Moving target segmentation method based on tensor decomposition and graph Laplacian regularization

Shudan Yuan[1], Hui Wang[2]
[1, 2]School of Artificial Intelligence, Hezhou University, Hezhou, China
[2]The Key Laboratory of Cognitive Radio and Information Processing, Guilin University of Electronic Technology, Guilin, China
[2]Corresponding author
E-mail: [1]*200500032@hzxy.edu.cn*, [2]*syswangxueleng@163.com*

Check for updates

Abstract. A new method for segmenting the target foreground from image frames is proposed by utilizing the theory of graph signal processing and the tensor decomposition model aiming at the problem that the segmentation results of the existing foreground segmentation methods in image frames under dynamic scenes are not high in accuracy. The intrinsic connection between image pixels in each frame of an image sequence is modeled as a graph, the image pixel intensities are modeled as graph signals, and the correlation between pixels is characterized by the graph model. According to the significant difference between the dynamic background and the target change in the moving foreground in the image sequence, the dynamic background region in each image frame is smoothed and suppressed, and the disturbing information of the dynamic background is transformed into the useful component information in the low-rank subspace. The connectivity between image pixels can be characterized by the graph Laplacian regularization term, and then the target foreground segmentation problem in the image sequence is equivalent to a constrained optimization problem with tensor decomposition and graph Laplacian regularization term. The alternating direction multiplier method is used to solve the optimization problem, and the simulation results on real scene data set verify the effectiveness of the algorithm.

Keywords: graph signal processing, moving object segmentation, dynamic background, tensor decomposition.

1. Introduction

Computer vision technology has been widely applied in fields such as automatic image surface defect detection [1-3] and video object segmentation [4-6], the main purpose of moving target segmentation is to effectively separate the foreground and background in an image. In practical application scenarios, the influence of lighting changes, rain and snow, camera jitter and camera defocus increases the difficulty of moving target detection, and how to quickly and accurately extract moving target information is a research hotspot. Currently, the moving object detection algorithms that have been studied more include inter-frame difference method [7], optical flow method [8], background subtraction method [9] and neural network method [10, 11]. Among them, the inter-frame difference algorithm is simple but easy to be disturbed by noise, which is mainly used in simple scenes. The optical flow method is more computationally intensive and easy to be affected by noise, illumination, etc. The key step of the background subtraction method is to model the background, and the modeling process usually assumes that the target foreground and the image background of image sequences obey the Gaussian distribution, and Kim et al. [12] proved that in indoor scenes, the image background and the foreground more closely obey the Laplace distribution, while for some complex scenes such as dynamic background modeling is more difficult. Video object segmentation methods based on deep learning have also been extensively studied. For example, Zhou et al. [13] studied a video object segmentation method based on semi-supervised transformer framework. Liang et al. [14] explored the task of language-guided video segmentation, they augment the transformer architecture with a finite memory so as to query the

entire video with the language expression in an efficient manner. However, deep learning methods require a large number of samples for model training.

Principal component analysis (PCA) [15] has a wide range of applications in motion target detection, which can simultaneously achieve the separation of the image background and the target foreground, but it also has major limitations, especially in the target detection in complex scenes is not ideal. Further, researchers have proposed robust principal component analysis (RPCA) algorithms [16] and their improved algorithms such as principal component pursuit based RPCA algorithm [17], approximated RPCA algorithm [18] and other improved RPCA algorithms [19], but none of these algorithms take advantage of the structural information features such as spatial-temporal continuity between each frame of an image sequence. In order to make full use of this spatial-temporal information, moving target detection methods under the tensor model have also been investigated. Typical algorithms such as the total variation regularized robust principal component analysis algorithm (TVRPCA) [20], which utilizes spatial-temporal continuity to detect the target foreground, and the algorithm further classifies the sparse components into the target foreground and the dynamic background, but the foreground and the background are prone to misclassification. Iterative block tensor singular value thresholding (IBTSVT) [21] algorithm that uses tensor principal component analysis to extract the target foreground, the algorithm loses the useful structural information of the tensor block when unfolding the tensor calculation. Incremental and multi-feature tensor subspace learning (IMTSL) algorithm [22] builds a robust low-rank background model based on multiple features, and the selection of features has a large impact on the results of the algorithm, while the algorithm computes with poor real-time performance. Patch-group-based tensor robust principal component analysis (PG-TenRPCA) algorithm [23] constructs a 4th order tensor by searching for non-local similarity information in each frame of an image sequence and then solves it optimally, which is too computationally intensive. Moving target detection algorithms utilizing RPCA in 3-D space have also been investigated [24, 25]. These algorithms are difficult to obtain satisfactory results when dealing with image sequences in the presence of a dynamic background.

In recent years, graph signal processing has a good performance in non-regular domain data processing [26-28], and is also widely used in image processing [29, 30]. Graph signal processing utilizes graph models to represent the connectivity and local structure of signals, and the intrinsic structural features of image data can be effectively obtained by using graph representations. For image sequences, the intensity of the pixel points in each image frame is defined as the signal value of the graph node, and the similarity between neighboring pixel points is characterized by the weights of the edges, which can sufficiently inscribe the intrinsic correlation between the pixel points in the image.

In order to improve the accuracy of the motion target detection results in dynamic scenes, this paper utilizes the motion change variability of the dynamic background and the target foreground in the image sequence under the spatio-temporal domain, and smooths and suppresses the dynamic background region in each frame of the image sequence, and the result can be used as the low-rank constrained component information of the 3D tensor data. The image sequence is modeled as a graph signal, and a motion foreground target detection method based on low-rank sparse decomposition and graph Laplacian regularization is proposed and solved for this optimization problem. To validate the effectiveness of the proposed method is compared with existing commonly used target detection methods. The experimental results show that the proposed algorithm is more effective for the problem of motion target segmentation in dynamic background. The major contributions of this work are summarized as follows.

A moving object segmentation model based on low-rank sparse decomposition and graph Laplacian regularization is constructed, and a graph Laplacian regularization constraint is added to the objective function to enhance the connectivity between the image pixels in each frame of the image sequence.

According to the similarity of dynamic background in each frame of image sequence, a method is proposed to transform dynamic background into low-rank constrained components to improve

the accuracy of low-rank decomposition results.

The use of tensor decomposition fully utilizes the complete structural information of each image frame in order to improve the effectiveness of the target segmentation results.

The remaining part of this paper is organized as follows. In Section 2, graph signal processing modeling, and the suppression of dynamic background algorithm and tensor decomposition algorithm are described in detail. The experimental method, procedure and results of this paper are presented in Section 3. The experiments were conducted on the public dataset ChangeDetection.net [31], and the proposed algorithm was compared with other commonly used moving target segmentation algorithms. The related details of the proposed algorithm are discussed in Section 4. The conclusion of this paper is given in Section 5.

