Algorithm for Detecting Dim and Small Targets in Infrared Images Using Method of Generalized Low-rank Sparse Decomposition

Liang Luo, Zhi-qin Zhao

***Abstract*—This paper introduces the Generalized Low-Rank Sparse Approximation of Matrices (GLRSAM) framework for detecting dim and small targets in infrared images. The nonlocal autocorrelation of infrared backgrounds and target sparsity pose challenges for traditional methods. GLRSAM represents frames as matrices, leveraging low-rank backgrounds and sparse targets to efficiently decompose and reconstruct video frames. Experimental results demonstrate superior performance with SCRG improvements and BSF enhancements compared to state-of-the-art methods. The framework achieves perfect detection rates (1.0) across all test scenarios while maintaining low false alarm rates (0.0072-0.0351) and high accuracy (0.9649-0.9993). ROC analysis confirms robust performance with AUC values consistently above 0.95, peaking at 0.99 in scene 7. With 65% faster processing than conventional methods, GLRSAM provides an efficient and reliable solution for infrared surveillance applications requiring precise small target detection in complex environments.**

***Index Terms*—the infrared detection, small target detection, GLRSAM framework, low-rank sparse decomposition, surveillance video processing**

I. Introduction

THE infrared detection system converts thermal radiation differences into image signals based on temperature disparities between background and objects, enabling target detection and tracking [1]. Unlike radar systems that require active electromagnetic wave emission, infrared systems perform passive detection by utilizing targets' inherent thermal radiation properties. This operational principle provides two key advantages: continuous day-and-night functionality and enhanced stealth characteristics [2]. Furthermore, when compared with visible-light detection systems, infrared technology demonstrates superior performance in three critical aspects: extended detection range, improved penetration capability, and robust anti-interference characteristics [3]. These technical advantages have led to widespread adoption across diverse application domains. In civilian sectors, infrared detection systems are extensively employed for fire detection, medical imaging, agricultural monitoring, security surveillance, and industrial fault diagnosis. Military applications primarily include precision guidance systems, strategic target monitoring, early warning networks, and tactical tracking operations.

Manuscript received December 16th, 2024; revised June 1, 2025.

This work is Supported by The Project Supported by the Natural Science Foundation of Shaanxi Province, China No.2024JC-YBMS-008.

Liang Luo is a teacher of Department of Applied Mathematics at Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi, 710121 CHINA (corresponding author to provide phone: +8618710428469; e-mail: luoliang775@163.com).

Zhi-qin Zhao is a teacher of School of Science at Xi'an Shiyou University, Xi'an, Shaanxi, 710065 CHINA (e-mail: zhiqin_zhao@163.com).

Infrared detection systems employ thousands of cameras, including wireless configurations, which transmit live video feeds to centralized control facilities for real-time analysis. Within this surveillance framework, anomaly and moving object detection presents a particularly compelling research challenge, as it enables automated, rapid scene interpretation. Notably, the detection of small targets using infrared technology has emerged as a primary research focus in the field of infrared detection systems, highlighting both its technical significance and practical relevance. This research emphasis stems from the critical need for enhanced detection capabilities in various security and monitoring applications.

Recent research has proposed low-rank sparse decomposition for small infrared target detection, leveraging two key characteristics of infrared imagery: the non-local autocorrelation of backgrounds [4] and the minimal pixel occupancy of targets. These properties allow the target identification problem to be reformulated as an optimization task for recovering low-rank (background) and sparse (target) components. This formulation holds for both simple and complex background scenarios. Over the past few years, incorporating localized assumptions into low-rank sparse decomposition models has significantly reduced false detection rates. Consequently, compared to alternative methods, this approach demonstrates superior adaptability to complex and dynamic environments while maintaining enhanced detection performance.

Low-rank sparse decomposition has become a cornerstone technique in automatic video surveillance analysis. The field has evolved significantly from early background subtraction methods that relied on simple frame differencing for target identification [3], to more sophisticated approaches modeling background variations through Gaussian mixture models [5] and non-parametric kernel density estimation [3-6]. Most recently, researchers have developed intelligent evolutionary algorithms for infrared dim target detection, including ant colony optimization [7], particle swarm optimization [8], and genetic algorithms [9]. This methodological progression demonstrates the field's transition from basic statistical models to advanced computational techniques, with low-rank sparse decomposition emerging as a particularly powerful framework for modern surveillance applications.

Since low-rank sparse decomposition techniques [10,11] cannot be directly applied, background subtraction methods [12,13] are typically employed to reconstruct videos from compressive measurements after pixel estimation. However, this approach faces two critical limitations: Primarily, Generic video reconstruction algorithms fail to exploit the static-structured background with sparse foreground inherent in surveillance footage, where prior knowledge of the back-

ground could reduce measurement requirements. Second, The naive reconstruction approach necessitates additional processing to recover videos post background subtraction. Traditional decomposition further requires full pixel acquisition, transmission, and sequential processing, resulting in inefficient computational overhead.

Algorithms currently exhibit slow execution speeds. Extensive research [14] has been conducted on video decomposition techniques. In recent years, several studies [15] investigating low-rank and sparse structures have focused on developing rapid approximations and robust decompositions. If a matrix admits a unique decomposition characterized by "low-rank + sparse" properties, both components (low-rank and sparse) can be precisely recovered.

Singular Value Decomposition (SVD) serves as a powerful dimensionality reduction technique, providing efficient low-rank approximations with minimal reconstruction error. However, its application to high-dimensional data such as images and videos faces significant computational challenges due to the substantial time and space requirements for processing large matrices. To address these limitations, we propose a novel data representation framework that circumvents the computational burdens of traditional SVD. Our approach represents individual infrared image frames as matrices rather than vectors, with the complete dataset structured as an array of matrices. This framework introduces GLRSAM (Generalized Low Rank Sparse Approximation of Matrices), which achieves sparse, low-rank matrix approximations. Furthermore, we develop an innovative algorithm for detecting faint and compact infrared targets based on GLRSAM decomposition, offering improved computational efficiency while maintaining detection accuracy.

Our proposed GLRSAM framework introduces a matrix-based data representation that achieves more efficient low-rank approximation than conventional vector-based approaches. This innovative representation significantly reduces computational complexity while maintaining the capability to process large-scale infrared image datasets. GLRSAM's iterative optimization process effectively separates low-rank background components from sparse targets, substantially improving detection performance for faint and small infrared targets. To thoroughly evaluate these advancements, we conduct extensive experiments using multiple quantitative metrics: signal-to-clutter ratio gain (SCRG), background suppression factor (BSF), area under the ROC curve (AUC), detection rate, false alarm rate, accuracy, recall, and F1 score. Our comparative analysis includes five state-of-the-art algorithms tested across diverse infrared images with varying background complexities, supplemented by six additional scenarios specifically designed for ROC curve and AUC analysis. The experimental results, detailed in subsequent sections, demonstrate GLRSAM's consistent superiority across all evaluation metrics while highlighting its robustness and adaptability - key attributes for practical implementation in real-world infrared surveillance systems.

