Moving Vehicle Detection for Remote Sensing Video Surveillance With Nonstationary Satellite Platform

Junpeng Zhang, *Student Member, IEEE*, Xiuping Jia, *Fellow, IEEE*, Jiankun Hu, *Senior Member, IEEE*, and Kun Tan, *Senior Member, IEEE*

Abstract—With satellite platforms gazing at a target territory, the captured satellite videos exhibit local misalignment and local intensity variation on some stationary objects that can be mistakenly extracted as moving objects and increase false alarm rates. Typical approaches for mitigating the effect of moving cameras in moving object detection (MOD) follow domain transformation technique, where the misalignment between consecutive frames is restricted to the image planar. However, such technique cannot properly handle satellite videos, as the local misalignment on them is caused by the varying projections from the 3D objects on the Earth's surface to 2D image planar. In order to suppress the effect of moving satellite platform in MOD, we propose a **M**oving-**C**onfidence-Assisted **M**atrix **D**ecomposition (MCMD) model, where foreground regularization is designed to promote real moving objects and ignore system movements with the assistance of a moving-confidence score estimated from dense optical flows. For solving the convex optimization problem in MCMD, both batch processing and online solutions are developed in this study, by adopting the alternating direction method and the stochastic optimization strategy, respectively. Experimental results on the videos captured by SkySat and Jilin-1 show that MCMD outperforms the state-of-the-art techniques with improved precision by suppressing effect of nonstationary satellite platforms.

Index Terms—Moving object detection, remote sensing video surveillance, satellite video, low rank matrix decomposition, moving satellite platform

1 Introduction

Satellite video surveillance has become an emerging technology for monitoring moving vehicles at city-scale. Since the first prototype video satellite launched by Google in 2013, two video satellite constellations with over ten video satellites, SkySat constellation [1] and Jilin-1 constellation [2], have been in operation. Moving Object Detection (MOD) on such videos captured from space serves as a most fundamental technique to achieve quantitative and automatic remote sensing surveillance. MOD methods have been broadly investigated for ground-based surveillance systems, where videos are commonly captured by a fixed camera with relatively high quality. They are challenged by the new issues in the satellite remote sensing videos –low spatial resolution (large ground sample distance), higher noise rate and nonstationary camera platform.

The ground sample distances of all the current satellite videos are no greater than 1 meter, so most of the vehicles are captured with only 5–20 pixels spatially. Satellite videos have limited spectral information as they are often monochromatic. Due to the lack of spectral information and the limited spatial features, such as shape and textures, detecting vehicles from satellite videos by deep learning-based detectors are challenging, despite of their success in natural images [3], [4], [5] or aerial high-resolution images [6], [7], [8], [9], [10]. In this case, the motion of moving vehicles is a more useful and important feature, which facilitates the background subtraction-based MOD approaches. To suppress the effect of random noise in MOD, additional constraints, such as smoothness on the edges [11], [12] and structured sparsity [13], [14], [15], are enforced on the foreground modelling.

Most importantly, satellite videos differentiate themselves from ground-based videos by the moving satellite platform. With satellite platforms orbiting around Earth and gazing at a target territory, the captured satellite videos exhibit a unique type of local misalignment —small and uneven motions on some stationary background objects. While the majority of the background exhibits sub-pixel and neglect motions [16], some apparent local misalignment can be observed on the tall structures, such as skyscrapers, as presented in Fig. 1. Such local misalignment is caused by the varying 2D projection from the complex 3D objects on the Earth's surface with changing camera locations, and typical domain transformation technique [17] cannot properly handle it. In addition to local misalignment, due to the surface roughness of different objects on the ground, the motion of satellite also leads to significant

- *Junpeng Zhang, Xiuping Jia, and Jiankun Hu are with the School of Engineering and Information Technology (SEIT), University of New South Wales, Canberra, NSW 2600, Australia. E-mail: zhangjunpeng9354@126.com, {x.jia, J.Hu}@adfa.edu.au.*
- *Kun Tan is with the Key Laboratory of Geographic Information Science, East China Normal University, Shanghai 200241, China. E-mail: tankuncu@gmail.com.*

Manuscript received 26 July 2020; revised 14 December 2020; accepted 11 March 2021. Date of publication 17 March 2021; date of current version 4 August 2022.
(Corresponding author: Xiuping Jia.)
Recommended for acceptance by K. Schindler.
Digital Object Identifier no. 10.1109/TPAMI.2021.3066696

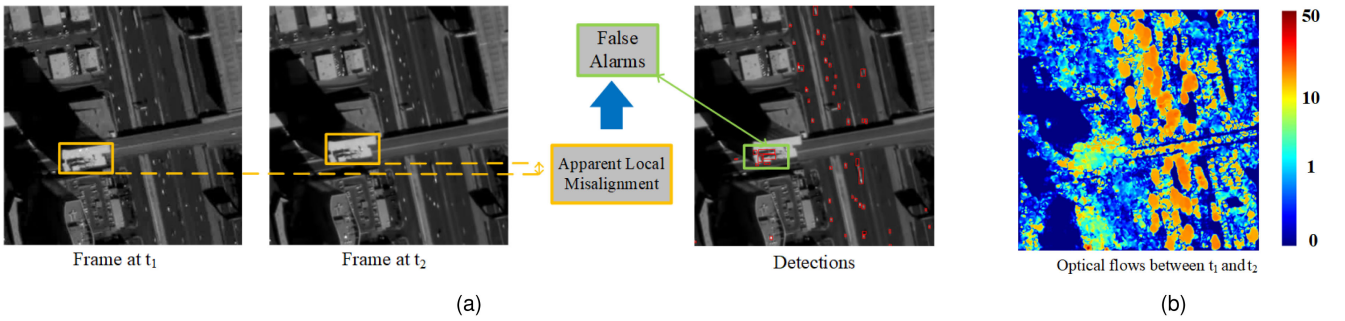


Fig. 1. Local misalignment in satellite videos can cause false alarms in MOD. (a) Local misalignment and false alarms. (b) Optical flows.

variations of the received signals, and subsequently results in local intensity variations on stationary objects, as shown in Fig. 1. The local misalignment and possible local intensity variations on stationary background objects make the motion of satellite platforms another primary source of false alarms in MOD. The dominating matrix decomposition-based MOD methods assume that the background is static. When they are applied to satellite video data, stationary background objects with such pixel shifts and/or intensity variations are recognised as moving objects, which negatively affects MOD performance by the reduced precision. To suppress the effect of moving camera, domain transformation [17] is introduced in matrix decomposition-based MOD approaches, and an optimal set of transformation is sought for conducting image registration, in which the misalignment is, however, restricted on the image planar [11], [18], [19]. The proper techniques for handling the 3D effect of moving satellite platform in MOD are barely explored so far but strongly needed.

In this paper, we propose a unified MOD framework where moving vehicles are separated from false alarms caused by the motion of satellites in video capturing. A moving-confidence score is introduced to indicate each pixel's likelihood of belonging to moving vehicles, and we estimate this score from the dense optical flows across a video. A foreground regularization is designed to exclude the pixels with small movements and promote the real object motions with the assistance of this moving-confidence score. The proposed low-rank matrix decomposition is named as **M**oving-**C**onfidence-Assisted **M**atrix **D**ecomposition (MCMD). For solving the optimization problem of MCMD, both batch and online optimization schemes are developed in this paper. We adopt the Alternating Direction Method of Multiplier (ADMM) approach for batch optimization, and provide an online algorithm based on stochastic optimization. To systematically evaluate MOD performance on satellite videos, two datasets were built from the videos captured by the two available video satellite constellations, SkySat and Jilin-1 satellites, respectively. Through experimental results on these datasets, we validate the improved detection precision of MCMD by eliminating the false alarms caused by nonstationary camera platform, and MCMD achieves superior overall performance when compared with state-of-the-art approaches.

In summary, the main contributions of this paper are four-fold:

1) To suppress the false alarms caused by the motion of satellite platforms in MOD on satellite videos, a moving-confidence-assisted foreground regularization is developed, where the moving confidence scores are estimated from the dense optical flows across a satellite video. We propose a novel low-rank and structured sparse matrix decomposition model by integrating the proposed moving-confidence-assisted foreground regularization for MOD in satellite videos, namely **M**oving-**C**onfidence-**A**ssisted **M**atrix **D**ecomposition (MCMD).
2) The proposed MCMD model is defined as a linear constrained convex optimization problem, and we adopt the Alternating Direction Method of Multiplier (ADMM) for solving it by batch optimization, which is denoted by B-MCMD. For scenarios where online processing is required, we also propose an online algorithm based on stochastic optimization, namely O-MCMD.
3) Two datasets containing the videos from SkySat and Jilin-1, the only two satellites which are currently available, were collected and annotated. Based on these datasets, experimental evaluation on both B-MCMD and O-MCMD methods were conducted against some state-of-the-art approaches. The datasets are valuable for remote sensing community in general to conduct the study on MOD in satellite videos in the coming years.

The remainder of this paper is organized as follows. Related work on matrix decomposition is first presented in Section 2. The proposed formulation for MOD is detailed in Section 3, which is followed by the batch and online optimization methods in Section 4. The analysis on the parameter settings of MCMD model and the experimental results are provided in Section 5. Finally, conclusions and suggestions for future research are given in Section 6.

2 Matrix Decomposition for MOD

In this work, we propose a novel background subtraction model for MOD, which is tied with low-rank matrix decomposition technique. Based on the stillness of background across a video, low-rank matrix decomposition technique separates the foreground matrix of encoded moving objects by modelling background objects with a low-rank matrix [20], [21], [22].

