Moving Object Detection in Satellite Videos via Spatial–Temporal Tensor Model and Weighted Schatten p-Norm Minimization

Qian Yin, Ting Liu, Zaiping Lin, Wei An, and Yulan Guo

***Abstract*— Low-rank matrix decomposition approaches have achieved significant progress in small and dim object detection in satellite videos. However, it is still challenging to achieve robust performance and fast processing under complex and highly heterogeneous backgrounds since satellite video data can neither adequately fit the foreground structure nor the background model in the existing matrix decomposition models. In this letter, we propose a novel object detection method based on a spatial–temporal tensor data structure. First, we construct a tensor data structure to exploit the inner spatial and temporal correlation within a satellite video. Second, we extend the decomposition formulation with bounded noise to achieve robust performance under complex backgrounds. This formulation integrates low-rank background, structured sparse foreground, and their noises into a tensor decomposition problem. For background separation, a weighted Schatten p-norm is incorporated to provide adaptive threshold to obtain the singular value of the background tensor. Finally, the proposed model is solved using the alternative direction method of multipliers (ADMM) scheme. Experimental results on various real scenes demonstrate the superiority of the proposed method against the compared approaches.**

***Index Terms*— Low-rank tensor recovery, moving object detection (MOD), satellite video, weighted Schatten p-norm.**

I. INTRODUCTION

As a new Earth observation technology, satellite video is able to provide a period of continuous observation over an area, providing rich dynamic information of an object, such as the moving trajectory, speed, and directions. Satellite video is important for numerous applications, such as space-based surveillance [1], traffic monitoring, and disaster rescue.

As an important task based on satellite videos, small and dim moving object detection (MOD) has attracted increasing attention in recent years. However, this task is highly challenging due to several facts.

1) *Low Spatial Resolution:* Due to the long distance between a target and the imaging platform, the object is extremely small. Besides, the appearance of objects changes significantly between the consecutive frames.
2) *Large Field of View:* Each frame in a satellite video is typically on the order of several to hundreds of megapixels, resulting in a large searching space and illumination variation.
3) *Heterogeneous Backgrounds and Complex Noise:* Objects are usually immersed and densely packed in highly heterogeneous and complex backgrounds.

Most state-of-the-art MOD methods for satellite videos follow a motion-based paradigm, that is, the background subtraction (BS) technique is used to separate a frame into foreground and background components. Typical BS methods include statistical models [2], [3] and robust principal component analysis (RPCA)-based models [4]–[6]. The RPCA-based methods can be categorized into batch-based methods [4], [5] and online methods [6].

Statistical methods (i.e., median (mean) model and statistical model (VIBE) [2]) usually compare each video frame with an adaptive background model (which is free of moving objects). Ahmadi *et al.* [3] employed a median background model to detect objects and used the nearest neighbor algorithm to produce trajectories. However, these statistical methods do not consider the structure knowledge of a video (e.g., temporal similarity of background and spatial contiguity of foreground). Consequently, their detection performance cannot be further improved, especially in complex and dynamic backgrounds.

To address this limitation, RPCA [4], [5], [7] was introduced to encode the temporal similarities of video backgrounds and mostly useful foreground prior structures (e.g., sparsity and spatial continuity). Pflugfelder *et al.* [6] and Zhang *et al.* [8] proposed several methods based on the low-rank and structured sparse decomposition (LSD) framework [5] to achieve MOD in satellite videos. However, these matrix RPCA-based methods can only convert the videos with a natural 3-D structure to a 2-D data, which can destroy the structure information and reduce the detection performance. In addition, these methods cannot achieve robust performance and fast processing speed in complex and highly heterogeneous backgrounds.

Motivated by the work for exploiting spatial–temporal and structural information in [9], [10], we incorporate a

---

Manuscript received June 22, 2021; revised August 20, 2021; accepted September 29, 2021. Date of publication October 1, 2021; date of current version January 10, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61972435 and Grant 61872379, and in part by the Natural Science Foundation of Guangdong Province under Grant 2019A1515011271. *(Corresponding author: Yulan Guo.)*

Qian Yin, Ting Liu, Zaiping Lin, and Wei An are with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: yinqian18@nudt.edu.cn).