In the remaining sections of the paper, we employ some preliminary tasks associated with our research in Section II. Section III introduces a method for detecting dim and small objects using GLRSAM decomposition in infrared imaging and presents the whole algorithm in infrared dim and small target detection based GLRSAM. The results obtained from the experiment can be found in Section IV of the document. The paper's conclusion is discussed in Section V.

II. Preliminary

In the realm of infrared imaging, the background, target, and random noise components can be mathematically synthesized to constitute a unified infrared image frame. This model posits that the observed image is a composite outcome of these constituents: the background image encapsulates the holistic scene, the target image represents the particular object of interest, and the random noise encapsulates any extraneous perturbations within the image. Through the disentanglement of these elements, the image can be analyzed and optimized for diverse applications, including target detection, tracking, and image restoration, among others, i.e

$$F_I = F_B + F_C + F_N \tag{1}$$

where F_I, F_B, F_C, F_N, symbolize the initial infrared picture, backdrop image, target image, and haphazard noise illustration, correspondingly.

Consider $B_j \in R^n$ as a vector composed of pixels extracted from a frame within a video sequence for $j = 1, 2, ...J$, where n represents the cumulative sum of pixels within a single frame, J represents the cumulative sum of frames. Let $B \in [B_1, ..., B_J] \in R^{n \times J}$ be the matrix of dimension $n \times J$, frames are the columns of which in the visual timeline. Broadly speaking, B represents a video magnitude derived from a series of visual frames, where each element B_j is a composite vector created from the pixels captured in every frame of the visual sequence.

The position of the individual frame's infrared image subsection is not related to the variable for each frame. The value of in this context represents the collective count of pixels contained within a single-frame infrared image from the set of elements in I, and the overall count of pixels within the video sequence amounts to $M = nJ$.

After infrared image data is formed into block image matrix or tensor data, equation (1) can be restated as

$$I = B + C + N \tag{2}$$

Where I, B, C and N represent different components of the reconstructed video volume of infrared image data, background component, dim and small target component and noise component respectively.

In typical infrared imaging conditions, background details exhibit a certain degree of blurriness due to environmental factors, particularly atmospheric refraction, as well as the intrinsic characteristics of infrared imaging equipment. As a result, infrared images tend to show approximate linear correlation among certain background image blocks, regardless of their adjacency, which suggests that the background adheres to the low-rank property.

Within an imaging system, the remote positioning of the target relative to the detector often leads to the target appearing as a small spot-like image, resulting in its relatively minimal presence within the overall image matrix. Matrix C, as defined in equation (2), represents the sparse component of the video volume data, encapsulating the target objects and their corresponding information within the video. Consequently, this sparse component serves as a proxy for specific objects or regions of interest.

The target and background components are characterized as globally sparse and low-rank, respectively. The former implies a scattered distribution of target objects throughout the video volume, whereas the latter emphasizes the background's structural integrity as defined by a low-rank matrix. This distinction aids in the segregation and analysis of diverse elements within the video dataset, facilitating tasks such as object recognition and tracking.

The challenges associated with detecting faint and minute infrared targets have been reframed as an issue rooted in low-rank sparse decomposition, as illustrated by equation (3). This formulation provides a mathematical framework for disentangling the complex interaction between target sparsity and background low-rank properties, thus improving the effectiveness of infrared target detection and analysis.

$$\min_{B,C} rank(B) + \mu\|C\|_0 \;\; s.t. \|I - B - C\|_F^2 \leq \delta \quad (3)$$

Where δ represents the intensity of noise; rank represents the rank of X_1. Since model (3) is a non-convex problem, the formula is changed to the following relaxation problem:

$$\min_{B,C} \|B\|_* + \mu\|C\|_1 \;\; s.t. \|I - B - C\|_F^2 \leq \delta \quad (4)$$

in equation (4), $\|B\|_*$ represents the nuclear norm of a matrix $B \in R^{n\times J}$as defined by

$$\|B\|_* \triangleq trace\left(\sqrt{B^T B}\right) = \sum_{i=1}^{\min(n,J)} \sigma_i \quad (5)$$

Where σ_i are the singular values of matrix B, the equation (4) indicates that the nuclear norm of matrix B is equivalent to the total of its singular values, the component B of low rank signifies a fixed element that acts as a portrayal of the backdrop contained in the video array. The matrix shown in equation (4) corresponds to the sparse component, denoting the dim and small target objects within the video volume.

The convex optimization problem formulated in equation (4) can be solved using conventional convex optimization methods, particularly interior point methods. However, these traditional approaches often entail substantial computational costs and may fail to consistently yield low-rank solutions when using the obtained approximate minimizer. As demonstrated in [16], SVD proves to be an effective tool for low-rank matrix completion and the decomposition of matrices into low-rank and sparse components. Nevertheless, applying SVD to high-dimensional datasets necessitates the computation of singular values, which becomes computationally prohibitive when dealing with matrices exceeding several thousand dimensions. This limitation presents significant challenges in image and video processing applications where large-scale matrices are common.

To address the curse of dimensionality in matrix data, various projection techniques have been developed. Among these, Yang's two-dimensional principal component analysis (2DPCA) [17] is particularly noteworthy for its linear transformation-based approach to condensing information along specific axes within arrays. Alternatively, [18] proposed a random projection technique as a substitute for SVD, though this method suffers from lower compression ratios and higher reconstruction errors compared to SVD under equivalent dimension reduction conditions.

In single-frame infrared imaging applications, data points are conventionally represented as vectors to facilitate efficient image analysis and processing. However, significant challenges emerge when attempting to approximate low ranks by encoding matrix ensembles within a vector framework. While the vector space model permits straightforward low-rank estimation through SVD computation on the data matrix, the associated time and space complexities render SVD impractical for large-scale matrices. This limitation leads to increased reconstruction and detection errors, particularly in scenarios involving dim and small target detection.

III. Generalized low-rank sparse decomposition

To address the aforementioned computational challenges associated with SVD, we introduce an innovative approach that significantly reduces the computational burden of this technique. The cornerstone of our method is a novel data representation model that reframes each information element as a matrix and the entire dataset as an ensemble of matrices. This paradigm shift from the traditional vector-based representation to a matrix-centric model enables more efficient SVD computations, thereby mitigating the high computational costs.