Let $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n] \in \mathbb{R}^{p \times n}$ denote an input video with n frames, where $\mathbf{d}_i \in \mathbb{R}^p$ is the vectorized ith frame, and $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n] \in \mathbb{R}^{p \times n}$, $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n] \in \mathbb{R}^{p \times n}$ and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] \in \mathbb{R}^{p \times n}$ are the background matrix, foreground matrix and model residual matrix, respectively. In general, matrix decomposition methods for MOD define an

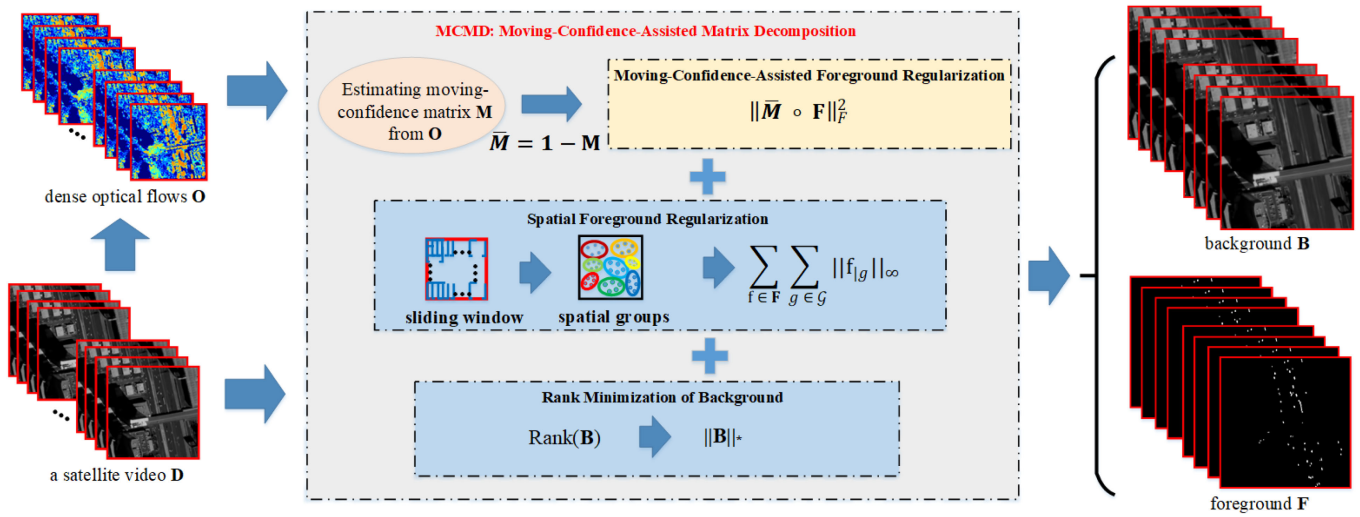


Fig. 2. Flowchart of the proposed MCMD model.

optimization as

$$\begin{aligned} &\arg\min_{\mathbf{B},\mathbf{F},\mathbf{E}} \operatorname{Rank}(\mathbf{B}) + \lambda_1 \Omega(\mathbf{F}) + \lambda_2 \|\mathbf{E}\|_F^2 \\ &\text{s.t.} \quad \mathbf{D} = \mathbf{B} + \mathbf{F} + \mathbf{E}, \end{aligned} \tag{1}$$

in which the background is commonly modeled by a low-rank matrix due to its global stillness.

The term $\Omega(\mathbf{F})$ refers to the regularization imposed on the foreground. Given that the moving objects are associated with a limited number of temporally changing pixels, the foreground is then modeled as sparse outliers to the low dimensional subspace, and ℓ_0 norm is deployed for promoting element-wise sparsity on the foreground [23], [24], [25], [26]. In practice, ℓ_0 sparsity foreground regularization is, however, prone to random noises in videos. In order to address this issue, spatial structural priors on the foreground are explored in matrix decomposition. Total Variation (TV) [27], [28] and Generalized Fused Lasso (GFL) [29] are imposed to enforce smoothness on the foreground, and the contiguous constraint on the edge of the moving objects is imposed on the foreground matrix by the first-order Markov Random Field (MRF) [11]. Structured sparsity-inducing norm [30], which defines the sparsity over groups of neighboring pixels, is also developed to suppress the effects of random noise in MOD [13], [14], [18], [31], since moving objects in a video are seldom in the form of isolate pixels. These matrix decomposition models assume the video for processing is well aligned. In this study, we extend them to address the obvious apparent local misalignment on stationary objects caused by nonstationary satellite platform to minimize false alarms in moving vehicle detection from satellite videos.

3 Moving-Confidence-Assisted Matrix Decomposition Model

Fig. 2 shows the proposed **M**oving-**C**onfidence-**A**ssisted **M**atrix **D**ecomposition Model (MCMD) which decomposes a data matrix to three matrices, corresponding to stationary background, moving objects and model residuals, respectively. Aside with the structured sparse regularization imposed on the foreground, the foreground is regularized by a moving-confidence matrix to consider each pixel's likelihood of belonging to moving objects. In this paper, we estimate the moving confidence scores from the dense optical flows across the video. In this way, the low-rank and structured sparse matrix decomposition model is enhanced to prevent the local misalignment from appearing in the foreground component.

3.1 Moving-Confidence Matrix Estimation

Dense optical flow measures pixel-wise displacement between a pair of frames [32], [33]. In satellite videos, moving satellite platform can cause small pixel displacements on stationary objects, whereas moving objects generally exhibit large displacements on their associated pixels, as illustrated in Fig. 1. This contrast implies that the pixels with larger optical flows are more likely to belong to moving objects, which is the rationale behind the generation of a moving confidence matrix in this paper.

For each frame $\mathbf{f}_i$, we utilize Farneback's algorithm [34] for extracting the dense optical flows between current frame and its next frame, and denote the magnitudes of the extracted optical flows by $\mathbf{o}_i \in \mathbf{p}$. A moving-confidence matrix $\mathbf{M}$ is introduced and its corresponding component $\mathbf{m}_i$ is obtained by converting the optical flows $\mathbf{o}_i$ using the following function

$$m_{i,j} = \frac{1}{1 + e^{-\alpha(o_{i,j} - \beta)}}, \tag{2}$$

where $m_{i,j}$ and $o_{i,j}$ refer to the jth elements in $\mathbf{m}_i$ and $\mathbf{o}_i$, respectively. By doing this, the optical flows are mapped to confidence scores in the range of $[0, 1.0]$. The other propriety of the introduced confidence scores is that they are not linearly related to the optical flow, instead, more sensitive to the presumed minimum speed of real moving objects, and insensitive to the speed lower or higher than that, by selecting the parameter α and β properly. β can be selected as the presumed minimum speed of real moving objects, which can be obtained from training video sequences or empirically. We propose to select a reasonable large α, such as $\alpha = 10$, so that the moving confidence emphasises the difference between

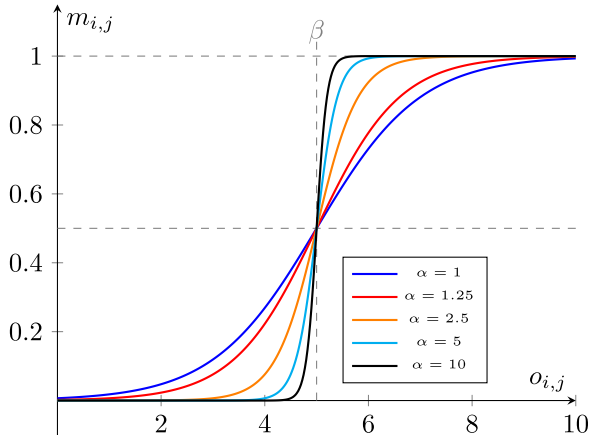


Fig. 3. Mapping from optical flow $o_{i,j}$ to confidence $m_{i,j}$ using difference α.

real moving objects and the movements due to the camera problem. As shown in Fig. 3, it promotes the first case by assigning the moving confidence toward 1, and suppresses the second case by assigning its moving confidence toward 0.

For preventing duplicated computation on consecutive and similar frames, a set of anchor frames can be selected from a video with a temporal stride, and the dense optical flows used in Eq. (2) are computed from each pair of consecutive anchor frames. The temporal stride for anchor frame selection is decided based on the frame rate of the video data and the understanding of the speed of moving objects. For videos with high frame rate, a large temporal stride is appropriate to adequately separate local misalignment from moving vehicles.

3.2 Problem Formulation

On top of this matrix decomposition model in Eq. (1), we impose a moving-confidence-assisted foreground regularization for suppressing the effect of nonstationary satellite platforms. The proposed model is called **M**oving-**C**onfidence-Assisted **M**atrix **D**ecomposition (MCMD) model. It defines an optimization problem as

$$\begin{aligned} &\arg\min_{\mathbf{B},\mathbf{F},\mathbf{E}} \operatorname{Rank}(\mathbf{B}) + \lambda_1 \Omega(\mathbf{F}) + \lambda_2 \Psi(\mathbf{F}) + \lambda_3 \|\mathbf{E}\|_F^2 \\ &\text{s.t.} \quad \mathbf{D} = \mathbf{B} + \mathbf{F} + \mathbf{E}, \end{aligned} \tag{3}$$

in which $\Omega(\mathbf{F})$ and $\Psi(\mathbf{F})$ denote the spatial foreground regularization term and the moving confidence-assisted foreground regularization term, respectively. The squared penalty term $\|\mathbf{E}\|_F^2$ handles the model residuals that do not fit either the background or the foreground. λ_1, λ_2 and λ_3 are the corresponding weights assigned to these regularization terms.

Spatial Foreground Regularization: To penalize the false alarms caused by random noises, we utilize the structured sparsity-inducing norm as spatial foreground regularization [13], [14], [15], [31], which is defined on the sparsity of the foreground over the spatial groups of neighboring pixels other than that over the individual pixels,

$$\Omega(\mathbf{F}) = \sum_{\mathbf{f}\in\mathbf{F}} \|\mathbf{f}\|_{\ell_1/\ell_\infty} = \sum_{\mathbf{f}\in\mathbf{F}} \sum_{g\in\mathcal{G}} \left\|\mathbf{f}_{|g}\right\|_\infty, \tag{4}$$

where $\mathcal{G}$ is the set of spatial groups of neighboring pixels of the foreground. $\mathbf{f}_{|g} \in \mathbb{R}^p$ refers to a sparse vector with non-zero elements at the indices represented in a group $g \in \mathcal{G}$. In this paper, the set of spatial groups $\mathcal{G}$ is defined by the image patches extracted using a sliding window approach.

Moving-Confidence-Assisted Foreground Regularization: The moving-confidence-assisted foreground regularization penalizes those foreground pixels with low moving confidence scores. Given that $\mathbf{M} \in \mathbb{R}^{p\times n}$ is the moving-confidence scores obtained by Section 3.1, this regularization is defined by

$$\Psi(\mathbf{F}) = \|(1-\mathbf{M}) \circ \mathbf{F}\|_F^2 = \left\|\overline{\mathbf{M}} \circ \mathbf{F}\right\|_F^2, \tag{5}$$

where $\overline{\mathbf{M}} \in \mathbb{R}^{p\times n}$ is the complementary confidence matrix to $\mathbf{M}$, $\overline{\mathbf{M}} = 1 - \mathbf{M}$, and $\overline{\mathbf{M}} \circ \mathbf{F}$ refers to the Hadamard product of $\overline{\mathbf{M}}$ and $\mathbf{F}$. Minimizing $\Psi(\mathbf{F})$ penalizes the occurrence of non-zero elements at the locations with low confidence in $\mathbf{M}$, by which the false alarms caused by local misalignment on stationary objects can be reduced.

