2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

A Novel Method for Segmenting Moving Objects in Aerial Imagery using Matrix Recovery and Physical Spring Model

Agwad ElTantawy, Mohamed S. Shehata
Department of Electrical and Computer Engineering
Faculty of Engineering and Applied Science
Memorial University of Newfoundland, Canada
(agwad.eltantawy, mshehata)@mun.ca

Abstract

Aerial imagery applications have gained a great interest especially in the area of comprehensive ground activities analysis. One of the key tasks in such applications is moving objects segmentation. Although many efforts have been presented in the literature that claim high true object detection rates, they still suffer from high false positive rates. This paper focuses on maintaining a high true positive detection rates while significantly reducing the false positive detection rates. To achieve this goal, this paper proposes a novel method that integrates matrix recovery concept with physical spring model to drastically reduce false detections. The proposed method segment all candidate moving objects by recovering the low rank matrix, which normally results high false positive detection. To reject false detections, each candidate moving object is modelled as a mass suspended by system of springs, such that the forces of springs attached to false detections is negligible whereas the forces of springs attached to a true moving object will be significant in response to the object motion. The results show that the proposed method, compared to other current state-of-the-art methods, achieved better true positive rates while drastically lowering the false positive rates.

1. Introduction

Aerial imagery applications, such as traffic monitoring [15], oil spill detection [10], search and rescue [17], reconnaissance missions [21], etc., have been flourished in the past few years. They provide robust and comprehensive analysis of complex activities by using aerial imagery which provides bird-eye view of the ground. The mobility of the platforms used for acquiring aerial imagery introduces a challenge facing many fundamental processes including automatic moving objects segmentation where static background elements can be falsely seen as moving. In spite of the efforts proposed [9] [3] in this regard, the false detection rate is still significant (it will be explained in the results section).

The paper focuses on maintaining a high true positive detection rates while significantly reducing the false positive detection rates. To achieve this goal, this paper proposes a novel method for segmenting moving objects in aerial imagery which have a very low false detection rates. Firstly, it utilizes the concept of matrix recovery [18] which considers moving objects as a noise corrupting original structure of a scene, i.e. underlying background. It segments all candidate moving objects in the scene by recovering underlying background using low-rank matrix optimization technique, i.e. inexact augmented Lagrange multiplier (IALM) [12]. Secondly, the resulted candidate moving objects are modelled as a mass connected to landmarks via system of springs. The motion of true moving objects compresses or stretches the springs causes a significant forces in the springs. In contrary, false detections do not make vibration on the spring consequently the forces of the springs are negligible. This evidence is used as indicator to eliminate false detections. The robustness of the proposed method is demonstrated on different sequences of DARPA VIVID dataset [5].

The rest of this paper is organized as follows: section 2 presents a literature review of matrix recovery methods for segmenting moving objects. The demonstration of the proposed method is presented in section 3. section 4 shows the experimental results of the proposed method compared with relative state-of-the-art. Section 5 concludes the paper and gives future directions.