Yulan Guo is with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China, and also with the School of Electronics and Communication Engineering, Sun Yat-sen University, Guangdong 510275, China (e-mail: yulan.guo@nudt.edu.cn).

Digital Object Identifier 10.1109/LGRS.2021.3117054

spatial–temporal tensor with RPCA (tensor RPCA) and employ the weighted Schatten p-norm minimization (WSNM) [11] to obtain the optimal results. In summary, the contributions of this letter can be summarized as follows.

1) We introduce a tensor representation to preserve the spatial–temporal information of pixels within a satellite video.
2) We propose a tensor RPCA analysis framework with bounded noise and a generalized WSNM to separate objects from the background by estimating the low-rank components. In addition, we adopt tensor singular value decomposition (t-SVD) for efficient inference.
3) We employ the alternating direction method of multipliers (ADMM) to solve the low-rank component recovery problem in our tensor RPCA analysis framework. Extensive experiments have demonstrated the superiority of our WSNM-STTN to the state-of-the-art methods.

II. Proposed Model

A. Matrix Decomposition Model for MOD

The extended matrix decomposition model (E-LSD) [8] considered foreground detection from a viewpoint of decomposition and optimization problem, which can be defined as

$$\mathbf{D} = \mathbf{B} + \mathbf{S} + \mathbf{E}. \tag{1}$$

Here, $\mathbf{D} \in \mathbb{R}^{s\times n}$ is an observed video, where s and n represent the number of pixels in a frame and the number of frames in a sequence, respectively. $\mathbf{B} \in \mathbb{R}^{s\times n}$, $\mathbf{S} \in \mathbb{R}^{s\times n}$, and $\mathbf{E} \in \mathbb{R}^{s\times n}$ are the estimated background, foreground, and residuals, respectively.

In E-LSD, an optimization problem is defined as

$$\begin{aligned}(\mathbf{B}^*, \mathbf{S}^*, \mathbf{E}^*) = \underset{\mathbf{B},\mathbf{S},\mathbf{E}}{\arg\min}\, &||\mathbf{B}||_* + \lambda_1||\mathbf{S}||_{\ell 1/\ell\infty} + \lambda_2||\mathbf{E}||_F^2\\ \text{s.t. } &\mathbf{D} = \mathbf{B} + \mathbf{S} + \mathbf{E}\end{aligned} \tag{2}$$

where $\lambda_1 > 0$ and $\lambda_2 > 0$ are the weights of sparsity term $||\mathbf{S}||_{\ell 1/\ell\infty}$ and the residual term $||\mathbf{E}||_F^2$, respectively. $||\mathbf{B}||_*$ means the nuclear norm of matrix $\mathbf{B}$, i.e., the sum of its singular values. $||\cdot||_{\ell 1/\ell\infty}$ is a norm to induce the structural sparsity, $||\cdot||_F$ represents the Frobenius norm.

However, the matrix decomposition model cannot preserve the structural information of the input video. It also cannot make good use of the spatiotemporal correlation prior to the background and spatiotemporal continuity of the foreground. In addition, E-LSD adopts convex nuclear norm minimization (NNM) to characterize the low-rank background, while NNM treats singular values equally. As a result, the accuracy of the estimated low-rank component is reduced in highly noisy scenarios [11], [12], and the low-rank component shrinks too much, which is called the over contraction problem [11].

B. Spatial–Temporal Tensor Model for MOD

Since a satellite video has a 3-D structure, a matrix extension of RPCA to Tensor RPCA can be used to address the aforementioned problem. Furthermore, we propose a tensor RPCA analysis framework with bounded noise to preserve the structure information in a satellite video and dig out interframe correlations within a satellite video. The problem of MOD in satellite videos can be formulated as

$$\mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{N} \tag{3}$$

where $\mathcal{D}, \mathcal{B}, \mathcal{T}, \mathcal{N} \in \mathbb{R}^{n_1\times n_2\times n_3}$ represent the original patch-tensor, background tensor, target tensor, and noise tensor, respectively.