2. Materials and methods

2.1. Graph representation of image sequences

The graph $G = (V, E)$ consists of a set of nodes (vertices) V and edges E, where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes and $E \subseteq V \times V$ is the set of edges between nodes. Define the Laplacian matrix corresponding to the graph G to be $L = D - A$, where A is an $n \times n$ adjacency matrix (n is the number of nodes in the graph). D denotes the degree matrix (D is a diagonal matrix with the i-th diagonal element D_{ii} denoting the degree of node v_i). The graph Laplace regular term [32] corresponding to the graph signal $s_G \in \mathbf{R}^N$ can be expressed as: $s_G{}^T L s_G = \sum_{(i,j) \in E} b(i,j) (s_G(i) - s_G(j))^2$.

The pixel points of each image frame in an image sequence are modeled as graph nodes, the signal value of a graph node corresponds to the gray value of a pixel point, each node is connected to its 4-neighborhood nodes, and the correlation between nodes can be characterized using the weights of edges. The weight between two nodes can be computed using a Gaussian kernel, i.e., $\omega(i,j) = \exp\left(-\frac{|s_G(i) - s_G(j)|^2}{\sigma^2}\right)$, where σ denotes the similarity kernel parameter, The adjacency matrix A can be obtained from $\omega(i,j)$ ($A = [\omega(i,j)]_{i,j \in V}$). It can be seen that the smaller the intensity difference between pixels the stronger the correlation, the smaller the value of σ, the more the graph focuses on local similarity, meaning that signal propagation in the graph is mainly concentrated between neighboring nodes, and the graph will exhibit more local structural features. Representing each vertex k as a signal component $a_G(k)$, the graph signal corresponding to each image frame can be represented as $a_G = [a_G(1), a_G(2), \ldots, a_G(k^2)]^T \in \mathbf{R}^{k^2 \times 1}$.

2.2. Tensor modeling

The dataset consisting of n frames of image sequence images can be represented in the following third order tensor form: $I = \{I^1, I^2, I^3, \cdots, I^n\}$, where the ith frame image data matrix can be expressed as $I^i \in \mathfrak{R}^{h \times w}$ ($i = 1, 2, 3, \cdots, n$), h and w represent the number of rows and columns of an image data matrix, respectively. The three dimensions of the third-order tensor are h, w and time. For an image sequence, I can be regarded as consisting of background $I_1 = \{I_1^1, I_1^2, I_1^3, \cdots, I_1^n\}$ and moving foreground $I_2 = \{I_2^1, I_2^2, I_2^3, \cdots, I_2^n\}$, i.e., $I = I_1 + I_2$. Along the time dimension, if each frame contains the same background, the background of the whole image sequence can be approximated as a subspace with low-rank characteristics. Meanwhile, the target foreground in each frame can be approximated as a subspace with sparse characteristics. Based on this, the tensor RPCA method is used to fully utilize the structural information between each frame in the image sequence to separate the moving target and background information.

The following combines the canoes dataset to analyze the change characteristics of the dynamic background in the spatio-temporal domain, as shown in Fig. 1(a) below, along the direction of the time dimension each frame of the image contains a static background (e.g., the wooded part of the forest), a dynamic background (e.g., the fluctuating lake surface), and a moving

target (e.g., the rowing boat). Along the time dimension, from the 1st frame to the nth frame, the grayscale change characteristics at a certain spatial location on each frame are shown in Fig. 1(b), i.e., the grayscale of the static region such as the spatial location point (x_1, y_1) of each frame remains basically unchanged, and the grayscale of the dynamic region such as the spatial location point (x_3, y_3) fluctuates repetitively within a small range, whereas the target of the motion there is a significant difference in the gray level of the passing spatial location such as (x_2, y_2). It can be seen that there is a large interference of dynamic background on target extraction when performing low-rank sparse decomposition of image sequences, and it is easy to misjudge the dynamic background as a moving target, thus suppressing the dynamic background in each image frame is a key link to improve the accuracy of moving target segmentation.
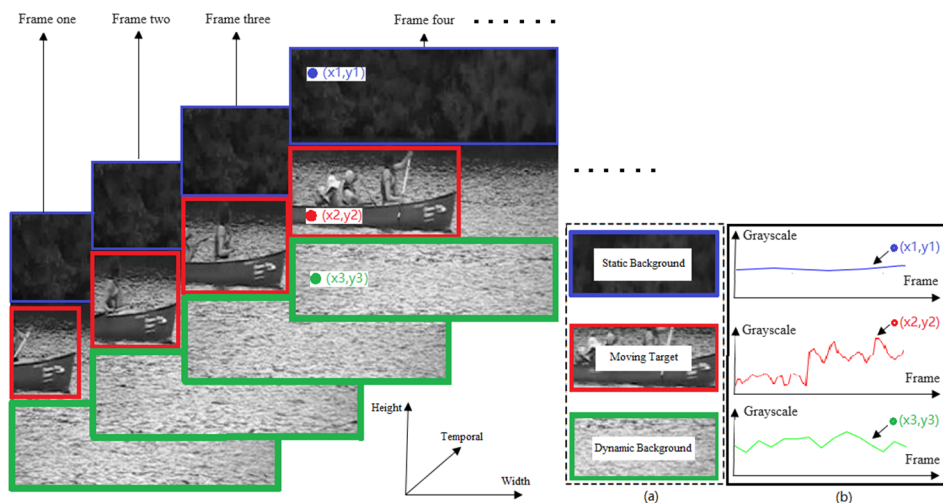

Fig. 1. Schematic diagram of the third-order tensor structure of an image sequence:
a) static, dynamic regions and moving targets in an image frame; b) schematic of gray
scale variation at spatial locations (x_1, y_1), (x_2, y_2) and (x_3, y_3)

2.3. Dynamic background suppression method

According to the above analysis, it can be seen that the dynamic background and the moving target in the image sequence along the direction of the time dimension in three-dimensional space have different change characteristics, i.e., when the moving target passes through the region of the dynamic background, there is a significant difference between the grayscale change of the target and the grayscale change of the dynamic background, and at the same time, the moving target has a continuity in the spatial-temporal domain. According to the above characteristics, the dynamic background region in each frame image is suppressed to improve the efficiency of moving object extraction.