To make the optimization problem in Eq. (3) more tractable, we relax the rank operator by its convex surrogate function, which is the nuclear norm $\|\mathbf{B}\|_*$. The relaxed optimization problem is then rewritten as

$$\begin{aligned} &\arg\min_{\mathbf{B},\mathbf{F},\mathbf{E}} \|\mathbf{B}\|_* + \lambda_1 \sum_{\mathbf{f}\in\mathbf{F}} \|\mathbf{f}\|_{\ell_1/\ell_\infty} + \lambda_2 \left\|\overline{\mathbf{M}} \circ \mathbf{F}\right\|_F^2 + \lambda_3 \|\mathbf{E}\|_F^2 \\ &\text{s.t.} \quad \mathbf{D} = \mathbf{B} + \mathbf{F} + \mathbf{E}, \end{aligned} \tag{6}$$

in which the scalar λ_2 controls the contribution of moving-confidence-assisted foreground regularization to the objective function in Eq. (6). As λ_2 increases from a minimal positive value, those foreground elements with low moving-confidence scores make more contribution to the objective function, and they are penalized more significantly when the foreground is updated. The next challenge is how to solve the formulated optimization problem. We develop two algorithms as presented in the following section, which give a batch optimization method (B-MCMD) and an online optimization method (O-MCMD) in Sections 4.1 and 4.2, respectively.

4 Optimization Method

4.1 Batch Optimization (B-MCMD)

The optimization problem defined in Eq. (6) couples two regularization term on $\mathbf{F}$, which makes it difficult to solve efficiently. To tackle this, we introduce an auxiliary variable $\mathbf{Z} \in \mathbb{R}^{p\times n}$, and define an equivalent optimization problem to Eq. (6) as

$$\begin{aligned} &\arg\min_{\mathbf{B},\mathbf{F},\mathbf{Z},\mathbf{E}} \|\mathbf{B}\|_* + \lambda_1 \sum_{\mathbf{f}\in\mathbf{F}} \|\mathbf{f}\|_{\ell_1/\ell_\infty} + \lambda_2 \left\|\overline{\mathbf{M}} \circ \mathbf{Z}\right\|_F^2 + \lambda_3 \|\mathbf{E}\|_F^2 \\ &\text{s.t.} \quad \mathbf{D} = \mathbf{B} + \mathbf{F} + \mathbf{E}, \\ &\qquad \mathbf{F} = \mathbf{Z}. \end{aligned} \tag{7}$$

We propose a batch optimization for solving this reformulated problem, named by B-MCMD. First, the linear constraints in Eq. (7) are removed using the Augmented Lagrangian Method (ALM) [35],

$$\begin{aligned}\arg\min_{\mathbf{B},\mathbf{F},\mathbf{Z},\mathbf{E},\mathbf{Y}} &\|\mathbf{B}\|_* + \lambda_1 \sum_{\mathbf{f}\in\mathbf{F}} \|\mathbf{f}\|_{\ell_1/\ell_\infty} + \lambda_2 \left\|\overline{\mathbf{M}} \circ \mathbf{Z}\right\|_F^2 \\ &+ \lambda_3 \|\mathbf{E}\|_F^2 + < \mathbf{Y}_1, \mathbf{D} - \mathbf{B} - \mathbf{F} - \mathbf{E} > \\ &+ \frac{\mu_1}{2} \|\mathbf{D} - \mathbf{B} - \mathbf{F} - \mathbf{E}\|_F^2 \\ &+ < \mathbf{Y}_2, \mathbf{F} - \mathbf{Z} > + \frac{\mu_2}{2} \|\mathbf{F} - \mathbf{Z}\|_F^2,\end{aligned} \tag{8}$$

where μ_1 and μ_2 are positive scalars, and $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are Lagrangian multipliers. In B-MCMD, we adopt the ADMM approach for minimizing the augmented Lagrangian function. The variables are updated alternatingly by solving their corresponding sub-problem in each iteration.

Algorithm 1. Proposed B-MCMD Method for MOD

Input: $\mathbf{D} \in \mathbb{R}^{p*n}$, $\overline{\mathbf{M}} \in \mathbb{R}^{p*n}$, $\lambda_1 > 0$, $\lambda_2 > 0$, $\lambda_3 > 0$, $\mu_1 > 0$, $\mu_2 > 0$, $\rho > 0$, $\bar{\mu}_1 = \mu_1 * 10^5$ and $\bar{\mu}_2 = \mu_2 * 10^5$.
Output: **B**, **F**, **Z** and **E**.

1: $\mathbf{B} = \mathbf{0}$, $\mathbf{F} = \mathbf{0}$, $\mathbf{E} = \mathbf{0}$, $\mathbf{X} = \mathbf{0}$ $\mathbf{Y}_1 = \mathbf{0}$ and $\mathbf{Y}_2 = \mathbf{0}$.
2: $k = 0$
3: **while** not converged **do**
4: Update $\mathbf{B}^{k+1}$ by Eq. (12).
5: Update $\mathbf{F}^{k+1}$ by Eq. (18).
6: Update $\mathbf{Z}^{k+1}$ by Eq. (20).
7: Update $\mathbf{E}^{k+1}$ by Eq. (21).
8: Update the Lagrangian multipliers by

$$\begin{cases} \mathbf{Y}_1^{k+1} = \mathbf{Y}_1^k + \mu_1(\mathbf{D} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1} - \mathbf{E}^{k+1}) \\ \mathbf{Y}_2^{k+1} = \mathbf{Y}_2^k + \mu_2(\mathbf{F}^{k+1} - \mathbf{X}^{k+1}) \end{cases}. \tag{9}$$

9: (Optionally) Update μ_1 and μ_2 by

$$\begin{cases} \mu_1 = \min\{\rho\mu_1, \bar{\mu}_1\} \\ \mu_2 = \min\{\rho\mu_2, \bar{\mu}_2\} \end{cases}. \tag{10}$$

10: Break if the termination conditions is satisfied.
11: $k = k + 1$.
12: **end while**
13: **return** $\mathbf{B}^{k+1}$, $\mathbf{F}^{k+1}$, $\mathbf{X}^{k+1}$ and $\mathbf{E}^{k+1}$.

Update B: Updating **B** in each iteration is achieved by solving the augmented optimization problem while fixing **F**, **Z**, $\mathbf{Y}_1$ and $\mathbf{Y}_2$:

$$\arg\min_{\mathbf{B}} \frac{1}{\mu_1}\|\mathbf{B}\|_* + \frac{1}{2}\left\|\mathbf{B} - (\mathbf{D} - \mathbf{F}^k - \mathbf{E}^k + \frac{\mathbf{Y}_1^k}{\mu_1})\right\|_F^2. \tag{11}$$

Let $\mathbf{G} = \mathbf{D} - \mathbf{F}^k - \mathbf{E}^k + \frac{1}{\mu_1}\mathbf{Y}_1^k$ and $\mathbf{U}\Sigma\mathbf{U}^T$ denote the Singular Value Decomposition (SVD) of **G**, the optimal solution to Eq. (11) is obtained by

$$\mathbf{B}^{k+1} = \mathbf{U}\left(\Sigma - \frac{1}{\mu_1}\mathbf{I}\right)_+ \mathbf{V}^T, \tag{12}$$

where $\left(\Sigma - \frac{1}{\mu_1}\mathbf{I}\right)_+$ is the element-wise soft-shrinking operator, $\left(\Sigma - \frac{1}{\mu_1}\mathbf{I}\right)_+ = \max\left(\Sigma - \frac{1}{\mu_1}\mathbf{I}, \mathbf{0}\right)$, and **I** is the identity matrix.

Update F: After obtaining $\mathbf{B}^{k+1}$, the foreground **F** is estimated by solving a structured sparse encoding problem,

$$\arg\min_{\mathbf{F}} \frac{\lambda_1}{\mu_1 + \mu_2}\sum_{\mathbf{f}\in\mathbf{F}} \|\mathbf{f}\|_{\ell_1/\ell_\infty} + \frac{1}{2}\|\mathbf{F} - \mathbf{H}\|_F^2, \tag{13}$$

with

$$\begin{cases} \mathbf{H} = \frac{\mu_1\mathbf{G}_1 + \mu_2\mathbf{G}_2}{\mu_1 + \mu_2} \\ \mathbf{G}_1 = \mathbf{D} - \mathbf{B}^{k+1} - \mathbf{E}^k + \frac{\mathbf{Y}_1^k}{\mu_1}. \\ \mathbf{G}_2 = \mathbf{Z}^k - \frac{\mathbf{Y}_2^k}{\mu_2} \end{cases} \tag{14}$$

As the spatial regularization term is frame-wise separable, this encoding problem can be solved by a series of frame-wise sub-problems. For the ith frame, the foreground component $\mathbf{f}_i$ is obtained as

$$\mathbf{f}_i = \arg\min_{\mathbf{f}} \frac{1}{2}\|\mathbf{h}_i - \mathbf{f}\|_2^2 + \lambda' \sum_{g\in\mathcal{G}} \left\|\mathbf{f}_{|g}\right\|_\infty, \tag{15}$$

where $\lambda' = \lambda_1/(\mu_1 + \mu_2)$, and the optimal solution is obtained from the dual problem of Eq. (15) according to the following lemma.

Lemma 4.1 (Dual of the structured sparsity encoding problem from [30], [36], [37]). *Given* $\mathbf{h} \in \mathbb{R}^p$ *and* $\mathbf{x} \in \mathbb{R}^p$, *a structured sparsity encoding problem on* **x** *with provided set of structured variables* $\mathcal{G}$ *is defined as*

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{h} - \mathbf{x}\|_2^2 + \lambda \sum_{g\in\mathcal{G}} \left\|\mathbf{x}_{|g}\right\|_\infty, \tag{16}$$

whose dual problem is provided as

$$\begin{aligned}\xi^* = \arg\min_{\xi} &\frac{1}{2}\left\|\mathbf{h} - \sum_{g\in\mathcal{G}} \xi^g\right\|_2^2 \\ &s.t.\ \forall g \in \mathcal{G}, \|\xi^g\|_1 \leq \lambda \text{ and } \xi_j^g = 0 \text{ if } j \notin g,\end{aligned} \tag{17}$$

where $\xi \in \mathbb{R}^{p\times|\mathcal{G}|}$ *is the dual variable. The dual problem defines a Quadratic Min-cost Network Flow problem, from whom the dual optimal* ξ^* *can be obtained. Then the primal optimal* $\mathbf{x}^*$ *is derived by*

$$\mathbf{x} = \mathbf{h} - \sum_{g\in\mathcal{G}} \xi^{*g}. \tag{18}$$

Update Z: With other variables in Eq. (8) fixed, the the auxiliary variable **Z** is updated by solving the following problem,

$$\arg\min_{\mathbf{Z}} \lambda_2\left\|\overline{\mathbf{M}} \circ \mathbf{Z}\right\|_F^2 + \frac{\mu_2}{2}\left\|\mathbf{Z} - (\mathbf{F}^{k+1} + \frac{\mathbf{Y}_2^k}{\mu_2})\right\|_F^2, \tag{19}$$

and its closed-form solution is derived from its first-order derivative,

$$\mathbf{Z}^{k+1} = \mu_2(\mathbf{F}^{k+1} + \frac{\mathbf{Y}_2^k}{\mu_2}) \oslash (2\lambda_2\overline{\mathbf{M}} \circ \overline{\mathbf{M}} + \mu_2), \tag{20}$$

where $\oslash$ refers to the element-wise divide operator.