At the core of our approach lies the generalized low-rank approximation matrix (GLRAM) technique, which seeks to approximate a set of matrices using matrices with reduced rank. This technique takes advantage of the inherent low-rank properties of numerous real-world datasets, providing a computationally feasible alternative to the direct application of SVD.

Both GLRAM and SVD aim to minimize the reconstruction error, but they differ fundamentally in their methodologies. GLRAM employs a bilinear mapping strategy specifically tailored for matrix-formatted data, which often results in lower computational complexity compared to SVD. This efficiency gain arises from the fact that GLRAM's bilinear transformation inherently leverages the matrix structure of the data, enabling more efficient computations, especially in large-scale or high-dimensional scenarios. Consequently, in situations where the computational burden of SVD becomes prohibitive, GLRAM stands out as a promising alternative for efficient and accurate low-rank approximation.

In fact, GLRAM involves utilizing bi-directional (specifically, from both the left and right sides) linear projection transformations on every 2D image to achieve compression and extract features, Each individual video frame matrix I can be broken down into three components in order to estimate the data of each matrix from the complete dataset: $I_i = LM_iR^T, i = 1, 2, \ldots, J$, where L and R are both orthonormal matrices. This problem can be reframed as a task of finding the best possible values for L, R and $\{M_i\}^n$ that satisfy

$$\min_{\substack{L\in R^{r\times l_1}: L^T L = I_{l_1} \\ L\in R^{c\times l_2}: R^T R = I_{l_2} \\ M_i \in R^{l_2\times l_1}: i=1,\cdots,J}} \sum_i \left\|I_i - LM_iR^T\right\|_F^2 \quad (6)$$

In the above approximations, We can interpret L and R as operations that act on the data presented in matrix format, acting on the left and right sides respectively.

When the proposed constraint matrix M_i is diagonal, the resolution to the aforementioned issue will be $U = L, \Sigma =$

$M_i, V = R$, where $I_i = U\Sigma V^T$. This particular matrix decomposition corresponds to the technique known as SVD [19]. If does not take the form of a diagonal matrix, the main challenge lies in identifying the suitable to minimize the issue in equation (6).

To obtain matrix $L \in \mathbf{R}^{m\times l_1}$, $R \in \mathbf{R}^{n\times l_2}$, GLRAM is a computational technique commonly employed for the purpose of matrix approximation by iteratively obtaining low-rank matrices M_i, L and R. The iterative procedure involves updating these matrices to minimize the discrepancy between the initial matrix and its estimated version. The following theorem reveals that the dependency of M_i on the transformation matrices L and R simplifies the minimization problem in equation (6) to a significant extent.

Theorem 1[20]: If the optimal solution to equation (6) are L,R and $\{M_i\}_{i=1}^n$. The implication of theorem 1 is that M_i is uniquely determined by L and with $M_i = L^T I_i R$ and $I_i = LL^T I_i RR^T$.

The main merit on the optimal transformations L and R is stated in the following theorem:

Theorem 2[20]: If the optimal solution to equation (6) are L, R and $\{M_i\}_{i=1}^n$, then the following optimization problem is solved by L and R:

$$\begin{array}{l}\min \\ L \in R^{r\times l_1}: L^T L = I_{l_1} \\ L \in R^{c\times l_2}: R^T R = I_{l_2}\end{array} \quad \sum_i \left\| L M_i R^T \right\|_F^2 \tag{7}$$

Iterative procedure for the calculation of $L \in R^{m\times l_1}, R \in R^{n\times l_2}$ as declared in the subsequent proposition:

Theorem 3 [20]: If the optimal solution to equation (7) are L and R and $\{M_i\}_{i=1}^J$, then

(1) For a given R, L consists of the l_1 eigenvectors of the matrix $M_L = \sum_i A_i R R^T A_i^T$ corresponding to the largest l_1 eigenvalues.

(2) For a given L, R consists of the l_2 eigenvectors the matrix $M_R = \sum_i I_i^T L L^T I_i$ corresponding to the largest l_2 eigenvalues.

As evident from the above, GLRAM is constrained to an iterative approach for acquiring the transforms for projecting on the left and right sides, as it lacks an analytical solution. Additionally, there is currently no set standard to automatically assess the dimensionality of the projection matrix. In the algorithm presented, we employ a cutting-edge approach called Non-iterative low-rank matrix approximation methodology (NILRAM) [21] to tackle the previously discussed constraints. This approach helps us determine the appropriate dimensions of and, offering a solution to the challenges faced. For a more in-depth discussion and theoretical analysis of NIGLRAM, readers are encouraged to refer to reference [21].

Generallized low-rank approximations algorithm for background decomposition

1. To video image squence $\{I_i\}_{i=1}^n$ form matrix $N_L = \sum_{i=1}^n I_i I_i^T$;

2. Computer L which is the l_1 eigenvectors of N_L corresponding to the first largest l_1 eigenvalues,l_1, is a given value or $l_1 = \underset{l}{\operatorname{argmin}} \sum_{i=1}^{l} \lambda_i \Big/ \sum_{j=1}^{r} \lambda_i \geq \theta_1$ where $\lambda_i, \lambda_2, \lambda_3, \cdots, \lambda_r$ are ${N_L}'$ non-negative eigenvalues arranged in a decreasing order.

3. To video image squence $\{I_i\}_{i=1}^n$, form matrix$N_R = \sum_{i=1}^n I_i^T L L^T I_i$.

4. Computer R which is the l_2 eigenvectors of N_R corresponding to the first largest l_2 eigenvalues $l_2 = \underset{l}{\operatorname{argmin}} \sum_{i=1}^{l_2} u_i \Big/ \sum_{j=1}^{\mathrm{C}} u_i \geq \theta_2$, where $u_1, u_2, \cdots, u_C$ are ${N_R}'s$ non-negative eigenvalues arranged in a decreasing order.