In order to recover the low-rank component more accurately and separate the object from background more perfectly, we incorporate WSNM [11] into the low-rank tensor approximation model. This is because the principle of WSNM is to assign different weights to the ℓ_p norm of singular values, which can adjust the power p to obtain a more suitable value to recover the background. The WSNM for a matrix is defined as

$$\|\mathcal{X}\|_{w,S_p} = \left(\sum_{i=1}^{\min\{n,m\}} w_i\sigma_i^p\right)^{\frac{1}{p}} \tag{4}$$

where $\mathcal{X} \in \mathbb{R}^{m\times n}$ represents the input matrix and $w = [w_1, \ldots, w_{\min\{n,m\}}]$ represents the weight values satisfying an nondescending order and the nonnegativity requirement. σ_i represents the ith singular value of $\mathcal{X} \in \mathbb{R}^{m\times n}$ and the value of power p satisfies $0 < p \le 1$. Both convex NNM and weighted NNM (WNNM) are the special cases of WSNM when $w = [1, \ldots, 1]$ and $w = [w_1, \ldots, w_{\min\{n,m\}}]$ with $p = 1$, respectively.

In our model, we generalize the definition of WSNM to tensor $\mathcal{B} \in \mathbb{R}^{n_1\times n_2\times n_3}$, that is

$$\|\mathcal{B}\|_{\mathcal{W},S_p}^p = \frac{1}{L}\sum_{i=1}^{r}\sum_{j=1}^{n_3}\left(\mathcal{W}(i,i,j)\big(\bar{\mathcal{S}}(i,i,j)\big)^p\right)^{\frac{1}{p}} \tag{5}$$

$$\mathcal{W}(i,i,j) = \frac{C\sqrt{mn}}{\bar{\mathcal{S}}(i,i,j)+\varepsilon} \tag{6}$$

where $r = \text{rank}_t(\mathcal{B})$ denotes the tensor tubal rank and $\mathcal{S}(i,i,1)$ is the entries on the diagonal of the first slice of $\mathcal{S}$ ($\mathcal{S} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is a diagonal tensor). The discrete Fourier transformation (DFT) of $\mathcal{S}$ is denoted as $\bar{\mathcal{S}} \in \mathbb{R}^{n_1\times n_2\times n_3}$. The entries on the diagonal of $\bar{\mathcal{S}}(:,:,j)$ are the singular values of $\bar{\mathcal{B}}(:,:,j)$. C is a tuning parameter, ε is a positive constant, and $\mathcal{W}$ denotes a weight tensor.

Then, the overall framework can be formulated as

$$\min_{\mathcal{B},\mathcal{T},\mathcal{N}} \|\mathcal{B}\|_{\mathcal{W},\mathcal{S}_p}^p + \lambda\|\mathcal{T}\|_1 + \beta\|\mathcal{N}\|_F^2 \quad \text{s.t. } \mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{N} \tag{7}$$

where λ and β represent the positive regularization parameters for the target and noise components, respectively.

C. Solution of the Proposed Model

To solve the proposed model, we adopt ADMM [13] and the inexact augmented Lagrangian multiplier (IALM) [14]. The problem in (7) can be rewritten by IALM as

$$\begin{aligned}L(\mathcal{B},\mathcal{T},\mathcal{N},y,\mu) = {}&\|\mathcal{B}\|_{\mathcal{W},\mathcal{S}_p}^p + \lambda\|\mathcal{T}\|_1 + \beta\|\mathcal{N}\|_F^2\\ &+\langle y, \mathcal{D}-\mathcal{B}-\mathcal{T}-\mathcal{N}\rangle\\ &+\frac{\mu}{2}\|\mathcal{D}-\mathcal{B}-\mathcal{T}-\mathcal{N}\|_F^2\end{aligned} \tag{8}$$

where $y \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ denotes the Lagrangian multiplier tensor, μ represents a penalty factor, and $\langle \cdot \rangle$ denotes the inner product operation. ADMM can decompose the problem in (8) into three optimization subproblems, including $\mathcal{B}$, $\mathcal{T}$, and $\mathcal{N}$. Since it is hard to optimize all three variables simultaneously, we approximately solve this optimization problem by alternatively minimizing one variable with the others fixed. The detailed process is given in the following.