Update E: The closed-form solution of **E** is derived by the first-order derivative of the objective function in Eq. (8) while fixing other variables:

$$\mathbf{E}^{k+1} = \frac{\mu_1}{2\lambda_2 + \mu_1}\left(\mathbf{D} - \mathbf{B}^{k+1} - \mathbf{F}^{k+1} + \frac{1}{\mu_1}\mathbf{Y}^k\right). \tag{21}$$

After updating all variables in each iteration, the relative termination criterion is examined by

$$\begin{aligned} &\frac{\left\|\mathbf{D}-\mathbf{B}^{k+1}-\mathbf{F}^{k+1}-\mathbf{E}^{k+1}\right\|_F}{\|\mathbf{D}\|_F} \leq \tau, \\ \& \quad &\frac{\left\|\mathbf{F}^{k+1}-\mathbf{X}^{k+1}\right\|_F}{\|D\|_F} \leq \tau, \end{aligned} \tag{22}$$

where $\tau = 10^{-7}$. The iteration continues till it reach the above termination criterion, and the entire procedure of B-MCMD is summarized in Algorithm 1.

4.2 Online Optimization (O-MCMD)

Like most canonical low-rank matrix decomposition approaches, the batch optimization B-MCAM relies on SVD over the entire video data $\mathbf{D}$, so it is not suitable for online processing systems. For a large satellite video, B-MCMD will also incur an extremely high computational cost. To overcome this problem, an online optimization method for solving MCMD is desired. In this paper, an online optimization method is developed by adopting the stochastic optimization techniques, and we denote the proposed online algorithm by O-MCMD.

First, the optimization problem of MCMD in Eq. (6) is rewritten to its unconstrained form as

$$\begin{aligned} \arg\min_{\mathbf{B},\mathbf{F},\mathbf{E}} \frac{1}{2}\|\mathbf{D}-\mathbf{B}-\mathbf{F}\|_F^2 + \lambda_1\|\mathbf{B}\|_* + \lambda_2 \sum_{\mathbf{f}\in\mathbf{F}} \|\mathbf{f}\|_{\ell_1/\ell_\infty} \\ + \lambda_3 \left\|\overline{\mathbf{M}} \circ \mathbf{F}\right\|_F^2, \end{aligned} \tag{23}$$

We can see minimizing $\|\mathbf{B}\|_*$ prohibits online processing, as it depends on SVD and couples all the frames in each iteration. This is addressed in the following.

We adopt the matrix factorization approximation of the nuclear norm by the sum of square penalties of its factorization [38], [39],

$$\|\mathbf{B}\|_* = \inf_{\mathbf{L}\in\mathbb{R}^{p\times r},\mathbf{R}\in\mathbb{R}^{r\times n}} \left\{ \frac{1}{2}\|\mathbf{L}\|_F^2 + \frac{1}{2}\|\mathbf{R}\|_F^2 : \mathbf{B} = \mathbf{L}\mathbf{R} \right\}, \tag{24}$$

where $\mathbf{L}$ is treated as the basis of a low-dimensional subspace, and $\mathbf{R}$ denotes the corresponding reconstruction coefficients with regards to $\mathbf{L}$. By substituting the nuclear norm in Eq. (23), the optimization problem is reformulated as

$$\begin{aligned} \arg\min_{\mathbf{L},\mathbf{R},\mathbf{F}} \frac{1}{2}\|\mathbf{D}-\mathbf{L}\mathbf{R}-\mathbf{F}\|_F^2 + \frac{\lambda_1}{2}\|\mathbf{L}\|_F^2 + \frac{\lambda_1}{2}\|\mathbf{R}\|_F^2 \\ + \lambda_2 \sum_{\mathbf{f}\in\mathbf{F}} \|\mathbf{f}\|_{\ell_1/\ell_\infty} + \lambda_3 \left\|\overline{\mathbf{M}} \circ \mathbf{F}\right\|_F^2, \end{aligned} \tag{25}$$

and this optimization problem is then rewritten in a frame-wise separable form as

$$\begin{aligned} \arg\min_{\mathbf{L},\mathbf{R},\mathbf{F}} \sum_{t=1}^{n} \Bigg\{ \frac{1}{2}\|\mathbf{d}_t - \mathbf{L}\mathbf{r}_t - \mathbf{f}_t\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{r}_t\|_2^2 \\ + \lambda_2 \sum_{g\in\mathcal{G}} \left\|\mathbf{f}_{t|g}\right\|_\infty + \lambda_3 \|\overline{\mathbf{m}}_t \circ \mathbf{f}_t\|_2^2 \Bigg\} \\ + \frac{\lambda_1}{2}\|\mathbf{L}\|_F^2, \end{aligned} \tag{26}$$

where $\mathbf{d}_t$ and $\overline{\mathbf{m}}_t$ denote the video frame and its corresponding motion confidence at a time instance t, respectively. $\mathbf{r}_t$ and $\mathbf{f}_t$ are the estimated coefficient and foreground for tth frame.

In Eq. (26), the basis $\mathbf{L}$ is shared across frames, which prevent it from being updated online. In order to achieve online process, we adopt the Stochastic Optimization (SA) for optimizing $\mathbf{L}$ in an online manner. For each new frame $\mathbf{d}_t$, the pair of $\mathbf{r}_t$ and $\mathbf{f}_t$ are estimated with the subspace basis $\mathbf{L}_{t-1}$ obtained from previous time instance $t-1$, then the basis $\mathbf{L}_t$ is updated by all existing pairs of $(\mathbf{r}_i, \mathbf{f}_i), \forall i \in \{1, 2, \ldots, t\}$. We summarize this online processing procedure in Algorithm 2.

Algorithm 2. Proposed O-MCMD Algorithm for MOD

Input: $\mathbf{d}_t \in \mathbb{R}^p$, $\overline{\mathbf{m}}_t \in \mathbb{R}^p$, $\mathbf{L}_{t-1} \in \mathbb{R}^{p\times r}$, $\mathbf{A}_{t-1}$ and $\mathbf{B}_{t-1}$,
Output: $\mathbf{b}_t$, $\mathbf{r}_t$, $\mathbf{s}_t$ and $\mathbf{L}_t$

```
r_t^0 = r_{t-1}, s_t^0 = s_{t-1} (with r_0 = 0 and s_0 = 0).
k = 0
while not converged do
    Update r_t^{k+1} by Eq. (29).
    Update f_t^{k+1} by Algorithm 3.
    if max{ ||r_t^{k+1} - r_t^k||_2 / p , ||f_t^{k+1} - f_t^k||_2 / p } ≤ τ then
        break
    end if
    k = k + 1.
end while
b_t = L_{t-1} r_t^{k+1}.
Update A_t and B_t by Eq. (33).
Update L_t by Eq. (32).
return b_t, r_t, s_t and L_t.
```

Estimate $\mathbf{r}_t$ and $\mathbf{f}_t$: With fixed $\mathbf{L}_{t-1}$, the reconstruction coefficient $\mathbf{r}_t$ and the foreground $\mathbf{f}_t$ are encoded by

$$\begin{aligned} (\mathbf{r}_t, \mathbf{f}_t) = \arg\min_{\mathbf{r},\mathbf{f}} \frac{1}{2}\|\mathbf{d}_t - \mathbf{L}_{t-1}\mathbf{r} - \mathbf{f}\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{r}\|_2^2 \\ + \lambda_2 \sum_{g\in\mathcal{G}(\mathbf{f})} \left\|\mathbf{f}_{|g}\right\|_\infty + \lambda_3 \|\overline{\mathbf{m}}_t \circ \mathbf{f}\|_2^2. \end{aligned} \tag{27}$$

We solve this encoding problem by the Block Coordinate Descent (BCD) method [40], in which $\mathbf{r}_t$ and $\mathbf{f}_t$ are updated alternatingly.

For the $(k+1)$-th iteration in BCD, $\mathbf{r}_t^{k+1}$ is obtained by

$$\mathbf{r}_t^{k+1} = \arg\min_{\mathbf{r}} \frac{1}{2}\left\|\mathbf{d}_t - \mathbf{L}_{t-1}\mathbf{r} - \mathbf{f}_t^k\right\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{r}\|_2^2, \tag{28}$$

whose closed-form solution is provided as

$$\mathbf{r}_t^{k+1} = \left(\mathbf{L}_{t-1}^T\mathbf{L}_{t-1} + \lambda_1\mathbf{I}\right)^{-1}\mathbf{L}_{t-1}^T\left(\mathbf{d}_t - \mathbf{f}_t^k\right). \tag{29}$$

The update of $\mathbf{f}_t^{k+1}$ is achieved by solving Eq. (27) with fixed $\mathbf{r}_t^{k+1}$ by

$$\begin{aligned} \mathbf{f}_t^{k+1} = \arg\min_{\mathbf{f}} \frac{1}{2}\|\mathbf{h} - \mathbf{f}\|_2^2 + \lambda_2 \sum_{g\in\mathcal{G}(\mathbf{f})} \left\|\mathbf{f}_{|g}\right\|_\infty \\ + \lambda_3 \|\overline{\mathbf{m}}_t \circ \mathbf{f}\|_2^2, \end{aligned} \tag{30}$$

where $\mathbf{h} = \mathbf{d}_t - \mathbf{L}_{t-1}\mathbf{r}_t^{k+1}$. As there is no closed-form solution available to Eq. (30), we introduce an auxiliary variable $\mathbf{z} \in \mathbb{R}^p$:

$$\begin{aligned}\mathbf{f}_t^{k+1} = \arg\min_{\mathbf{f}} \frac{1}{2}\|\mathbf{h}-\mathbf{f}\|_2^2 + \lambda_2 \sum_{g\in\mathcal{G}(\mathbf{f})} \|\mathbf{f}_{|g}\|_\infty \\ + \lambda_3 \|\overline{\mathbf{m}}_t \circ \mathbf{z}\|_2^2 \\ \text{s.t.} \quad \mathbf{f} = \mathbf{z},\end{aligned} \tag{31}$$

which is solved by an ADMM method, as described in Algorithm 3.

Algorithm 3. Proposed ADMM Method for Encoding Foreground in O-MCMD

Input: $\mathbf{d}_t \in \mathbb{P}^p$, $\mathbf{m}_t \in \mathbb{R}^p$, $\mathbf{h} \in \mathbb{R}^p$, $\lambda_2 > 0$, $\lambda_3 > 0$, $\mu > 0$, $\rho > 0$ and $\bar{\mu} = \mu * 10^5$.

Output: $\mathbf{f}$.

