LOW-RANK MATRIX DECOMPOSITION WITH SUPERPIXEL-BASED STRUCTURED SPARSE REGULARIZATION FOR MOVING OBJECT DETECTION IN SATELLITE VIDEOS

Junpeng Zhang, Xiuping Jia*, Jiankun Hu

School of Engineering and Information Technology, University of New South Wales, Canberra 2610, Australia

Commission III, WG II/III

KEY WORDS: Satellite Video, Moving Object Detection, Low-rank Matrix Decomposition, Background Subtraction, Structured Sparsity Inducing Norm

ABSTRACT:

With new accessibility to satellite videos, retrieving the dynamic information of moving objects over a vast territory becomes possible with the development of advanced video processing and machine learning techniques. Detecting moving objects can be based on the structures of both background and foreground of a satellite video, and the background is assumed to lay in a low dimensional subspace. As the moving objects in satellite videos are groups of neighbouring pixels other than isolated pixels, Low-rank and Structured Sparse Decomposition (LSD) with structured sparsity regularization on the foreground can suppress the false alarms caused by isolated outliers. However, in LSD, the groups of neighbouring pixels are extracted by a fixed sliding window over each video frame, which ignores the coherence on the appearance of a moving object. For example, a moving object can be in an irregular shape and arbitrary orientation. In this paper, we argue that the spatial groups on the foreground can be defined using the concept of superpixels, where each superpixel is formed by a group of spatially connected similar pixels obtained from over-segmentation. We conduct low-rank matrix decomposition at superpixel level, which is named as Superpixel-based LSD (S-LSD). To handle the variation in moving objects, we combine the superpixels at a range of scales in the superpixel-based spatial regularization on the foreground. With the reduction in the number of spatial groups, S-LSD presents reduced computation complexity. The results on two satellite videos show a satisfactory performance with a significant saving in processing time when the proposed S-LSD approach is applied.

1. INTRODUCTION

Recently, the cube satellites Jilin-1 (Luo et al., 2017) and SkySat (Team, 2016) can produce satellite videos over a large territory. Unlike previous still images with low revisiting frequency, a satellite video is a sequence of 2-D spatial frames captured by the satellite with a high frame rate. The abundant temporal information in these videos is helpful for retrieve motion information on objects of interest over a larger territory, which facilitates a wide range of applications including target tracking (Mou, Zhu; Du et al., 2018; Zhang et al., 2018; Uzkent et al., 2018) and traffic monitoring (Kopsiaftis, Karantzalos). Detecting moving objects from satellite videos plays a vital role in these applications. Contemporary object detectors achieve state-of-the-art detection performance by learning a image-based detector from manually annotated training images (Long et al., 2017; Li et al., 2017; Ding et al., 2018; Liu et al., 2018). However, in satellite videos, the applicability of these approaches is limited by the accessibility to the sufficient annotations for training such over-parameterized models. Alternatively, unsupervised methods for Moving Object Detection (MOD) can separate moving objects from the background scene by making use of the temporal information.

The canonical approaches for MOD assume each frame in a video is constructed by a foreground and a background. The background part of a frame is considered temporally stable and similar, while the temporally changing foreground part contains the moving objects. Based on this assumption, the dominating set of MOD approaches are based on the low-rank matrix decomposition, where the background data lay in a low dimensional subspace and the moving objects in the foreground are considered as the sparse outliers (Bouwmans, Zahzah; Bouwmans et al., 2017, 2018). Robust Principle Component Analysis (RPCA), as a fundamental method in this set, imposes pixel-wise sparsity regularization term on the foreground in the low-rank matrix decomposition problem (Candès et al., 2011), whose solution can be obtained by Principle Component Pursuit (RPCA-PCP) (Lin et al., 2011; Candès et al., 2011; Wright et al., 2009) and Fast Low Rank Approximation (GoDec) (Zhou, Tao). However, RPCA is prone to the false alarms caused by the isolated outliers in satellite videos.

To suppress these false alarms, the spatial regularization terms are imposed on the foreground in low-rank matrix decomposition. Total Variation (TV) regularization is deployed to enforce the smoothness on the foreground in the matrix decomposition (Xu et al., 2017). The first-order Markov Random Field (MRF) is also integrated into low-rank matrix decomposition to constrain the moving objects to be contiguous (Zhou et al., 2013; Shakeri, Zhang). In satellite videos where spatial resolution is low and color information is limited, these approaches have limited improvement in MOD performance, as they risk merging neighbouring targets.

Another set of spatial prior on the foreground is defined on the sparsity over groups of spatial neighboring pixels other than independent pixels. The structured sparsity-inducing norm (Jenatton et al., 2011) is then introduced to regularize the foreground (Liu et al., 2015; Xu et al., 2013; Zhang et al., 2019a). In