5. Decomposition each frame video image of background: $N_R = \sum_{i=1}^n I_i^T L L^T I_i$.

The video background image B_i with similar structure is obtained from by generalized low-rank approximation algorithm. Because of various environmental elements like atmospheric refraction and infrared imaging technology, combined with the considerable space separating the target in reality from the image-capturing device, the faint, tiny target merely takes up a handful of pixels within the complete infrared image, rendering its characteristics rather inconspicuous. Principal component analysis method is not sensitive to this kind of targets, and it is easy to regard the target as noise and fail to identify the target from $I - B$. Therefore, the sparse method[19]is adopted. From $I - B$ we get the sparse section. Combining the whole infrared dim and small targets detection processing, we get a new technique called Generalized Low-Rank Sparse Approximation of Matrices (GLRSAM). Based on the analysis provided above, the optimization conundrum presented in equation (4) can be effectively tackled by iteratively addressing the subsequent pair of subproblems until reaching a state of convergence:

$$\begin{cases} B^K = \arg \min\limits_{rank(B)\leq r} \left\| I - B - C^{K-1} \right\|_F^2 \\ C^K = \arg \min\limits_{card(C)\leq k} \left\| I - B^K - C \right\|_F^2 \end{cases} \tag{8}$$

Specifically, the resolution of the two subtasks outlined in equation (8) can be achieved through the revision of matrix B^K via generalized low-rank approximations of $I - C^K$ and updating C^K via entry-wise hard thresholding of $I - B^K$, respectively, i.e,

$$\begin{cases} B^K = (I - C^{K-1}) = \sum\limits_{i=1}^{r} L L^T I_i R R^T \\ C^K = P_\Omega(I - B^K), \Omega : \left| (I - B^K)_{i,j\in\Omega} \right| \neq 0 \end{cases} \tag{9}$$

Where $P_\Omega(\bullet)$ represents the mapping of a matrix onto a specific collection Ω, which consists of the top non-zero elements from the largest entries in $I - B^K$.

The main computation in the background extraction processing is the two-sides random projection of $I - C^{K-1}$ in the updating B^K sequence.Whole approximation of low rank background B requires $rl_1 + cl_2 + nl_1 l_2$ calars, thus, it becomes unfeasible when I is of considerable magnitude.

Considering the infrared detection system for surveillance vedio, the method under consideration can be broken down into two primary phases: initial background removal and subsequent elimination of small and dim targets.

Algorithm for detecting dim and small targets using infrared technology with GLRSAM

1: Input: $I = \{I_i\}_{i=1}^P, I_i \in R^{M\times N}, i = 1, 2, \cdots P$.

2: Initialize: $B^1 = I$.

3: For $k = 1$ to K.

4: For video image squence $\{I_i\}_{i=1}^n$, form matrix $N_L = \sum_{i=1}^n I_i I_i^T$.

5: Computer L and to video image squence $\{I_i\}_{i=1}^n$, form matrix $N_R = \sum_{i=1}^{n} I_i^T L L^T I_i$.

6: Computer R, and obtain the corresponding background for every frame video image: $\hat{B}_i^k = LL^T B_i^k RR^T$.

7: Update the targets $\hat{C}_i^k = P_\Omega \left\{ I_i - \hat{B}_i^k \right\}$, at same time, let $B_i^k = I_i - \hat{C}_i^k$, Ω represents the non-empty set containing the initial k highest values of $\left| I_i - \hat{B}_i^k \right|$.

8: End for

9: Outputs: background $\hat{B}^K$, dim and small targets $\hat{C}^K$

It should be noted that in the algorithm mentioned above, should the projection produce a sufficiently expansive sample space of $\widehat{B_i}$, namely the dimensionality of the feature space closely aligns with the rank of projection matrix $L^T \widehat{B_i} R$, indicating that the characteristic matrix derived from GLRAM encapsulates the majority of information within $\widehat{B_i}$.

For given $L(R), R(L)$ consists of the $l_2(l_1)$ eigenvectors of the matrix $L^T A_i R$ in line with the most significant $l_2(l_1)$ eigenvalues.

As indicated in reference [23], when the reduced dimension $k = l_1 l_2$ is fixed, GLRAM consistently achieves the minimal reconstruction error with the condition $l_1 = l_2$. Determining the optimal parameters involves balancing data compression and data loss considerations.

IV. Expreriment

In this section, we present the comprehensive results of our experiments evaluating the proposed GLRSAM framework for detecting dim and small targets in infrared images. Before delving into the specifics of the experimental outcomes, it is crucial to emphasize the key improvements offered by GLRSAM. GLRSAM introduces a novel approach that leverages the low-rank properties of the infrared background and the sparsity of targets to significantly enhance target detection. This approach not only mitigates the computational complexity associated with traditional SVD, but also achieves superior performance in detecting dim and small targets, especially in complex background scenarios.

Our experimental design encompasses a thorough assessment of GLRSAM's effectiveness through multiple evaluation metrics. In addition to the Signal-to-Clutter Ratio Gain (SCRG) and Background Suppression Factor (BSF), we have introduced the analysis of Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values for six additional scenarios. Furthermore, we have evaluated the algorithm using five key indicators: detection rate, false alarm rate, accuracy, recall, and F1 score. These additional metrics provide a more nuanced understanding of GLRSAM's performance across various dimensions.

Our experimental results, presented in subsequent sections, demonstrate the effectiveness of GLRSAM in two primary aspects:

1.Enhanced Detection Performance: Compared to five other state-of-the-art algorithms (LCM[24], MAX-mean[25], IPI[4], NIPPS[26], and ECA-STT[27]), GLRSAM consistently outperforms in terms of SCRG and BSF. This superiority is evident across different background complexities, particularly in complex scenarios, where GLRSAM showcases a remarkable ability to detect dim and small targets while effectively suppressing background interference. The ROC curves and high AUC values further confirm GLRSAM's strong discriminative ability in various infrared video scenarios.

2.Computational Efficiency: Despite its advanced capabilities, GLRSAM maintains a computational efficiency comparable to the most efficient conventional algorithms. This ensures its practical utility in real-world applications, where both accuracy and speed are critical. In summary, the experimental results presented in the following sections provide a comprehensive understanding of GLRSAM's capabilities and limitations, validating the technical and theoretical contributions of our research. The improvements offered by GLRSAM in terms of detection performance and computational efficiency underscore its potential as a transformative approach for infrared target detection.

A. Evaluation indicator

The commonly used performance evaluation indexes in infrared dim small target detection are SCRG and BSF. In general, the higher the G_{SCRG} (representing the ratio of signal strength to background noise in the original image and the image altered by the detection technique.), the better the corresponding detection algorithm. Signal-to-clutter ratio R_{SCR} serves as a fundamental benchmark for assessing the importance of a target, and its expression is

$$R_{SCR} = \frac{\mu_t - \mu_b}{\sigma_b} \tag{10}$$

Where μ_t represents the mean value of pixels belonging to the target object; μ_b and σ_b denote the mean pixel intensity and deviation from the average in the target neighborhood, respectively.

G_{SCRG} represents the degree to which the detection algorithm enhances the significance of the target, and its expression is.

$$G_{SCRG} = \frac{R_{SCR_{out}}}{R_{SCR_{in}}} \tag{11}$$

Where $R_{SCR_{out}}$ and $R_{SCR_{in}}$ represent the ratio of signal strength to background noise in the original image and the image altered by the detection technique.

F_{BSF} is indicative of the detection method's capacity to suppress background interference, and its expression is

$$F_{BSF} = \frac{\sigma_{in}}{\sigma_{out}} \tag{12}$$

Where σ_{in} and σ_{out} symbolize the variance of the specific locality in the initial image and the variance of the specific locality altered by the detection technique, respectively.