The key problem in suppressing the dynamic background is how to distinguish the foreground and dynamic background regions in a frame, for which the solution is computed in terms of pixels based on the significant difference between the moving target and the dynamic background in terms of gray scale changes. As shown in Fig. 2, a frame of an image with any coordinate point at a spatial location such as (x_p, y_q), where $1 \leq x_p \leq w$, $1 \leq y_q \leq h$. Taking the grayscale $g_{p,q,k}(x_p, y_q)$ of each image frame along the time dimension direction at the spatial location point (x_p, y_q) yields a one-dimensional array of length n, denoted as Vec. Where k denotes the kth frame of the image sequence, $k = 1, 2, \ldots, n$ $(n = N)$. Iterating through all the spatial locations on an image frame yields $w \times h$ one-dimensional arrays, and the dynamic background suppression operation is performed on each one-dimensional array to obtain the image sequence after dynamic background suppression.
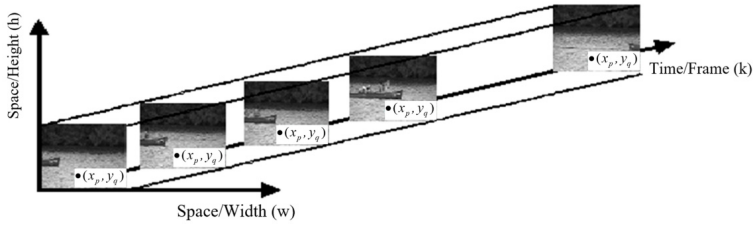
Fig. 2. Schematic of data processing in pixels

The steps to achieve suppression of the dynamic background in pixels are as follows:

Step 1. Set the static background region determination threshold th_{min}, and the comparison and determination threshold of dynamic background and moving target are set as th_{mean} and th_D respectively, initialize $p = 1$, $q = 1$, and we take the empirical value $th_{min} \in [5,15]$, $th\_mean \in [0.4, 0.6]$, $th_D \in [0.2, 0.3]$.

Step 2. Calculate the difference between the maximum grayscale value L_{max} and the minimum grayscale value S_{min} in the array Vec and denote it as d. In order to deal with the random noise interference, the averaging method is used for finding L_{max} and S_{min}.

Step 3. Compare d and th_{min} to quickly determine the static background region in the image, i.e., if $d < th_{min}$ then the spatial location (x_p, y_q) is determined to be the static background, and its gray value $g_{p,q,k}(x_p, y_q)$ remains unchanged, otherwise perform step 4 .

Step 4. Taking $k = 1,2,\ldots,n$, and $temp = |g_{p,q,k}(x_p, y_q) - th_{mean}|$ is calculated sequentially, $temp$ and th_D are compared to detect the dynamic background region and the moving target region, i.e., if $temp > th_D$ then determine that the spatial position (x_p, y_q) in the kth frame image is the dynamic background and take $g_{p,q,k}(x_p, y_q) = th_{mean}$, otherwise determine that it is a moving target and keep the gray scale value $g_{p,q,k}(x_p, y_q)$ unchanged. It should be noted that, we take $g_{p,q,k}(x_p, y_q) = th_{mean}$, on the one hand, to realize the suppression operation of the dynamic background, on the other hand, it takes the same value at (x_p, y_q) for each image frame, which realizes the transformation of the dynamic background region into a low-rank constrained component for the whole image sequence.

Step 5. Traversing the region $1 \leq x_p \leq w$, $1 \leq y_q \leq h$, and steps 2-4 respectively are performed respectively to obtain 3D image sequence data for suppressing the dynamic background.

As shown below, Fig. 3(a) is a frame from the original image sequence, and Fig. 3(b) shows a frame processed using the dynamic background suppression algorithm. Comparison of Figs. 3(a) and 3(b) shows that the dynamic background has been suppressed to a large extent. The presence of dynamic interference information in the red labeled region in Fig. 3(b) is mainly due to the bias arising from the selection of the empirical threshold th_D, which will be further dealt with when performing the tensor sparse component decomposition in Section 2.4.

Next, we will further analyze the value ranges for th_{min}, th_{mean}, and th_D. th_{min} is used to determine the static background area in the image. The grayscale variation values of the static background area in sequence images from different datasets can be used to obtain it. Based on the experimental results from the data set in this paper, $th_{min} \in [5,15]$ is chosen. th_{mean} is determined based on the inter-frame target grayscale variations calculated from the ground truth. For the experimental dataset in this paper, $th_{mean} \in [0.4, 0.6]$ is selected. Based on the determination of th_{min} and th_{mean}, the value range of th_D can be further obtained, to quantitatively analyze th_D, we use SSIM (Structural Similarity Index) to assess the reasonableness of its value. The statistical results are shown in Fig. 4.

The value of SSIM typically ranges from 0 to 1, with a value closer to 1 indicating higher visual similarity between the two images. The curves shown in Fig. 4 are the SSIM values calculated with the 20th frame image from six datasets as a reference. The fall, fountain01, and

fountain02 datasets show a turning point around $th_D = 0.2$, while the boats, canoe, and overpass datasets show a turning point around $th_D = 0.3$.

Fig. 3. Schematic diagram of the results of suppressing dynamic background:
a) original image, b) result of suppressing dynamic background
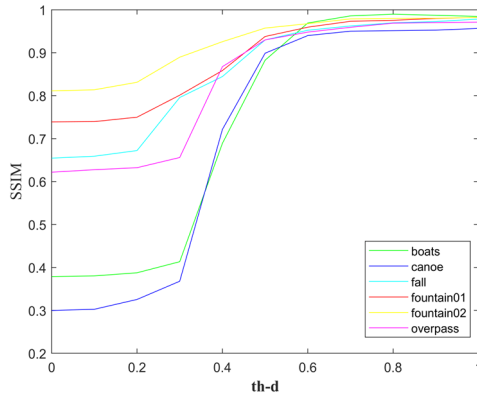
Fig. 4. Analysis of parameter th_D

2.4. Graph regularization constraints and tensor sparse component decomposition

Separating the image background from the motion foreground from the image sequence using the tensor RPCA algorithm can be described as solving the following optimization problem:

$$\min_{l,l_1,l_2} \quad \lambda\psi_2(l_2) + \psi_1(l_1),$$
$$s.t. \quad l = l_2 + l_1, \tag{1}$$

where λ denotes the weight parameter, $\psi_2(l_2)$ denotes motion foreground, $\psi_1(l_1)$ denotes background information, l, l_2 and l_1 denote the vectorization of tensors I, I_2 and I_1, respectively.

As shown in Fig. 3(b) above, although the dynamic background is largely suppressed, some perturbation information still exists. Whereas for an image sequence, the background information can be approximated as a low-rank subspace, the moving foreground can be approximated as a sparse subspace. In this regard, a low-rank sparse decomposition can be performed, and we further decompose the background information $\psi_1(l_1)$ in Eq. (1) into the sum of the low-rank component ℓ and the perturbation component e. Eq. (1) can be rewritten as:

$$\min_{l,e,\ell,l_2} \quad \lambda\psi_2(l_2) + \kappa\phi(e) + \Omega(\ell),$$
$$s.t. \quad l = l_2 + e + \ell, \tag{2}$$

where κ denotes the weight parameter. To solve Eq. (2), the a priori features of the moving foreground and the image background are analyzed separately. The foreground is sparse and the graph Laplacian regularization term is used to characterize the spatial graph structure of the image frame, then Eq. (2) can be rewritten as:

$$\min_{l,e,\ell,s} \quad \lambda\|s\|_1 + \Upsilon tr(s^T L s) + \kappa\phi(e) + \Omega(\ell),$$
$$s.t. \quad l = s + e + \ell, \tag{3}$$

where $\|s\|_1$ is the L_1 norm, representing the sparsity of the tensor, $tr(s^T L s)$ is the graph Laplacian regularization term, which describes the smoothness of the sparse components on the graph, L is the graph Laplacian matrix, and Υ is the weight parameter.

