

# JSTR: Joint Spatio-Temporal Reasoning for Event-based Moving Object Detection

Hanyu Zhou, Zhiwei Shi, Hao Dong, Shihan Peng, Yi Chang, and Luxin Yan\*

**Abstract**—Event-based moving object detection is a challenging task, where static background and moving object are mixed together. Typically, existing methods mainly align the background events to the same spatial coordinate system via motion compensation to distinguish the moving object. However, they neglect the potential spatial tailing effect of moving object events caused by excessive motion, which may affect the structure integrity of the extracted moving object. We discover that the moving object has a complete columnar structure in the point cloud composed of motion-compensated events along the timestamp. Motivated by this, we propose a novel joint spatio-temporal reasoning method for event-based moving object detection. Specifically, we first compensate the motion of background events using inertial measurement unit. In spatial reasoning stage, we project the compensated events into the same image coordinate, discretize the timestamp of events to obtain a time image that can reflect the motion confidence, and further segment the moving object through adaptive threshold on the time image. In temporal reasoning stage, we construct the events into a point cloud along timestamp, and use RANSAC algorithm to extract the columnar shape in the cloud for peeling off the background. Finally, we fuse the results from the two reasoning stages to extract the final moving object region. This joint spatio-temporal reasoning framework can effectively detect the moving object from motion confidence and geometric structure. Moreover, we conduct extensive experiments on various datasets to verify that the proposed method can improve the moving object detection accuracy by 13%.

## I. INTRODUCTION

Event camera [1] is a neuromorphic vision sensor with the advantage of low latency and high dynamic range, asynchronously triggers events due to illumination change, applied in space situational awareness [2], [3], [4] and robot obstacle avoidance [5]. Event-based moving object detection is a challenging task. When static camera platform, event camera only senses moving objects. When dynamic camera platform, it is difficult to detect moving objects since the background events generated by ego motion are mixed with the moving object events.

Existing methods mainly take optimization strategy [6], [7], [8], [9], [10] or inertial measurement unit (IMU) [5], [11], [12] to align background events with different timestamps to the image spatial coordinate at the same timestamp via motion compensation to distinguish moving objects. For example, Mitrokhin *et al.* [8] formulated 3D geometric relationships within event stream as an optimization model, and

\*Corresponding author.

Hanyu Zhou, Zhiwei Shi, Hao Dong, Shihan Peng, Yi Chang and Luxin Yan are with National Key Lab of Multispectral Information Intelligent Processing Technology, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China. Email: {hyzhou, shizhiwei, donghao0205, pengshihan, yichang, yanluxin}@hust.edu.cn

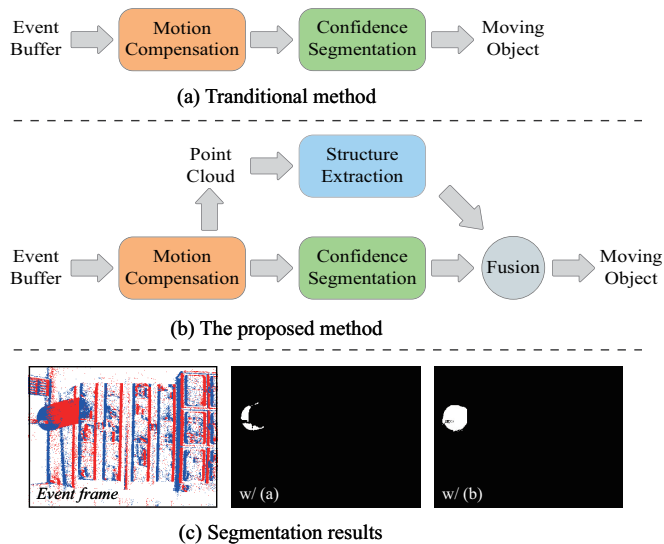


Fig. 1: Illustration of different moving object detection frameworks. (a) Traditional method. (b) The proposed joint spatio-temporal reasoning framework. (c) Corresponding segmentation results. Traditional method reasons the motion confidence distribution from spatial dimension for moving object, while suffering from the incomplete contour. Joint spatio-temporal reasoning can further exploit the structure characteristic of moving object from the temporal event point cloud for structure integrity of moving object.

