposition via the truncated nuclear norm and a sparse regularizer," *Vis. Comput.*, vol. 35, no. 11, pp. 1549–1566, Nov. 2019.

[43] Z. Hu, Y. Wang, R. Su, X. Bian, H. Wei, and G. He, "Moving object detection based on non-convex RPCA with segmentation constraint," *IEEE Access*, vol. 8, pp. 41026–41036, 2020.

[44] Z. Xu, B. Min, and R. C. C. Cheung, "A robust background initialization algorithm with superpixel motion detection," *Signal Process., Image Commun.*, vol. 71, pp. 1–12, Feb. 2019.

[45] Y. Li, G. Liu, Q. Liu, Y. Sun, and S. Chen, "Moving object detection via segmentation and saliency constrained RPCA," *Neurocomputing*, vol. 323, pp. 352–362, Jan. 2019.

YONGLI WANG received the B.S. degree in operations research and control theory from the Shandong University of Science and Technology, Qingdao, China, in 2001, and the Ph.D. degree in applied mathematics from Shanghai Jiao Tong University, Shanghai, China, in 2006. Her research interests include optimization theory and method, low-rank representation, image processing, and distributed computation.

HONGCHAO WEI received the B.S. degree in information and computing science from Qingdao Agricultural University, China, in 2018. He is currently pursuing the M.S. degree with the College of Mathematics and Systems Science, Shandong University of Science and Technology. His research interests include image segmentation and clustering algorithms.

XIAOYUN DING received the B.S. degree in applied mathematics from Binzhou University, China, in 2019. She is currently pursuing the M.S. degree with the College of Mathematics and Systems Science, Shandong University of Science and Technology. Her research interests include image processing, moving object detection, and image denoising.

JULIANG TAO received the B.S. degree in information and computing science from the Lanzhou University of Technology, China, in 2019. She is currently pursuing the M.S. degree with the College of Mathematics and Systems Science, Shandong University of Science and Technology. Her research interests include nonconvex low-rank matrix decomposition, moving object detection, and image processing.

• • •