As for the image background, along the time dimension, the background of an image sequence can be approximated as a low-rank component plus perturbation information, and the tucker decomposition can be used to solve for the low-rank background component in 3D space:

$$I_1 = \varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3 + \varepsilon, \tag{4}$$

where U_1 and U_2 are orthogonal in the spatial dimension and orthogonal to the time dimension U_3. ς denotes the kernel tensor, ε denotes the perturbation information. Further, the optimization problem shown in Eq. (3) can be written as follows:

$$\min_{l,s,e,\varsigma,U_1,U_2,U_3} \quad \lambda\|s\|_1 + \Upsilon tr(s^T L s) + \frac{1}{2}\|e\|^2,$$
$$s.t. \quad l = s + e + Vec(\varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3), \tag{5}$$

where Vec denotes the vectorization of the tensor. Using the augmented Lagrange function to solve the above optimization problem, the Lagrange function of Eq. (5) can be expressed as:

$$\begin{aligned}
L\mu(l,s,\varsigma,U1,U2,U3,e) &= \lambda\|s\|_1 + \Upsilon tr(s^T L s) + \frac{1}{2}\|e\|^2 \\
&\quad -\langle \lambda^1, l - s - e - Vec(\varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3)\rangle \\
&\quad +\frac{\beta^1}{2}\|l - s - e - Vec(\varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3)\|^2,
\end{aligned} \tag{6}$$

where λ^1 denote Lagrange multiplier vector, β^1 denote the penalization factors. The following uses alternating direction method of multipliers (ADMM) [33-35] to solve Eq. (6).

Step 1. Updating l, keeping the other terms constant:

$$l = \frac{\beta^1(l_2 + e + Vec(\varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3))}{\beta^1}. \tag{7}$$

Step 2. Updating $[\varsigma, U1, U2, U3]$, since $l_1 = l - s - \varepsilon - T(\lambda^1/\beta^1)$, where T denotes the tensor transformation. The Tucker decomposition is applied to l_1 in three dimensions, and then the low-rank component l_1 is obtained by an iterative operation.

Step 3. Updating e, keeping the other terms constant:

$$e = \frac{\beta^1\left(l - s - Vec(\varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3) - \frac{\lambda^1}{\beta^1}\right)}{1 + \beta^1}. \tag{8}$$

Step 4. Updating the sparse part can be done in two sub-steps.
i) Updating graph Laplacian regularization term:

$$s^{k+1} = \underset{s}{argmin} \frac{\lambda}{2} \left\| s - s_{temp}^{k+1} \right\|_1^2 + \Upsilon tr(s^T L s). \tag{9}$$

This process can be realized by a linear transformation, i.e., $s^{k+1} = (\lambda I + 2\Upsilon L)^{-1}(\lambda s_{temp}^{k+1})$.

(ii) Updating Sparse component:

$$s = \frac{\beta^1(l - e - Vec(\varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3)) - \lambda^1}{\beta^1 I}. \tag{10}$$

Step 5. Updating h, keeping the other terms constant. The optimization of h can be equated to:

$$\underset{h}{min} \quad \lambda \|h\|_1 + \frac{\beta^1}{2} \left\| h - \left(Dl_2 + \frac{\lambda^1}{\beta^1} \right) \right\|^2. \tag{11}$$

This optimization problem can be solved by using soft shrinkage operators.

Step 6. The parameter λ^1 is updated using the following equation:

$$\lambda^1 = \lambda^1 - \upsilon \beta^1(l - Vec(\varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3) - e - s), \tag{12}$$

where υ denotes the adjustment parameter. In this paper, we take $\upsilon = 1.2$, the penalty factor is taken as $\beta^1 = 10^{-5}/\eta$, and η takes the mean of l.

2.5. Complexity analysis

Assume there are N frames in the sequence, with each frame having M pixels, resulting in a total of $N \times M$ nodes. The time complexity for constructing the adjacency matrix is typically $O((N \times M)^2)$. The time complexity for computing the Laplacian matrix L is linearly related to the size of the graph, i.e., $O(N \times M)$. For tensor decomposition, it typically requires $O(r^3)$ matrix multiplications, so the computational complexity is $O(r^3)$, where r represents the rank of the tensor decomposition. The computational complexity for each update in ADMM is $O(r^2 d)$, where d is the total dimension of the tensor. If the maximum number of iterations for ADMM is T, the overall time complexity is $O((N \times M)^2 + T \times r^2 \times d)$.

3. Results and analysis

The experimental image sequence data in this paper are selected from six sets of real scenes containing dynamic backgrounds in the publicly available dataset ChangeDetection.net, including: boats, canoe, fall, fountain01, fountain02 and overpass.

3.1. Ablation experiment

3.1.1. Dynamic background suppression

The selection of threshold th_D will have a certain impact on the dynamic background region segmentation and its suppression results in pixels. If th_D is too small, it may misclassify the target as dynamic background; if th_D is too large, it may misclassify the dynamic background as the target, thus failing to effectively suppress the dynamic background. In this paper, we take the empirical values $th_{min} = 10$, $th_{mean} = 0.5$, $th_D = 0.25$. In order to objectively analyze the effectiveness of the proposed algorithm in suppressing the dynamic background, we conducted ablation experiments. The comparison results of the proposed algorithm with the unsuppressed image sequence dynamic background method for segmenting the target are shown in Fig. 5.

