Event-Based Motion Segmentation of Small Objects in the Wild

Zhezheng Sun,[1] Chuantong Hou,[1] Longguang Wang,[1] Xinpu Deng,[1] and Miao Li[1]

[1] *School of Electronic Science, National University of Defense Technology, Changsha, China*

E-mail: lm8866@foxmail.com

Event-based cameras are sensitive to brightness changes and can capture rich temporal information with very high temporal resolution, which has great potential for motion segmentation of moving objects. Under static background, events are only triggered by motion of objects, thereby moving objects can be easily segmented. However, in many real-world applications, events are also be triggered by the motion of camera or background and submerge the ones corresponding to moving objects. In this letter, we propose an event-based motion segmentation method to segment moving small objects in events obtained from the wild. First, motion estimation is performed to align the events triggered by the background. Then, candidate events corresponding to moving objects or moving backgrounds are detected. Finally, motion information is adopted to segment the events of moving small objects from the ones triggered by the background. In addition, we develop the first dataset for event-based motion segmentation of small objects, namely EMSS. Experimental results demonstrate the effectiveness of our method and show that our method can achieve robust motion segmentation of small moving objects in the wild.

Introduction: Event-based cameras are bio-inspired vision sensors [1] with four unique advantages, including high temporal resolution, high dynamic range, low power consumption, and low amount of data. Event-based camera output asynchronous spatio-temporal pulse signal (i.e., event) when intensity changes at the time they occur. Each event is defined as $e = (x, y, t, p)$, where x and y are spatial locations, t is a timestamp, and $p \in \{-1, 1\}$ indicates whether the intensity is increased or decreased [2].

The attractive properties of event cameras have enabled numerous computer vision tasks, such as feature tracking [3, 4], simultaneous localization and mapping (SLAM) [5, 6] , and computational photography [7, 8]. Since event cameras are sensitive to moving objects and can capture rich temporal information, it is natural to segment moving objects using event data.

Mitrokhin et al. [9] proposed a neural network for event-based motion segmentation in indoor scenes. Zhou et al. [10] cast the motion segmentation problem as an energy minimization and use spatio-temporal features of the event data for motion segmentation. Lu et al. [11] proposed a cascaded multi-model fitting method to segment moving objects, which achieves superior performance. However, these motion segmentation methods focus on large objects (e.g., cars and pedestrians) and cannot be directly applied to small objects.

Due to the lack of texture and shape information, it is challenging to segment small moving objects from the event data, especially in the wild.

Recently, Shu et al. [12] first studied the motion segmentation problem of small objects and proposed to conduct segmentation in the projected image of events. Nevertheless, this method inevitably introduces information loss in temporal domain and brings redundant computation.

In this letter, we propose an event-based motion segmentation method to segment small moving objects directly from spatio-temporal events acquired in the wild.

Our method consists of three stages, as illustrated in Fig 1. First, event registration is performed to align events triggered by the motion of camera or background. Then, events corresponding to the motion of objects are detected from the aligned events via multi-scale progressive clustering. Since event registration aligns events produced by the background, the events corresponding to the moving objects can be then segmented.

The main contributions of this letter are summarized as follows:

1. We propose an event-based motion segmentation method to segment small moving objects in the wild. To the best of our knowledge, our method is the first attempt to achieve motion segmentation of small objects from event data acquired from the wild.
2. We develop the first dataset, namely EMSS, for event-based motion segmentation of small objects. Our dataset provides a benchmark for performance evaluation.
3. Experiments show that our method produces promising motion segmentation results on real-world event data.

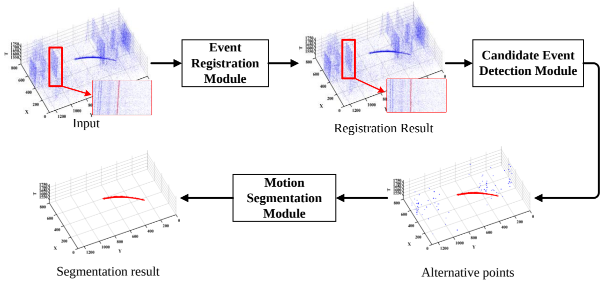


Fig 1 *An overview of our method.*

The Proposed Approach: The main idea of our method is to segment moving small objects directly from spatio-temporal events. Our method consists of an event registration stage, a candidate event detection stage and a motion segmentation stage.