2. Literature Review

Matrix recovery using convex program has flourished with robust segmentation of moving objects. Candes et al. proposed principal component pursuit (PCP) [4] which decomposes a video into moving objects and background through sparsity estimation and low-rank matrix recovery. Sparsity estimation is achieved by minimizing $l_1 norm$ while low-rank matrix is recovered by minimizing nuclear norm. PCP uses equality constraint for the minimization where it assumes the video is exactly equal to moving objects and the background. In 3-term decomposition method [13], the equality constraint is more fixability to considere two type of sparsity in a video, i.e. moving objects and turbulence. So, the video is decomposed using inexact augmented Lagrange multiplier into background, moving objects and turbulence. However, 3-term decomposition method is still susceptible to high false detection in case of the existence of high contrast spot in the background.
Wohlberg et al. [19] used the fact of moving objects should be connected across the frames to reduce the false detections. Total variation penalty term was introduced to the minimization problem which ensures temporal connectivity of moving objects. On the other side, Xin et al. in [20] depend on the fact of moving objects are correspond to meaningful objects such as human, cars, etc., therefore moving objects' pixels are spatially connected. For this purpose, Fused Lasso term is added to the minimization problem to penalize the spatial connectivity between moving object pixels. However, The former methods suffer from very high false positive detections when moving camera platforms is used for acquiring the video as the background elements appear falsely moving. Also, the matrix recovery concept requires that the original structure of the matrix to be linearly correlated.
RASL [14] introduced a term represents transformation domain between the consecutive frames to reduce the effect of moving camera platform and ensure that the original structure of the matrix are linearly correlated. The transformation domain is calculated using iterative first order Taylor series [7] to alignment the video's images. Therefore, the video is decomposed into background, moving objects and the transformation domain. In RASL, the decomposition is achieved by augmented Lagrange multiplier. DEtecting Contiguous Outliers in the Low-rank Representation (DECOLOR) [22] calulates the transformation domain same as RASL, while uses SOFT-IMPUTE algorithm video decomposition. Eltantawy et. al in method [6] achieved a better performance for segmenting small-size moving object from high altitude moving camera platforms, i.e. airplane, by utilizing IALM for video decomposition. Although, RASL, DECOLOR and method in [6] achieved efficient segmentation with moving camera platforms, it is still trade of between the true positive detections and false positive detections.
The motivation of this paper is segmenting moving objects from aerial imagery such that maintaining high true positive detections and dramatically reducing false positive detections. To the best of the authors knowledge, the proposed method in this paper is the first moving object segmentation method that utilizes the physical spring model with matrix recovery concept to reduce the false positive detections without sacrificing the performance. The results outperform current state-of-the-art methods as shown in the results section.

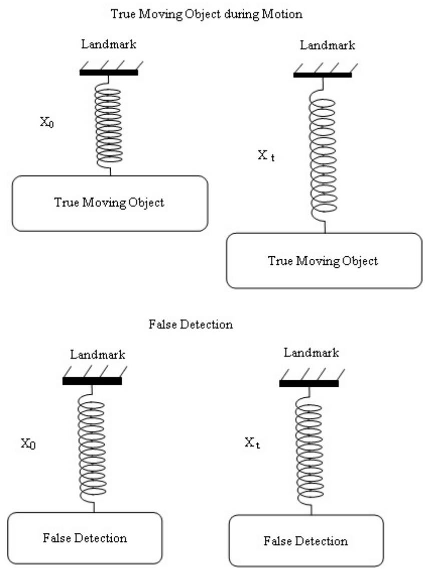

Figure 1. Spring model for true moving object and false detection

3. The Proposed Method

3.1. Problem Formulation

The proposed method segments moving objects in aerial imagery by integrating matrix recovery concept with physical spring model. In matrix recovery, segmentation process is tackled as a low-rank matrix optimization problem in which moving objects are considered as sparse noise that corrupts the videos original structure (the underlying background). However, matrix optimization normally results high false detection rate. The physical spring model differentiates between the false detections and true moving objects. In this model, a moving object is considered as a mass suspended to a spring, and the spring base is attached to a landmark. The mass of true moving objects should make this spring vibrating, i.e compressed or stretched, overtime due to object motion, as shown in Figure 1. On the other side, if false detection is examined using the same physics spring model, the vibration of the spring will negligible. Hence, the segmentation problem can be formulated as follows:

$$SM(\min_{B,O_{Cndt}} \|B\|_* + \lambda\|O_{Cndt}\|_1, s.t.\ F = B + O_{Cndt}) \tag{1}$$

where SM denotes spring model that rejects all possible false detections in the candidate moving objects O_{Cndt}, B denotes the underlying background, F is the frames matrix which contains set of frames stacked as vector column, $\|X\|_*$ is the nuclear norm, and finally $\|X\|_1$ denotes $L_1 - norm$ such that $\|X\|_1 = \Sigma_{ij}(|X_{ij}|)$.
For completeness the underlying background B must be linearly correlated which is not the case in aerial imagery. Thus, a transformation domain T is added to Eq.(1) as follows:

$$SM(\min_{B,O_{Cndt}} \|B\|_* + \lambda\|O_{Cndt}\|_1, s.t.\ FoT = B + O_{Cndt}) \tag{2}$$