The results of the ablation experiments of three sets of sequence images, canoe, fountain01

and fountain02, respectively, are shown in Fig. 5, which shows that the proposed algorithm removes the dynamic interference information well.
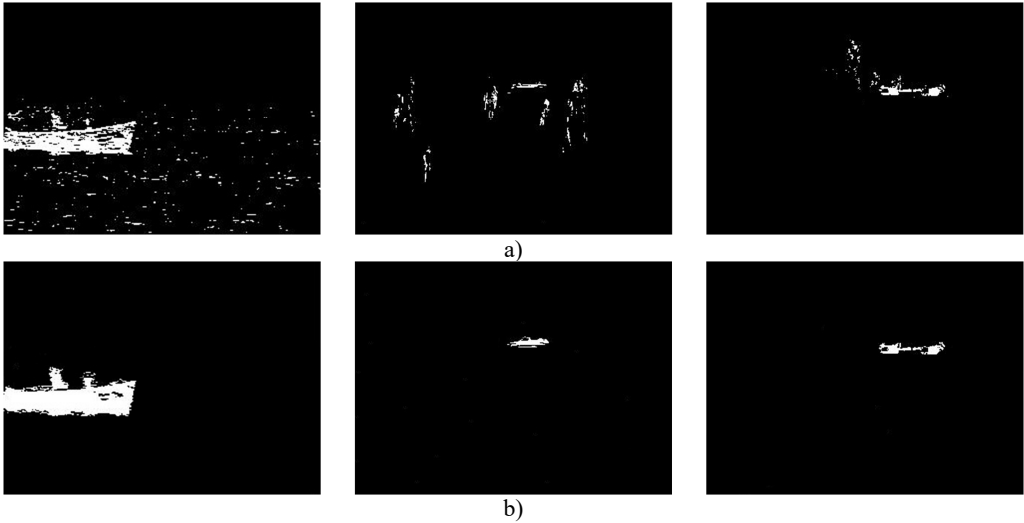
Fig. 5. Dynamic background suppression comparison results:
a) dynamic background not removed, b) dynamic background removal

3.1.2. Graph Laplacian regularization constraints

In order to verify the effectiveness of segmenting the target using graph regularization constraints, the graph Laplacian regularization method is compared with the three-dimensional total variation (3D-TV) method under the suppression of dynamic background. For the moving foreground, as shown in Fig. 2 above, the motion target in two adjacent frames of an image sequence in a 3D spatial structure is continuous in its spatial position and undergoes only small displacements. Therefore, 3D-TV can be used to model and analyze the moving target, combined with the three-dimensional space shown in Fig. 2 above, let i denote the width direction, j denote the height direction, and t denote the time dimension direction, then the 3D-TV model can be described by the following equation:

$$TV_{i,j,t}(l_2) = |l_2(i,j,t) - l_2(i+1,j,t)| + |l_2(i,j,t) - l_2(i,j+1,t)|$$
$$+|l_2(i,j,t) - l_2(i,j,t+1)|. \tag{13}$$

The gray level at spatial location (i,j,t) is denoted by $g(i,j,t)$, let $g_w(i,j,t) = g(i+1,j,t) - g(i,j,t)$, $g_h(i,j,t) = g(i,j+1,t) - g(i,j,t)$, $g_k(i,j,t) = g(i,j,t+1) - g(i,j,t)$, and the above tensor is vectorized, and set $Dl = [g_w, g_h, g_k]$, where g_w, g_h, g_k denote the result of vectorization of $g_w(i,j,t)$, $g_h(i,j,t)$, $g_k(i,j,t)$ respectively. Rewriting Eq. (13) as L_1 norm:

$$||l_2||_{3D-TV} = ||Dl_2||_1. \tag{14}$$

Combining the 3D-TV model and the previous method of suppressing dynamic background processing, i.e., plugging Eqs. (4) and (14) into Eq. (13), the optimization problem corresponding to Eq. (2) can be rewritten as:

$$\min_{l,l_2,e,h,\varsigma,U_1,U_2,U_3} \lambda\|h\|_1 + \frac{1}{2}\|e\|^2,$$

$$s.t. \quad h = Dl_2,$$

$$l = l_2 + e + Vec(\varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3), \tag{15}$$

where Vec denotes the vectorization of the tensor. Using the augmented Lagrange function to solve the above optimization problem, the Lagrange function of Eq. (15) can be expressed as:

$$L\mu(l, l_2, \varsigma, U1, U2, U3, e, h) = \lambda\|h\|_1 + \frac{1}{2}\|e\|^2 - <\lambda^1, h - Dl_2> + \frac{\beta^1}{2}\|h - Dl_2\|^2$$
$$-\langle\lambda^2, l - l_2 - e - Vec(\varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3)\rangle \tag{16}$$
$$+\frac{\beta^2}{2}\|l - l_2 - e - Vec(\varsigma \times_1 U_1 \times_2 U_2 \times_3 U_3)\|^2.$$

where λ^1, λ^2 denote Lagrange multiplier vector, β^1, β^2 denote the penalization factors. Eq. (16) can also be optimized by ADMM algorithm, i.e., update the l, $[\varsigma, U1, U2, U3]$, e, l_2, h, λ^1 and λ^2 parameters separately.

The results of the TV algorithm and the graph regularization algorithm for segmenting the foreground target respectively are shown in Fig. 6.

a)

b)
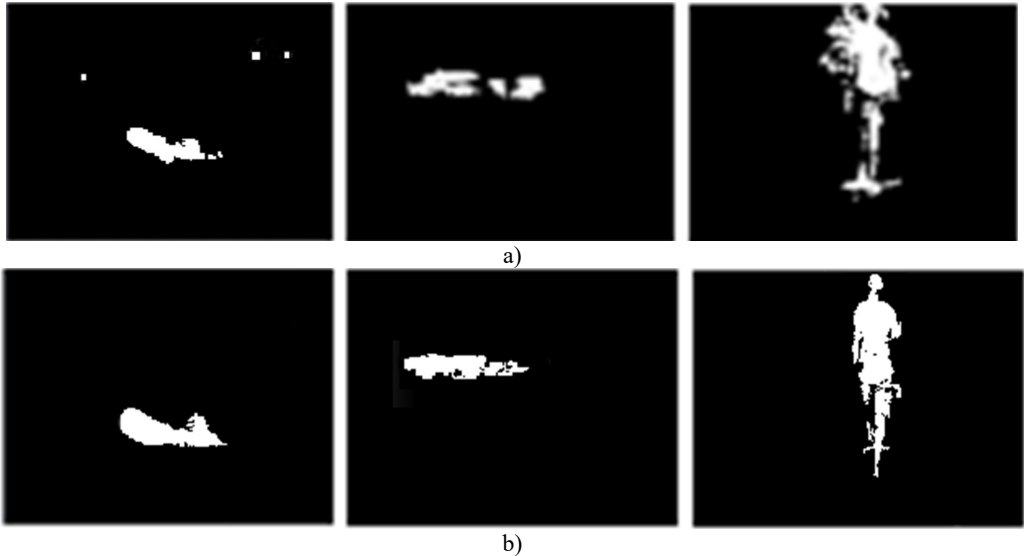
Fig. 6. Foreground target segmentation results: a) TV algorithm, b) graph regularization algorithm

Comparing the segmentation results shown in Fig. 6(a) and (b), it can be seen that the foreground targets segmented by the graph regularization algorithm have better continuity and integrity in terms of subjective visual effects. It will be further discussed and analyzed quantitatively in Section 3.2.

Further, as can be seen from Eq. (6), Υ is the regularization parameter used to penalize the connectivity of target foreground pixels. Taking the 30th frame image from the boats dataset as an example to analyze the regularization parameter, the results of the sensitivity analysis of r are shown in Fig. 7. It can be seen that the optimal segmentation result is obtained when $\Upsilon = 0.6$.

