Detection of Multiple Small Moving Targets Against Complex Ground Background

Junhua Yan[1,2](✉), Jingchun Qi[2], Xuyang Cai[2], Yin Zhang[1,2], Kun Zhang[2], and Yue Ma[2]

[1] Key Laboratory of Space Photoelectric Detection and Perception (Nanjing University of Aeronautics and Astronautics), Ministry of Industry and Information Technology, No. 29 Yudao Street, Nanjing 210016, China
yjh9758@126.com

[2] College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, Jiangsu, China

Abstract. To tackle the problem that it is difficult to detect small moving targets accurately against complex ground background, a target detection algorithm that combines target motion information and trajectory association is proposed. To tackle the problem of small target size, firstly, background motion compensation is performed to obtain the background motion parameters. Then, forward and backward motion history maps are calculated to fuse continuous difference images for enhanced motion information of small targets. Finally, morphology processing is used to obtain the area of small moving targets. To tackle the problem of complex background, the Kalman predictor is used to predict the target position, and the Hungarian matching algorithm is used to correlate targets to obtain the target trajectory. Then, based on the target trajectory, targets missed by detection are supplemented to improve the target recall rate and false alarm targets are filtered out to improve the target precision rate. Experimental results show that the proposed algorithm has good detection performance, with the recall rate higher than 93%, the precision rate higher than 92%, and the F-measure higher than 93%.

Keywords: Small target detection · Complex background · Target motion information · Trajectory association · Trajectory feature

1 Introduction

In this paper, we aim to investigate the detection of multiple small moving targets, such as flying UAVs (Unmanned Aerial Vehicle), against complex ground background. When the altitude of the photodetector platform is far above that of UAVs, the targets in the image have few pixels, making them small targets that lack morphology information. When UAVs and the photodetector platform fly above a terrain with complex features, the obtained images may have a complex background. In this paper, we focus on the above difficulties in order to accurately detect multiple small moving targets against complex ground background.

X. Gao et al. (Eds.): MobiHealth 2021, LNICST 440, pp. 293–315, 2022.
https://doi.org/10.1007/978-3-031-06368-8_20

Small moving target detection against complex background has always been a challenging issue, which have drawn extensive research attention.

In 2012, M. Hofmann et al. [1] proposed a method following a non-parametric background modeling paradigm, which adjusted foreground judgment threshold and model update rate according to background complexity. This method performs well when the background is steady but performs badly when targets are small. Siam M et al. [2] extracted FAST (Features from accelerated segment test) corners, and classified targets' optical flow through a clustering algorithm - DBSCAN (Density based spatial clustering of applications with noise) to detect moving targets. This method has difficulty extracting corner points and easily misses targets when the targets are small.

In 2013, Kwang Moo Yi et al. [3] proposed a pixel-based method modeling the background through dual-mode single Gaussian model (SGM) and compensating the motion of the camera by mixing neighboring models. This method is able to detect small targets, but the false alarm rate was high when background is complex. Shen Hao et al. [4] proposed a novel hierarchical moving target detection method based on spatiotemporal saliency. Temporal saliency based on Forward-Backward Motion History Image and spatial saliency is combined to get refined detection results. When the background is complex, the method is prone to miss detection, because the spatial saliency of the target is not significant.

In 2014, Shakeri M et al. [5] applied a two-level registration to estimate the effect of camera motion for motion compensation, extracted target pixels by Gaussian mixture model, refined noisy results using component-based and pixel-based methods, and improved the detection accuracy through the temporal coherence of foreground motion. This method performs well in complex environments, but it easily misses small targets. Sadeghitehran et al. [6] extracted BRISK [7] (Binary Robust Invariant Scalable Key points) corner optical flow features and classified targets' optical flow through ELM (Evolving Local Means) algorithm to detect moving targets. The method can adapt to complex backgrounds, but the detection effect is poor when lacking target texture information.

In 2015, Wang Z et al. [8] computed the 2-dimensional histogram of entropy flow field to estimate background motion, obtained the difference image through background motion compensation, and detected targets by spatial-temporal association. This method omits targets easily when targets are small. Wei Liu et al. [9] used an improved Oriented FAST and Rotated BRIEF (Binary Robust Independent Elementary Features) algorithm to achieve an accurate background moving model and then detected small moving targets in aerial video by multiplying four continuous difference images with morphology processing. The method has good real-time performance, but the detection results depend on the accuracy of the ORB (Oriented FAST and Rotated BRIEF) feature matching results. The target detection effect is poor when the background is complex. Artem Rozantsev et al. [10] used boosted trees algorithm for motion compensation, obtained spatio-temporal image cubes by stacking motion-stabilized image windows over several consecutive frames and detected targets in spatio-temporal image cubes through AdaBoost classifier. In 2017, they substituted CNN (Convolutional Neural Networks) [11] for boosted trees to better adapt to complex background. This method has difficulty obtaining accurate detection results when targets are small.

In 2016, Junhua Yan et al. [12] proposed a detection algorithm for small moving targets based on adaptive threshold segmentation. In this method, the background motion was compensated by pyramid Lucas-Kanade optical flow of feature points, so the difference images were binary segmented using an adaptive segmentation threshold to detect moving targets. This method has poor detection performance in case of complicated background because it is difficult to accurately compensate for background motion. Meng Yi et al. [13] proposed a detection algorithm for small moving targets based on multi-view aerial registration system. This method extracted global Harris feature points, then compensated for background motion through Delaunay triangulation match and accumulated motion energy to detect small targets. The disadvantage of this method is that background noise is easily mis-detected as moving targets in complex background. Yang T et al. [14] obtained target motion information by background model, built the motion heat map by target motion accumulation, and detected targets in the hot regions based on saliency-based background model. Although this method is suitable for small targets, it has high false alarm rate in complex background. Li Y et al. [15] proposed a novel spatio-temporal saliency approach, which calculated the spatial saliency map and the temporal saliency map on the spatial domain and the temporal domain, depicted the motion consistency characteristic of the moving target by continuous multi-frame video sequence, and obtained spatio-temporal saliency map by fusing them to detect moving targets. However, it is difficult for this method to detect targets in complex background.

In 2017, Lou J et al. [16] raised an approach for small targets detection through region stability and saliency. The stability map was generated by a set of locally stable regions derived from sequential Boolean maps. The saliency map was obtained by comparing the color vector of each pixel with its Gaussian blurred version. Both the stability and saliency maps were integrated in a pixel-wise multiplication manner for small targets detection. This method has high false alarm rate and bad detection performance when the targets are inconspicuous in complex background. Yan J et al. [17] put forward a moving target detection algorithm, which obtained background compensated images based on a nonlinear transformation model. This method can cope with image distortion caused by severe rotation of the detection platform and realize the detection of slowly moving targets in complex rotating background. However, the miss detection rats of this method is high when target size is small. Gao J et al. [18] proposed a method to detect small targets, which combined the self-correlation features of backgrounds and the commonality features of targets in the spatio-temporal domain. The method proposed a dense target extraction model based on nonlinear weights, and a sparse target extraction model based on entry-wise weighted robust principal component analysis to detect small targets, and suppressed background clutters based on target trajectory to improve the detection precision. This method has good detection result for infrared images, but the detection effect is poor for visible light images, where it is difficult to distinguish small targets from the background.