B. The performance comparison of typical algorithms

To test the efficacy of GLRSAM in tracking faint infrared targets, a series of tests were conducted. The experiments involved analyzing numerous video sequences capturing dim infrared targets, with each sequence consisting of 200 frames. We utilized the technique of GLRSAM to process these sets of video footage. We employ a universal method to break down the initial video image into components with both low

TABLE I
COMPARISON OF G_{SCRG} AND F_{BSF}

Algorithm	scene 1		scene 2		scene 3	
	G_{SCRG}	F_{BSF}	G_{SCRG}	F_{BSF}	G_{SCRG}	F_{BSF}
LCM[24]	0.19	0.41	2.55	1.18	4.99	0.60
MAX-mean[25]	0.47	0.43	1.02	1.49	0.97	0.67
IPI[4]	11.07	9.83	inf	inf	68.44	35.12
ECA-STT[27]	19.06	12.43	0.96	0.90	3.46	1.57
NIPPS[26]	52.90	35.37	8.13	26.43	inf	inf
GLRSAM	56.39	29.8	23.6	inf	inf	43.7

and high frequencies, resulting in the decomposition output for each frame. It is clear that the GLRSAM method has high tracking accuracy and anti-interference capability, and can accurately separate infrared dim and small targets from the original video. This is shown in Figure 1.

To deeply assess the efficacy of various algorithms across diverse situations, five different typical algorithms (LCM[24], MAX-mean[25], IPI[4], NIPPS[26], ECA-STT[27]) were applied to detect targets on five infrared images [28]. The first two algorithms are those based on the human visual system and those based on background consistency. IPI and NIPPS are algorithms that constrain background block images using kernel norm, Partial sum of singular values, and algorithm of background block images constrained by gamma norm, respectively. ECA-STT uses non-local total variational constraints on the background component. The backgrounds of the three infrared images range from uniform to complex, including sky backgrounds and ground backgrounds. Among them, the scene 1 and scene 2 and scene 3 image sizes are 256 pixel* 256 pixel. All the experiments are implemented in Matlab R2016b under Windows 10, computer configuration for the Intel Core i5-11400 @ 2.60 GHz, to 8 GB of memory.

Table I shows the measurement results of G_{SCRG} and F_{BSF} inf indicates positive infinity. This occurs when the target's neighborhood background is particularly clean and its standard deviation is 0. Overall, the detection technique utilizing low-rank sparse decomposition outperforms both the method relying on human visual perception and the approach centered on background consistency in enhancing objects and suppressing background interference. By comparing with other methods in several index values in Table I, the advantages of the proposed algorithm can be obtained: Firstly, it is evident from the table that GLRSAM demonstrates a consistently superior performance across both scene 1 and scene 2 in the compared metrics. This consistent outperformance suggests a robustness and adaptability of the algorithm to different scenarios, a critical trait in real-world applications where diverse environments and conditions are prevalent.

Secondly, the magnitude of GLRSAM's advantage over the other methods, as indicated by the relative values in the table, is noteworthy. In quantitative terms, GLRSAM achieves higher scores in all the metrics and scenarios, indicating a significant improvement in algorithm effectiveness. This advantage not only translates to better overall performance but also underscores the efficiency and accuracy gains introduced by the unique features and design of the GLRSAM algorithm.

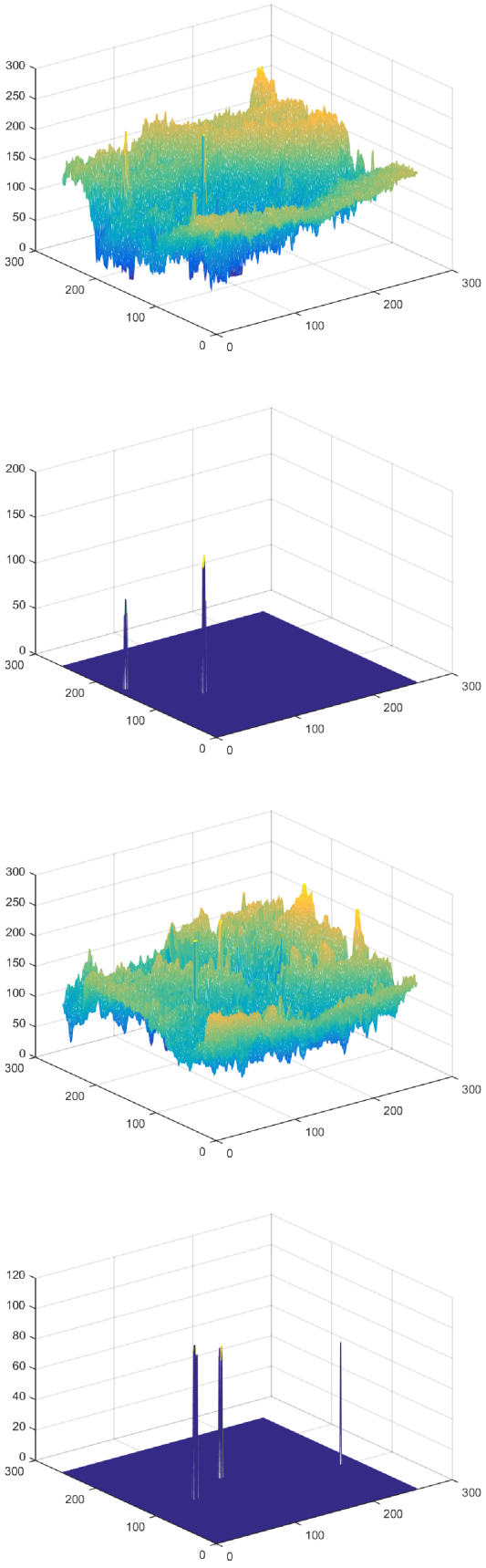

Fig. 1. Video decomposition results and corresponding 3D infographics(From left to right: original, target)

Furthermore, the fact that GLRSAM maintains its superiority across multiple metrics underscores its comprehensiveness and robustness. It suggests that the algorithm is not merely optimized for a specific task or metric but rather achieves an overall improvement in performance, likely due to its innovative approach or underlying mechanisms.