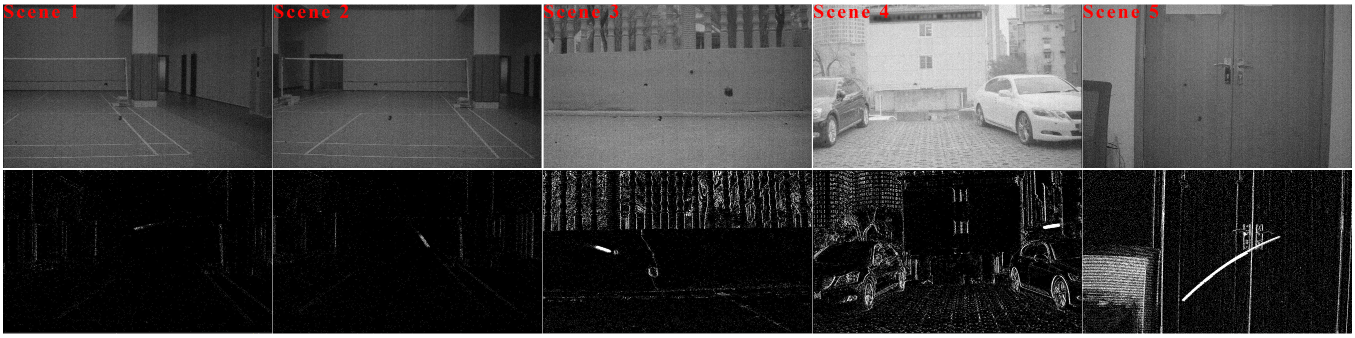
Event Registration: Given an event stream acquired in the wild, the events are triggered mainly by three different sources: the motion of objects, the motion of camera or background, and the intensity changes caused by the thermal noise of the camera. The events produced by the thermal noise are randomly scattered in the spatio-temporal space and quite different from the trajectories of moving objects. However, the events triggered by the motion of camera or background are usually similar with object trajectories. Consequently, we are motivated to perform event registration to align these events. For events produced within a narrow temporal window (t_1, t_2), the motion of the camera or background can be considered as rigid motion. Consequently, our goal is to estimate the offset $(\triangle x, \triangle y)$ from time t_1 to time t_2 and then the events produced within this window can be aligned. To achieve efficient motion estimation, a simple yet effective method is adopted. First, events triggered at t_1 and t_2 are projected onto images to produce I_{t_1} and I_{t_2}. Then, an image registration based on pyramid method is used to estimate offsets in the frequency domain [13].

Candidate Event Detection: As analyzed above, the moving of objects and background triggers events in the event stream. The faster the objects move, the higher density the event data is. Motivated by this, we detect candidate events based on their local density in a spatio-temporal neigborhood. For any event $e = [x_1, y_1, t_1]^T$ in the event stream, its k nearest neighbors are used to calculate their average distance to event e. The lower the distance is, the higher density the local neighborhood is. Then, events with local distance lower than the threshold b are detected as candidate ones. Since events triggered by thermal noise are random and sparse, these events can be easily excluded after this stage. Our method is summarized as follows:

Algorithm 1 Candidate event detection

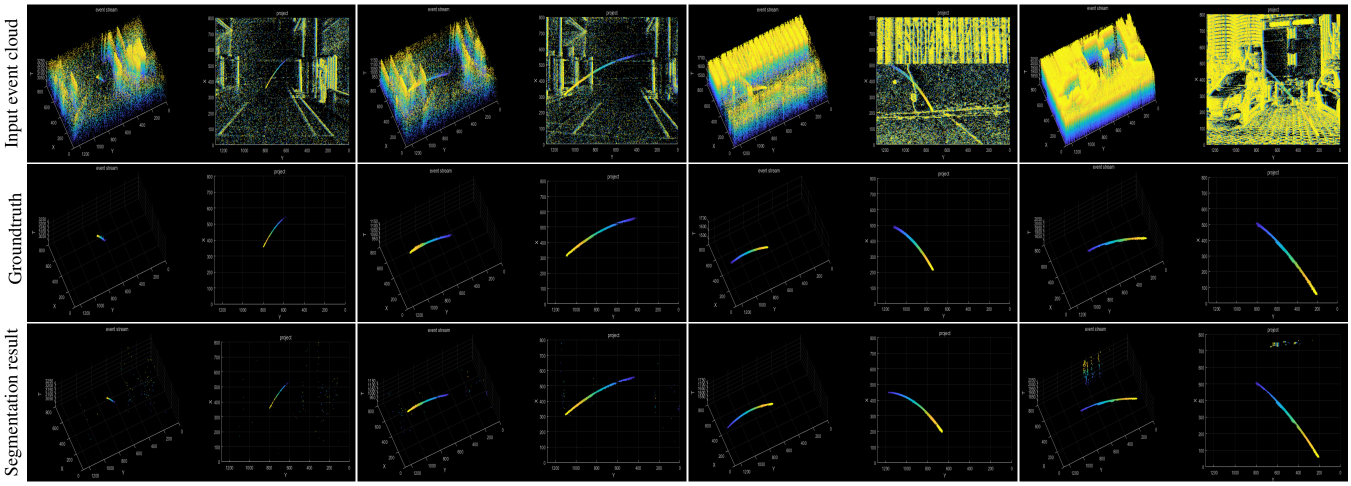
1: **Input:** event stream $A = [X, Y, T]^T$,neighbor parameter k, segmentation threshold $[a, b]$
2: **for** each event $a = [x, y, z]^T \in A$ **do**
3: Calculate the Euclidean distance between event a and other events: dis_1, dis_2, dis_3 ...
4: Find k events with the nearest Euclidean distance.
5: Calculate the average distance of k events: dis_{mean}
6: **end for**
7: **Find** all events which $dis_{mean} \in [a, b]$
8: **Output:**The index of events that meet the conditions.

Motion Segmentation: Although events triggered by moving objects can be detected after candidate event detection, events produced by moving background are also detected as candidates. To handle these false alarms,



(a) (b) (c) (d) (e)

Fig 2 *Visualization of examples in the proposed dataset. (a), (b), (c), and (d) are real scenes while (e) is a synthetic scene.*



(a) (b) (c) (d)

Fig 3 *Motion segmentation results produced by our method. (a) and (b) are two indoor scenes while (c) and (d) are two outdoor scenes.*

motion information is further exploited to segment moving objects from moving backgrounds. After event registration, the motion of events triggered by backgrounds are compensated. Consequently, the events triggered by moving objects can be segmented. First, for each event e, its k nearest neighbors along the temporal dimension are selected and ranked as a queue. Then, n points at the begining and the end are used to calculate two centroids, respectively:

$$Centroid_{begin} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} y_i\right) \tag{1}$$

$$Centroid_{end} = \left(\frac{1}{n}\sum_{i=k-n}^{k} x_i, \frac{1}{n}\sum_{i=k-n}^{k} y_i\right) \tag{2}$$

Next, a motion amplitude is calculated using these two centroids:

$$Amplitude = \left\|Centroid_{begin} - Centroid_{end}\right\|_2 \tag{3}$$

Finally, events with motion amplitudes larger than a threshold d are separated as the events corresponding to moving objects. After event registration, events triggered by the motion of backgrounds are aligned such that their motion amplitudes are relatively small. In contrast, events corresponding to moving objects have much larger motion amplitudes.

Experiments: In this section, we first introduce the datasets and experimental details. Then, we conduct performance evaluation on our dataset. Finally, ablation experiments are conducted to validate the effectiveness of our method.

Datasets and Experimental Settings:

Dataset: We develop the first dataset for event-based motion segmentation of small objects, namely EMSS.[1] Our dataset consists of 11 video sequences and covers indoor and outdoor scenes with both synthetic and real objects. Specifically, we first used the CeleX5 event camera to acquire numerous video sequences in the wild. Then, synthetic moving objects were added to sequences without moving objects to generate synthetic sequences.

[1] https://github.com/as513621939/EMSS

Evaluation Metrics: We adopt the widely used Intersection over Union (IoU) metric for quantitative evaluation. For each event stream, the IoU score is calculated as:

$$IoU = \frac{TP}{FP + TP + FN} \times 100\% \tag{4}$$

where TP, FP, and FN represent true positive, false positive, and false negative points, respectively. In addition, F_1 score is also adopted as the evaluation metric:

$$F_1 = \frac{2P \cdot R}{P + R} \tag{5}$$

where P and R represent the precision and recall, respectively.