* Corresponding author

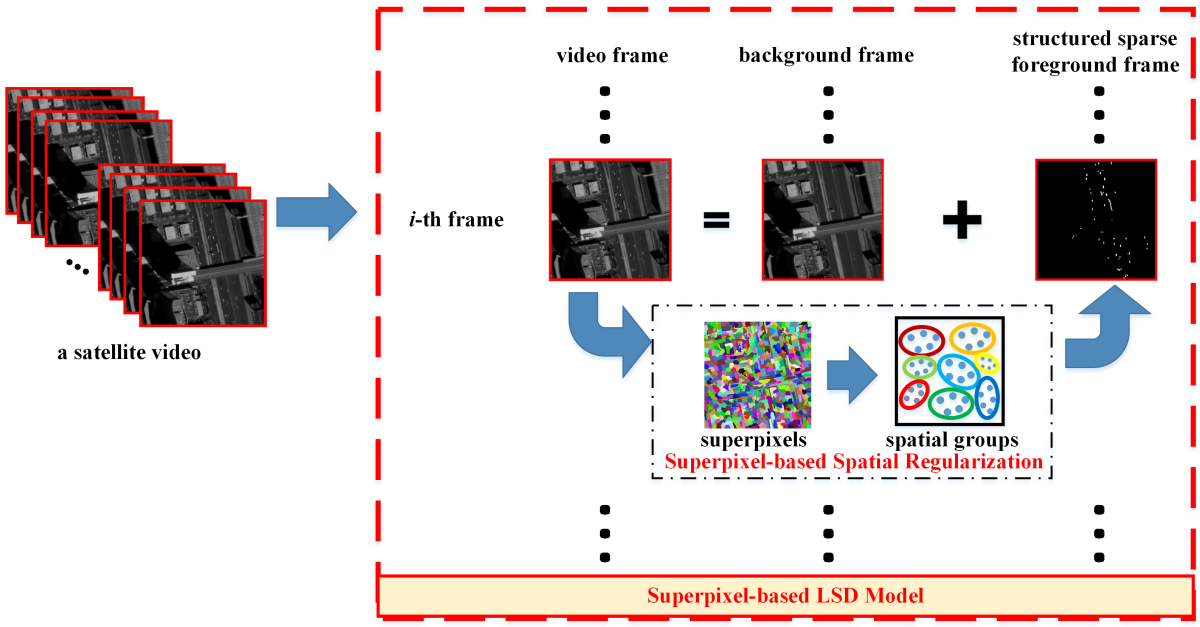


Figure 1. Illustration of S-LSD for moving object detection in satellite videos.

satellite videos, an Extended Low-rank and Structured Sparse Matrix Decomposition (E-LSD) model is proposed for boosting the MOD performance by imposing structured sparse regularization on the foreground (Zhang et al., 2019a,b). However, the spatial groups of neighbouring pixels in these approaches are extracted by a fixed sliding window over each video frame, which ignores the irregular shapes and arbitrary orientations of a moving object. Another disadvantage of using sliding window approach is that many spatial windows are processed unnecessarily, which leads to increased processing time.

In this paper, we argue that the spatial groups in the spatial regularization can be constructed from superpixels, where each superpixel is formed by a group of spatially connected similar pixels obtained from over-segmentation. In satellite videos, it is reasonable that we assume a moving object is commonly composed of one or more coherent regions and each of them can be extracted by over-segmentation. Inspired by this observation, we propose to conduct low-rank and structured sparse matrix decomposition with spatial groups defined by superpixels, which is named as **S**uperpixel-based **L**ow-rank and **S**tructured Sparse Matrix **D**ecomposition (S-LSD) in this paper. To handle the moving objects in various sizes, we also combine spatial groups from multiple sets of superpixels at a range of scales. In S-LSD, the number of spatial groups is less than it in LSD or E-LSD, which helps reduce the computation complexity of S-LSD in practice. We compared the proposed S-LSD with the state-of-the-art algorithms on two satellite videos and the experimental results validate the significant reduced processing time by S-LSD with satisfactory MOD performance.

The remainder of this paper is organized as follows. The proposed S-LSD is presented in Section 2. The experimental results and performance comparison against state-of-the-art approaches are presented in Section 3. Finally, conclusions is given in Section 4.

2. PROPOSED METHOD

2.1 Problem Formulation

The proposed Superpixel-based LSD (S-LSD) is defined as a low-rank matrix decomposition problem, where the rank of the background is minimized. In order to suppress false alarms caused by isolated outliers in the foreground, S-LSD imposes a superpixel-based structured sparse regularization on the foreground.

Given a sequence of n video frames and each frame contains p pixels, S-LSD decomposes its corresponding matrix $\mathbf{D} \in \mathbb{R}^{p\times n}$ to a low-rank background matrix $\mathbf{B} \in \mathbb{R}^{p\times n}$ and a structured sparse foreground matrix $\mathbf{S} \in \mathbb{R}^{p\times n}$. The optimization problem of S-LSD is formulated as

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

where $\Omega(\mathbf{S})$ refers to the spatial regularization term on the foreground $\mathbf{S}$, and $\mathbf{E}$ is introduced to handle the noise in the model. λ_1 and λ_2 are the weights assigned to the spatial regularization term and the noise term, respectively.

In this paper, we assume the moving objects are sparse groups of neighboring non-zero pixels in the foreground, and the structured sparsity-inducing norm (Jenatton et al., 2011, 2010; Jia et al., 2012) is adopted to regularize the foreground as

$$\Omega(\mathbf{S}) = \sum_{\mathbf{s}\in\mathbf{S}} \|\mathbf{s}\|_{\ell_1/\ell_\infty} = \sum_{\mathbf{s}\in\mathbf{S}} \sum_{g\in\mathcal{G}(\mathbf{s})} \eta_g \left\|\mathbf{s}_{|g}\right\|_\infty, \tag{2}$$

where $\mathcal{G}(\mathbf{s})$ refers to the set of spatial groups of neighboring pixels of the foreground $\mathbf{s}$, and $\mathbf{s}_{|g} \in \mathbb{R}^p$ is a sparse vector with non-zero elements at the indices represented in a group $g \in \mathcal{G}$. For each group of pixels $g \in \mathcal{G}(\mathbf{s})$, η_g is the weight for a group of the pixels. Applying the structured sparsity-inducing norm on the foreground data as the spatial regularization term tends to assign zeros to the pixels in a group, thus the isolated outliers

on the foreground are suppressed. In S-LSD, no temporal relationship on the foreground is defined in Equation. (1), so the structured sparse penalty is frame-wise independent.