In 2018, Zhang Z et al. [19] proposed a novel flying target detection algorithm based on the spatial and temporal context. This method used multi-frame video sequences to calculate forward and backward motion history maps to extract temporal context information, used conditional random fields to extract spatial contexts, then fused spatial context and temporal context to detect flying targets. This method has poor detection results in complex background, where it is difficult to extract spatial contexts, resulting in low recall rate. Yan D et al. [20] used the ORB operator to extract global feature points, compensated the global motion model through affine transformation and calculated the difference image, then accumulated the multi-frame difference images to obtain the target motion energy map to accurately detect small moving targets in UVA videos. This method accumulates motion energy and has good detection result for small targets, but the false alarm rate is high in the case of complex background.

In 2019, Yi et al. [21] raised a method for fast small moving target detection guided by visual saliency (TDGS), which extracted visual salient regions including small targets according to the differences in global features between the targets and the background, and detected small targets through their temporal relativity in multi-frames. This method has difficulty detecting targets when the color texture of the small targets resembles that of the complex background, resulting in low recall rate.

Therefore, this paper focuses on solving the problem of small target size and complex ground background. To tackle the problem of small target size, firstly, background motion compensation is performed to obtain the background motion parameters. Then, forward and backward motion history maps are calculated to fuse continuous difference images for enhanced motion information of small targets. Finally, morphology processing is used to obtain the area of small moving targets. To solve the problem of complex background, the Kalman predictor is used to predict the target position, and the Hungarian matching algorithm is used to correlate targets to obtain the target trajectory. Then, based on the target trajectory, targets missed by detection are supplemented to improve the target recall rate and false alarm targets are filtered out to improve the target precision rate. The block diagram of the proposed algorithm is shown in Fig. 1:

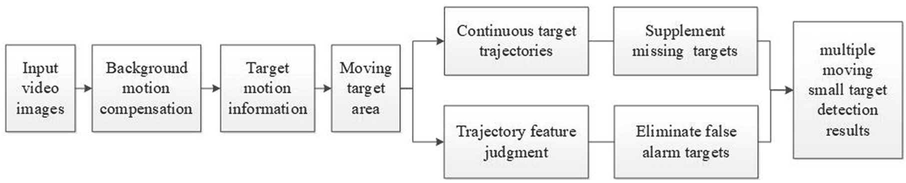


Fig. 1. Block diagram of the algorithm for multiple small moving target detection in complex ground background

2 Detection of Multiple Small Moving Targets

Small targets have a few pixels and lack shape information. Therefore, the motion information of small targets is used for detection, which are more suitable for small targets in the scene.

2.1 Background Motion Compensation

In this paper, regional random points and the Lucas–Kanade (LK) optical flow method [22] are used to obtain random optical flow tracking points. Regional random points can well represent regional characteristics and do not require much gradient calculations. Regional random points are uniformly extracted from the image I(t), and corresponding random optical flow tracking points are obtained from the image using the LK optical flow method, as shown in Fig. 2.

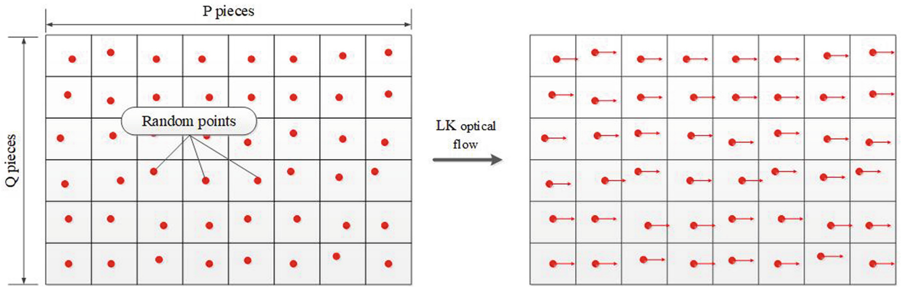


Fig. 2. Regional random optical flow tracking points

The RANSAC (Random Sample Consensus) algorithm is used to fit the 8-parameter homography matrix P_t^{t+1}, which is the background motion estimation parameters from frame $I(t)$ to frame $I(t+1)$, as shown in Eq. (1):

$$\begin{bmatrix} x_i^{t+1} \\ y_i^{t+1} \\ 1 \end{bmatrix} = P_t^{t+1} \begin{bmatrix} x_i^t \\ y_i^t \\ 1 \end{bmatrix} \tag{1}$$

where (x_i^t, y_i^t) is the coordinate of the regional random point in frame $I(t)$, and (x_i^{t+1}, y_i^{t+1}) is the coordinate of the corresponding optical flow tracking point. Background motion compensation is done based on background motion parameter P, as shown in Eq. (2):

$$I^{'}(t \mp 1) = P_{t \mp 1}^t I(t \mp 1) \tag{2}$$

where I′ is the motion compensated image, "−" means forward motion compensation and "+" means backward motion compensation.

2.2 Target Motion Information

In order to enhance the motion information of small targets, the Forward Backward Motion History Image (FBMHI [23]) is used to fuse continuous difference images with background motion compensation in order to obtain the complete motion information of small targets.

Forward Motion History Image (FMHI) is used to extract forward motion information of targets, as shown in Eq. (3):

$$H_F(t) = \begin{cases} \max\left(0, P_{t-1}^t H_F(t-1) - d\right) & if\ D_F(t) < T \\ 255 & if\ D_F(t) \geq T \end{cases} \tag{3}$$

where $H_F(t)$ is forward motion information map, P_{t-1}^t is the background motion parameter from frame(t − 1) to frame(t), d is attenuation parameter ranged from 0 to 255. In order to form the span of pixel intensity values within continuous L frames, d is set as 255/L. L represents the effective number of layers of forward moving images within the FMHI, and L is set as 3 in this paper. $D_F(t)(D_F(t) = \left|I(t) - I^{'}(t-1)\right|)$ is the forward difference image, and $I^{'}(t-1)$ is the forward motion compensated image. The adaptive threshold T is determined using the OTSU theory [24].

Backward Motion History Image (BMHI) is used to extract backward motion information of targets, as shown in Eq. (4):

$$H_B(t) = \begin{cases} \max\left(0, P_{t+1}^t H_B(t+1) - d\right) & if\ D_B(t) < T \\ 255 & if\ D_B(t) \geq T \end{cases} \tag{4}$$

where $H_B(t)$ is backward motion information map, P_{t+1}^t is the background motion parameter from frame(t + 1) to frame(t), and d and T are the same as in formula (3). $D_B(t)(D_B(t) = \left|I(t) - I^{'}(t+1)\right|)$ is the backward difference image, $I^{'}(t+1)$ is the backward motion compensated image.