1) Updating $\mathcal{B}$ with other variables fixed, the formulation (8) can be defined as

$$\mathcal{B}^{k+1} = \arg\min_{\mathcal{B}} \|\mathcal{B}\|^{p}_{\mathcal{W},\mathcal{S}_p} + \frac{\mu^k}{2}\left\|\mathcal{D} - \mathcal{B}^k - \mathcal{T}^k - \mathcal{N}^k + \frac{y^k}{\mu^k}\right\|_F^2. \tag{9}$$

To solve the problem in (9), we incorporate the generalized soft-thresholding (GST) method [11] into tensor singular value thresholding (t-SVT) [15], [16]. Consequently, (9) can be rewritten as

$$\mathcal{B}^{k+1} = \mathcal{D}_{\mathcal{W},\mathcal{S}_p(\mu^k)^{-1}}\left(\mathcal{D} - \mathcal{T}^k - \mathcal{N}^k + \frac{y^k}{\mu^k}\right) \tag{10}$$

where $\mathcal{D}_{\mathcal{W},\mathcal{S}_p(\mu^k)^{-1}}(\cdot)$ denotes the ADMM algorithm. It should be noticed that the weights $w = [w_1, \ldots, w_r]$ are in a nondescending order, and the singular values satisfy a nonascending order: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$.

2) Updating $\mathcal{T}$ with other variables fixed, the formulation can be defined as

$$\mathcal{T}^{k+1} = \arg\min_{\mathcal{T}} \lambda\|\mathcal{T}\|_1 + \frac{\mu^k}{2}\left\|\mathcal{D} - \mathcal{B}^{k+1} - \mathcal{T} - \mathcal{N}^k + \frac{y^k}{\mu^k}\right\|_F^2. \tag{11}$$

The problem in (11) is a typical l_1 regularized minimization problem. Therefore, we can obtain the overall optimal solution through an elementwise shrinkage operation [17]

$$\mathcal{T}^{k+1} = \mathcal{F}_{\lambda/\mu^k}\left(\mathcal{D} - \mathcal{B}^{k+1} - \mathcal{N}^k + \frac{y^k}{\mu^k}\right) \tag{12}$$

where $\mathcal{F}_{\lambda/\mu^k}(\cdot)$ represents the elementwise shrinkage operator.

3) Updating $\mathcal{N}$ with other variables fixed, the formulation can be defined as

$$\mathcal{N}^{k+1} = \arg\min_{\mathcal{N}} \beta\|\mathcal{N}\|_F^2 + \frac{\mu^k}{2}\left\|\mathcal{D} - \mathcal{B}^{k+1} - \mathcal{T}^{k+1} - \mathcal{N} + \frac{y^k}{\mu^k}\right\|_F^2. \tag{13}$$

The solution of the above problem can be obtained by

$$\mathcal{N}^{k+1} = \frac{\mu\left(\mathcal{D} - \mathcal{B}^{k+1} - \mathcal{T}^{k+1}\right) + y^k}{2\beta + \mu^k}. \tag{14}$$

4) Updating multipliers y with other variables fixed

$$y^{k+1} = y^k + \mu^k\left(\mathcal{D} - \mathcal{B}^{k+1} - \mathcal{T}^{k+1} - \mathcal{N}^{k+1}\right). \tag{15}$$

5) Updating μ^{k+1} by the following equation:

$$\mu^{k+1} = \min\left(\rho\mu^k, \mu_{\max}\right). \tag{16}$$

Finally, the proposed method is summarized in Algorithm 1.