3.2. Matrix Recovery

The proposed method uses IALM to solve the optimization problem in Eq.(2) by minimizing Eq.(3-5) to recover the values for transformation domain T, background B and candidate moving objects O_{Cndt}:

$$T = arg\min_{T} \ell(T, B, O_{Cndt}, Y) \tag{3}$$

$$B = arg\min_{B} \ell(T, B, O_{Cndt}, Y) \tag{4}$$

$$O_{Cndt} = arg\min_{O_{Cndt}} \ell(T, B, O_{Cndt}, Y) \tag{5}$$

where Y is Lagrange multiplier. To calculate the transformation domain T, a linearisation of non-linear constraint in Eq.(2) is needed. Therefore, Taylor series approximation is used to reformulate this constraint in a linear form as shown in Eq.(6)

$$FoT + \Sigma_i^n J\triangle t\varepsilon_i\varepsilon_i^T = B + O_{Cndt}, \tag{6}$$

Then, The value of T is iteratively refined by estimating the value of $\triangle t$.

$$T = T + arg\min_{\triangle t} \ell(\triangle t, B, O_{Cndt}, Y) \tag{7}$$

The closed-form solution of $\triangle t$ is:

$$\begin{aligned}\triangle t =& \langle Y, \Sigma_i^n J\triangle t\varepsilon_i\varepsilon_i^T\rangle + \\ & \frac{\mu}{2}\|FoT + \Sigma_i^n J\triangle t\varepsilon_i\varepsilon_i^T - B - O_{Cndt}\|_f^2\end{aligned} \tag{8}$$

The background is recovered by singular value decomposition algorithm [8] which leads to the following closed-form solution:

$$(U, S, V) = svd(FoT + \mu^{-1}Y - O_{Cndt}) \tag{9}$$

$$B = US_{\mu^{-1}}[S]V^T \tag{10}$$

Convex optimization theory [2] is used to obtain a closed-form solution for calculating the candidate moving objects O_{Cndt}:

$$O_{Cndt} = S_{\lambda\mu^{-1}}[FoT + Y - B] \tag{11}$$

The output of the optimization process is candidate moving objects in each frame which may contain false detections, so the physical spring model is applied to eliminate the false detections. Algorithm 1 summarizes the calculation of candidate moving objects.

Algorithm (1)

Input: F, μ and ρ
Initialization:

$$\begin{aligned} T_0 &= InitialWarping(F) \\ FoT_0 &= Alignment(F, T) \\ Y_0 &= sgn(FoT_0)/J(FoT_0) \end{aligned}$$

//**while** loop below estimates B, O_{Cndt}, and T
While not converged **do**

$$\begin{aligned} \triangle t &= \Sigma_i^n J^{\xi}(O_{Cndt,t} + B_t - FoT_t - \mu_t^{-1}Y_t)\varepsilon_i\varepsilon_i^T \\ T_{k+1} &= T_k + \triangle t \\ FoT_{k+1} &= UpdateAlignment(FoT_k, T_{k+1}) \\ (U, S, V) &= svd(FoT_{k+1} + \mu^{-1}Y_k - O_{Cndt,k}) \\ B_{k+1} &= US_{\mu^{-1}}[S]V^T \\ O_{Cndt,k+1} &= S_{\lambda\mu^{-1}}[FoT_{k+1} + Y_k - B_{k+1}] \\ Y_{k+1} &= Y_k + \mu_t(FoT_{k+1} - B_{k+1} - O_{Cndt,k+1}) \\ \mu_{k+1} &= \rho\mu_k \\ k &= k + 1 \end{aligned}$$