Through FMHI and BMHI, target moving information map is obtained as shown in Eq. (5):

$$H_{FB}(t) = \min(blur(H_F(t)), blur(H_B(t))) \tag{5}$$

where $blur(\bullet)$ is a smoothing filter which effectively reduces the impact of background noise. $\min(\bullet)$ operation can effectively suppress the trail of the motion history map to guarantee that the detected pixels are those within the boundary of moving targets. The target motion information map $H_{FB}(t)$ is shown in Fig. 3.

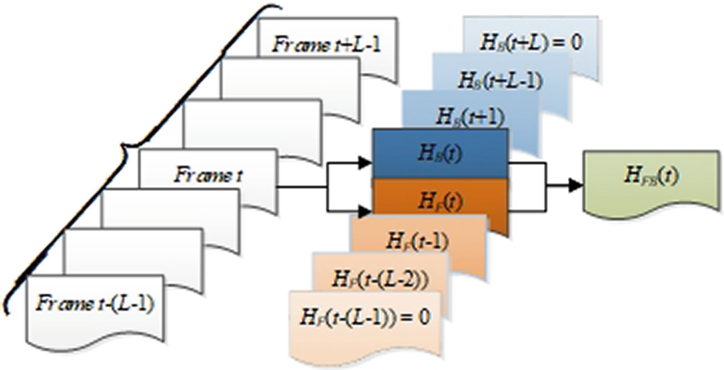


Fig. 3. Target motion information extraction

2.3 Extraction of Moving Target Area

Double thresholds are calculated on the target motion information map $H_{FB}(t)$ with the Otsu method, and the lower threshold δ is used to retain the recall rate of targets. $H_{FB}(t)$ is binarized to get the binary map $M_{FB}(t)$, as shown in Eq. (6).

$$M_{FB}(t) = \begin{cases} 255 & H_{FB}(t) > \delta \\ 0 & H_{FB}(t) \leq \delta \end{cases} \tag{6}$$

One erosion and two dilation operations are performed on the binary map $M_{FB}(t)$ to get the moving target area map $M_{RGE}(t)$, as shown in Fig. 4.

$$M_{REG}(t) = ((M_{FB}(t) \Theta b_{erode}) \oplus b_{dilate}) \oplus b_{dilate} \tag{7}$$

where Θ and $\oplus$ respectively represents erosion and dilation operation, and b_{erode} and b_{dilate} respectively represents rhombus structure element whose R = 1 and R = 3, with R being the radius.

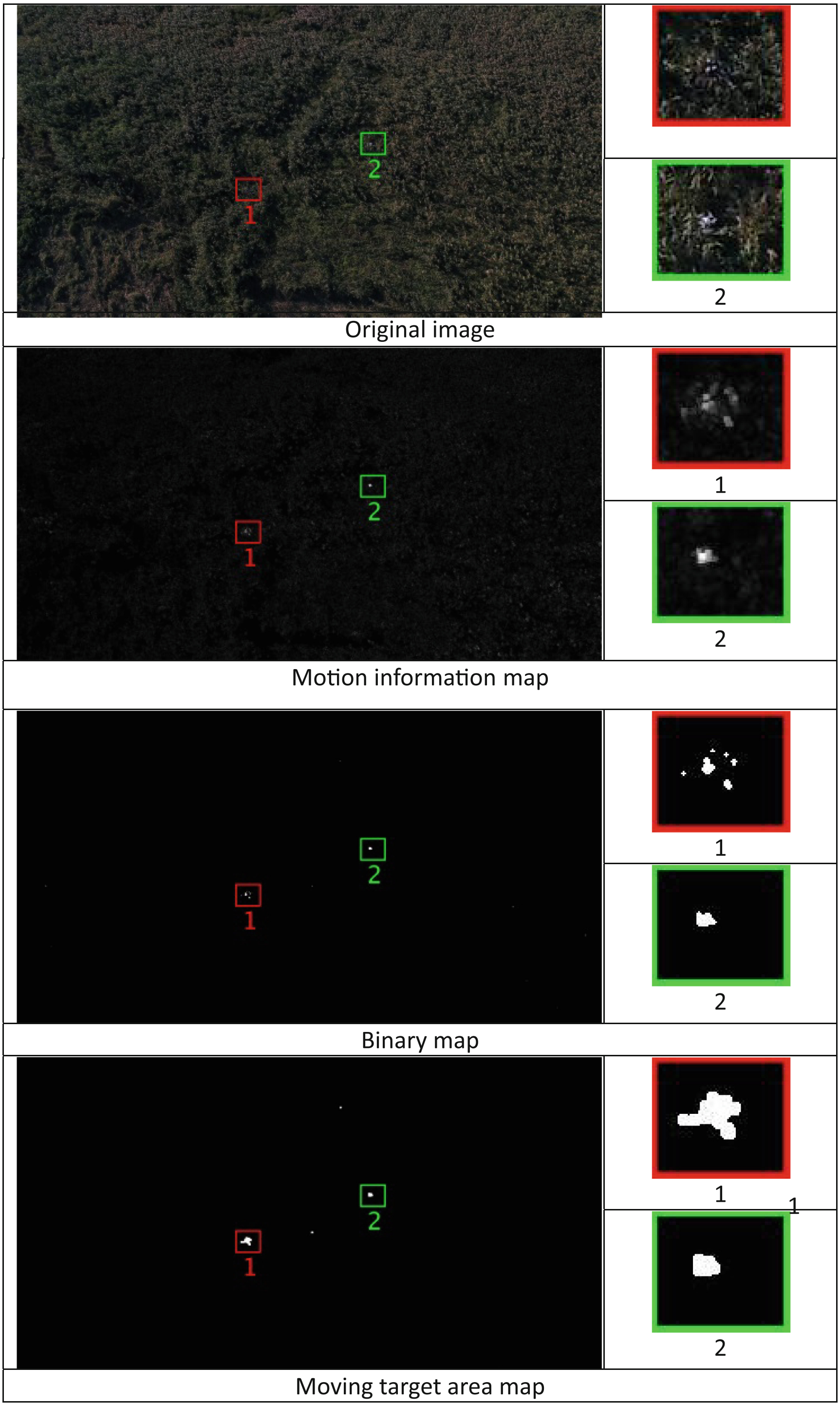


Fig. 4. Moving target area map extraction

3 Detection of Multiple Small Moving Targets Against Complex Ground Background

There are some missed detections and false alarms in the moving target area map $M_{RGE}(t)$. When the background is complex, the false alarm rate increases, and the contrast between small targets and the background is inconspicuous, making it difficult to detect small targets. The Kalman predictor is used to predict each target's position. Based on the Euclidean distance between the detected target position and the predicted target position, the Hungarian matching algorithm is used to correlate targets to obtain multiple target trajectories. If no detected target has the position that matches the predicted target position, which indicates missed detection, then the missed target is supplemented at the predicted target position to improve the recall rate. If the false alarm trajectory features are different from the target trajectory features, then the trajectory features are used to filter out false alarm target trajectories to improve precision rate.