Algorithm 1 Process of WSNM-STTN

```
Input: The image sequence $d_1, \ldots, d_P \in \mathbb{R}^{n_1 \times n_2}$,
number of frames L, tunning parameter $H$, parameters
$\lambda, \beta, p, \mu > 0$
Initialize: Transform the image sequence
$d_1, \ldots, d_P \in R^{n_1 \times n_2}$
into the tensor $\mathcal{D}, \mathcal{B}^0 = \mathcal{T}^0 = \mathcal{N}^0 = 0 \in R^{n_1 \times n_2 \times n_3}$,
$y^0 = 0$, $\mu_0 = 1e\text{-}2$, $\mu_{\max} = 1e7$, $k = 0$,
$\rho = 1.5$, $\zeta = 1e\text{-}6$, $\beta = 100$.
While : not converged do
1 : Update $\mathcal{B}^{k+1}$ according to Eq. 10.
2 : Update $\mathcal{T}^{k+1}$ according to Eq. 11.
3 : Update $\mathcal{N}^{k+1}$ according to Eq. 14.
4 : Update multipliers $y$ according to Eq. 15.
5 : Update $\mu^{k+1}$ according to Eq. 16.
6 : Check the convergence conditions
    $\frac{\|\mathcal{D}-\mathcal{B}^{k+1}-\mathcal{T}^{k+1}-\mathcal{N}^{k+1}\|_F^2}{\|\mathcal{D}\|_F^2} \leq \zeta$.
7 : Update $k = k + 1$.
end While
Output : $\mathcal{B}^{k+1}, \mathcal{T}^{k+1}, \mathcal{N}^{k+1}$.
```

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Dataset and Metrics

We evaluated the proposed WSNM-STTN on nine satellite video datasets (as listed in Table I). The first two videos (i.e., Video 001 and Video 002) were captured by SkySat.[1] Their spatial resolution is 1.0 m, while their frame rate is 30 frames per second (FPS). Videos 003–009 are provided by Chang Guang Satellite Technology Company Ltd.[2] Their spatial resolution is 1.0 m and their frame rate is 10 FPS. All these datasets mainly cover traffic scenarios of urban areas. Note that, MOD in videos 003–009 is a challenge due to the complex background. In contrast, the backgrounds of videos 001 and 002 captured by SkySat are mainly composed of roads, which is relatively easy to achieve good detection performance. In our experiments, moving cars are selected as the targets of interest.

We use three evaluation metrics, including precision, recall, and F_1 score [18], to evaluate the performance of our WSNM-STTN algorithm.

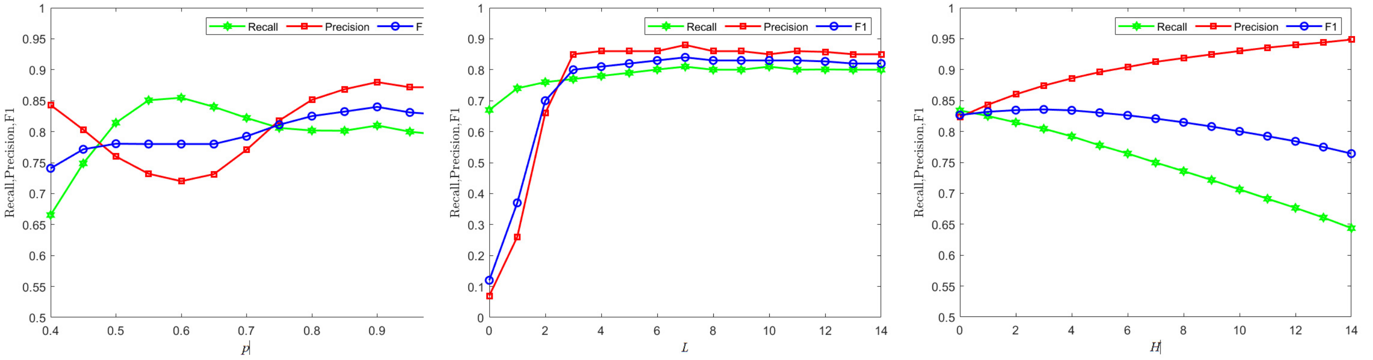
B. Parameter Setting

In the proposed WSNM-STTN algorithm, parameters are properly set to achieve good object detection performance. The regularized parameter λ in (8) represents the influence of the object tensor. λ is set to $(H/(\max(m, n) \times L)^{1/2})$, where m and n are the width and the height of the input image,

---

[1] https://www.youtube.com/watch?v=lKNAY5ELUZY/

[2] http://www.charmingglobe.com/index.aspx/



(a) (b) (c)

Fig. 1. Recall, precision, and $F1$ results achieved by our model with different values of (a) p, (b) L, and (c) H.