2.2 Superpixel-based Structured Sparse Regularization

In the spatial regularization term $\Omega(\mathbf{S})$, the groups of neighboring pixels $\mathcal{G}(\mathbf{s})$ are commonly constructed by the patches extracted by a fixed sliding windows over each frame (Liu et al., 2015; Xu et al., 2013; Zhang et al., 2019a,b), which leads to the increased number of generated spatial groups and hurts the efficiency in solving Equation. (1). In this paper, we propose to build the spatial groups $\mathcal{G}(\mathbf{s})$ from the superpixels extracted by over-segmentation. With the superpixels extracted from a frame $\mathbf{s}$ at a given scale, a spatial group is constructed by the pixels in a superpixel. In order to handle the variation in moving objects, $\mathcal{G}(\mathbf{s})$ combines the superpixels extracted at a range of scales.

Let $\mathcal{M} = \{m_1, m_2, \cdots, m_{|\mathcal{M}|}\}$, $m_1 < m_2 < \cdots < m_{|\mathcal{M}|}$, note a selected set of superpixel scales, and $\mathcal{G}_m(\mathbf{s})$ is referred to the groups constructed from the superpixels extracted at a given scale $m \in \mathcal{M}$. Given a set of scales $\mathcal{M}$, we define the entire set of spatial groups as

$$\mathcal{G}(\mathbf{s}) = \mathcal{G}_{m_1}(\mathbf{s}) \cup \mathcal{G}_{m_2}(\mathbf{s}) \cup \cdots \cup \mathcal{G}_{m_{|\mathcal{M}|}}(\mathbf{s}). \tag{3}$$

With the spatial groups defined above, the proposed Superpixel-based Structured Sparse Regularization on a foreground frame $\mathbf{s}$ is defined as

$$\|\mathbf{s}\|_{\ell_1/\ell_\infty} = \sum_{m\in\mathcal{M}} \sum_{g\in\mathcal{G}_m(\mathbf{s})} \eta_g \left\|\mathbf{s}_{|g}\right\|_\infty, \tag{4}$$

where η_g is the weight for a group of neighboring pixels. In this paper, we assign different weights to different superpixels. For spatial groups constructed from the superpixels at coarse scales, the small objects in the foreground would be suppressed, as the pixels in such groups may be forced to be zero. Based on this understanding, the weights for spatial groups at larger scales are decreased,

$$\eta_g = \frac{|\mathcal{G}_m(\mathbf{s})|}{|\mathcal{G}_{m_1}(\mathbf{s})|}\eta_0, \forall g \in \mathcal{G}_m(\mathbf{s}) \text{ and } m \in \mathcal{M}, \tag{5}$$

where $|\mathcal{G}_m(\mathbf{s})|$ is the number of spatial groups in $\mathcal{G}_m(\mathbf{s})$, and η_0 is the initial weight for the spatial regularization term. Since we always have $|\mathcal{G}_{m_1}| > |\mathcal{G}_{m_2}| > \cdots > |\mathcal{G}_{m_{|\mathcal{M}|}}|$, the weights of spatial groups at large scales are always less than η_0. The same weight is assigned to the spatial groups constructed from the superpixels of the same scale. For simplicity, we assign 1.0 to the initial weight η_0. In this paper, the superpixels are extracted by the SEEDS approach (Van den Bergh et al., 2015), where the initial size of a superpixel in SEEDS corresponds to the scale m in the Superpixel-based Structured Sparse Regularization.

2.3 Solution to S-LSD

To make the problem in Equation. (1) more tractable, the nuclear norm $\|\mathbf{B}\|_*$, which is the convex relation of the $\mathrm{Rank}(\mathbf{B})$, is utilized to replace the rank minimization, and the linear constraint is removed by the Augmented Lagrangian method. Then

Algorithm 1 The Alternating Direction Method of Multipliers (ADMM) for S-LSD

Input: $\mathbf{D} \in \mathbb{R}^{p*n}$, $\lambda_1 > 0$, $\lambda_2 > 0$, $\mu > 1.0$, $\bar{\mu} = \mu \times 1.0\mathrm{e}5$, $\rho > 1.0$ and $\tau = 1.0\mathrm{e}{-7}$
Output: $\mathbf{B}$, $\mathbf{S}$ and $\mathbf{E}$
1: $\mathbf{B}^0 = \mathbf{0}$, $\mathbf{S}^0 = \mathbf{0}$, $\mathbf{E}^0 = \mathbf{0}$ and $k = 0$
2: **while** $\frac{\|\mathbf{D}-\mathbf{B}^k-\mathbf{S}^k-\mathbf{E}^k\|_F}{\|\mathbf{D}\|_F} \leq \tau$ **do**
3: Update $\mathbf{B}^{k+1}$ by solving

$$\mathbf{B}^{k+1} = \arg\min_{\mathbf{B}} \frac{1}{\mu}\|\mathbf{B}\|_* + \frac{1}{2}\left\|(\mathbf{D}-\mathbf{S}^k-\mathbf{E}^k+\frac{1}{\mu}\mathbf{Y}^k)-\mathbf{B}\right\|_F^2. \tag{7}$$