3.1 Target Trajectories Association

Within the moving target area map $M_{RGE}(t)$, the target's location in the next frame $M_{RGE}(t+1)$ can be predicted. Based on the Euclidean distance between the detected target position and the predicted target position, corresponding targets are associated to form multiple target trajectories.

1) Predict targets. Eight-connected domain algorithm is used to detect targets and their locations, then the Kalman predictor is used to predict each target's position in the next frame $M_{RGE}(t+1)$.
2) Detect targets. Eight-connected Domain algorithm is used to detect targets and their locations in $M_{RGE}(t+1)$.
3) Associate targets. In $M_{RGE}(t+1)$, the detected position and the predicted position of each target is matched. If the detected position of the target matches the predicted position, they are associated as the same target. If the predicted position of the target matches none of the detected positions, which indicates missed detection, then the missed target is supplemented at the predicted position in $M_{RGE}(t+1)$. If the detected position of the target matches none of the predicted positions, then this indicates that a new target appears.
4) Determine target trajectory. All the associated targets are determined as target trajectories, if the target in one trajectory is supplemented at the predicted position for five consecutive frames, the target is considered to have disappeared, and the trajectory will be deleted. Then the target trajectories are determined, as shown in Fig. 5.

In Fig. 5, the targets which are associated in the 1^{st}, 2^{nd} and 3^{rd} frames are determined as three target trajectories. Their labels are ID1, ID2 and ID3. A new target appears in the 4th frame and is labeled as ID4. In the consecutive 5^{th}, 6^{th}, 7^{th}, 8^{th} and 9^{th} frames, the target trajectory ID3P is supplemented at the predicted position, which suggests that the target has disappeared, so the target ID3 is deleted from the 5^{th} frame. In the 11^{th} frame, the target ID1 is supplemented at the predicted position, which is marked as ID1P. In the 12^{th} frame, the missed target ID1P is associated as ID1 again.

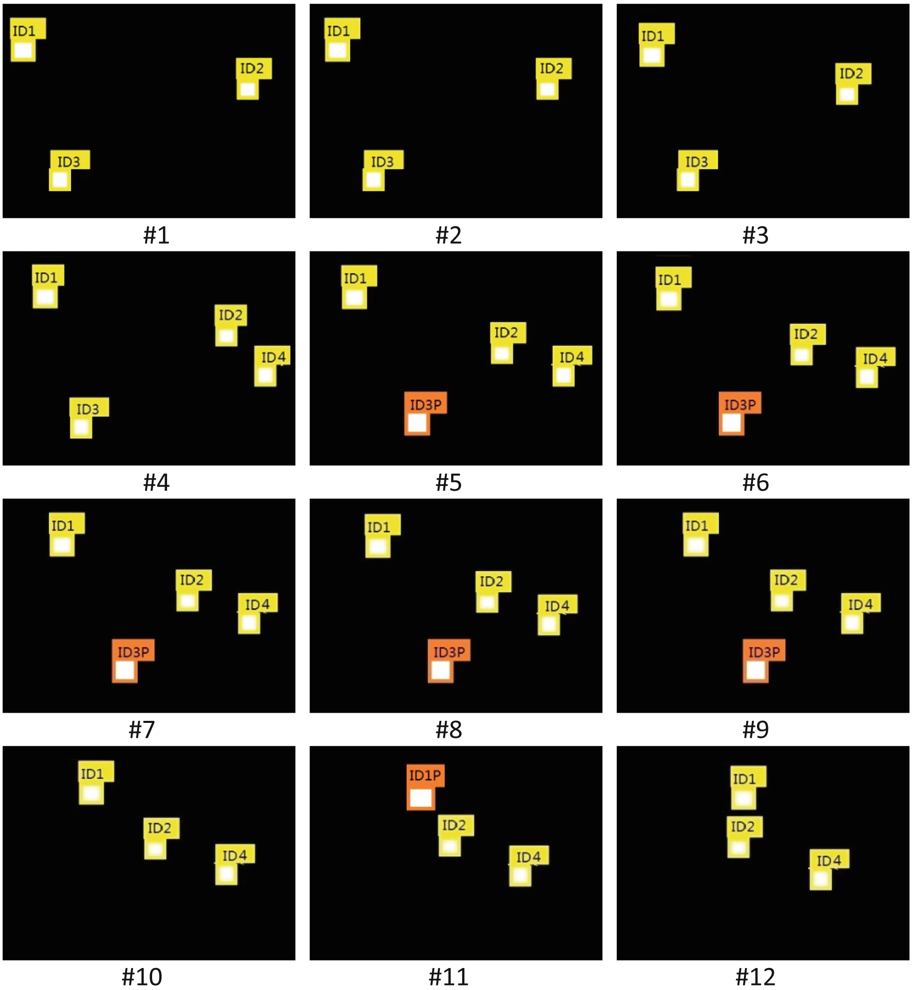


Fig. 5. Result of target trajectory correlation

3.2 Extraction of Target Trajectory Features

The trajectory features of the real targets are different from that of the false alarm targets, hence the trajectory features can be used to filter out the false alarm target trajectories. The target trajectory features include target area feature A, target position feature V and target movement direction feature D. The mean square error of the target area in the target trajectory is calculated to obtain the target area feature A. The target's area in the real trajectory changes little, so the value of A is small, while the target's area in the false alarm target trajectory changes greatly, giving a large A value. The mean square error of the moving pixels of the targets between adjacent frames in the target trajectories is calculated to obtain the target position feature V. The target's position in the real target trajectory changes little, so the value of V is small, while the target's position in the false alarm target trajectory changes greatly, giving a large V value. The mean square error of the angle between the target movement direction and the vertical direction in the target trajectory is calculated to obtain the target movement direction feature D. The target's movement direction is ordered in the real target trajectory, so the value of D is small, the target moving direction is disordered in the false alarm target trajectory, giving a large D value.

Due to the movement of the target, the target trajectory features change greatly during the existence of trajectories. Therefore, the target trajectories need to be segmented. Target trajectory features in each segment change little, which can improve the accuracy of false alarm target trajectories filtration. In this paper, 10 consecutive frames are used as a segment to extract the target trajectory features. As shown in Eq. (8–10), 'area' represents the size of the target area, 'speed' represents the moving pixels between adjacent frames, and 'direction' represents the angle between the direction of the target's movement direction and the vertical direction.

$$A = Var(area_t, area_{t+1}, ..., area_{t+9}) \tag{8}$$

$$V = Var(speed_t, speed_{t+1}, ..., speed_{t+9}) \tag{9}$$

$$D = Var(direction_t, direction_{t+1}, ..., direction_{t+9}) \tag{10}$$

Combining the target area feature A, target position feature V, and target movement direction feature D, the target trajectory feature vector S = [A,V,D] is obtained.