solved the ego motion parameters via iterative optimization for aligning background events. Sanket *et al.* [10] fitted the homography transformation matrix between adjacent event frames using CNN, and warped the current event frame to the reference for distinguishing motion. These optimization-based methods rely heavily on high-quality event data, while suffering degradation when applied in adverse conditions (*e.g.*, nighttime scene). Falanga *et al.* [5] and Zhao *et al.* [12] utilized IMU to directly compensate background motion, avoiding the problem of degraded events. However, in Fig. 1 (a), these moving object detection methods neglect the potential spatial tailing effect of the moving object due to excessive motion, which affects the structure integrity of the extracted moving object in Fig. 1 (c). In this work, our goal (seeing Fig. 1 (b)) is to further improve the structure of the extracted moving object from event point cloud in Fig. 1 (c).

To address this issue, we map the motion-compensated events into the event frame in Fig. 2 (a), and transform the motion-compensated events as the point cloud along the timestamps in Fig. 2 (b). We observe that, in the spatial

event frame, the background is sharp while the moving object has a serious tailing effect; in the temporal point cloud, the background is globally discrete while the moving object shows a complete columnar structure. Therefore, we suggest that extracting the regular structure from event point cloud benefits to promoting event-based moving object detection.

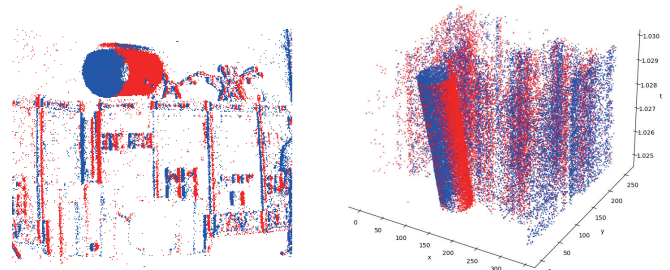
In this work, we propose a joint spatio-temporal reasoning method for event-based moving object detection, as shown in Fig. 3. Specifically, we first calculate the rotation matrix from the IMU via Rodrigues algorithm [13], and then use the rotation matrix to project events with different timestamps to the current timestamp through motion compensation. In the spatial reasoning stage, we map the compensated events to the same image coordinate, discretize the event timestamps to generate a motion confidence map (namely time image), and then segment the time image for the moving object via adaptive threshold. In the temporal reasoning stage, we transform the events into a point cloud along the timestamp, and take the RANSAC algorithm to extract the regular columnar structure in the point cloud for the moving object. Finally, we propose a contour-based fusion module to fuse the results of the spatial and temporal reasoning strategies to obtain the final moving object region. This joint spatio-temporal reasoning method can effectively detect the moving object from motion confidence and geometry structure. Moreover, we propose an event dataset to verify the effectiveness of the proposed method under various illumination conditions. Overall, the main contributions are as follows:

- We propose a joint spatio-temporal reasoning framework (JSTR) for event-based moving object detection, which can effectively extract the moving object with the complete regular structure.
- We discover that the moving object has a significant columnar structure in the event point cloud, which motivates us to design a RANSAC-based cloud structure extraction module for the moving object.
- We propose an event dataset for moving object detection, and verify the superiority of the proposed method on both the proposed and public event datasets.