4: Update $\mathbf{S}^{k+1}$ by solving

$$\mathbf{S}^{k+1} = \arg\min_{\mathbf{S}} \frac{\lambda_1}{\mu}\Omega(\mathbf{S}) + \frac{1}{2}\left\|(\mathbf{D}-\mathbf{B}^{k+1}-\mathbf{E}^k+\frac{1}{\mu}\mathbf{Y}^k)-\mathbf{S}\right\|_F^2. \tag{8}$$

5: $\mathbf{E}^{k+1} = \frac{\mu}{2\lambda_2+\mu}(\mathbf{D}-\mathbf{B}^{k+1}-\mathbf{S}^{k+1}+\frac{1}{\mu}\mathbf{Y}^k)$.
6: $\mathbf{Y}^{k+1} = \mathbf{Y}^k + \mu(\mathbf{D}-\mathbf{B}^{k+1}-\mathbf{S}^{k+1}-\mathbf{E}^{k+1})$.
7: $\mu = \min\{\rho\mu, \bar{\mu}\}$, $k = k+1$.
8: **end while**
9: **return** $\mathbf{B}^{k+1}$, $\mathbf{S}^{k+1}$ and $\mathbf{E}^{k+1}$

we obtain the reformulated optimization problem as

$$\begin{aligned}(\mathbf{B}^*, \mathbf{S}^*, \mathbf{E}^*) = \arg\min_{\mathbf{B},\mathbf{S}} &\|\mathbf{B}\|_* + \lambda_1\|\mathbf{S}\|_{\ell_1/\ell_\infty} \\ &+\lambda_2\|\mathbf{E}\|_F^2 + \langle\mathbf{Y}, \mathbf{D}-\mathbf{B}-\mathbf{S}-\mathbf{E}\rangle \\ &+\frac{\mu}{2}\|\mathbf{D}-\mathbf{B}-\mathbf{S}-\mathbf{E}\|_F^2,\end{aligned} \tag{6}$$

where $\mathbf{Y} \in \mathbb{R}^{p*n}$ is the Lagrangian multiplier and $\mu > 0$ is a positive scalar. This augmented problem can be solved by alternating direction method of multipliers (ADMM) (Boyd et al., 2011) or Block Coordinate Descent (BCD) (Wright, 2015). In this paper, we adopt the ADMM approach for solving this augmented problem in Equation. (6), since the sufficient guarantee on the convergence has been provided for the set of optimization problems that minimizes the sum of three functions with uncoupled variables under a three-block linear constraint (Cai et al., 2017; Zhang et al., 2019a).

As summarized in Algorithm 1, the procedure for solving Equation. (6) is to alternatingly solve sub-problems with respect to $\mathbf{B}$, $\mathbf{S}$ and $\mathbf{E}$ with two remaining variables fixed until it is converged.

2.3.1 Update B At each iteration, $\mathbf{B}^{k+1}$ is updated by the Singular Value Thresholding approach (Wright et al., 2009; Cai et al., 2010). Let $\mathbf{G} = \mathbf{D}-\mathbf{S}^k-\mathbf{E}^k+\frac{1}{\mu}\mathbf{Y}^k$, and $\mathbf{U}\Sigma\mathbf{V}^T = \mathbf{G}$ is the Singular Value Decomposition (SVD) of $\mathbf{G}$. Then B is updated by

$$\mathbf{B}^{k+1} = \mathbf{U}\mathcal{S}_{\frac{1}{\mu}}(\Sigma)\mathbf{V}^T, \tag{9}$$

where $\mathcal{S}_{\frac{1}{\mu}}(\Sigma)$ conducts the element-wise soft-shrinkage on the diagnose matrix Σ by

$$\mathcal{S}_{\frac{1}{\mu}}(\Sigma) = \max(\Sigma - \frac{1}{\mu}, \mathbf{0}), \Sigma \succ 0, \mu > 0. \tag{10}$$

Table 1. Information on the evaluation datasets

Video	Frame Size	Cross Validation		Performance Evaluation	
		#Frames	#Vehicles	#Frames	#Vehicles
001	400 × 400	200	9306	500	18167
002	600 × 400	200	13443	500	39362

2.3.2 Update S The sub-problem defined in Equation. (8) is decomposed to a set of frame-wise optimization problems, since the spatial regularization term on the foreground is frame-wise independent.

Given a frame $\mathbf{d} = \mathbf{D}_i$, $\forall i \in \{1, \cdots, n\}$, the decomposed optimization problem for the foreground frame $\mathbf{s} = \mathbf{S}_i^{k+1}$ is rewritten as

$$\arg\min_{\mathbf{s}} \frac{1}{2} \|\mathbf{h} - \mathbf{s}\|_2^2 + \lambda' \sum_{g \in \mathcal{G}} \eta_g \left\|\mathbf{s}_{|g}\right\|_\infty , \tag{11}$$

where $\mathbf{h} = \mathbf{d} - \mathbf{B}_i^{k+1} - \mathbf{E}_i^{k+1} + \frac{1}{\mu}\mathbf{Y}_i^{k+1}$, and $\lambda' = \frac{\lambda_1}{\mu}$. When the spatial groups of pixels in $\mathcal{G}$ are non-overlapped, the problem in Equation. (11) can be solved by the Group-LASSO method (Yuan, Lin). However, as $\mathcal{G}$ combines the spatial groups at different scales, overlapped groups of variables are observed, and Equation. (11) cannot be solved directly. Instead, the solution is obtained by its dual problem as a Quadratic Min-cost Network Flow problem,