End while

Output: O_{Cndt}

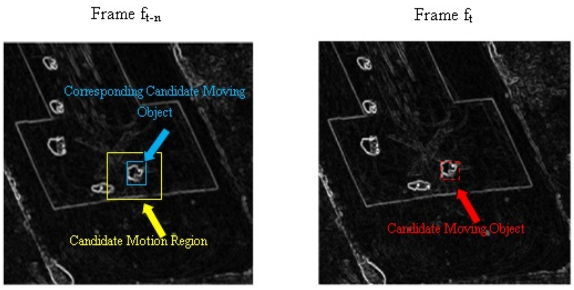

Figure 2. Gradient template matching

3.3. Spring Model

The spring model consists of two steps: objects tracking step and springs forces calculation step. The object tracking step is matching-based where the candidate moving objects in frame f_t are matched to their correspondences in frame f_{t-n}. In case of aerial imagery, matching the candidate moving objects among the frames using features, such as SURF [1], FAST [16], etc., is very difficult as these objects appear very small. Generally, small objects suffer from lack of strong features consequently wrong matching is very frequent. Therefore, the matching in the proposed method is accomplished using gradient template matching, as shown in Figure 2, in which cross-normalization [11] is applied between the gradient of the candidate moving objects and the gradient of frame f_{t-n}. To avoid ambiguous matching, cross-normalization is applied on candidate motion region of frame f_{t-n}. The candidate motion region is an area where the moving object is expected to be in frame f_{t-n}.

In springs forces calculation step, the candidate moving object in frame f_t and its correspondence in frame f_{n-t} modelled as a single mass, i.e. $ObjectMass$, suspended to The system of springs, shown in Figure 3. Such system consists of eight springs arranged in all directions to sense the motion of $ObjectMass$ in any

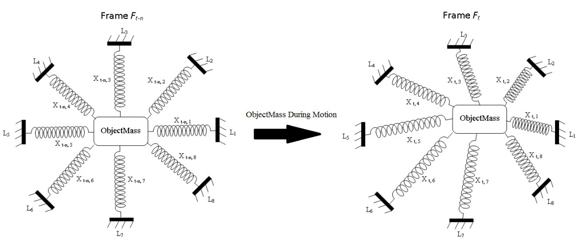

Figure 3. The system of springs for object mass

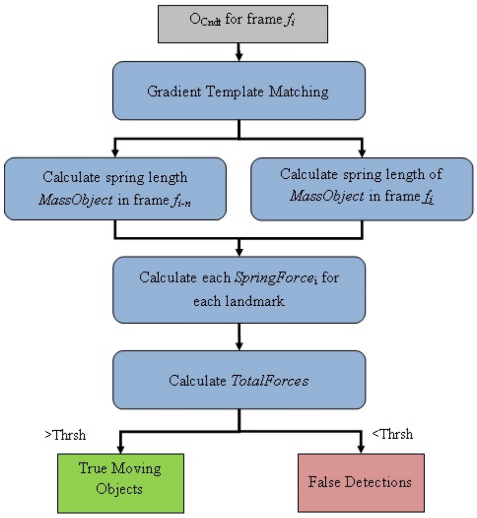

Figure 4. Spring model algorithm

direction, and each spring's base is attached to a landmark (which is manually selected for each sequence). The springs vibrations, compressing and stretching, are tracked for each $ObjectMass$ to calculate the total system of spring forces, $TotalForces$, shown in Eq. 12.

$$TotalForces = \Sigma_{i=1}^{8} SpringForce_i \tag{12}$$

The $SpringForce_i$ refers to a spring force that is calculated as:

$$SpringForce_i = K(X_{t-n} - X_t) \tag{13}$$

X_{t-n} and X_t represent the length of a spring in the system of spring frame f_{t-n} and frame f_t, respectively. K is spring constant and it is equal to be one, in our case. Other forces that may affect the system of springs are ignored, such as gravity force, damping force, etc., as they are irrelevant to moving object segmentation. $TotalForces$ is used to determine if $ObjectMass$ refers to true moving object. in case of true moving objects $TotalForces$ is significant due to the motion of the objects. otherwise, $TotalForces$ are very small, i.e. negligible. Mathematically, the determination of true moving objects can be formulated as:

$$O = \begin{cases} 1, & TotalForces\ of\ ObjectMass\ > Thrsh \\ 0, & TotalForces\ of\ ObjectMass\ < Thrsh \end{cases} \tag{14}$$

where O is the true moving objects at each frame, and $Thrsh$ denotes threshold value that is empirical selected for each sequence. Figure 4 summarizes the detection of true moving objects using spring model.