3.2. Moving target extraction results and analysis

On the basis of suppressing the dynamic background, the tensor decomposition and graph

Laplacian regularization algorithm is used to extract the moving targets, and the proposed algorithm is compared with five other typical algorithms for moving target detection, including the RPCA algorithm, the RPCA optimization algorithm, the TVRPCA algorithm, the IBTSVT algorithm, and the IMTSL algorithm. The experimental results are shown in Fig. 8.
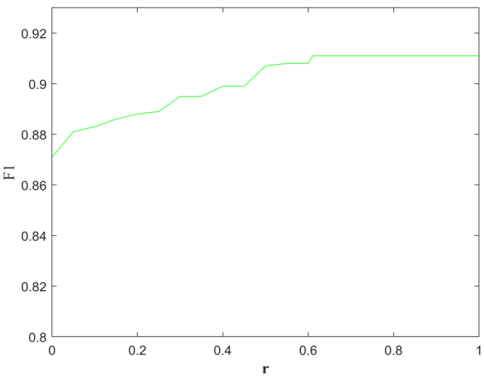

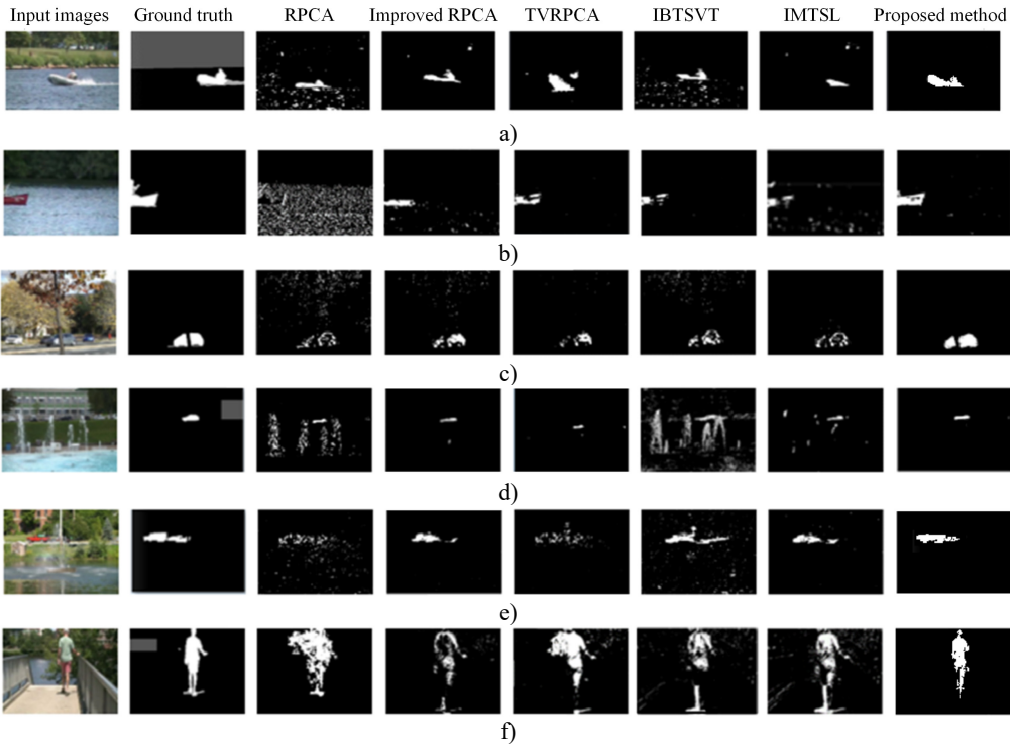**Fig. 7.** Sensitivity analysis of the regularization parameter

Fig. 8. Results of different algorithms for extracting moving targets:
a) boats, b) canoe, c) fall, d) fountain01, e) fountain02, f) overpass

According to the results shown in Fig. 6, in terms of subjective visual effects, the moving object detected by the proposed algorithm is the clearest and most complete. Both the RPCA and RPCA optimization algorithms use the method of matrix decomposition, which does not take advantage of the structural information between each frame. In addition, RPCA optimization algorithm performs dynamic background suppression on the basis of matrix decomposition, which improves the segmentation accuracy to a certain extent. However, the final results obtained by this

method largely depend on the effect of using RPCA algorithm to perform low-rank sparse decomposition of image sequences. However, RPCA decomposition under dynamic background is difficult to obtain satisfactory results, which directly leads to the low accuracy of the final extraction target of the algorithm. TVRPCA, IBTSVT, and IMTSL algorithms are based on the tensor model for the extraction of the motion target, but when every frame of the image there is dynamic background information, especially when the dynamic background region occupies a large proportion in one frame, it is difficult to solve the problem of misclassification between the dynamic background and the motion foreground.

Further, in order to provide an objective evaluation of the segmentation results, a comprehensive indicator F metric is used to measure the segmentation results, which is calculated by the following formula:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F = \frac{2P \cdot R}{P + R}. \tag{17}$$

where P is precision, R is recall, TP, FP and FN indicate true positives, false positives and false negatives respectively. F is an important comprehensive index to evaluate the results of target segmentation. Its value is between 0 and 1, and the closer the value is to 1, the better the performance of the object segmentation model. The results of statistically different algorithms for calculating F-values are shown in Table 1.

Table 1. F-values calculated by different algorithms

Data set	Algorithm					
	RPCA algorithm	RPCA optimization algorithm	TVRPCA algorithm	IBTSVT algorithm	IMTSL algorithm	Proposed method
Boats	0.23	0.56	0.69	0.47	0.52	0.91
Canoe	0.12	0.20	0.42	0.41	0.31	0.86
Fall	0.21	0.85	0.28	0.33	0.58	0.92
Fountain01	0.02	0.67	0.62	0.39	0.36	0.79
Fountain02	0.22	0.71	0.38	0.46	0.68	0.88
Overpass	0.37	0.53	0.48	0.45	0.71	0.85

According to the results of statistical F in Table 1, the proposed algorithm has obtained optimal segmentation results on the six data sets of boats, canoe, fall, fountain01, fountain02 and overpass. Further, the two datasets of boats and canoe have most of the dynamic background region in each frame, and the target foreground is always moving in the dynamic background region (the boat rowing on the fluctuating lake), the results obtained by the proposed algorithm on these two datasets are much better than the other five algorithms (24 % and 51 % higher than the sub-optimal algorithms, respectively), which also shows the effectiveness of the proposed algorithm for suppressing the dynamic background algorithm is effective. In addition, the algorithm also obtains good segmentation results on the fall, fountain01 and fountain02 datasets, but the F-score on the fountain01 data set is only 0.79, which is mainly due to the fact that the target is small and there is a large proportion of occlusion in the image sequence, making segmentation of the foreground target less effective. Overall, the proposed algorithm is more advantageous for the problem of detecting moving targets under dynamic background in image sequences.