1: $\mathbf{f}^0 = \mathbf{0}$, $\mathbf{z}^0 = \mathbf{0}$ and $\mathbf{y}^0 = \mathbf{0}$.
2: $k = 0$.
3: **while** not converged **do**
4: Update $\mathbf{f}^{k+1}$:

$$\mathbf{f}^{k+1} = \arg\min_{\mathbf{f}} \frac{\lambda_2}{1+\mu} \sum_{g\in\mathcal{G}(\mathbf{f})} \|\mathbf{f}_{|g}\|_\infty + \frac{1}{2}\left\|\mathbf{f} - \frac{\mathbf{h}+\mu\mathbf{z}^k - \mathbf{y}^k}{1+\mu}\right\|_2^2,$$

which is solved by Lemma 4.1.
5: $\mathbf{z}^{k+1} = \mu(\mathbf{f}^{k+1} + \frac{1}{\mu}\mathbf{y}^k) \oslash (2\lambda_3 \overline{\mathbf{m}}_t \circ \overline{\mathbf{m}}_t + \mu)$.
6: $\mathbf{y}^{k+1} = \mathbf{y}^k + \mu(\mathbf{f}^{k+1} - \mathbf{z}^{k+1})$.
7: (Optionally) $\mu_1 = \min\{\rho\mu, \bar{\mu}\}$.
8: **if** $\frac{\|\mathbf{f}^{k+1} - \mathbf{z}^{k+1}\|_2}{\|\mathbf{d}_t\|_2} \le \tau$ **then**
9: **break**
10: **end if**
11: $k = k + 1$.
12: **end while**
13: **return** $\mathbf{f}^{k+1}$.

Update $\mathbf{L}_t$: The update of the basis $\mathbf{L}_t$ is achieved by minimizing a cost function based the estimated pairs of $(\mathbf{r}_i, \mathbf{f}_i), \forall i \in \{1, \ldots, t\}$,

$$\begin{aligned}\mathbf{L}_t &= \arg\min_{\mathbf{L}} \frac{1}{t}\sum_{i=1}^{t}\left\{\frac{1}{2}\|\mathbf{d}_i - \mathbf{L}\mathbf{r}_i - \mathbf{s}_i\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{r}_i\|_2^2 \right. \\ &\quad \left. + \lambda_2 \sum_{g\in\mathcal{G}(\mathbf{f})} \|\mathbf{f}_{i|g}\|_\infty + \lambda_3 \|\overline{\mathbf{m}}_i \circ \mathbf{f}_i\|_2^2\right\} + \frac{\lambda_1}{2t}\|\mathbf{L}\|_F^2 \\ &= \arg\min_{\mathbf{L}} Tr(\mathbf{L}(\lambda_1\mathbf{I} + \mathbf{A}_t)\mathbf{L}^T) - 2Tr(\mathbf{L}^T\mathbf{B}_t),\end{aligned} \tag{32}$$

where $Tr(\cdot)$ is the trace of a matrix. Two auxiliary accumulation matrices, $\mathbf{A}_t \in \mathbb{R}^{p\times r}$ and $\mathbf{B}_t \in \mathbb{R}^{r\times r}$, are introduced to remove duplicated calculations at each time instance:

$$\begin{cases}\mathbf{A}_t &= \mathbf{A}_{t-1} + \mathbf{r}_t\mathbf{r}_t^T \\ \mathbf{B}_t &= \mathbf{B}_{t-1} + (\mathbf{d}_t - \mathbf{s}_t)\mathbf{r}_t^T\end{cases}. \tag{33}$$

TABLE 1
Information on the Evaluation Datasets

Dataset	No.	Frame Size × #Frames	Usage
SkySat	0	400 × 400 × 700	Parameter Selection
	1	600 × 400 × 700	Performance Evaluation
	2	500 × 400 × 600	Performance Evaluation
	3	600 × 400 × 500	Performance Evaluation
	4	400 × 400 × 500	Performance Evaluation
	5	500 × 600 × 500	Performance Evaluation
Jilin-1	6	400 × 500 × 300	Performance Evaluation
	7	400 × 600 × 300	Performance Evaluation
	8	700 × 400 × 300	Performance Evaluation

Similar to [12], [39], [41], the optimization problem defined in Eq. (32) is solved by a Block Coordinate Descent Method for avoiding matrix inverse of large matrix.

Prior to starting O-MCMD, we initialize $\mathbf{L}_0$ by random values, and for O-MCMD τ is set by a very small number, such as 1×10^{-5}.

5 Experiments and Results

The MOD performance of the proposed MCMD model was inclusively validated on nine satellite videos as listed in Table 1, where the first six (No. 0 to No.5) were captured by SkySat and the rest three (No. 6 to No. 8) were obtained from Jilin-1. These satellites are the only two video satellite constellations in operation. The SkySat dataset has ground sample distances of around 1.0 meter. These videos are in gray-scale, and their frame rate is 30 frames per second. For Jilin-1 dataset, the ground sample distance is also around 1.0 meter, and their frame rate is 10 frames per second. Gray-scale was used in the experiments to match the video type of SkySat. The raw video data captured by both satellites is geometrically corrected by using the Rational Polynomial Coefficient (RPC) transformation that is derived from the sensor model and the corresponding terrain information. A boundary box was manually annotated for each moving vehicles in both datasets as groundtruth. In this paper, we used Video No.0 in SkySat dataset for discussion on parameter settings, and the remaining videos were utilized for performance evaluation against existing state-of-the-art methods.[1]

The detection performance of MOD on satellite videos was evaluated on recall, precision and F_1 scores given by

$$\begin{aligned}\text{Rcll} &= TP/(TP+FN) \\ \text{Prec} &= TP/(TP+FP) \\ F_1 &= \frac{2\times\text{Rcll}\times\text{Prec}}{\text{Rcll}+\text{Prec}},\end{aligned} \tag{34}$$

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 complement the vehicles in small size in satellite videos, the threshold of IoU is set as 0.3.[2]

1. The annotation is public available through https://github.com/zhangjunpeng9354/satellite_video_mod_groundtruth.git.

2. The estimated foreground is built by contiguous values rather than binary value, so we deploy threshold segmentation as post-processing for extracting the foreground mask and the moving objects [42].

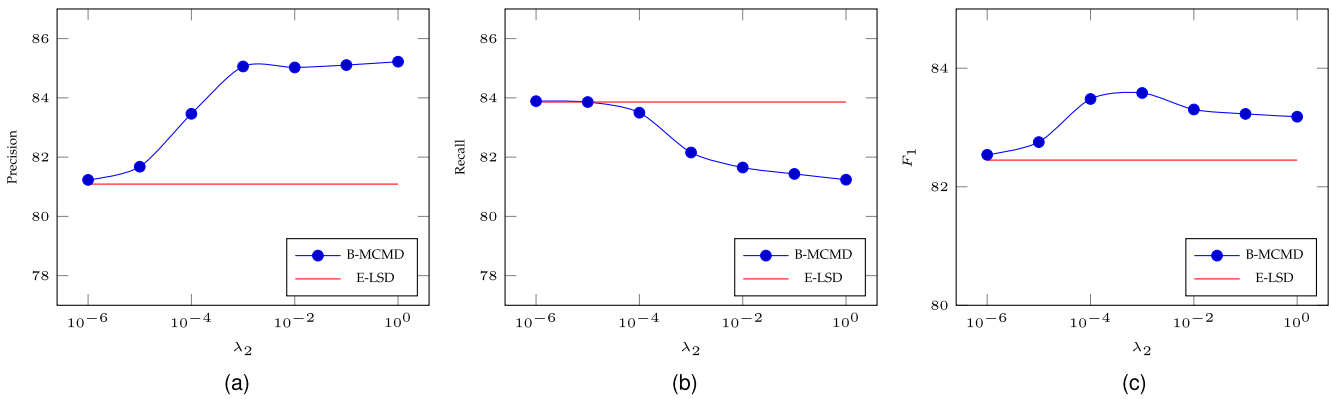


Fig. 4. MOD performance by B-MCMD with varying λ_2 with fixed λ_1 and λ_3 on Video No.0 in Skysat dataset. (a) Recall. (b) Precision. (c) F_1.

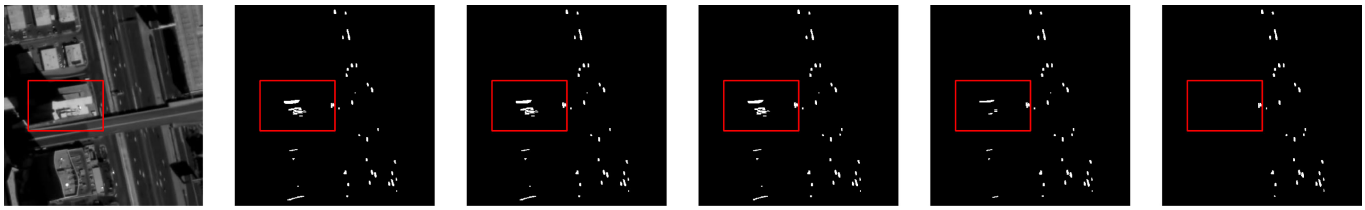


Fig. 5. Visualization on the foreground mask of Video No.0 in SkySat dataset by B-MCMD with increasing λ_2.

5.1 Effect of Moving-Confidence-Assisted Foreground Regularization

In the MCMD model, the moving-confidence-assisted regularization is imposed on the foreground, and its contribution to objective function in Eq. (6) is controlled by the weight λ_2. To experimentally examine the effectiveness of the proposed foreground regularization $\Psi(\mathbf{F})$, we gradually increase λ_2 from 10^{-6} to 1 in B-MCAM while fixing $\lambda_1 = 1/\sqrt{\max\{p,n\}}$ and $\lambda_3 = \lambda_1/5$, to make the moving-confidence foreground regularization contribute more to the objective function. As presented in Fig. 4, as λ_2 increases from 10^{-6} to 10^{-3} , the precision by B-MCMD is improved by about 3 percent, while there is a drop less than 1.0 percent in term of the recall rate. The improved precision can be attributed to the reduced false alarms caused by moving satellite platform. As highlighted in the red regions in Fig. 5, while λ_2 is increased, less foreground objects are mistakenly extracted on the stationary background objects with local misalignment. Further increase of λ_2 may lead to a limited gain on precision while it would cause a small drop in term of recall, which should be owning to over-penalizing slow moving objects with low moving-confident scores. Therefore, thanks to the proposed foreground regularization, B-MCMD can achieve better precision than E-LSD with little loss on the recall rate with a wide range of selection of λ_2.

As observed in Figs. 4 and 5, when λ_2 decreases toward zero, like $\lambda_2 = 10^{-6}$, the same outputs are generated by B-MCMD and E-LSD. This is because MCMD is degenerated to E-LSD [14] when $\lambda_2 = 0$, and in this case the confidence-assisted regularization no longer makes contribution to the objective function in Eq. (6).

5.2 Performance Evaluation

To verify the superior performance by MCMD, four batch-based state-of-the-art approaches, which are RPCA [24], DECOLOR [11], LSD [13] and E-LSD [14], and four state-of-the-art online approaches, GRASTA [43], OR-PCA [41], GOSUS [31] and O-LSD [15], were selected for comparison.