3.3 Filtration of False Alarm Target Trajectories Based on the Trajectory Feature Vector

In the trajectory feature vector S, when A, V, and D are less than the corresponding threshold, the trajectory is determined to be a real target trajectory; otherwise, it is determined to be the false alarm target trajectory, as shown in Eq. (11).

$$flag = \begin{cases} 1 & if\ A < m\_area, V < m\_speed, D < m\_direction \\ 0 & otherwise \end{cases} \tag{11}$$

where m_area, m_speed, and m_direction are thresholds corresponding to A, V, and D, which are related to the target motion in the video. The mean square errors of A, V, and D values are normalized to get the minimum values of A, V, and D. The thresholds corresponding to A, V, and D are determined by experiments to be double the minimum values. If flag is 1, the trajectory is determined to be the real target trajectory, otherwise, the trajectory is determined to be the false alarm target trajectory, as shown in Fig. 6.In Fig. 6, there are four target trajectories from the 1st frame to the 10th frame, which are respectively labeled as ID1, ID2, ID3, and ID4. The changes in area, position, and movement direction of the target ID1 are small from the 1st frame to the 10th frame. A, V, and D are all smaller than the corresponding thresholds, so ID1 is determined to be a real target trajectory. The change in the area of the target ID2 is large from the 1st frame

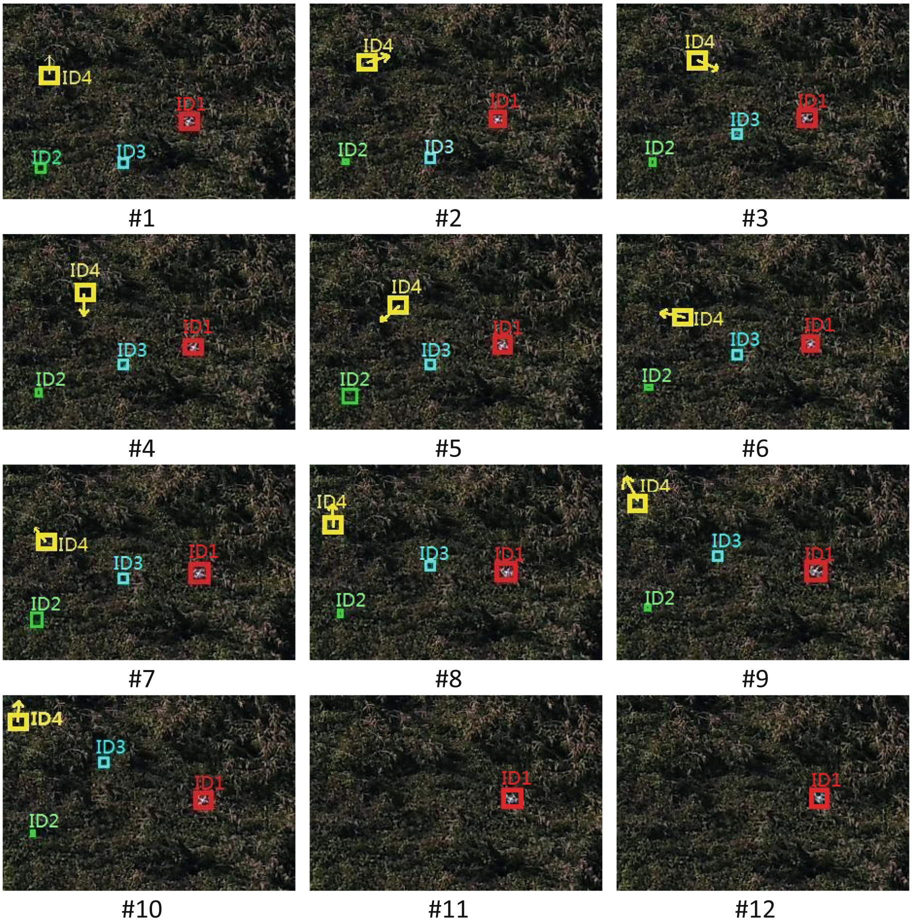


Fig. 6. Target trajectories

to the 10th frame. The minimum area is about 9 pixels in the 2nd frame and the maximum area is about 200 pixels in the 5th frame. A is greater than the corresponding threshold, so ID2 is determined to be a false alarm target trajectory. The change in position of the target ID3 is large from the 1st frame to the 10th frame. The target moves about 2 pixels from the 3rd frame to the 4th frame, and it moves about 30 pixels from the 9th frame to the 10th frame. V is greater than the corresponding threshold, so ID3 is determined to be a false alarm target trajectory. The change in movement direction of the target ID4 is large from the 1st frame to the 10th frame. The target movement direction is to the right in the 2nd frame, downward in the 4th frame and to the left in the 6th frame. D is greater than the corresponding threshold, so ID3 is determined to be a false alarm target trajectory. Therefore, ID2, ID3, and ID4 are filtered out.