Implementation Details: All experiments in this paper were conducted on a PC with an AMD Ryzen 7 3700X CPU using PyTorch platform.

Performance Evaluation: We compare our method with three representative methods, including an event-based motion segmentation method EV-Segment [14], a filter-based method [16], and a cluster-based method [15]. Note that, EV-Segment is developed for large object motion segmentation. Quantitative results are presented in Table 1, while visual results are shown in Fig 3. From Table 1, we can see that our method outperforms other approaches by notable margins on most scenarios. Specifically, on the indoor scene (scene 2), our method produces much higher IoU and F_1 scores (80.18/0.8899) as compared to [23]. Moreover, it can be observed from Fig 3 that our method can accurately segment the trajectories of small moving objects even under complicated environments.

As can be seen from the Table 1, our method has achieved good results in the real scene, especially in the indoor scene, it can accurately segment the small moving objects, but there will be more false alarms in part scene. Performance has some disadvantages compared with the method [16].

Ablation Experiments and Parameter Analyses:

Ablation Experiments: To exclude events triggered by the background during motion segmentation, event registration is performed to align the events produced by background. To demonstrate its effectiveness, we

Table 1. IoU and F_1 scores achieved by different methods.

Method	Statistical filter [14]		EV-Motion [15]		Euclidean cluster [16]		Ours	
	IoU	F_1 score	IoU	F_1 score	IoU	F_1 score	IoU	F_1 score
Scene1(indoor)	2.62	0.0511	5.89	0.1111	31.96	0.4836	**62.28**	**0.7675**
Scene2(indoor)	14.74	0.2568	9.72	0.1771	40.43	0.5757	**80.18**	**0.8899**
Scene3(outdoor)	3.16	0.0612	5.38	0.1019	48.87	0.5785	**61.85**	**0.7635**
Scene4(outdoor)	6.15	0.1159	11.78	0.2103	**80.01**	**0.8877**	77.06	0.8704
Scene5(synthetic)	13.3	0.2241	20.25	0.3096	56.44	0.6960	**73.12**	**0.8090**
Average	8.00	0.1419	10.60	0.1820	51.54	0.6443	**70.90**	**0.8201**

removed the event registration stage and directly perform motion segmentation on raw event streams.

Table 2. mIoU and F_1 score achieved by our method with different settings.

Cluster	Reistration	Segmentation	$mIoU$	F_1 score
✓			66.38	0.7964
✓	✓		67.96	0.8047
✓		✓	66.53	0.7974
✓	✓	✓	**70.34**	**0.8228**

It can be observed from Table 2 that event registration introduces notable performance improvements, with $mIoU$ being improved from 66.38 to 67.96. This demonstrates that event registration is beneficial to the motion segmentation. It can be seen that the horizontal offset decision without registration is not good for the segmentation gain, and the $mIoU$ is increased by 0.15, which is because the background is not registered and also has a large offset in the horizontal direction, while the offset decision after registration shows an $mIoU$ gain of 3.62.