TABLE I
DETAILS OF NINE SATELLITE VIDEO DATASETS

Module	Video 001	Video 002	Video 003~009
Image Size	400×400	600×400	1024×1024
Vehicles	27,473	52,807	157,525
Frames	700	700	2,250

TABLE II
DETECTION PERFORMANCE ACHIEVED BY OUR MODEL AND BATCH-BASED ALGORITHMS ON SKYSAT SATELLITE VIDEOS. THE BEST RESULTS ARE SHOWN IN **RED** AND THE SECOND BEST RESULTS ARE SHOWN IN BLUE (RE: RECALL, PRE: PRECISION)

Method	Video 001			Video 002			Avg(F1)↑
	Re ↑	Pre ↑	$F1$ ↑	Re ↑	Pre↑	$F1$↑	
RPCA [19]	0.94	0.41	0.57	0.90	0.78	0.84	0.70
GoDec [4]	0.95	0.36	0.52	0.90	0.81	0.85	0.69
DECOLOR [7]	0.77	0.59	0.67	0.80	0.81	0.80	0.73
LSD [5]	0.87	0.71	0.78	0.82	0.91	0.86	0.82
E-LSD [8]	0.85	0.79	0.82	0.80	0.94	0.86	0.84
WSNM-STTN	0.94	0.76	0.84	0.95	0.80	0.87	0.86

TABLE III
DETECTION PERFORMANCE ACHIEVED BY OUR MODEL AND THE OTHER METHODS (I.E., E-LSD AND O-LSD) ON SKYSAT SATELLITE VIDEOS

Method	Video 001				Video 002				Avg(F1)↑
	Re ↑	Pre ↑	$F1$ ↑	FPS ↓	Re ↑	Pre↑	$F1$↑	FPS ↓	
O-LSD [6]	0.65	0.64	0.64	6.57s	0.73	0.90	0.81	10.75s	0.73
E-LSD [8]	0.85	0.79	0.82	17s	0.80	0.94	0.86	33s	0.84
WSNM-STTN	0.94	0.76	0.84	0.23s	0.95	0.80	0.87	0.3s	0.86

respectively, and L represents the number of input frames used to dig out the interframe information in the model. We use tuning parameter H to control λ. Fig. 1(a)–(c) shows the recall, precision, and $F1$ curves with respect to the power p, the number of frames L, and the tuning parameter H on the test datasets, respectively. Based on the tuning results, we set $p = 0.9$, $L = 8$, and $H = 4$ in the following experiments to obtain the optimal detection performance.

C. Comparison With the State-of-the-Art Methods

We conduct extensive experiments to demonstrate the robustness of our method to various scenarios in real applications:

1) SkySat Dataset: To test the effectiveness of our WSNM-STTN on Skysat satellite videos, following [6] and [8], we compare our method with five batch-based state-of-the-art approaches (i.e., RPCA [19], GoDec [4], DECOLOR [7], LSD [5], and E-LSD [8]) and one state-of-the-art online approaches (i.e., O-LSD [6]). As shown in Tables II and III, the WSNM-STTN method achieves the highest overall performance among these batch methods and online method, with an average $F1$ (Avg-$F1$) of 0.86 being achieved. This is because the tensor RPCA in our model can dig out interframe information in consecutive frames to boost the detection performance. In addition, comparing to the state-of-the-art online approaches O-LSD, WSNM-STTN achieves a comparable detection performance with significantly reduced processing time, that is, the processing time for each frame of WSNM-STTN is 30 times shorter than O-LSD. This is because the t-SVD operation in our method can speed up the inference process.