For B-MCMD, we set $\lambda_1 = 1/\sqrt{\max\{p,n\}}$, $\lambda_2 = 0.001$ and $\lambda_3 = \lambda_1/5$, and for O-MCMD, we set $\lambda_1 = 1/\sqrt{\max\{p,n\}}$, $\lambda_2 = 10 \times \lambda_1$ and $\lambda_3 = 10^3$. When computing the dense optical flows, we empirically selected the temporal stride for extracting the anchor frames as 10 and 3 for SkySat dataset and Jilin-1 dataset, respectively, based on their frame rates. We set $\alpha = 10$, and β was selected as 1.87 and 1.67 for SkySat dataset and Jilin-1 dataset, respectively, which roughly correspond to a ground speed of 20 km/h.[3]

SkySat Dataset: Compared with the state-of-the-art MOD algorithms, the proposed B-MCMD and O-MCMD methods both achieve improved performance with reduced false alarm rates. As presented in Fig. 6, the detection results by existing methods exhibit a high rate of false alarms on stationary background objects. By introducing the moving-confidence-assisted foreground regularization by B-MCMD and O-MCMD few false alarms are generated on background objects. Quantitatively, B-MCMD and O-MCMD achieve the highest overall performance with $F_1 = 79.54\%$ and $F_1 = 73.79\%$ among the batch methods and online methods, respectively, as presented in Table 2. Compared with the existing best performer E-LSD, B-MCMD has a small reduction in terms of recall, which should be owing to the overlapping in optical flow between the real slow moving objects and the local misalignment on stationary background objects. For O-MCMD, the introduced regularization helps improve the online estimation of the low-rank subspace, which leads to improved recall rates. Compared with the batch methods,

3. It is worth noting that fine-tuning on these parameters may improve the MOD performance further.

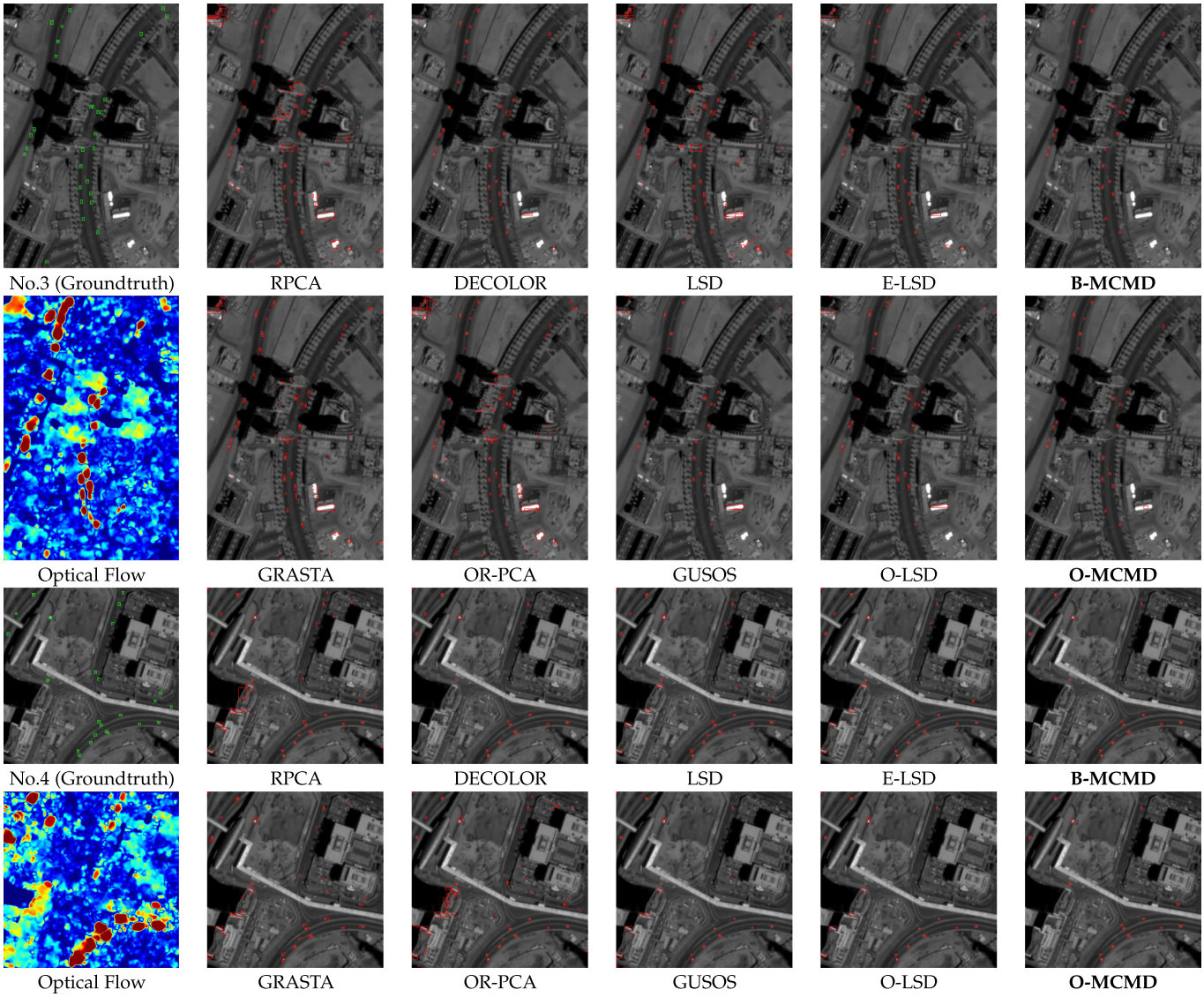


Fig. 6. Optical flow results and the detection results obtained by the proposed B-MCMD, O-MCMD and eight existing state-of-the-art methods on selected videos from SkySat dataset.

O-MCMD even outperforms the batch method LSD by about 10 percent in term of F_1.

Jilin-1 Dataset: B-MCMD and O-MCMD achieve superior MOD performance against the state-of-the-art batch and online methods on Jilin-1 dataset in term of precision and F_1. As illustrated in Fig. 7, false alarms are generated by the existing methods on stationary background objects, which is owing to the local misalignment caused by moving satellite platform. With proposed moving-confidence-assisted regularization on the foreground, B-MCMD and O-MCMD are able to suppress these false alarms. Quantitatively, as shown in Table 3, B-MCMD achieves the highest F_1 score of 71.57 percent, and it boosts the MOD performance by around 10 percent from E-LSD, which achieves the second highest F_1. O-MCMD achieves the highest F_1 among the online approaches, and its performance is close to the current best batch performer, E-LSD.

In summary, the proposed MCMD model with the moving-confidence-assisted foreground regularization boosts the MOD performance by both batch and online optimization algorithms presented in this paper.

5.3 An Ablation Study

The proposed MCMD model is built upon two foreground regularizations —a structured sparse foreground regularization and a moving-confidence-assisted foreground regularization. An ablation study is performed to validate the effectiveness of each part. The sparse foreground regularization can be element-wised or structure-based. Both versions are tested, so five different combinations are formed.

The ablation study results are given in Table 4. Model (A), Model (B) and Model (C) are three baseline models where only the element-wise sparse foreground regularization or the moving-confidence-assisted foreground regularization is adopted, or both are adopted, respectively. The proposed MCMD model adopts the structured sparse foreground regularization and the moving-confidence-assisted foreground regularization jointly. We can see the structured sparse regularization improves the MOD performance than element-wised, as it can address the noise issue in MOD better. Together with the moving-confidence-assisted foreground regularization, the proposed MCMD achieves the highest F1 scores for both datasets.

TABLE 2
Detection Performance Comparison Against the State-of-the-Art Algorithms on SkySat Dataset

Method		No.1			No.2			No.3		
		Rcll	Prec	F_1	Rcll	Prec	F_1	Rcll	Prec	F_1
Batch	RPCA	91.65%	74.50%	82.19%	84.02%	24.65%	38.12%	92.44%	20.96%	34.17%
	DECOLOR	74.69%	80.39%	77.44%	60.28%	68.71%	64.21%	77.86%	75.02%	76.41%
	LSD	86.80%	70.79%	77.98%	68.68%	52.43%	59.46%	84.55%	37.78%	52.23%
	E-LSD	81.71%	93.30%	87.12%	64.56%	79.52%	71.27%	78.43%	76.96%	77.69%
	B-MCMD	79.69%	95.27%	86.78%	60.65%	89.62%	72.35%	73.74%	92.04%	81.88%
Online	GRASTA	78.58%	51.27%	62.05%	66.11%	20.56%	31.36%	67.45%	14.25%	23.52%
	OR-PCA	89.14%	60.06%	71.77%	75.29%	19.93%	31.52%	81.08%	20.64%	32.90%
	GUSOS	70.77%	68.55%	69.64%	58.42%	33.28%	42.41%	54.19%	47.66%	50.72%
	O-LSD	74.71%	91.24%	82.15%	51.12%	51.03%	51.08%	50.63%	68.06%	58.07%
	O-MCMD	72.64%	93.91%	81.92%	59.56%	84.79%	69.97%	67.74%	90.72%	77.57%
Method		No.4			No.5			Avg		
		Rcll	Prec	F_1	Rcll	Prec	F_1	Rcll	Prec	F_1
Batch	RPCA	94.90%	31.09%	46.84%	87.21%	42.60%	57.24%	**90.04**%	38.76%	51.71%
	DECOLOR	61.42%	80.67%	69.74%	69.82%	79.22%	74.22%	68.81%	76.84%	72.40%
	LSD	79.84%	51.85%	62.87%	78.47%	60.66%	68.43%	79.63%	54.70%	64.19%
	E-LSD	75.16%	69.29%	72.09%	74.24%	81.86%	77.86%	74.82%	80.19%	77.14%
	B-MCMD	69.30%	90.03%	78.32%	71.16%	87.23%	78.38%	70.91%	**90.84**%	**79.54**%
Online	GRASTA	63.02%	16.17%	25.74%	55.06%	20.59%	29.97%	66.04%	24.57%	34.53%
	OR-PCA	86.64%	28.12%	42.46%	82.00%	37.50%	51.46%	**82.83**%	33.25%	46.02%
	GUSOS	55.72%	40.55%	46.94%	37.63%	34.58%	36.04%	55.35%	44.92%	49.15%
	O-LSD	41.21%	57.39%	47.97%	53.55%	70.31%	60.80%	54.24%	67.61%	60.01%
	O-MCMD	59.44%	84.85%	69.91%	61.81%	79.58%	69.58%	64.24%	**86.77**%	**73.79**%

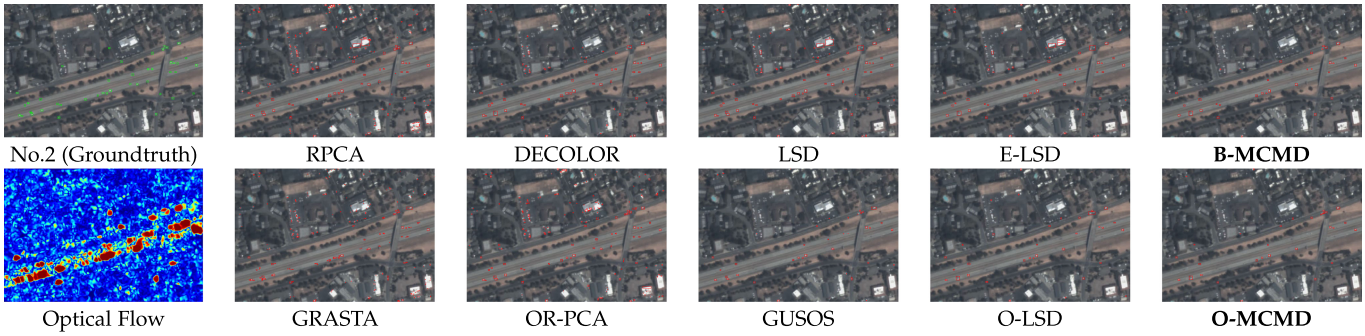


Fig. 7. Optical flow results and the detection results obtained by the proposed B-MCMD, O-MCMD and eight existing state-of-the-art methods on selected videos from Jilin-1 dataset.