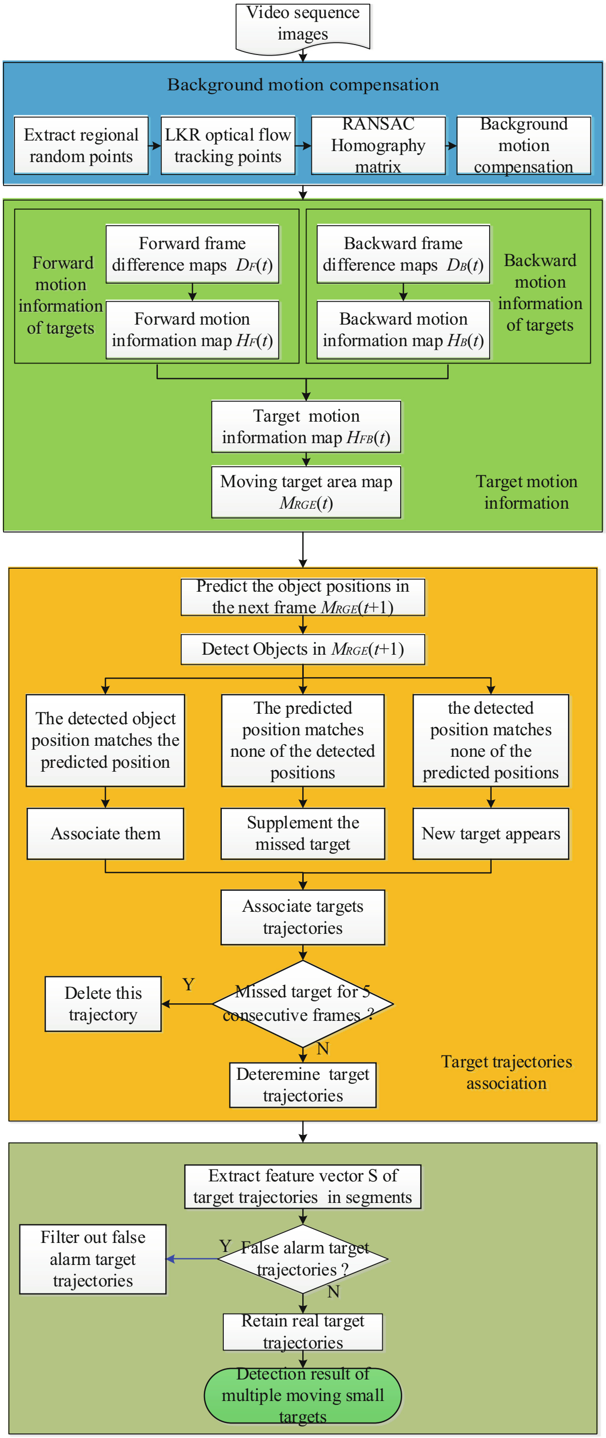


Fig. 7. Flow chart of multiple small moving target detection algorithm in complex ground background

4 Flowchart of the Proposed Algorithm

Figure 7 shows the flowchart of the algorithm for the detection of multiple small moving targets against complex ground background (DMSMT-CGB). Firstly, random points in the area and their corresponding optical flow tracking points are uniformly extracted, and the homography matrix is calculated using RANSCA to compensate for background motion. Secondly, multiple difference images are used to extract the forward motion information map $H_F(t)$ and the backward motion information map $H_B(t)$. $H_F(t)$ and $H_B(t)$ are fused to obtain the target motion information map $H_{FB}(t)$, and the target area map $M_{REG}(t)$ is obtained through adaptive thresholding and morphology processing. Then, the Kalman predictor is used to predict the target position in $M_{REG}(t+1)$, and targets are detected. The Hungarian matching algorithm is used to correlate targets. If the detected position of the target matches the predicted position, they are associated as the same target; If the predicted position of the target matches none of the detected positions, a missed target is supplemented at the predicted position. If the detected position of the target matches none of the predicted positions, it is determined that a new target appears. All the associated targets are determined as multiple target trajectories. If a target in one trajectory is supplemented at the predicted position for 5 consecutive frames, the target is considered to have disappeared, and the trajectory will be deleted. Finally, the trajectory features are extracted in segments, and whether they are false alarm target trajectories is determined based on their changes. False alarm target trajectories are filtered out and true target trajectories are retained to obtain the detection result of multiple small moving targets in complex background.

5 Experimental Results and Analysis

In order to verify the detection effect of the proposed algorithm for detecting multiple small moving targets in complex ground background. The algorithm DMSMT-CGB in this paper is compared with four other state-of-the-art algorithms for multiple moving target detection, which are a detection algorithm based on spatio-temporal saliency (ST saliency), a detection algorithm based on dual-mode Gaussian background modeling (DGM), a detection algorithm based on clustering algorithm density based spatial clustering of applications with noise (DBSCAN) and a detection algorithm based on the clustering algorithm evolutionary local mean (ELM).

All experimental results were obtained with the same data and initialization conditions. The experiment environment: VS2010, Matlab2016. The experiment platform: 3.60 Ghz-Intel i7 processor, 64-bit win7 system and 8 GB memory.

5.1 Evaluation Index

The recall rate (R), precision rate (P) and F-measure (F) are used to quantitatively evaluate the multiple target detection algorithms. R represents the proportion of targets that are correctly detected among all the targets. P represents the proportion of targets that are correctly detected among all the detection results. F is the harmonic weighted average of the two, which is calculated is as follows:

$$R = TP\Big/(TP + FN) \tag{12}$$

$$P = TP \big/ (TP + FP) \tag{13}$$

$$F = \frac{\left(1 + \beta^2\right) \times P \times R}{\beta^2 \times P + R} \tag{14}$$

where TP represents the number of targets that are correctly detected as targets, FN represents the number of targets that are incorrectly detected as background, and FP represents the number of backgrounds that are incorrectly detected as targets. β determines the relative significance of the recall and precision rate. $\beta = 1$ means that the significance is equal. $\beta > 1$ means that the recall rate is more significant, $\beta < 1$ means that the precision rate is more (significant). After overall consideration of recall and precision rate, in this paper β is set to 1.

5.2 Experimental Data

For the purpose of the current experiment, the experimental video has to satisfy a number of conditions, such as background motion, small target scale, weak contrast, and complex background. It is difficult to obtain an experimental video that meets all the conditions above. In this paper, we obtained three experimental videos by field shooting.

The first experimental video (Translational grass background) contains 2000 frames of images with the image size of 1920 × 1080 pixels. The background of the images is grass, and the targets include a small white quad-rotor UAV (Unmanned Aerial Vehicle) and a black one flying in a straight line close to the ground from near to far. The camera platform is set on another UAV during shooting, which follows the two targets in translation motion. From the whole video, 500 consecutive images containing the target are selected to obtain the experimental video VD1.

The second experimental video (Rotational road background) contains 2400 frames of images with the image size of 1920 × 1080 pixels. The background of the images is city roads. The targets include a small white quad-rotor UAV and a black one slowly flying in a curve close to the ground. The camera platform is set on another UAV during shooting, which follows the two targets in rotation motion. From the whole video, 500 consecutive images containing the targets are selected to obtain the experimental video VD2.

The third experimental video (Pitching road background) contains 1000 frames of images with the image size of 1920 × 1080 pixels. The background of the images is city roads, and the targets include a small white quad-rotor UAV and a black one flying in a straight line close to the ground from far to near. The camera platform is set on another UAV during shooting, which is in pitching movement. From the whole video, 200 consecutive images containing the targets are selected to obtain the experimental video VD3.

In the above three experimental videos, the black UAV is the key target because it is similar in color to the background, and has a small size, which makes it easily obscured by the background. The characteristics and main detection difficulties of each video are shown in Table 1.

Table 1. Experiment videos

Video	Frame	Resolution (pixels)	Minimum target size (pixels)	Major detection difficulties
VD1	500	1920 × 1080	6 × 5	Translational background, scene motion, small target scale, weak contrast, black UAV obscured by background
VD2	500	1920 × 1080	20 × 15	Rotational background, complex background, black UAV obscured by background, the white and the black UAVs' slow relative motion to background
VD3	200	1920 × 1080	20 × 15	Pitching Background, complex background, black UAV obscured by background

5.3 Result Analysis

The proposed algorithm and other comparison algorithms are tested on the three experimental videos to compare the detection performance. Experimental results are compared and analyzed using the aforementioned evaluation indicators.