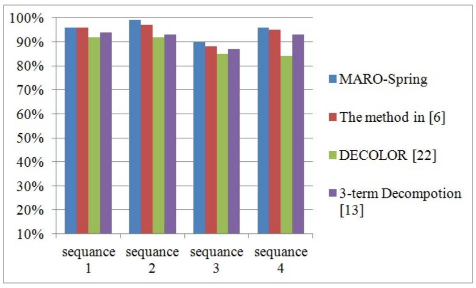

Figure 5. True positive rate

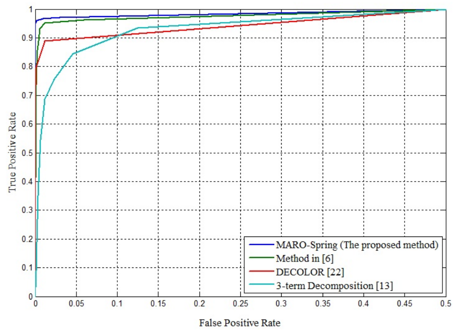

Figure 7. Receiver operating characteristic curve

4. Results

The performance of the proposed method is evaluated using DAPRA VIVID dataset [5] which contains aerial images from different environments, i.e. a runway, dirt road and wooded area. DAPRA VIVID dataset reflects different types of motion including translation, rotation, scaling, or combination. The moving objects sizes in this dataset vary from 15x15 to 50x50 pixels.
Fig. 8 illustrates the results of the proposed method on sample frames compared with other three current state-of-the-art methods, namely: The method in [6], DECOLOR [22] and 3-term decomposition [13]. It is clear that these current state-of-the-art methods suffer in DARPA VIVID dataset while the proposed method performs significantly better especially in reducing the false detections dramatically without sacrificing the true detection rate.
To quantitatively evaluate the proposed method, region-based measure is used as follows: First, connected component labelling is applied on the binary mask resulting from the methods under evaluation to label detected areas. Then, if it overlaps with the ground-truth, then it is considered a correct detection. Finally, True positive rate (TPR) and false positive rate (FPR) are used as accuracy measurement.
By examining Fig. 5, the proposed method provides the best overall accuracy in all sequences where it achives an average of 96% TPR. The method in [6] comes second with an average TPR of 94%. Finally, the 3-term decomposition and DECOLOR methods have TPR of 92% and 88% respectively.

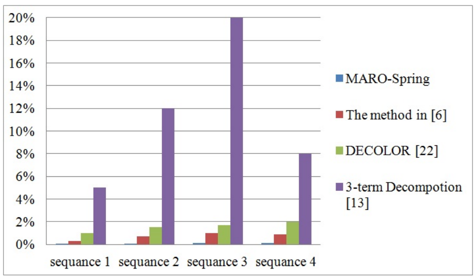

Figure 6. False positive rate

Fig. 6. supports the main claim of this paper where it shows the corresponding FPR for each method. It is evident that the proposed method has the lowest false positive rates, while false detection rates are increased with DECOLOR and the method in [6]. The highest false detections is occurred by 3-term decomposition method.
The receiver operating characteristic (ROC) curve, shown in Fig. 7, depicts the complete relationship between TPR and FPR for each method being evaluated. The proposed method reduces FPR without sacrificing TPR. The mothod in [6] has lower TPR with higher FPR. For DECOLOR, it has low FPR with very low TPR. The 3-term decomposition method shows that it is very sensitive to the false detections, however, its TPR is better than DECOLOR. Overall, Figure 7, shows that the proposed method has the best overall performance with an area under the curve of 0.985.

5. Conclusion

In this paper, a novel method is proposed to segmenting moving objects in aerial imagery. It utilizes matrix recovery concept and physical spring model to eliminate the false detections. The results show that the proposed method has the best accuracy with the lowest false detection compared to relevant current-state-of-the-art methods. Future work includes enhancing the proposed method to meet real-time constrains. Specifically, using IALM gave an advantage to easily parallalize the methods and hence can be implemented on parallel architectures.