5.4 Compute Complexity and Running Time

For processing a video with n frames, the computation complexity for each iteration in B-MCMD is $\mathcal{O}(n(p^2+p+\sum_{g\in\mathcal{G}}|g|))$, where $|g|$ is the number of elements in g. If B-MCMD takes k iterations to converge, the overall computation complexity sums up to $\mathcal{O}(k[n(p^2+p+\sum_{g\in\mathcal{G}}|g|)])$. In

TABLE 3
Detection Performance Comparison Against the State-of-the-Art Algorithms on Jilin-1 Dataset

Method	No. 6			No. 7			No. 8			Avg		
	Rcll	Prec	F_1	Rcll	Prec	F_1	Rcll	Prec	F_1	Rcll	Prec	F_1
RPCA	82.40%	10.08%	17.97%	90.17%	17.57%	29.41%	88.45%	15.45%	26.31%	**87.01**%	13.37%	24.56%
DECOLOR	60.73%	27.50%	37.86%	75.51%	47.39%	58.24%	73.18%	69.96%	71.54%	69.81%	48.28%	55.88%
LSD	73.53%	36.49%	48.78%	80.97%	47.24%	59.67%	79.13%	55.19%	65.03%	77.88%	46.31%	57.83%
E-LSD	71.86%	45.40%	55.65%	76.31%	50.15%	65.26%	71.99%	62.32%	66.81%	73.39%	52.62%	62.57%
B-MCMD	57.21%	69.46%	62.74%	71.33%	80.18%	75.50%	68.45%	86.59%	76.46%	65.66%	**78.74**%	**71.57**%
GRASTA	58.72%	8.68%	15.12%	70.87%	14.84%	24.55%	73.22%	23.08%	35.10%	67.60%	15.53%	24.92%
OR-PCA	67.99%	10.45%	18.12%	86.30%	21.25%	34.10%	86.15%	13.62%	23.52%	**80.15**%	15.11%	25.25%
GUSOS	60.79%	21.55%	31.82%	65.39%	59.29%	62.19%	60.83%	49.78%	54.75%	62.34%	43.54%	49.59%
O-LSD	53.25%	33.95%	41.47%	56.79%	66.25%	61.16%	56.85&	65.98%	61.08%	55.63%	55.39%	54.57%
O-MCMD	48.54%	54.44%	51.32%	57.10%	77.00%	65.57%	55.86%	73.16%	63.34%	53.83%	**68.20**%	**60.08**%

TABLE 4
Ablation Study in Terms of the F_1 Score for MOD on the SkySat and Jilin-1 Datasets

Model	Configuration		Dataset	
	Sparse Foreground Regularization	Moving-Confidence-Assisted Foreground Regularization	SkySat	Jilin-1
(A)	✓ (Element-wise)	-	55.91%	24.83%
(B)	-	✓	60.89%	46.73%
(C)	✓ (Element-wise)	✓	76.56%	60.75%
(D)	✓ (Structured)	-	77.14%	62.57%
B-MCMD	✓ (Structured)	✓	**79.54%**	**71.57%**

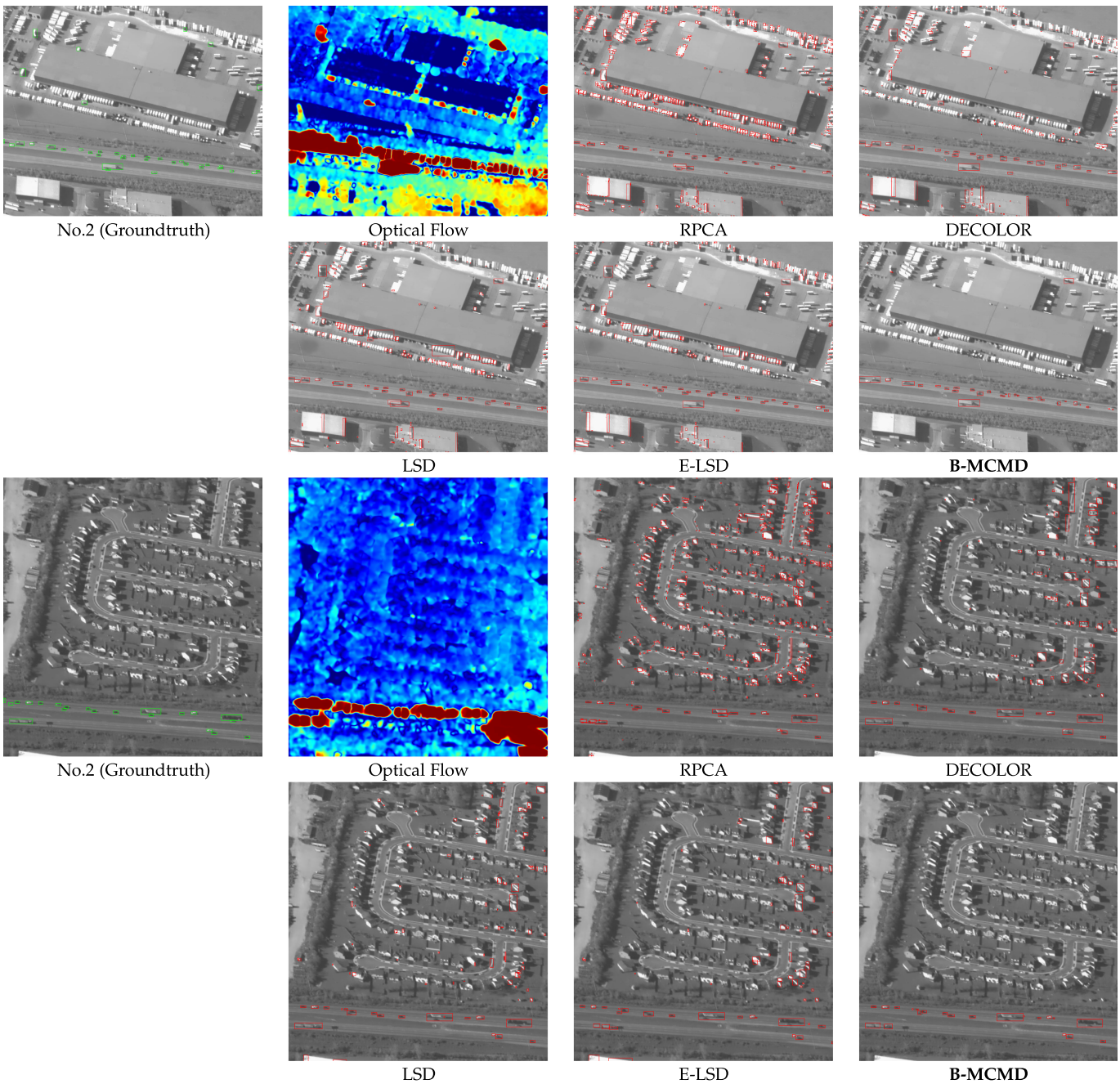


Fig. 8. Optical flow results and the detection results obtained by the proposed B-MCMD method and four existing state-of-the-art methods on selected videos from CLIF 2006 dataset.

terms of the processing time, in our experiments, B-MCMD took 3600 seconds to process a video with 500 frames in size of 400×400, whereas E-LSD and LSD took 2736 seconds and 13030 seconds for processing the same video, respectively. As the computation complexity of B-MCMD grows quadratic with respect to p, B-MCMD is suitable for off-board processing for frames in moderate size.

For O-MCMD, the computation complexity is determined by the iterative optimization in Algorithm 2. For each iteration in Algorithm 3, the complexity for estimating foreground and background by Eq. (27) is $\mathcal{O}(k_1[r^3 + k_2(p + \sum_{g \in \mathcal{G}} |g|])$, where k_1 is the number of iterations in BCD for solving Eq. (27) and k_2 is the number of iterations for Algorithm 3 to converge. The complexity for updating the subspace basis is $\mathcal{O}(r^2 + pr + pr^2)$. This sums up to $\mathcal{O}(r^2 + pr + pr^2 + k_1[r^3 + k_2(p + \sum_{g \in \mathcal{G})|} |g|))$ for processing a frame in Algorithm 2. In practice, k_1 usually takes a number smaller than 5, thus encoding the foreground by Eq. (30) occupies the majority of computation complexity in O-MCMD. For O-MCMD, it took 33 seconds on average for processing a frame in the size of 400×400, and the iterative optimization contributes the majority of the time for solving Eq. (30). O-MCMD is more suitable for online applications and large frames.

5.5 Experiments on Wide Area Surveillance Videos

In this paper, the proposed MCMD model was also tested on the challenging CLIF 2006 dataset that is captured by Wide Area Surveillance (WAS) system. In WAS, the videos are collected from a high-altitude aerial platform and exhibit global platform motion. Detecting moving objects from them is severely affected by the local misalignment on stationary background objects.

The CLIF 2006 dataset[4] is collected at the rate of 2 frames per second. The frames are pre-processed through intensity adjustment, mosaic and frame-to-frame alignment [44], [45]. From the pre-processed videos of CLIF 2006, two videos in size of 650×800 and 700×650 are selected for performance evaluation. They contain 40 and 35 frames, respectively.[5] The boundary boxes for moving vehicles on them are manually annotated.

As presented in Fig. 8, the non-zero optical flow on stationary background objects indicates the occurrence of local misalignment between frames, and false alarms are generated on them by RPCA, DECOLOR, LSD and LSD methods. With the introduced moving-confidence-assisted foreground regularization, the proposed B-MCMD method succeeds in suppressing those false alarms due to their low moving-confidence scores. Quantitatively, MCMD achieves improved MOD performance when compared with state-of-the-art approaches. As shown in Table 5, the precision by B-MCMD is over 50 and 40 percent higher than the existing best performer E-LSD on both videos, and the F_1 scores are greatly improved by B-MCMD. This improvement should be attributed to the introduced moving-confidence-assisted regularization on the foreground.