Detection of Multiple Moving Targets Against the Translational Grass Background. The performance of the proposed algorithm for detecting multiple moving targets against the translational grass background is verified on VD1. Experimental results of the proposed algorithm and other comparison algorithms are shown in Table 2.

Table 2 shows that the proposed algorithm has the highest TP, the lowest FP, the most correctly detected targets and the lowest false alarm rate, indicating superiority over the other four algorithms. The DBSCAN algorithm and the ELM algorithm are based on target feature points. In VD1, the small target size makes it difficult to extract feature points, hence the two algorithms fail to detect targets.

Detection of Multiple Moving Targets Against the Rotating Road Background. The performance of the proposed algorithm for detecting multiple moving targets against the rotating road background is verified on VD2. Experimental results of the proposed algorithm and other comparison algorithms are shown in Table 3.

Table 3 shows that the proposed algorithm has the highest TP, and the most correctly detected targets. Although the DBSCAN algorithm and the ELM algorithm have lower FP, the TP of these two algorithms are both very low, which means a large number

Table 2. Results of multiple moving target detection in the translational grass background

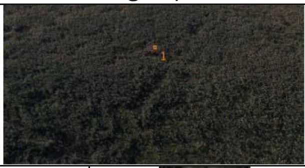
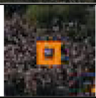
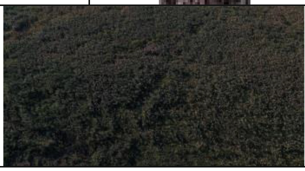
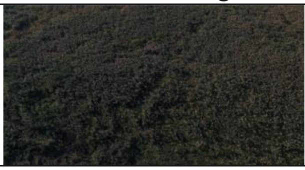
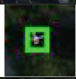
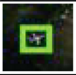
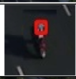
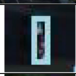
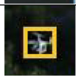
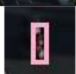
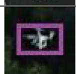
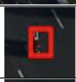
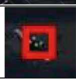
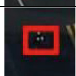
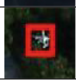
Algorithm	135^{th} frame (2 moving targets)		385^{th} frame (2 moving targets)		Whole video
ST saliency					TP=488 FP=122
	1		1		
DGM					TP=388 FP=5581
	1		1		
DBSCAN					TP=0
	Undetected targets		Undetected targets		
ELM					TP=0
	Undetected targets		Undetected targets		
DMSMT-CGB（OURS）					TP=933 FP=70
	1		1		
	2		2		

Table 3. Results of multiple moving target detection in the rotating road background

<table>
<tr><th>Algorithm</th><th colspan="2">15th frame (3 moving targets)</th><th colspan="2">220th frame (2 moving targets)</th><th>Whole video</th></tr>
<tr><td rowspan="2">ST saliency</td><td colspan="2"></td><td colspan="2"></td><td rowspan="2">TP=518
FP=745</td></tr>
<tr><td>1</td><td></td><td>1</td><td></td></tr>
<tr><td rowspan="2">DGM</td><td colspan="2"></td><td colspan="2"></td><td rowspan="2">TP=509
FP=385</td></tr>
<tr><td>1</td><td></td><td>1</td><td></td></tr>
<tr><td rowspan="3">DBSCAN</td><td colspan="2"></td><td colspan="2"></td><td rowspan="3">TP=309
FP=12</td></tr>
<tr><td>1</td><td></td><td rowspan="2">1</td><td rowspan="2"></td></tr>
<tr><td>2</td><td></td></tr>
<tr><td rowspan="3">ELM</td><td colspan="2"></td><td colspan="2"></td><td rowspan="3">TP=324
FP=10</td></tr>
<tr><td>1</td><td></td><td rowspan="2">1</td><td rowspan="2"></td></tr>
<tr><td>2</td><td></td></tr>
<tr><td rowspan="4">DMSMT-CGB （OURS）</td><td colspan="2"></td><td colspan="2"></td><td rowspan="4">TP=1098
FP=89</td></tr>
<tr><td>1</td><td></td><td>1</td><td></td></tr>
<tr><td>2</td><td></td><td rowspan="2">2</td><td rowspan="2"></td></tr>
<tr><td>3</td><td></td></tr>
</table>

of missed detections. The proposed algorithm has lower FP while ensuring the highest TP, which which means that the most number of targets are correctly detecting while ensuring few false alarms. Therefore, the performance of the proposed algorithm is better than the other four algorithms.

Detection of Multiple Moving Targets Against the Pitching Road Background. The performance of the proposed algorithm for detecting multiple moving targets against the pitching road background is verified on VD3. Experimental results of the proposed algorithm and other comparison algorithms are shown in Table 4.

Table 4 shows that the proposed algorithm has the highest TP and the most correctly detected targets. Although the FP of the ST saliency algorithm, the DBSCAN algorithm and the ELM algorithm are lower than that of the current algorithm, the TP of these three algorithms are very low, which means a large number of missed detections. The proposed algorithm has lower FP while ensuring the highest TP, which means that the most number of targets are correctly detected while ensuring the least number of missed detections and few false alarms. Therefore, the performance of the proposed algorithm is better than the other four algorithms.

Performance Comparison Among Multiple Moving Target Detection Algorithms. The performance of the proposed DMSMT-CGB algorithm is compared with the other four multiple target detection algorithms, and their performance is evaluated by recall rate (R), precision rate (P), and F measure (F) indicators. Experimental results are shown in Table 5.

Table 5 shows that the proposed algorithm has the highest recall rate R and F, indicating that our algorithm performs better than the other four algorithms. It is observed that in the experimental video VD2 and VD3, the precision rate P of the ELM algorithm is higher than that of the proposed algorithm by 4.5% and 3.4% respectively, and that the precision rate P of the DBSCAN algorithm is higher than that of the proposed algorithm by 3.8% and 2.6% respectively. However, the recall rate R of the ELM algorithm is 27.8% in VD2 and 46.5% in VD3, which is much lower than that of the proposed algorithm. Likewise, the recall rate R of the DBSCAN algorithm is 26.2% in VD2 and 50.5% in VD3, which is much lower than that of the proposed algorithm. In VD1, the DBSCAN algorithm and the ELM algorithm cannot detect the targets, so that the recall rate (R), precision rate (P), and F-measure (F) are all 0. Experimental results show that the proposed algorithm has the best detection performance for detecting multiple small moving targets against complex ground background.