4. Discussion

In this paper, we investigate how to effectively segment the target foreground from an image sequence disturbed by a dynamic background using tensor decomposition and graph signal processing methods. The dynamic background is suppressed by smoothing, so that each frame in the image sequence can be described as a superposition of two components, i.e., the background

component with smooth change characteristics and the foreground component with sparsity and connectivity, and a graph model is established to optimize the solution.

Graph Laplacian regularization is based on graph structure, which can better express the local neighborhood structure of image data. For each frame in an image sequence, the graph Laplacian constraint term is able to describe the similarity between neighboring pixels, and by introducing similarity constraints on neighboring pixels, the graph Laplacian regularization effectively reduces isolated noise points in the foreground region. In fact this local structure is particularly important because neighboring pixels in an image tend to belong to the same region thus effectively smoothing the segmentation of the target foreground. Especially when dealing with image sequences of dynamic scenes, the graph Laplacian regularization can suppress the interference of noise and make the foreground more coherent, thus showing the contours of the moving target more clearly.

The ablation experiments in Section 3.1 of the paper and the comparison experiments of typical motion target segmentation algorithms in Section 3.2 verify the effectiveness of this paper's method in suppressing dynamic background interference noise and completely segmenting foreground targets. The proposed method provides a new way to accurately segment the target foreground in image sequences in complex natural scenes.

It should be noted that the method in this paper processes the image sequence as a whole. For video signals, the sequence must first be frame-by-frame segmented, and the segmentation results are output frame by frame, which introduces a certain amount of delay. The overpass dataset in this paper contains the largest number of images, a total of 473 frames, and the average processing time for one image is 1.03 s (The experiment is conducted on a PC, the CPU is Intel core i5 2.27 GHz, 2G memory, and win10 operating system). Additionally, each frame in the six datasets used in this paper contains only a single target. For multi-object segmentation, especially when the targets overlap (occlusion issue), further discussion is needed. Incorporating domain knowledge, such as using geometric or semantic priors of the objects, could be an effective approach. Typically, this requires a combination of various techniques, which also increases computational complexity.

5. Conclusions

When each frame of an image sequence contains dynamic background information, it is challenging to realize the segment of moving targets by using low-rank sparse decomposition, especially when there is a large region of dynamic background in a frame which limits the low-rank constraint ability and makes it easy to mis detect or miss detection. To address this problem, this paper firstly adopts the dynamic background suppression algorithm based on the difference between the target foreground and the dynamic background motion characteristics in the spatial-temporal domain, which realizes the conversion of the dynamic background into a low-rank constraint component and lays the foundation for the effective realization of the low-rank sparse decomposition. The tensor decomposition and graph Laplace regularization algorithms are then used to take full advantage of the spatio-temporal continuity between two adjacent frames as well as the similarity of neighboring pixels in each frame image, and the alternating direction method of multipliers iterative optimization method is used to complete the effective separation of the background and the moving foreground. The experimental results prove that the proposed algorithm is feasible and has more advantages than other algorithms. In the future, the use of deep learning methods to address target segmentation under small sample conditions requires further research. On one hand, data augmentation should be combined with generative adversarial networks (GAN), and on the other hand, the spatiotemporal consistency between inter-frame data should be leveraged to design 3D convolutional neural network models to extract more comprehensive target features, effectively addressing the issue of target occlusion.

Acknowledgements

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Shudan Yuan: conceptualization, formal analysis, funding acquisition, investigation, validation, writing-original draft preparation, visualization, supervision. Hui Wang: conceptualization, data curation, funding acquisition, methodology, resources, writing-review and editing software, validation, visualization, supervision.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Z. Wu, Y. Tang, B. Hong, B. Liang, and Y. Liu, "Enhanced precision in dam crack width measurement: leveraging advanced lightweight network identification for pixel-level accuracy," *International Journal of Intelligent Systems*, Vol. 2023, No. 1, Sep. 2023, https://doi.org/10.1155/2023/9940881

[2] K. Hu, Z. Chen, H. Kang, and Y. Tang, "3D vision technologies for a self-developed structural external crack damage recognition robot," *Automation in Construction*, Vol. 159, p. 105262, Mar. 2024, https://doi.org/10.1016/j.autcon.2023.105262

[3] Y. Tang et al., "Novel visual crack width measurement based on backbone double-scale features for improved detection automation," *Engineering Structures*, Vol. 274, p. 115158, Jan. 2023, https://doi.org/10.1016/j.engstruct.2022.115158

[4] Y. Z. Yang, P. F. Wang, X. L. Hu, N. Kuang, and Z. And Yang, "Moving object detection optimization algorithm based on robust principal component analysis," (in Chinese), *Journal of Electronics and Information technology*, Vol. 40, pp. 1309–1315, 2018, https://doi.org/10.19650/j.cnki.cjsi.j1904640

[5] J. Zhou, Z. Pang, and Y.-X. Wang, "RMem: restricted memory banks improve video object segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18602–18611, Jun. 2024, https://doi.org/10.1109/cvpr52733.2024.01760

[6] Z. Shen, Y. He, X. Du, J. Yu, H. Wang, and Y. Wang, "YCANet: target detection for complex traffic scenes based on camera-LiDAR fusion," *IEEE Sensors Journal*, Vol. 24, No. 6, pp. 8379–8389, Mar. 2024, https://doi.org/10.1109/jsen.2024.3357826

[7] Q. Tan, Z. Du, and S. Chen, "Moving target detection based on background modeling and frame difference," *Procedia Computer Science*, Vol. 221, pp. 585–592, Jan. 2023, https://doi.org/10.1016/j.procs.2023.08.026

[8] M. Alipour Sormoli, M. Dianati, S. Mozaffari, and R. Woodman, "Optical flow based detection and tracking of moving objects for autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 25, No. 9, pp. 12578–12590, Sep. 2024, https://doi.org/10.1109/tits.2024.3382495

[9] X. Lu, C. Xu, L. Wang, and L. Teng, "Improved background subtraction method for detecting moving objects based on GMM," *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 13, No. 11, pp. 1540–1550, May 2018, https://doi.org/10.1002/tee.22718

[10] M. Raimondi, G. Ciattaglia, A. Nocera, L. Senigagliesi, S. Spinsante, and E. Gambi, "mmDetect: YOLO-based processing of mm-wave radar data for detecting moving people," *IEEE Sensors Journal*, Vol. 24, No. 7, pp. 11906–11916, Apr. 2024, https://doi.org/10.1109/jsen.2024.3366588

[11] S. Yan, F. Zhang, Y. Fu, W. Zhang, W. Yang, and R. Yu, "A deep learning-based moving target detection method by combining spatiotemporal information for ViSAR," *IEEE Geoscience and Remote Sensing Letters*, Vol. 20, pp. 1–5, Jan. 2023, https://doi.org/10.1109/lgrs.2023.3326205