## II. RELATED WORK

### A. Event-based Moving object Detection

Event moving object detection aims to extract the independent moving object from dynamic event stream. An intuitive method is object detection [6], [14], which used the ground truth to train a deep model for object detection. However, this method cannot distinguish whether the motion of the object is caused by the object itself or the ego motion of the camera. Another method is motion compensation-based moving object detection [6], [7], [8], [9], [10], [5], [11], [12], which mainly aligned the background events into the same image spatial coordinate to distinguishing the moving object using the motion obtained by optimization or assistant hardware. For example, Sanket *et al.* [10] utilized a deep model to fit the homography matrix. Mitrokhin *et al.* [8] constructed a motion energy function to iteratively



(a) Compensated event frame (b) Compensated event point cloud

Fig. 2: Visualization of event frame and point cloud. (a) Compensated event frame. (b) Compensated event point cloud. Moving object has an obvious tailing effect in the event frame, while has a significant columnar structure in the event cloud. This motivates us to improve moving object detection from the structure integrity of moving object.

optimize motion parameters. Stoffregen *et al.* [7] designed a local motion clustering algorithm to obtain the motion of different moving objects. These optimization-based methods rely heavily on high-quality event data, but may suffer degradation when applied in adverse conditions (*e.g.*, nighttime scenes). Recently, IMU-based motion compensation methods have received attention. Forrai *et al.* [11] and Zhao *et al.* [12] calculated the rotation matrix from IMU via Rodrigues algorithm for compensating the background motion, relieving the burden on high-quality event data. However, these motion compensation-based moving object detection approaches neglect the potential spatial tailing effect of moving object events, limiting the structure integrity of the extracted moving object. To overcome this, we further exploit the structure characteristic of the moving object in the event cloud to improve moving object detection.

### B. Point Cloud Regular Structure Extraction

The moving object has a columnar structure in the event point cloud, while the background is globally discrete. Our goal is to fit the regular structure and peel off the outliers in the event point cloud. A novel method is RANSAC [15], which randomly samples some points from the data set and iteratively optimizes the parameters of the pre-defined objective function of the regular structure to obtain the optimal model, whose output is the points from the regular structure. Tao *et al.* [16] designed a registration model to map the correspondence between different point clouds via RANSAC. Liu *et al.* [17] used RANSAC to fit the circle shape. Zhang *et al.* [18] removed the mismatched points during video stabilization using RANSAC. In this work, we are the first to introduce RANSAC to extract the regular structure of the moving object in the event point cloud.

## III. EVENTS DIFFERENCE ANALYSIS

Event-based moving object detection is to distinguish the moving object from the background, and the key is to model their difference. We map the compensated events to the same spatial image coordinate, and transform the compensated events to the temporal point cloud coordinate