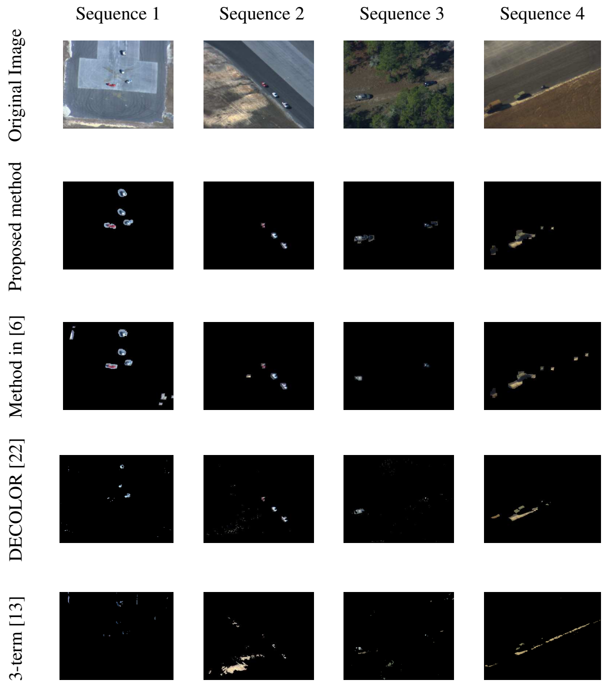

Figure 8. Sample results of the compared method

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *Computer vision–ECCV 2006*, pages 404–417. Springer, 2006.

[2] D. P. Bertsekas. *Convex optimization theory*. Athena Scientific Belmont, MA, 2009.

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

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

[5] R. Collins, X. Zhou, and S. K. Teh. An open source tracking testbed and evaluation web site. In *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, pages 17–24, 2005.

[6] A. ElTantawy and M. S. Shehata. Moving object detection from moving platforms using lagrange multiplier. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 2586–2590. IEEE, 2015.

[7] W. H. Foy. Position-location solutions by taylor-series estimation. *IEEE Transactions on Aerospace and Electronic Systems*, pages 187–194, 1976.

[8] G. H. Golub and C. Reinsch. Singular value decomposition and least squares solutions. *Numerische mathematik*, 14(5):403–420, 1970.

[9] K. A. Joshi and D. G. Thakore. A survey on moving object detection and tracking in video surveillance system. *IJSCE, ISSN*, pages 2231–2307, 2012.

[10] R. Kristiansen, E. Oland, and D. Narayanachar. Operational concepts in uav formation monitoring of industrial emissions. In *Proceedings of the 3rd IEEE International Conference on Cognitive Infocommunications. Kosice, Slovakia*, 2012.

[11] J. Lewis. Fast normalized cross-correlation. In *Vision interface*, volume 10, pages 120–123, 1995.

[12] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.

[13] O. Oreifej, X. Li, and M. Shah. Simultaneous video stabilization and moving object detection in turbulence. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(2):450–462, 2013.

[14] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2233–2246, 2012.

[15] A. Puri, K. Valavanis, and M. Kontitsis. Statistical profile generation for traffic monitoring using real-time uav based video data. In *Control & Automation, 2007. MED'07. Mediterranean Conference on*, pages 1–6. IEEE, 2007.

[16] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1508–1515. IEEE, 2005.

[17] A. Ryan and J. K. Hedrick. A mode-switching path planner for uav-assisted search and rescue. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, pages 1471–1476. IEEE, 2005.

[18] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 853–860. IEEE, 2012.

[19] B. Wohlberg, R. Chartrand, and J. Theiler. Local principal component pursuit for nonlinear datasets. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 3925–3928. IEEE, 2012.

[20] B. Xin, Y. Tian, Y. Wang, and W. Gao. Background subtraction via generalized fused lasso foreground modeling. *arXiv preprint arXiv:1504.03707*, 2015.

[21] W.-L. Yang, L. Lei, and J.-S. Deng. Optimization and improvement for multi-uav cooperative reconnaissance mission planning problem. In *Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2014 11th International Computer Conference on*, pages 10–15. IEEE, 2014.

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