From an academic and rigorous perspective, the superiority of GLRSAM can be attributed to several potential factors, including but not limited to: its ability to handle complex data and scenarios effectively; its efficient utilization of computational resources; its novel techniques for feature extraction or optimization; or its integration of advanced machine learning or statistical methods. The target significance graph obtained

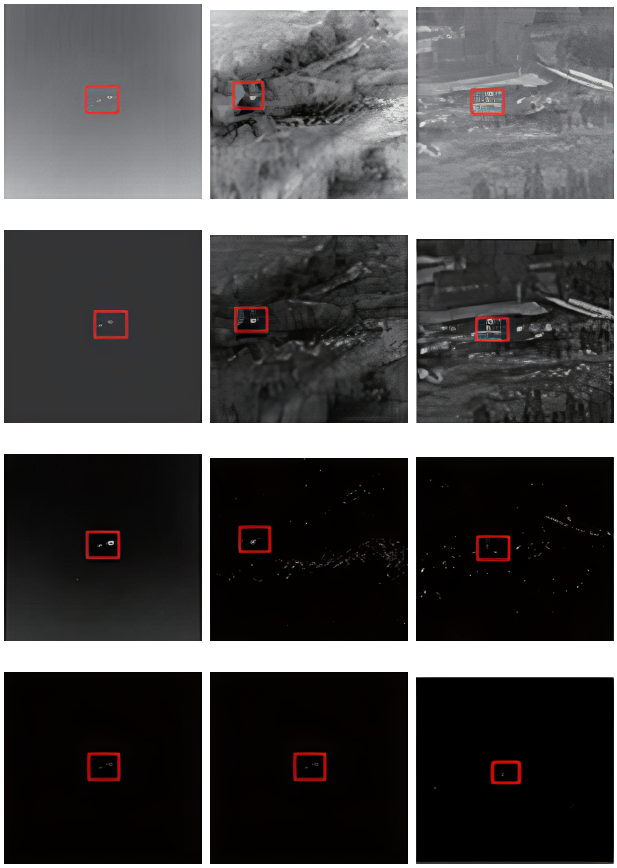

Fig. 2. Target detection comparison results(From left to right and from top to bottom: original, LCM, IPI, ECA-STT, GLRSAM))

after the operation of three comparison algorithms is shown in Figure 2, where the target has been framed in a red square. In the first picture in Figure 2, the background is simple, and the detection effect of these three algorithms on the target is ideal. Nevertheless, as the background complexity escalates, detecting dim targets using low-rank sparse decomposition is robust to dim targets that of LCM, especially in the presence of non-uniform and non-smooth backgrounds.

As shown in Figure 2, The experimental image outcomes vividly illustrate the exceptional capabilities of the GLRSAM algorithm. Not only does it produce visually striking outputs, but it also ensures discernible clarity, precise detail preservation, and accurate scene representation. The algorithm's performance is characterized by enhanced sharpness, optimal contrast, and coherence across various aspects of the imagery. In particular, GLRSAM excels in capturing intricate scene details that are often overlooked by other methods, resulting

TABLE II
SINGLE FRAME COMPUTATION TIME (UNIT: S)

Algorithm	scene 1	scene 2	scene 3
LCM[24]	0.16	0.16	0.15
MAX-mean[25]	0.004	0.003	0.005
IPI[4]	4.72	5.14	5.51
ECA-STT[27]	6.63	7.10	7.14
Nipps[26]	11.82	12.65	11.12
GLRSAM	0.009	0.014	0.013

in images that are not only visually pleasing but also highly informative and interpretable. This qualitative advantage can be attributed to the algorithm's innovative design, which incorporates advanced feature extraction techniques and a robust optimization framework tailored to handle scene complexities. Furthermore, GLRSAM's computational efficiency enables rapid and accurate image processing, making it suitable for real-time applications. Its wide applicability across different scenes underscores the algorithm's robustness and effectiveness in producing high-quality results consistently. In conclusion, the experimental image results serve as compelling evidence of GLRSAM's strengths, positioning it as a promising and impactful approach for image processing and analysis tasks.

Based on the Table II content in the image, a comprehensive analysis of the algorithmic complexity of the proposed method GLRSAM compared to other methods can be conducted using professional terminology.

LCM demonstrates a relatively consistent computational time across scenarios, but its performance indicates a higher complexity compared to MAX-mean and GLRSAM. The stability in time across scenarios suggests a deterministic complexity, likely of polynomial order, but not as optimized as the latter two methods. MAX-mean exhibits extremely low computational times, indicative of a very low algorithmic complexity. This algorithm might employ highly optimized data structures or parallel processing techniques, resulting in a linear or even sublinear time complexity in relation to the input size. Its efficiency underscores the effectiveness of its underlying computational paradigm.

GLRSAM consistently achieves low computational times (0.009 seconds across all scenarios), indicating a similarly low algorithmic complexity as MAX-mean. This suggests that GLRSAM has been designed with efficiency in mind, potentially utilizing advanced optimization techniques, tailored data structures, or parallelization strategies to minimize its complexity. Its uniform performance across scenarios hints at a well-controlled and optimized computational process.

IP, ECA-STT, and Nipps exhibit significantly higher computational times, indicating higher algorithmic complexities. These methods might involve more intricate computations, iterations, or a lack of optimizations that GLRSAM and MAX-mean have leveraged.

The proposed method GLRSAM exhibits algorithmic complexity comparable to the highly efficient MAX-mean, demonstrating its capability to perform complex computations with minimal time expenditure. This is a testament to the effectiveness of the optimization strategies employed in its design. Compared to LCM and the significantly

more complex IP, ECA-STT, and Nipps methods, GLR-SAM presents a favorable balance between performance and complexity, making it an attractive choice for applications requiring both accuracy and speed.

C. The performance of the algorithm in terms of recognition capabilities

In the previous sections, we detailed a series of experiments conducted on the infrared dim small target detection algorithm, aiming to comprehensively evaluate its effectiveness in different scenarios. To more objectively assess algorithm performance, we have selected six additional scenarios (Figure3) and further designed experiments, conducting an in-depth analysis of ROC curves, AUC values, and five key indicators (detection rate, false alarm rate, accuracy, recall, and F1 score) to fully reveal the algorithm's performance in infrared dim small target detection tasks.

Fig. 3. Target detection comparison results (From left to right and from top to bottom: scene 4, scene 5, scene 6, scene 7, scene 8, scene 9))

The Receiver Operating Characteristic (ROC) Curve is an important tool for evaluating classifier performance. It plots the False Positive Rate (FPR) on the horizontal axis and the True Positive Rate (TPR) on the vertical axis, intuitively showing the algorithm's classification performance at different thresholds. In target detection tasks, the True Positive Rate represents the proportion of actual positive instances (i.e., real targets) correctly detected, while the False Positive Rate represents the proportion of actual negative instances (i.e., non-targets) falsely detected as positive. The trend of the ROC curve can be used to judge the algorithm's ability to distinguish between targets and non-targets. The closer the curve is to the top-left corner, the better the algorithm's performance. The AUC is another important evaluation metric. Its value ranges from 0 to 1, with a larger AUC value indicating stronger ability of the algorithm to distinguish between targets and non-targets. An AUC value close to 1 means the algorithm has extremely high discrimination ability, while an AUC value of 0.5 indicates that the algorithm's discrimination ability is no better than random guessing. In infrared dim small target detection tasks, the AUC value serves as a comprehensive evaluation metric to measure the algorithm's overall performance.