[12] H. Kim, "Robust foreground extraction technique using background subtraction with multiple thresholds," *Optical Engineering*, Vol. 46, No. 9, p. 097004, Sep. 2007, https://doi.org/10.1117/1.2779022

[13] W. Zhou et al., "TSDTVOS: Target-guided spatiotemporal dual-stream transformers for video object segmentation," *Neurocomputing*, Vol. 555, p. 126582, Oct. 2023, https://doi.org/10.1016/j.neucom.2023.126582

[14] C. Liang, W. Wang, T. Zhou, J. Miao, Y. Luo, and Y. Yang, "Local-global context aware transformer for language-guided video segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 8, pp. 10055–10069, Aug. 2023, https://doi.org/10.1109/tpami.2023.3262578

[15] M. Waqas, S. Kidera, and T. Kirimoto, "PCA-based detection algorithm of moving target buried in clutter in doppler frequency domain," *IEICE Transactions on Communications*, Vol. E94-B, No. 11, pp. 3190–3194, Jan. 2011, https://doi.org/10.1587/transcom.e94.b.3190

[16] 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, Jun. 2011, https://doi.org/10.1145/1970392.1970395

[17] T. Bouwmans and E. H. Zahzah, "Robust PCA via principal component pursuit: a review for a comparative evaluation in video surveillance," *Computer Vision and Image Understanding*, Vol. 122, pp. 22–34, May 2014, https://doi.org/10.1016/j.cviu.2013.11.009

[18] T. Zhou and D. Tao, "GoDec: Randomized low-rank and sparse matrix decomposition in noisy case," in *International Conference on Machine Learning*, pp. 33–40, 2011, https://doi.org/http://dx.doi.org/

[19] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40, No. 7, pp. 1726–1740, Jul. 2018, https://doi.org/10.1109/tpami.2017.2732350

[20] X. Cao, L. Yang, and X. Guo, "Total variation regularized RPCA for irregularly moving object detection under dynamic background," *IEEE Transactions on Cybernetics*, Vol. 46, No. 4, pp. 1014–1027, Apr. 2016, https://doi.org/10.1109/tcyb.2015.2419737

[21] L. Chen, Y. Liu, and C. Zhu, "Iterative block tensor singular value thresholding for extraction of lowrank component of image data," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1862–1866, Mar. 2017, https://doi.org/10.1109/icassp.2017.7952479

[22] A. Sobral, C. G. Baker, T. Bouwmans, and E.-H. Zahzah, "Incremental and multi-feature tensor subspace learning applied for background modeling and subtraction," in *Image Analysis and Recognition*, pp. 94–103, Oct. 2014, https://doi.org/10.1007/978-3-319-11758-4_11

[23] W. Cao et al., "Total variation regularized tensor RPCA for background subtraction from compressive measurements," *IEEE Transactions on Image Processing*, Vol. 25, No. 9, pp. 4075–4090, Sep. 2016, https://doi.org/10.1109/tip.2016.2579262

[24] M. Shakeri and H. Zhang, "Moving object detection in time-lapse or motion trigger image sequences using low-rank and invariant sparse decomposition," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 5133–5141, Oct. 2017, https://doi.org/10.1109/iccv.2017.548

[25] L. Zhu, Y. Hao, and Y. Song, "L1/2 norm and spatial continuity regularized low-rank approximation for moving object detection in dynamic background," *IEEE Signal Processing Letters*, Vol. 25, No. 1, pp. 15–19, Jan. 2018, https://doi.org/10.1109/lsp.2017.2768582

[26] W. Hu, J. Pang, X. Liu, D. Tian, C.-W. Lin, and A. Vetro, "Graph signal processing for geometric data and beyond: theory and applications," *IEEE Transactions on Multimedia*, Vol. 24, pp. 3961–3977, Jan. 2022, https://doi.org/10.1109/tmm.2021.3111440

[27] H. Gao et al., "Graph convolutional network with connectivity uncertainty for EEG-based emotion recognition," *IEEE Journal of Biomedical and Health Informatics*, Vol. 28, No. 10, pp. 5917–5928, Oct. 2024, https://doi.org/10.1109/jbhi.2024.3416944

[28] W. Cai, Y. Li, X. Zhang, T. Jiang, and Q. Qi, "Impulsive noise suppression based on ℓq – norm and graph Laplacian regularization for NOMA system," *IEEE Wireless Communications Letters*, Vol. 13, No. 8, pp. 2050–2054, Aug. 2024, https://doi.org/10.1109/lwc.2024.3398121

[29] J. Pang and G. Cheung, "Graph Laplacian regularization for image denoising: analysis in the continuous domain," *IEEE Transactions on Image Processing*, Vol. 26, No. 4, pp. 1770–1785, Apr. 2017, https://doi.org/10.1109/tip.2017.2651400

[30] J. Zeng, Q. Zhu, T. Tian, W. Sun, L. Zhang, and S. Zhao, "Deep unrolled weighted graph Laplacian regularization for depth completion," *International Journal of Computer Vision*, Vol. 133, No. 1, pp. 190–210, Jul. 2024, https://doi.org/10.1007/s11263-024-02188-3

[31] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, pp. 1–8, Jun. 2012, https://doi.org/10.1109/cvprw.2012.6238919

[32] W. Qi, S. Guo, and W. Hu, "Generic reversible visible watermarking via regularized graph Fourier transform coding," *IEEE Transactions on Image Processing*, Vol. 31, pp. 691–705, Jan. 2022, https://doi.org/10.1109/tip.2021.3134466

[33] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, Vol. 3, No. 1, pp. 1–122, Jan. 2010, https://doi.org/10.1561/2200000016

[34] J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, and M. Puschel, "D-ADMM: a communication-efficient distributed algorithm for separable optimization," *IEEE Transactions on Signal Processing*, Vol. 61, No. 10, pp. 2718–2723, May 2013, https://doi.org/10.1109/tsp.2013.2254478

[35] Y. Chen, Y. Wang, M. Li, and G. He, "Augmented Lagrangian alternating direction method for low-rank minimization via non-convex approximation," *Signal, Image and Video Processing*, Vol. 11, No. 7, pp. 1271–1278, Apr. 2017, https://doi.org/10.1007/s11760-017-1084-9

Shudan Yuan received M.S. degrees in the School of Computer Science and Engineering from Guilin university of technology in 2015. She is currently a Lecturer in the Department of Artificial Intelligence, Hezhou University, Hezhou, China. Her research interests include computer application technology and artificial intelligence.

Hui Wang received Ph.D. degree in information and communication engineering from the University of Guilin University of Electronic Technology, Guilin, in 2022. He is currently a Associate Professor in the Department of Artificial Intelligence, Hezhou University, Hezhou, China. His research interests include intelligent information processing and artificial intelligence.