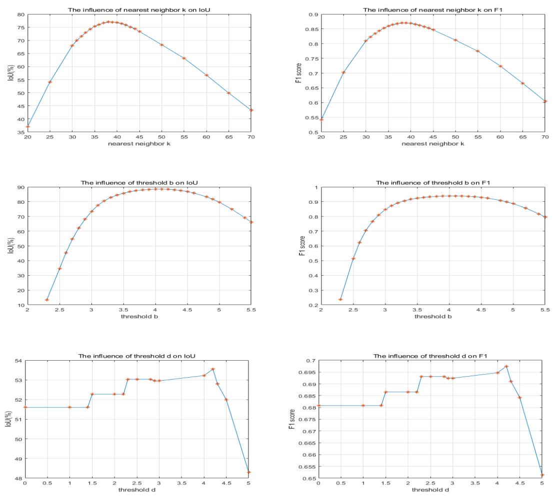


Fig 4 *Results achieved by our method with different parameters.*

Parameter Analyses: The key parameters in our method are the number of nearest neighbors k, the segment threshold b and the offset threshold d. We further conduct experiments to investigate their effects to the performance of our method. From Fig 4 we can see that the IoU and F_1 scores of our method first increases as the number of nearest neighbors k increases. Then, when k is larger than 38, increasing k cannot obtain further gain and introduce notable performance drop. Consequently, k is set to 35 as the default setting of our method. The segment threshold b shares a similar observation with k and $b = 4.1$ achieves the best performance. Therefore, we use $b = 4$ as our default setting. The increase of threshold d can significantly inhibit the generation of false alarm points, so with the increase of d, IoU and F_1 scores increases gradually, and reaches the optimal value when it exceeds 4.2. Therefore, we use $d = 4$ as our default setting.

Conclusion: This letter introduces an event-based motion segmentation method for small objects in the wild. Our method first conducts event registration to align events triggered by the background. Then, detection and segmentation model are used to directly segment the trajectories of small moving objects in the spatio-temporal event data. In addition, we propose the first dataset for motion segmentation of small objects. Experiments demonstrate the effectiveness of our method and show the promising performance produced by our method.

Acknowledgments: This work was supported in part by the National Natural Science Foundation of China (No. 61921001 and No. 62001478)

© 2022 The Authors. *Electronics Letters* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

Received: 2 June 2022 *Accepted:* X XX 202X

References

1. Brandli, C., et al.: A 240 × 180 130 db 3 µs latency global shutter spatiotemporal vision sensor. IEEE Journal of Solid-State Circuits 49(10), 2333–2341 (2014). doi:10.1109/jssc.2014.2342715
2. Gallego, G., et al.: Event-based vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(1), 154–180 (2020)
3. Lagorce, X., et al.: Asynchronous event-based multikernel algorithm for high-speed visual features tracking. IEEE Transactions on Neural Networks and Learning Systems 26(8), 1710–1720 (2014)
4. Valeiras, D.R., et al.: An asynchronous neuromorphic event-driven visual part-based shape tracking. IEEE Transactions on Neural Networks and Learning Systems 26(12), 3045–3059 (2015)
5. Kim, H., Leutenegger, S., Davison, A.J.: Real-time 3d reconstruction and 6-dof tracking with an event camera, , pp. 349–364. (2016)
6. Rebecq, H., et al.: Evo: A geometric approach to event-based 6-dof parallel tracking and mapping in real time. IEEE Robotics and Automation Letters 2(2), 593–600 (2016)
7. Tulyakov, S., et al.: Time lens: Event-based video frame interpolation pp. 16150–16159. (2021). doi:10.1109/CVPR46437.2021.01589
8. Rebecq, H., et al.: High speed and high dynamic range video with an event camera. IEEE Transactions on Pattern Analysis and Machine Intelligence 43(6), 1964–1980 (2019)
9. Mitrokhin, A., et al.: Ev-imo: Motion segmentation dataset and learning pipeline for event cameras pp. 6105–6112. (2019). doi:10.1109/IROS40897.2019.8968520
10. Zhou, Y., et al.: Event-based motion segmentation with spatio-temporal graph cuts. IEEE Transactions on Neural Networks and Learning Systems pp. 1–13. (2021). doi:10.1109/TNNLS.2021.3124580
11. Lu, X., Zhou, Y., Shen, S.: Event-based motion segmentation by cascaded two-level multi-model fitting pp. 4445–4452. (2021). doi:10.1109/IROS51168.2021.9636307
12. Shu, Y., et al.: Small moving object detection and tracking based on event signals pp. 792–796. (2021). doi:10.1109/ICCC54389.2021.9674247
13. Stoffregen, T., et al.: Event-based motion segmentation by motion compensation pp. 7243–7252. (2019). doi:10.1109/ICCV.2019.00734
14. Zhang, M., et al.: Maize point cloud data filtering algorithm based on vehicle 3d lidar. Transactions of the Chinese Society for Agricultural Machinery 50(4), 170–178 (2019)
15. Li, Y., Wang, J., Yao, K.: Modified phase correlation algorithm for image registration based on pyramid. Alexandria Engineering Journal 61(1), 709–718 (2022)
16. Liu, H., et al.: Point cloud segmentation based on euclidean clustering and multi-plane extraction in rugged field. Measurement Science and Technology 32(9), 095106 (2021)