2) Jilin-1 Dataset: To test the effectiveness of our WSNM-STTN method on Jilin-1 satellite videos, we compare our method with three batch RPCA-based state-of-the-art approaches (i.e., GoDec [4], DECOLOR [7], and E-LSD [8]) and three statics modeling-based methods (i.e., MDTT [3], VIBE [2], and D&T [18]). As shown in Table IV, WSNM-STTN achieves the highest overall performance against other methods, with an average precision of 0.90 and an average $F1$ of 0.83 being reported. Compared to the matrix decomposition method E-LSD, the proposed method even improves the performance by 0.12 and 0.13 in terms of average precision and average $F1$, respectively.

In summary, the proposed WSNM-STTN model can achieve robust performance and fast processing in complex and highly heterogeneous backgrounds.

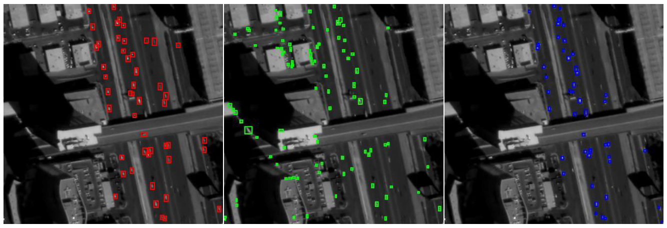
D. Ablation Study

We have demonstrated the effectiveness of introducing bounded noises $\mathcal{N}$ in (8). In this section, we also conduct ablation experiments and visualize the results of WSNM-STTN with noises (i.e., WSNM-STTN w/noises) and without noises

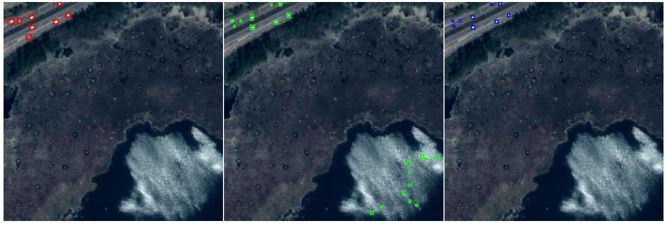
TABLE IV

QUANTITATIVE RESULTS ACHIEVED BY DIFFERENT METHODS ON JILIN-1 SATELLITE VIDEOS. THE BEST RESULTS ARE SHOWN IN **RED** AND THE SECOND BEST RESULTS ARE SHOWN IN BLUE (RE: RECALL AND PRE: PRECISION)

Method	Video 003			Video 004			Video 005			Video 006			Video 007			Video 008			Video 009			Average		
	Re↑	Pre ↑	F1 ↑	Re ↑	Pre ↑	F1 ↑	Re ↑	Pre ↑	F1 ↑	Re↑	Pre ↑	F1 ↑	Re ↑	Pre ↑	F1↑	Re ↑	Pre↑	F1 ↑	Re↑	Pre ↑	F1 ↑	Re ↑	Pre ↑	F1↑
VIBE [2]	0.61	0.34	0.44	0.82	0.61	0.70	0.68	0.59	0.63	0.65	0.52	0.58	0.72	0.65	0.69	0.60	0.42	0.49	0.45	0.44	0.44	0.65	0.51	0.57
GoDec [4]	0.92	0.51	0.65	0.73	0.81	0.77	0.93	0.53	0.68	0.72	0.38	0.50	0.72	0.74	0.73	0.81	0.42	0.55	0.93	0.25	0.39	0.82	0.52	0.61
DECOLOR [7]	0.24	0.92	0.38	0.77	0.88	0.82	0.89	0.83	0.86	0.44	0.93	0.60	0.74	0.84	0.79	0.71	0.80	0.75	0.30	0.69	0.42	0.58	0.84	0.66
MTTP [3]	0.74	0.67	0.70	0.67	0.84	0.74	0.71	0.84	0.77	0.64	0.86	0.73	0.62	0.77	0.69	0.55	0.73	0.62	0.25	0.49	0.33	0.60	0.74	0.65
D&T [18]	0.71	0.91	0.80	0.69	0.86	0.76	0.84	0.84	0.84	0.75	0.85	0.80	0.63	0.82	0.71	0.64	0.76	0.70	0.83	0.43	0.56	0.73	0.78	0.74
E-LSD [8]	0.71	0.83	0.77	0.75	0.88	0.81	0.64	0.67	0.65	0.61	0.86	0.72	0.57	0.92	0.70	0.55	0.82	0.66	0.58	0.61	0.60	0.63	0.80	0.70
WSNM-STTN	0.79	0.93	0.85	0.79	0.91	0.84	0.92	0.93	0.92	0.63	0.92	0.75	0.75	0.89	0.81	0.76	0.89	0.82	0.82	0.82	0.82	0.78	0.90	0.83