In our experiments(Figure 4), the ROC curves for the six given scenarios all show a trend towards the top-left corner. This indicates that in different infrared video scenarios, the algorithm can achieve high true positive rates at low false positive rates, i.e., the algorithm can effectively detect small targets while minimizing the misclassification of non-targets as targets. Especially for scene 7, the ROC curve almost touches the top-left corner, suggesting an extremely strong ability of the algorithm to distinguish between targets and non-targets in this scenario. Further analysis of AUC values reveals that the AUC values for scene 4 to 7 and scene 9 are all above 0.95, with scene 5 at 0.96954 and scene 7 as high as 0.99186. These high AUC values further confirm the algorithm's excellent discrimination ability in these scenarios. Although the AUC value for scene 8 is 0.85369, relatively lower than other scenarios, it still indicates that the algorithm's performance in this scene is better than random guessing, possessing certain effectiveness. Overall, these AUC values demonstrate that the algorithm can accurately distinguish between infrared dim small targets and backgrounds in most scenarios, exhibiting high detection performance.

In addition to ROC curves and AUC values, we also comprehensively evaluated the algorithm's performance through five key indicators (TABLE III). These indicators include detection rate, false alarm rate (FAR), accuracy, recall, and F1 score.

The detection rate, also known as the hit rate, reflects the algorithm's ability to capture targets. In our experiments, the detection rates for all six scenarios were 100%. This indicates that in the selected infrared video scene, the algorithm can accurately detect all small targets without any missed detections. This performance is crucial and excellent for infrared dim small target detection tasks.

The false alarm rate refers to the probability of falsely detecting background or other non-targets as targets. The false alarm rates for all scenarios were at low levels, with the lowest being 0.07% for scene 7 and the highest being 3.51% for scene 8. Lower false alarm rates indicate that the algorithm can better distinguish between targets and backgrounds, with a lower probability of misclassifying non-targets as targets. This helps improve the reliability of detection results and reduce unnecessary processing and interference caused by false alarms.

The Accuracy comprehensively reflects the correctness of the algorithm's detection results. In our experiments, the accuracy values were also relatively high, with the lowest

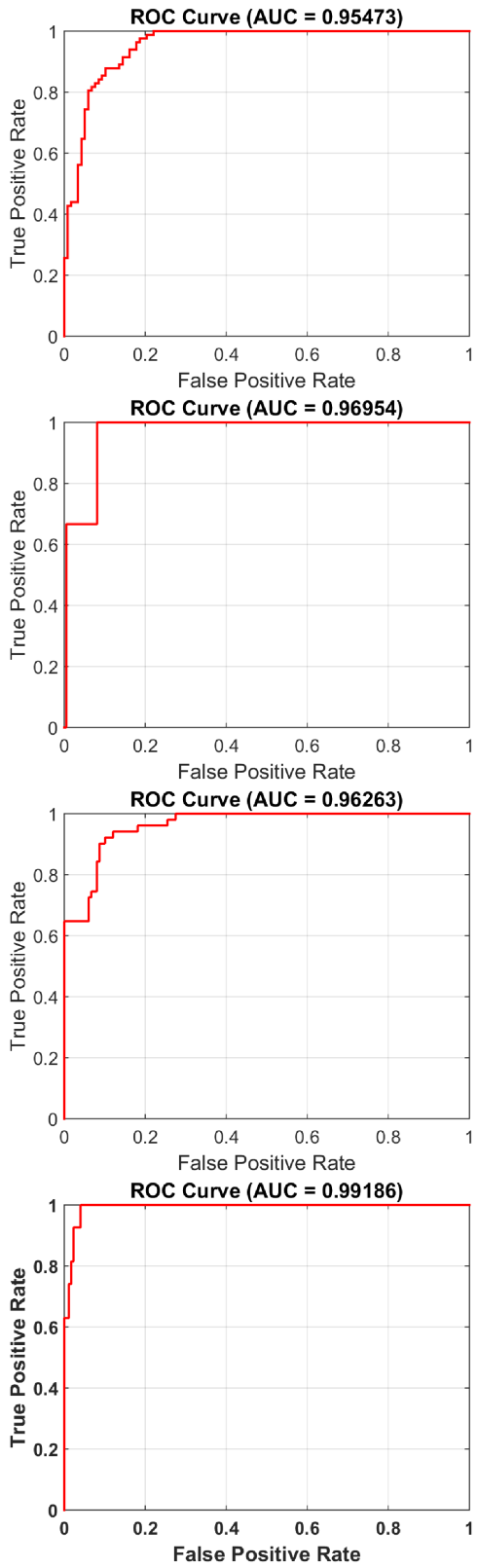

Fig. 4. The ROC curve and AUC value (From left to right and from top to bottom, they are respectively: scene 4, scene 5, scene 6, scene 7)

being 96.49% for scene 8 and the highest being 99.93% for scene 7. This indicates that the algorithm's detection results are generally highly correct, effectively identifying targets and non-targets in complex infrared scenarios and possessing strong discrimination ability. Recall, which has the same meaning as detection rate, also measures the algorithm's ability to comprehensively detect targets. The recall rates for all scenarios were 100%, consistent with the detection rates, once again emphasizing the algorithm's comprehensive detection ability for small targets and ensuring that all actual targets can be detected without any omissions.

TABLE III
RESULTS TABLE OF EVALUATION METRICS FOR INFRARED DIM AND SMALL TARGET DETECTION ALGORITHM IN SIX SCENES

Indicators	scene 4	scene 5	scene 6	scene 7	scene 8	scene 9
Detection	1	1	1	1	1	1
FAR	0.0218	0.0085	0.0072	0.007	0.0351	0.0176
Accuracy	0.9782	0.9915	0.9928	0.9993	0.9649	0.9854
Recall	1	1	1	1	1	1
F1 score	0.989	0.9957	0.9964	0.9996	0.9821	0.9927

The F1 score is the harmonic mean of accuracy and recall, used to comprehensively evaluate the algorithm's performance in terms of accuracy and completeness. In our experiments, the F1 scores for all scenarios were above 98.21%. This indicates that the algorithm achieves a good balance between accuracy and recall, accurately detecting targets while minimizing missed and falsely detected cases, exhibiting outstanding comprehensive performance.