$$\begin{aligned} \xi^* = \arg\min_{\xi} & \frac{1}{2} \left\| \mathbf{h} - \sum_{g \in \mathcal{G}} \xi^g \right\|_2^2 \\ & \text{s. t. } \mathbf{h} - \mathbf{s} + \sum_{g \in \mathcal{G}} \xi^g = 0, \\ & \forall g \in \mathcal{G}, \|\xi^g\|_1 \leq \lambda' \eta_g \text{ and } \xi_j^g = 0 \text{ if } j \notin g, \end{aligned} \tag{12}$$

where $\xi \in \mathbb{R}^{p \times |\mathcal{G}|}$ is the dual variable. This Quadratic Min-cost Network flow problem is defined and solved in (Mairal et al., 2010). After solving the dual problem, the foreground $\mathbf{S}_i = \mathbf{s}$ is obtained by

$$\mathbf{s} = \mathbf{h} - \sum_{g \in \mathcal{G}} \xi^{*g}, \tag{13}$$

in which ξ^* refers to the optimal solution to Equation. (12).

2.4 Computation Complexity

For processing a video in the length of n, the computation complexity for solving Equation. (1) is related to the number of spatial groups in $\mathcal{G}$, $\mathcal{O}(n(p^2 + \sum_{g \in \mathcal{G}} |g|))$. Compared with constructing $\mathcal{G}$ by sliding window, the superpixel-based spatial regularization usually has a reduced number of spatial groups. In case that the processing time is critical, S-LSD with spatial groups constructed at a single proper scale may achieve both reduced processing time and moderate MOD performance at the same time. Combining spatial groups of multiple scales in S-LSD may improve the MOD performance by handling moving objects in different sizes, which, on the contrary, may increase the processing time. In practice, by reducing the number of spatial groups, S-LSD can help reduce the time consumption for processing large satellite videos with satisfactory performance.

3. EXPERIMENTS

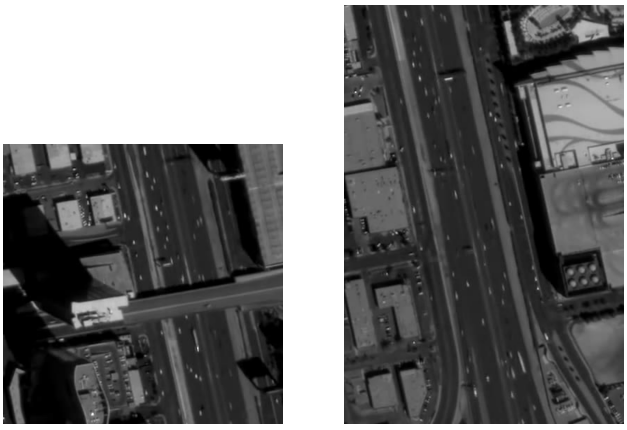
3.1 Dataset

The detection performance of S-LSD was evaluated on two satellite videos. They were captured over Las Vagas, USA on March 25, 2014, whose spatial resolution is 1.0 meter and the frame rate is 30 frames per second. Both videos contain 700 frames with boundary boxes for moving vehicles as groundtruth, and details on both videos are listed in Table. 1 [1].

The MOD performance on these videos is evaluated on recall, precision and F_1 scores given by

$$\begin{aligned} \text{recall} &= TP/(TP + FN) \\ \text{precision} &= TP/(TP + FP) \\ F_1 &= \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \end{aligned}, \tag{14}$$

where TP denotes the number of correct detections, FN and FP are the numbers of missed detections and false alarms, respectively. In this paper, we define a correct detection with maximum Intersection over Union (IoU) against the groundtruth greater than a threshold. To accommodate the vehicles in small size in satellite videos, the threshold is set as 0.3 [3].



(a) Video 001 (b) Video 002

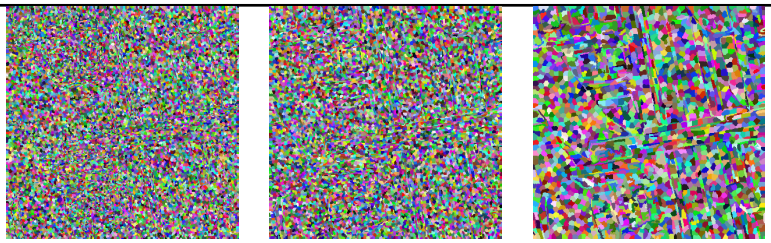
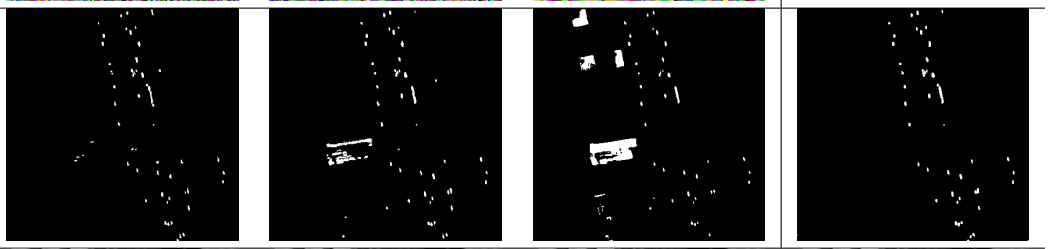
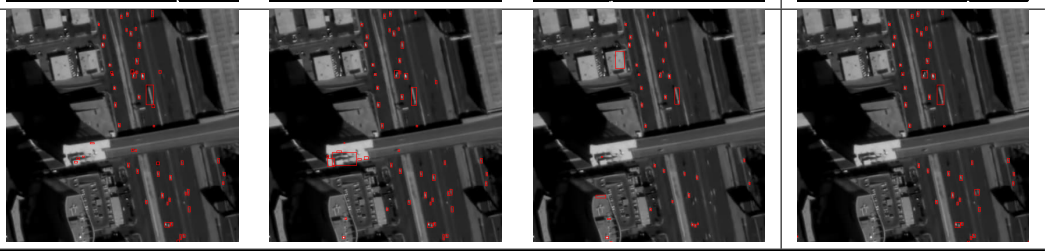
Figure 3. Exemplar frames from Video 001 and Video 002.