4. CLIF 2006 dataset is available from https://www.sdms.afrl.af.mil/index.php?collection=clif2006.

5. Given the short length of both videos, the experiments in this part are conducted by batch MOD methods.

TABLE 5
Detection Performance Comparison Against Four State-of-the-Art Methods on the CLIF 2006 Dataset

	Video Size	Method	Rcll	Prec	F_1
a	$650 \times 800 \times 40$	RPCA	63.89%	2.83%	5.42%
		DECOLOR	55.38%	7.93%	13.88%
		LSD	74.39%	10.97%	19.12%
		E-LSD	73.75%	12.56%	21.47%
		B-MCMD	61.59%	64.39%	62.96%
b	$700 \times 650 \times 35$	RPCA	67.38%	1.86%	3.61%
		DECOLOR	66.67%	6.91%	12.53%
		LSD	72.82%	8.86%	15.80%
		E-LSD	69.38%	10.73%	18.59%
		B-MCMD	62.95%	55.35%	58.90%

6 Conclusion

In satellite videos, MOD performance of matrix decomposition methods is severely challenged by the motion of satellite platforms, in particular. In this paper, we cope with the unaddressed false alarm issue caused by the motion of satellites. To achieve such goal, we introduce a moving-confidence matrix using dense optical flow and develop a moving-confidence-assisted foreground regularization, by which the foreground elements with low moving confidences are penalized. By imposing such regularization on the foreground, a novel low-rank decomposition model, which is named as **M**oving **C**onfidence-Assisted **M**atrix **D**ecomposition (MCMD), is established. Two optimization solutions are developed to the proposed model for batch handling (B-MCMD) and online processing (O-MCMD), respectively. B-MCMD can produce better detection performance as expected. The online implementation, O-MCMD, can meet the requirement for on-board processing and online processing, to save data storage and transmission cost. O-MCMD can also be applied offline for large satellite videos, when batch optimization scheme exceeds the capacity of current computation systems. Compared with the dominating supervised object detectors, the proposed MCMD model in this paper does not require for manual annotations. The results from MCMD have the potential to be used as training samples for training a powerful Deep Convolutional Neural Network-based object detector, which may alleviate the burden of manually annotating training samples.

Acknowledgments

This work was supported by Australia Research Council (ARC) grants with projects IDs: LP180100663, DP190103662, and DP190103660. The authors would like to thank Planet Team for providing the data in this research [1].

References

[1] "Planet application program interface: In space for life on Earth", Planet Team, 2017. Accessed: Jul. 1, 2020. [Online]. Available: https://api.planet.com

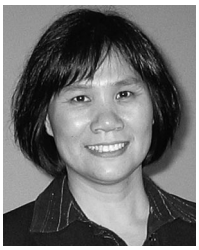
[2] Y. Luo, L. Zhou, S. Wang, and Z. Wang, "Video satellite imagery super resolution via convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2398–2402, Dec. 2017.

[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Advances Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[4] W. LiuD. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
[6] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.
[7] K. Li, G. Cheng, S. Bu, and X. You, "Rotation-insensitive and context-augmented object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2337–2348, Apr. 2018.
[8] P. Ding, Y. Zhang, W.-J. Deng, P. Jia, and A. Kuijper, "A light and faster regional convolutional neural network for object detection in optical remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 141, pp. 208–218, 2018.
[9] W. Liu, L. Ma, and H. Chen, "Arbitrary-oriented ship detection framework in optical remote-sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 6, pp. 937–941, Jun. 2018.
[10] Y. Xu *et al.*, "Gliding vertex on the horizontal bounding box for multi-oriented object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1452–1459, Apr. 2021.
[11] 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.
[12] M. Shakeri and H. Zhang, "Corola: A sequential solution to moving object detection using low-rank approximation," *Comput. Vis. Image Understanding*, vol. 146, pp. 27–39, 2016.
[13] 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.
[14] 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.
[15] J. Zhang, X. Jia, J. Hu, and J. Chanussot, "Online structured sparsity-based moving-object detection from satellite videos," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6420–6433, Sep. 2020.
[16] 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, pp. 1944–1957, 2020.
[17] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2233–2246, Nov. 2012.
[18] S. E. Ebadi and E. Izquierdo, "Foreground segmentation with tree-structured sparse RPCA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2273–2280, Sep. 2018.
[19] K. Gilman and L. Balzano, "Panoramic video separation with online grassmannian robust subspace estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2019, pp. 643–651.
[20] T. Bouwmans and E. H. Zahzah, "Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance," *Comput. Vis. Image Understanding*, vol. 122, pp. 22–34, 2014.
[21] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Comput. Sci. Rev.*, vol. 23, pp. 1–71, 2017.
[22] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, "On the applications of robust PCA in image and video processing," *Proc. IEEE*, vol. 106, no. 8, pp. 1427–1457, Aug. 2018.
[23] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Advances Neural Inf. Process. Syst.*, 2011, pp. 612–620.
[24] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, 2011, Art. no. 11.
[25] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Advances Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.
[26] T. Zhou and D. Tao, "GoDec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 33–40.
[27] X. Cao, L. Yang, and X. Guo, "Total variation regularized RPCA for irregularly moving object detection under dynamic background," *IEEE Trans. Cybern.*, vol. 46, no. 4, pp. 1014–1027, Apr. 2016.
[28] Y. Xu, Z. Wu, J. Chanussot, M. Dalla Mura, A. L. Bertozzi, and Z. Wei, "Low-rank decomposition and total variation regularization of hyperspectral video sequences," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1680–1694, Mar. 2018.
[29] B. Xin, Y. Tian, Y. Wang, and W. Gao, "Background subtraction via generalized fused lasso foreground modeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 4676–4684.
[30] R. Jenatton, J.-Y. Audibert, and F. Bach, "Structured variable selection with sparsity-inducing norms," *J. Mach. Learn. Res.*, vol. 12, no. Oct, pp. 2777–2824, 2011.
[31] J. Xu, V. K. Ithapu, L. Mukherjee, J. M. Rehg, and V. Singh, "Gosus: Grassmannian online subspace updates with structured-sparsity," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 3376–3383.
[32] S. S. Beauchemin and J. L. Barron, "The computation of optical flow," *ACM Comput. Surv.*, vol. 27, no. 3, pp. 433–466, 1995.
[33] K. R. Aires, A. M. Santana, and A. A. Medeiros, "Optical flow using color information: preliminary results," in *Proc. ACM Symp. Appl. Comput.*, 2008, pp. 1607–1611.
[34] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Proc. Scand. Conf. Image Anal.*, 2003, pp. 363–370.
[35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[36] J. Mairal, R. Jenatton, F. R. Bach, and G. R. Obozinski, "Network flow algorithms for structured sparsity," in *Proc. Advances Neural Inf. Process. Syst.*, 2010, pp. 1558–1566.
[37] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, "Proximal methods for hierarchical sparse coding," *J. Mach. Learn. Res.*, vol. 12, pp. 2297–2334, 2011.
[38] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
[39] J. Shen, P. Li, and H. Xu, "Online low-rank subspace clustering by basis dictionary pursuit," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 622–631.
[40] S. J. Wright, "Coordinate descent algorithms," *Math. Program.*, vol. 151, no. 1, pp. 3–34, 2015.
[41] J. Feng, H. Xu, and S. Yan, "Online robust PCA via stochastic optimization," in *Proc. Advances Neural Inf. Process. Syst.*, 2013, pp. 404–412.
[42] Z. Gao, L.-F. Cheong, and M. Shan, "Block-sparse RPCA for consistent foreground detection," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 690–703.
[43] J. He, L. Balzano, and A. Szlam, "Incremental gradient on the grassmannian for online foreground and background separation in subsampled video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1568–1575.
[44] V. Reilly, H. Idrees, and M. Shah, "Detection and tracking of large number of targets in wide area surveillance," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 186–199.
[45] I. Saleemi and M. Shah, "Multiframe many–many point correspondence for vehicle tracking in high density wide area aerial videos," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 198–219, 2013.

Junpeng Zhang (Student Member, IEEE) received the BSc degree from the China University of Mining and Technology, Xuzhou, China, in 2013, and the master's degree in surveying engineering from the China University of Mining and Technology, Xuzhou, China, in 2016. He is currently working toward the PhD degree in electrical engineering from the University of New South Wales, Australia. His research interests include object detection and tracking in remote sensing imaginary. He was the winner of "DSTG Best Contribution to Science Award" in Digital Image Computing: Techniques and Applications 2018 (DICTA 2018). He served as the chair of IEEE GRSS UNSW Canberra Student Chapter during 2020.



Xiuping Jia (Fellow, IEEE) received the BEng degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982, and the PhD degree in electrical engineering from the University of New South Wales, Australia, in 1996. Since 1988, she has been with the School of Engineering and Information Technology, The University of New South Wales at Canberra, Australia, where she is currently an associate professor. Her research interests include remote sensing, image processing and spatial data analysis. She has authored or coauthored more than 270 referred papers, including more than 160 journal papers, addressing various topics including data correction, feature reduction and image classification using machine learning techniques. She has co-authored of the remote sensing textbook titled Remote Sensing Digital Image Analysis [Springer-Verlag, 3rd (1999) and 4th eds. (2006)]. She is a subject editor for the Journal of Soils and Sediments and serves as an associate editor-in-chef of the *IEEE Transactions on Geoscience and Remote Sensing*.

Jiankun Hu (Senior Member, IEEE) receive the PhD degree in control engineering from the Harbin Institute of Technology, China, in 1993, and the master's degree in computer science and software engineering from Monash University, Australia, in 2000. He was a research fellow with Delft University, The Netherlands, from 1997 to 1998, and The University of Melbourne, Australia, from 1998 to 1999. He is currently a full professor of cyber security with the School of Engineering and Information Technology, the University of New South Wales at Canberra, Australia. His main research interest include the field of cyber security, including biometrics security, where he has published many papers in high-quality conferences and journals including the *IEEE Transactions on Pattern Analysis and Machine Intelligence*. He has served on the editorial boards of up to seven international journals and served as a Security Symposium Chair of the IEEE Flagship Conferences of IEEE ICC and IEEE GLOBECOM. He has obtained nine Australian Research Council (ARC) Grants. He served at the prestigious Panel of Mathematics, Information and Computing Sciences, ARC ERA (The Excellence in Research for Australia) Evaluation Committee 2012.

Kun Tan (Senior Member, IEEE) received the BS degree in information and computer science from Hunan Normal University, Hunan, China, in 2004, and the PhD degree in photogrammetric and remote sensing from the China University of Mining and Technology, Jiangsu, China, in 2010. He was a Joint PhD candidate in Remote Sensing at Columbia University, USA during September 2008 to September 2009. From 2010 to 2018, he has been with the Department of Surveying, Mapping and Geoinformation, China University of Mining and Technology, Xuzhou City, Jiangsu Province, China. He is currently a professor at East China Normal University, Shanghai, China. His research interests include hyperspectral image classification and detection, spectral unmixing, quantitative inversion of land surface parameters, and urban remote sensing.