Video 001 (Groundtruth) WSNM-STTN w/o noises WSNM-STTN w/ noises



Video 008 (Groundtruth) WSNM-STTN w/o noises WSNM-STTN w/ noises

Fig. 2. Demonstration on the importance of N in WSNM-STTN.

(i.e., WSNM-STTN w/o noise) in Fig. 2. It can be observed that our WSNM-STTN method achieves a high detection rate and low false alarms rate by introducing bounded noises.

IV. CONCLUSION

In this letter, we propose a WSNM-STTN model to detect dim and small moving objects in satellite video. With the STTN model, the proposed model can dig out temporal information within a sequence. Besides, we propose an extended tensor RPCA with bounded noise and incorporate WSNM to solve the overshrink problem in low-rank estimation, which is superior to noiseless modeling methods. Then, we optimize our model by ADMM to detect objects. Extensive experiments show that WSNM-STTN can achieve a high detection rate and a low false alarms rate under complex background with heavy noise. In addition, WSNM-STTN converges faster than the matrix decomposition approach by a large margin.

REFERENCES

[1] J. Shao, B. Du, C. Wu, and L. Zhang, "Can we track targets from space? A hybrid kernel correlation filter tracker for satellite video," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8719–8731, Aug. 2019.

[2] O. Barnich and M. Van Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2010.

[3] S. A. Ahmadi, A. Ghorbanian, and A. Mohammadzadeh, "Moving vehicle detection, tracking and traffic parameter estimation from a satellite video: A perspective on a smarter city," *Int. J. Remote Sens.*, vol. 40, no. 22, pp. 8379–8394, Nov. 2019.

[4] T. Zhou and D. Tao, "Godec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 33–40.

[5] X. Liu, G. Zhao, J. Yao, and C. Qi, "Background subtraction based on low-rank and structured sparse decomposition," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2502–2514, Aug. 2015.

[6] R. Pflugfelder, A. Weissenfeld, and J. Wagner, "On learning vehicle detection in satellite video," 2020, *arXiv:2001.10900*. [Online]. Available: http://arxiv.org/abs/2001.10900

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

[8] J. Zhang, X. Jia, and J. Hu, "Error bounded foreground and background modeling for moving object detection in satellite videos," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2659–2669, Apr. 2020.

[9] M. Zhao, L. Li, W. Li, R. Tao, L. Li, and W. Zhang, "Infrared small-target detection based on multiple morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 6077–6091, Jul. 2021.

[10] M. Zhao, W. Li, L. Li, P. Ma, Z. Cai, and R. Tao, "Three-order tensor creation and Tucker decomposition for infrared small-target detection," *IEEE Trans. Geosci. Remote Sens.*, early access, Feb. 18, 2021, doi: 10.1109/TGRS.2021.3057696.

[11] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, and L. Zhang, "Weighted schatten p-norm minimization for image denoising and background subtraction," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4842–4857, Oct. 2016.

[12] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

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

[15] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, Jan. 2020.

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

[17] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[18] W. Ao, Y. Fu, X. Hou, and F. Xu, "Needles in a Haystack: Tracking city-scale moving vehicles from continuously moving satellite," *IEEE Trans. Image Process.*, vol. 29, no. 7, pp. 1944–1957, Oct. 2019.

[19] J. Wright, A. Ganesh, S. Rao, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," Coordinated Sci. Lab., Urbana, IL, USA, Tech. Rep. UILU-ENG-09-2210, 2009.