3.2 Selecting Proper Scales for Spatial Regularization

The scale of the spatial groups in the superpixel-based spatial regularization term plays an important role in S-LSD. We first

[1] Moving vehicles are manually labelled by the Computer Vision Annotation Tool (CVAT), and a boundary box is provided for each moving object on each frame.

[3] The estimated foreground is built by contiguous values rather than binary value, so we adopt the threshold segmentation as post-processing for extracting the foreground mask and the moving objects (Gao et al., 2012).

Method	S-LSD			**E-LSD**
Superpixels				-
Foreground mask				
Detection				

Method	LSD			**E-LSD**
	$\mathcal{M}=\{4\}$	$\mathcal{M}=\{16\}$	$\mathcal{M}=\{64\}$	
$\lvert\mathcal{G}\rvert$	15616	9268	2486	39204
Recall ↑	90.3%	87.6%	71.7%	81.5%
Precision ↑	77.7%	72.8%	74.9%	85.9%
F_1 ↑	83.54%	79.52%	73.22%	83.64%
Time (s) ↓	369.83	350.04	351.57	1054.85

Figure 2. Demonstration on the effects of different scales in spatial regularization in S-LSD.

evaluate the performance of S-LSD with single scale spatial regularization. When selecting a small scale, a considerable number of spatial groups will be constructed in $\mathcal{G}$, which thus increases the processing time. However, when selecting an enlarged scale, the small moving objects in the foreground may be suppressed, as pixels in a spatial group tends to be zero together. As presented in Figure. 2, when the scale of the spatial group increases from 4 to 64, the recall rate of the moving object drops from 90.3% to 71.7%. At the same time, as less spatial groups are constructed in $\mathcal{G}$ for larger scales, the processing time is reduced.

$\mathcal{M}$	Recall	Precision	F_1	Time (s)
$\{4\}$	90.3%	77.7%	83.54%	369.83
$\{4, 16\}$	84.7%	86.2%	86.48%	800.10
$\{4, 16, 64\}$	83.6%	87.4%	85.42%	942.62

Table 2. Information on the evaluation datasets

In this paper, we combine spatial groups of different scales to handle moving objects in different sizes. As shown in Table. 2, when combining two scales of spatial groups in the regularization term, $\mathcal{M}=\{4, 16\}$, the MOD performance is improved. When more scales are introduced, $\mathcal{M}=\{4, 16, 64\}$, the MOD performance drops a bit, as some small moving objects may be improperly suppressed by the large spatial groups at the scale of 64. A moderate number of scales is recommended.

In the following experiments, we select S-LSD with a single scale with $\mathcal{M}=\{4\}$, and, for S-LSD with multiple-scale spatial regularization, we set $\mathcal{M}=\{4, 16\}$. The weights λ_1 and λ_2 are selected by cross validation, and further fine-tunes on $\mathcal{M}$, λ_1 and λ_2 may improve the MOD performance by S-LSD more.

3.3 Comparison with Other Methods

To verify the effectiveness of S-LSD, we compare the detection performance against three state-of-the-art approaches, which are RPCA (Candès et al., 2011), LSD (Liu et al., 2015) and E-LSD (Zhang et al., 2019a). RPCA is a low-rank matrix decomposition method without spatial constraints on the foreground,and is solved by Principal Component Pursuit. LSD and E-LSD both impose the structured sparse regularization on the foreground, where the spatial groups are constructed by a sliding window.

As presented in Table. 3 and Figure. 4, S-LSD with the single scale spatial regularization $\mathcal{M}=\{4\}$ achieves comparable performance in term of recall with significantly reduced processing time. Compared with the RPCA, the superpixel-based structured sparse regularization in S-LSD helps reduce in false alarms due to noises in the data, which leads to improved detection precision. Compared with LSD and E-LSD where the structured sparse regularization is based on a sliding window, S-LSD reduces the processing time with comparable MOD performance. The reduction in time consumption by S-LSD shows it is more applicable for the applications where processing time is critical. S-LSD improves the detection precision to the extend

that the remaining false alarms are caused by other factors. As presented in Figure. 4, moving objects are mistakenly recognized on the top of buildings in left bottom part of the video. Theses false alarms should be owing to the motion of the satellite in capturing the videos, and suppressing these false alarms is beyond the topic of paper.

When multiple-scale spatial regularization is imposed, S-LSD improves the detection precision with moderate increase in time consumption, as shown in Table. 3. Compared with LSD and E-LSD, S-LSD ($\mathcal{M} = \{4, 16\}$) achieves the highest precision on both videos with lower time costs. As Video 002 contains fewer large moving vehicles, applying spatial regularization with larger scale may suppress small moving vehicles, which leads to the small drop in the recall rate by S-LSD. The difference scales are selected for different videos because the performance of S-LSD is related to the over-segmentation performance, which is affected by the complexity of video as well as the object size.