Table 4. Results of multiple moving target detection in the pitching road background

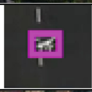
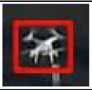
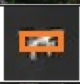
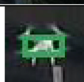
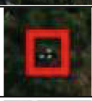
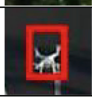
Algorithm	10^{th} frame (2 moving targets)		136^{th} frame (2 moving targets)		Whole video
ST saliency					TP=236 FP=14
	1		1		
DGM					TP=350 FP=186
	1		1		
DBSCAN					TP=202 FP=4
	1		1		
ELM					TP=186 FP=2
	1		1		
DMSMT-CGB （OURS）					TP=389 FP=18
	1		1		
	2		2		

Table 5. Performance evaluation of multiple small moving target detection algorithms in complex ground background)

	Algorithm	*R*	*P*	*F*
VD1	ST saliency	48.8	80.0	60.6
	DGM	38.8	6.5	11.1
	DBSCAN	0	0	0
	ELM	0	0	0
	DMSMT-CGB (OURS)	93.3	93.0	93.2
VD2	ST saliency	44.5	41.0	42.7
	DGM	43.7	56.9	49.4
	DBSCAN	26.2	96.3	41.2
	ELM	27.8	97.0	43.2
	DMSMT-CGB (OURS)	94.2	92.5	93.4
VD3	ST saliency	59.0	94.3	72.8
	DGM	87.5	65.3	74.8
	DBSCAN	50.5	98.1	66.7
	ELM	46.5	98.9	63.3
	DMSMT-CGB (OURS)	96.7	95.5	96.1

6 Conclusion

This paper proposes a multiple small moving target detection algorithm against complex ground background, which solves the problem of small targets, which have few pixels and lack topographical information, making them difficult to be accurately detected against complex background. In the proposed algorithm, multiple forward-backward target motion information is fused based on the FBMHI algorithm to improve the recall rate. Target trajectories are correlated to supplement missed targets at the predicted position and reduce missed targets. Target trajectory features are extracted in segment to filter out false alarm target trajectories, further reducing the false alarm rate. Experimental results show that the proposed algorithm has higher recall rate, precision rate and F-measure. Future research will be focused on conducting in-depth research on the detection of multiple small moving targets against severe rotational background, which will help solve the problem of accurately detecting targets by the photodetector platform under large-range UAV movement.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (61471194 and 61705104), the Fundamental Research Funds for the Central Universities (NJ2020021, NT2020022), the Natural Science Foundation of Jiangsu Province (BK20170804), and National Defense Science and Technology Special Innovation Zone Project.

References

1. Hofmann, M., Tiefenbacher, P., Rigoll, G.: Background segmentation with feedback: the pixel-based adaptive segmenter. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 38–43. IEEE (2012)
2. Siam, M., ElSayed, R., ElHelw, M.: On-board multiple target detection and tracking on camera-equipped aerial vehicles. In: 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 2399–2405. IEEE (2012)
3. Moo Yi, K., Yun, K., Wan Kim, S., Jin Chang, H., Young Choi, J.: Detection of moving objects with non-stationary cameras in 5.8 ms: bringing motion detection to your mobile device. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 27–34 (2013)
4. Shen, H., Li, S., Zhu, C., Chang, H., Zhang, J.: Moving object detection in aerial video based on spatiotemporal saliency. Chin. J. Aeronaut. **26**(5), 1211–1217 (2013)
5. Shakeri, M., Zhang, H.: Detection of small moving targets using a moving camera. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014), pp. 2777–2782. IEEE (2014)
6. Sadeghi-Tehran, P., Clarke, C., Angelov, P.: A real-time approach for autonomous detection and tracking of moving objects from UAV. In: 2014 IEEE Symposium on Evolving and Autonomous Learning Systems (EALS), pp. 43–49. IEEE (2014)
7. Leutenegger, S., Chli, M., Siegwart, R.Y.: BRISK: binary robust invariant scalable keypoints. In: 2011 International Conference on Computer Vision, pp. 2548–2555. IEEE (2011)
8. Wang, Z., Wang, C., Huang, F., Liu, J.: Based on spatial-temporal multiframe association infrared target detection. In: MIPPR 2015: Automatic Target Recognition and Navigation, vol. 9812, p. 98121A. International Society for Optics and Photonics (2015)
9. Wei, L., Weijie, Z., Cheng, L., Zhonglin, X., Kaiqiao, T.: Small moving object detection based on the improved ORB feature matching method. Optoelectron. Eng. **42**(10), 13–20 (2015)
10. Rozantsev, A., Lepetit, V., Fua, P.: Flying objects detection from a single moving camera. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4128–4136 (2015)
11. Rozantsev, A., Lepetit, V., Fua, P.: Detecting flying objects using a single moving camera. IEEE Trans. Pattern Anal. Mach. Intell. **39**(5), 879–892 (2016)
12. Junhua, Y., He, D., Qianqian, X., Yong, Y.: Slowly moving small object detection in complex moving background based on adaptive threshold segmentation. Electron. Des. Eng. **24**(06), 77–80+84 (2016)
13. Meng, Y., Yan, C.: Detection and tracking of small moving objects based on multi-view aerial photography registration. Comput. Eng. Appl. **52**(14), 27–31 (2016)
14. Yang, T., et al.: Small moving vehicle detection in a satellite video of an urban area. Sensors **16**(9), 1528 (2016)
15. Li, Y., Zhang, Y., Yu, J.G., Tan, Y., Tian, J., Ma, J.: A novel spatio-temporal saliency approach for robust dim moving target detection from airborne infrared image sequences. Inf. Sci. **369**, 548–563 (2016)
16. Lou, J., Zhu, W., Wang, H., Ren, M.: Small target detection combining regional stability and saliency in a color image. Multimed. Tools. Appl. **76**(13), 14781–14798 (2017).
17. Yan, J., Xu, Q., Duan, H., Yang, Y., Xiao, Y.: Slow ground moving object detection in rotating complex background. J. Imaging Sci. Technol. **61**(2), 20507–20511 (2017)
18. Gao, J., Wen, C., Liu, M.: Robust small target co-detection from airborne infrared image sequences. Sensors **17**(10), 2242 (2017)
19. Zhang, Z., Cao, Y., Ding, M., Zhuang, L., Wang, Z.: Spatial and temporal context information fusion based flying objects detection for autonomous sense and avoid. In: 2018 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 569–578. IEEE (2018)

20. Yan, D., Sun, W.: Small moving object detection based on sequence confidence method in UAV video. In: Krömer, P., Zhang, H., Liang, Y., Pan, J.-S. (eds.) ECC 2018. AISC, vol. 891, pp. 676–683. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-03766-6_76
21. Yi, X., Wang, B., Zhou, H., Qin, H.: Dim and small infrared target fast detection guided by visual saliency. Infrared Phys. Technol. **97**, 6–14 (2019)
22. Lucas, B.D.: An iterative image registoration techniques with an application to stereo vision. Proc. IJCAI **81**(3), 674–679 (1981)
23. Yin, Z., Collins, R.: Moving object localization in thermal imagery by forward-backward MHI. In: 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW 2006), p. 133. IEEE (2006)
24. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)