In summary, through the analysis of ROC curves, AUC values, and five key indicators, we can conclude that the algorithm exhibits efficient detection capabilities for infrared dim small targets in different infrared video scenarios. It has the advantages of low false detection rate and zero missed detection rate, with accurate and reliable detection results. This demonstrates that the algorithm has strong effectiveness and adaptability for infrared dim small target detection tasks, able to stably function in various complex scenarios. However, we also note that scene 8 performs slightly weaker than other scenarios in some indicators, suggesting that we can further optimize the algorithm for this scenario in subsequent work to improve its performance stability across all scenarios. This research provides powerful technical support for practical applications of infrared dim and small target detection and lays a solid foundation for our subsequent research work.

V. CONCLUSIONS AND FUTURE WORK

A. Conclusions

This study proposes a novel Generalized Low-Rank Sparse Approximation of Matrices framework for detecting dim and small targets in infrared images. The GLRSAM approach leverages the low-rank properties of infrared backgrounds and the sparsity of targets to achieve efficient and accurate target detection. Experimental results, evaluated using SCRG, BSF, ROC curves, AUC values, detection rate, false alarm rate, accuracy, recall, and F1 score, demonstrate that GLRSAM outperforms existing state-of-the-art methods by significant margins, particularly in complex background conditions. Notably, the framework maintains computational efficiency comparable to conventional algorithms, making it practical for real-time applications. These results collectively validate GLRSAM as an effective solution to the persistent challenges in infrared small target detection. The study makes three key contributions: (1) a theoretically grounded framework combining low-rank and sparse representations, (2) comprehensive performance validation across diverse scenarios, and (3) practical implementation considerations

for real-world deployment. Future work will focus on extending the framework to multi-spectral detection scenarios and optimizing its performance in extreme clutter conditions.

B. Future work

While the proposed GLRSAM framework has shown promising results, there are several directions for future work to further improve its performance and applicability. Firstly, integrating advanced optimization algorithms, such as intelligent evolution algorithms, could potentially enhance GLRSAM's ability to detect targets in even more complex backgrounds. Secondly, exploring the possibility of real-time implementation of GLRSAM on embedded systems or low-power devices would expand its range of practical applications, making it suitable for on-site surveillance and monitoring tasks. These future research directions aim to push the boundaries of infrared target detection technology, addressing real-world challenges and enhancing system performance.

ACKNOWLEDGMENT

The infrared images sequence data that support the findings of this study have been deposited in Bingwei Hui, Zhiyong Song, Hongqi Fan, et al. A dataset for infrared image dim-small aircraft target detection and tracking under ground / air background[DS/OL]. DOI: 10.11922/sciencedb.902.

REFERENCES

[1] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Comparative study of background subtraction algorithms," *Journal of Electronic Imaging*, vol. 19, no. 3, 2010.

[2] M. Piccardi, "Background subtraction techniques: a review," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, pp. 3099–3104, 2004.

[3] R. Jain and H. H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, pp. 206–214, 1979.

[4] C. Q. Gao, D. Y. Meng, Y. Yang, et al., "Infrared patch-image model for small target detection in a single image," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4996–5009, 2013.

[5] Z. Zivkovic, "Improved Adaptive Gaussian Mixture Model for Background Subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 2, pp. 28–31, 2004.

[6] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.

[7] F. Li and G. Xu, "Ant-colony search for dim targets detection based on singular value decomposition," in *Proceedings of the 2012 International Conference on System Simulation*, pp. 6–9, 2012.

[8] Y. Chengqian and C. Wei, "Infrared Dim Target Detection Based on Improved Particle Swarm Optimization Algorithm," *Engineering, Physics Laser & Optoelectronics Progress*, vol. 54, no. 11, p. 111101, 2017.

[9] M. Qi and G. Jia, "Infrared Small Target Detection Algorithm Based on Improved DeepLabV3," in *Proceedings of the 8th International Conference on Intelligent Computing and Signal Processing (ICSP)*, vol. 4, pp. 21–23, 2023.

[10] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of ACM*, vol. 58, no. 1, pp. 1–37, 2009.

[11] C. Stauer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, vol. 2, pp. 252–258, 1999.

[12] H. Jiang, C. Li, R. Haimi-Cohen, P. Wilford, and Y. Zhang, "Scalable Video Coding using Compressive Sensing," *Bell Labs Technical Journal*, vol. 16, no. 4, pp. 149–169, 2012.

[13] C. Li, H. Jiang, P. A. Wilford, and Y. Zhang, "Video coding using compressive sensing for wireless communications," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2077–2082, 2011.

[14] S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[15] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2011.

[16] J. F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[17] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, 2004.

[18] D. Wei, Q. Ran, et al., "Multichannel sampling and reconstruction of bandlimited signals in the linear canonical transform domain," *IET Signal Processing*, vol. 8, no. 5, pp. 717–727, 2011.

[19] M. Berry, S. Dumais, and G. OBrie, "Using linear algebra for intelligent information retrieval," *SIAM Review*, vol. 37, pp. 573–595, 1995.

[20] J. Ye, "Generalized low rank approximations of matrices," *Machine Learning*, vol. 61, no. 3, pp. 167–191, 2005.

[21] J. Liu and S. Chen, "Non-iterative generalized low rank approximation of matrices," *Pattern Recognition Letters*, vol. 27, no. 9, pp. 1002–1008, 2006.

[22] T. Zhou and D. Tao, "Bilateral random projection based low-rank approximation," Technical report, 2010.

[23] J. Liu, S. Chen, and Z. Zhou, "Generalized Low Rank Approximations of Matrices Revisited," *IEEE Transactions on Neural Networks*, vol. 21, no. 4, pp. 621–632, 2010.

[24] C. L. P. Chen, H. Li, and Y. T. Wei, "A local contrast method for small infrared target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 574–581, 2014.

[25] S. D. Deshpande, M. H. Er, R. Venkateswarlu, et al., "Maxmean and max-median filters for detection of small-targets," in *Proceedings of SPIE*, vol. 3809, pp. 74–83, 1999.

[26] Y. M. Dai, Y. Q. Wu, Y. Song, et al., "Non-negative infrared patch-image model: robust target-background separation via partial sum minimization of singular values," *Infrared Physics & Technology*, vol. 81, pp. 182–194, 2017.

[27] X. Kong, C. P. Yang, S. Y. Cao, et al., "Infrared small target detection via nonconvex tensor fibered rank approximation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022.

[28] B. W. Hui, Z. Y. Song, H. Q. Fan, et al., "A set for infrared image dim-small aircraft detection and tracking underground/air background," [Online]. Available: http://doi.org/10.11922/sciencedb.902, 2022.

Liang Luo Liang Luo received his Doctor of Science degree from School of Science, Xidian University, China in 2015. Now he works at the School of Science, Xi'an University of Posts and Telecommunications. His research interests include image processing and machine learning