4. CONCLUSION

In this paper, we propose a **S**uperpixel-based **L**ow-rank and **S**tructured Sparse **D**ecomposition (S-LSD) algorithm for moving object detection, where superpixel-based structured sparse regularization is imposed on the foreground. We show that S-LSD with single-scale spatial regularization reduces the time consumption greatly with moderate detection performance, which makes it more applicable for processing large satellite videos. S-LSD with multiple-scale spatial regularization offers good detection performance, which is more suitable for application with high requirement for precision. With improved over-segmentation approaches for satellite videos, the MOD performance of S-LSD would be further improved.

ACKNOWLEDGEMENTS

This work is partially supported by China Scholarship Council. The authors would like to thank Planet Team for providing the data in this research (Team, 2016).

References

Bouwmans, T., Javed, S., Zhang, H., Lin, Z., Otazo, R., 2018. On the applications of robust PCA in image and video processing. *Proceedings of the IEEE*, 106(8), 1427–1457.

Bouwmans, T., Sobral, A., Javed, S., Jung, S. K., Zahzah, E.-H., 2017. Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset. *Computer Science Review*, 23, 1–71.

Bouwmans, T., Zahzah, E. H., 2014. Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122, 22–34.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J. et al., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1), 1–122.

Cai, J.-F., Candès, E. J., Shen, Z., 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4), 1956–1982.

Cai, X., Han, D., Yuan, X., 2017. On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function. *Computational Optimization and Applications*, 66(1), 39–73.

Candès, E. J., Li, X., Ma, Y., Wright, J., 2011. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3), 11.

Ding, P., Zhang, Y., Deng, W.-J., Jia, P., Kuijper, A., 2018. A light and faster regional convolutional neural network for object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 141, 208–218.

Du, B., Sun, Y., Cai, S., Wu, C., Du, Q., 2018. Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm. *IEEE Geoscience and Remote Sensing Letters*, 15(2), 168–172.

Gao, Z., Cheong, L.-F., Shan, M., 2012. Block-sparse rpca for consistent foreground detection. *European Conference on Computer Vision*, Springer, 690–703.

Jenatton, R., Audibert, J.-Y., Bach, F., 2011. Structured variable selection with sparsity-inducing norms. *Journal of Machine Learning Research*, 12(Oct), 2777–2824.

Jenatton, R., Mairal, J., Obozinski, G., Bach, F. R., 2010. Proximal methods for sparse hierarchical dictionary learning. *ICML*, 1, Citeseer, 2.

Jia, K., Chan, T.-H., Ma, Y., 2012. Robust and practical face recognition via structured sparsity. *European conference on computer vision*, Springer, 331–344.

Kopsiaftis, G., Karantzalos, K., 2015. Vehicle detection and traffic density monitoring from very high resolution satellite video data. *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 1881–1884.

Li, K., Cheng, G., Bu, S., You, X., 2017. Rotation-insensitive and context-augmented object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 56(4), 2337–2348.

Lin, Z., Liu, R., Su, Z., 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. *Advances in neural information processing systems*, 612–620.

Liu, W., Ma, L., Chen, H., 2018. Arbitrary-oriented ship detection framework in optical remote-sensing images. *IEEE Geoscience and Remote Sensing Letters*, 15(6), 937–941.

Liu, X., Zhao, G., Yao, J., Qi, C., 2015. Background subtraction based on low-rank and structured sparse decomposition. *IEEE Transactions on Image Processing*, 24(8), 2502–2514.

Long, Y., Gong, Y., Xiao, Z., Liu, Q., 2017. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5), 2486–2498.

Luo, Y., Zhou, L., Wang, S., Wang, Z., 2017. Video satellite imagery super resolution via convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 14(12), 2398–2402.

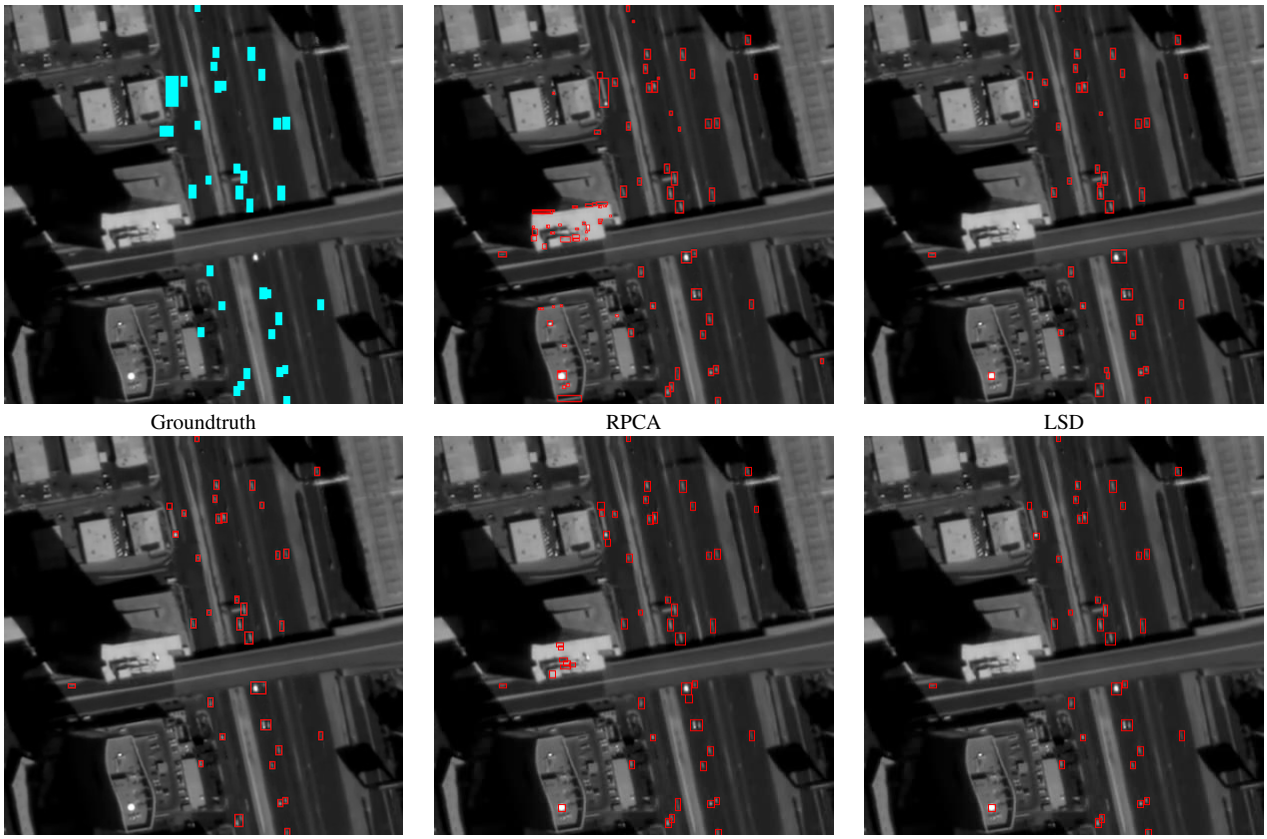


Figure 4. Detection results by different methods on Video 001.

Video	001				002			
	Recall ↑	Precision ↑	F_1 score ↑	Time (s) ↓	Recall ↑	Precision ↑	F_1 score ↑	Time (s) ↓
RPCA	94.4%	40.8%	56.98%	1256.50	90.2%	78.1%	83.72%	2019.76
LSD	86.8%	70.8%	77.98%	12591.39	82.2%	90.9%	86.31%	21215.65
E-LSD	85.0%	78.9%	81.82%	2960.56	79.6%	93.8%	86.13%	4417.95
S-LSD $\mathcal{M} = \{4\}$	90.5%	66.0%	76.32%	**1133.30**	84.3%	91.5%	**87.76%**	**1783.07**
S-LSD $\mathcal{M} = \{4, 16\}$	86.6%	**79.8%**	**83.07%**	1769.79	78.7%	**94.0%**	85.70%	2749.89

Table 3. Detection Performance Evaluation.

Mairal, J., Jenatton, R., Bach, F. R., Obozinski, G. R., 2010. Network flow algorithms for structured sparsity. *Advances in Neural Information Processing Systems*, 1558–1566.

Mou, L., Zhu, X. X., 2016. Spatiotemporal scene interpretation of space videos via deep neural network and tracklet analysis. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 1823–1826.

Shakeri, M., Zhang, H., 2016. COROLA: a sequential solution to moving object detection using low-rank approximation. *Computer Vision and Image Understanding*, 146, 27–39.

Team, P., 2016. Planet application program interface: In space for life on earth. san francisco, ca.

Uzkent, B., Rangnekar, A., Hoffman, M. J., 2018. Tracking in aerial hyperspectral videos using deep kernelized correlation filters. *IEEE Transactions on Geoscience and Remote Sensing*, 1–13.

Van den Bergh, M., Boix, X., Roig, G., Van Gool, L., 2015. Seeds: Superpixels extracted via energy-driven sampling. *International Journal of Computer Vision*, 111(3), 298–314.

Wright, J., Ganesh, A., Rao, S., Peng, Y., Ma, Y., 2009. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. *Advances in neural information processing systems*, 2080–2088.

Wright, S. J., 2015. Coordinate descent algorithms. *Mathematical Programming*, 151(1), 3–34.

Xu, J., Ithapu, V. K., Mukherjee, L., Rehg, J. M., Singh, V., 2013. Gosus: Grassmannian online subspace updates with structured-sparsity. *Proceedings of the IEEE International Conference on Computer Vision*, 3376–3383.

Xu, Y., Wu, Z., Chanussot, J., Dalla Mura, M., Bertozzi, A. L., Wei, Z., 2017. Low-rank decomposition and total variation regularization of hyperspectral video sequences. *IEEE Transactions on Geoscience and Remote Sensing*, 56(3), 1680–1694.

Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67.

Zhang, J., Jia, X., Hu, J., 2019a. Error bounded foreground and background modeling for moving object detection in satellite videos. *IEEE Transactions on Geoscience and Remote Sensing*, 1-11.

Zhang, J., Jia, X., Hu, J., Chanussot, J., 2019b. Online structured sparsity-based moving object detection from satellite videos. *arXiv preprint arXiv:1911.12989*.

Zhang, J., Jia, X., Hu, J., Tan, K., 2018. Satellite multi-vehicle tracking under inconsistent detection conditions by bilevel k-shortest paths optimization. *2018 Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, 1–8.

Zhou, T., Tao, D., 2011. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. *International Conference on Machine Learning*, Omnipress.

Zhou, X., Yang, C., Yu, W., 2013. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3), 597–610.