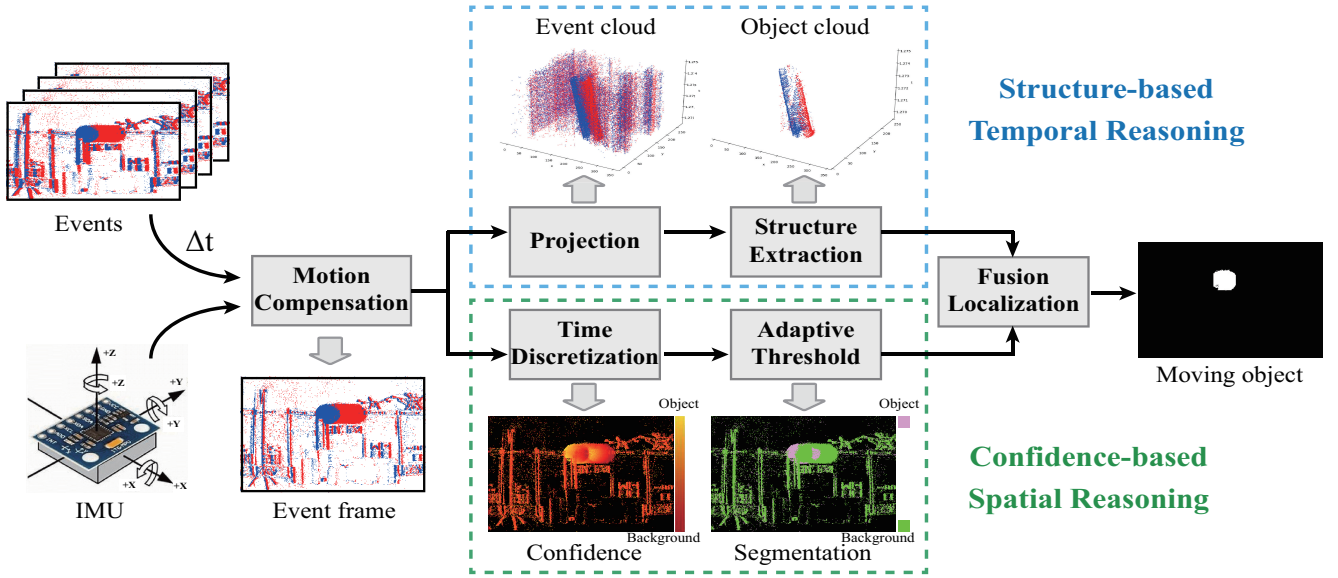


Fig. 3: The framework of joint spatio-temporal reasoning (JSTR) mainly contains confidence-based spatial reasoning and structure-based temporal reasoning. Motion compensation aligns the background events. Confidence-based spatial reasoning segments the moving object region via adaptive threshold segmentation on time image (namely motion confidence map). Structure-based temporal reasoning further extracts the complete contour of the moving object from the event cloud.

along the timestamp. As shown in Fig. 2 (a), motion-compensated background events show a clear boundary, while there exists an obvious tailing effect of the moving object events. The main reason is that, motion compensation is to align the background events caused by ego motion, not the independent moving object. Event camera with high temporal resolution can trigger a large number of moving object events, resulting in the tailing effect that affects the structure integrity of the extracted moving object. As shown in Fig. 2 (b), in the event point cloud, the background events are globally discrete, while the moving object has a significant 3D columnar structure. This is because events of moving object can still maintain their inertial motion in a short period, resulting in regular structure in the point cloud. Extracting such regular shape is beneficial to improving the structure integrity of the extracted moving object. Therefore, in Fig. 3, we predict the position of the moving object from the spatial event frame, and fine-tune the structure of the moving object from the temporal event point cloud.

#### IV. JOINT SPATIO-TEMPORAL REASONING

As shown in Fig. 3, we input a dynamic event stream and align the background events via motion compensation. Then, we segment the region of the moving object from the compensated events with the confidence-based spatial reasoning, and refine the structure of the moving object from the compensated events with the structure-based temporal reasoning. In this section, we will introduce the necessary data association and the specific module construction.

##### A. IMU-based Motion Compensation

The essence of motion compensation is to calculate the pose transformation between adjacent timestamps (target and reference), and warp the pixel coordinate of the target

timestamp to the reference timestamp via the geometry projection relationship [19]. Given a dynamic event stream  $C$  in a certain period, we can obtain the event  $e_t = [\mathbf{x}, t, p]$  and the corresponding angular velocity  $w_t$  from IMU, where  $\mathbf{x} = \{x, y\}$ ,  $t$ ,  $p$  are the pixel coordinate, timestamp, and polarity of the event, respectively. We first calculate the average angular velocity  $\bar{w} = \sum_{\delta t} w_t$ , and then calculate the three-axis Euler angles  $\alpha$ ,  $\beta$ ,  $\gamma$  of the event  $e_t$  relative to the reference timestamp  $t_0$ , where  $\alpha = \bar{w}_x(t - t_0)$ ,  $\beta = \bar{w}_y(t - t_0)$ ,  $\gamma = \bar{w}_z(t - t_0)$ . We obtain the rotation matrix  $R$  of the event  $e_t$  relative to the reference timestamp  $t_0$  via Rodrigues equation:

$$R = R_z(\gamma) \cdot R_y(\beta) \cdot R_x(\alpha), \quad (1)$$

where  $R_x$ ,  $R_y$  and  $R_z$  denote the rotation matrix with the Euler angles  $\alpha$ ,  $\beta$  and  $\gamma$ , respectively. Then, we transform the pixel coordinate  $\mathbf{x}$  of the event  $e_t$  to the reference timestamp via the geometry projection:

$$\hat{\mathbf{x}} = KRK^{-1}\mathbf{x}, \quad (2)$$

where  $K$  is camera internal parameter. Note that, camera translation is the same as the above transformation process, while the input is acceleration value rather than angular velocity. Thus, the compensated event stream is as follows:

$$\hat{C} = \{\hat{x}, \hat{y}, t_0\} \forall \{x, y, t\} \in C. \quad (3)$$

##### B. Confidence-based Spatial Reasoning

In the compensated event stream, object motion includes ego motion and independent motion, and there exists an obvious event density difference between the moving object and the background. That is to say, the higher the event density, the higher the moving object confidence. This motivates us to design a confidence-based spatial reasoning module.

**Time Discretization.** To model the event density difference between the moving object and the background, we follow

Falanga *et al.* [5] to discretize the timestamps of the events. We first count the number of compensated events at different pixels from target timestamp  $t$  to reference timestamp  $t_0$ :

$$\Omega_{ij} = \{\{\hat{x}, \hat{y}, t\} : \{\hat{x}, y, t_0\} \in \hat{C}, i = \hat{x}, j = \hat{y}\}, \quad (4)$$

where  $i, j$  denote the x,y-axis pixel coordinates of the events. Then, we construct the count image  $I_{ij} = |\Omega_{ij}|$ , and calculate the time image as follows:

$$T_{ij} = \frac{1}{I_{ij}} \sum t : t \in \Omega_{ij}. \quad (5)$$

Since the pixel value of the time image will become larger as the larger timestamp, the subsequent image process will be blocked. To this end, we normalize the time image to  $[-1, 1]$  for motion confidence as follows:

$$\rho(i, j) = T(t : t \in \Omega_{ij}) - \phi(T)/\delta t, \quad (6)$$

where  $\phi(T)$  denotes the mean operation.

**Adaptive Threshold Segmentation.** The normalized motion confidence map can reflect the probability of the moving object, and the key is how to reasonably set a threshold to segment the moving object. An intuitive strategy is to set a fixed threshold for segmentation. However, this strategy is not efficient. The main reason is that various ego motion can change the probability of the background, disturbing the impact of the segmentation threshold on the moving object extraction. Therefore, we introduce a linear function to adaptively adjust the segmentation threshold according to the IMU data as follows:

$$\tau(\omega) = a \cdot \|\omega\| + b, \quad (7)$$

where  $\omega$  denotes the square root of three-axis angular velocity, and  $a, b$  are the pre-defined parameters.  $a$  is a weight that adapts to the angular velocity, and  $b$  adjusts the classification threshold on moving object and background.

**Morphological Filter.** Adaptive threshold can allow the initial segmentation result, while there may exist the potential background events mixed with the moving object. To further remove the background events, we introduce a contour-based morphological filter strategy, which models the spatial shape difference between the moving object and the residual background. Specifically, we first use Sobel operator to extract the contours of the segmented image. Then, we construct a fixed-size sliding window with all one values and use the window to traverse all contours, where the inner product is utilized to measure the pixel distribution discrepancy between each contour and the sliding window. The smaller the distribution discrepancy, the more likely the moving object and should be retained; conversely, the background should be filtered. After contour-based morphological filter, the segmentation image basically retains the independent moving object regions.

### C. Structure-based Temporal Reasoning

Confidence-based spatial reasoning could promise a preliminary segmentation result of the moving object, while the structure integrity of the extracted moving object is limited by the object tailing effect. Combined with our observation in Fig. 2, we propose a structure-based temporal reasoning in the event point cloud. Next, we will describe the columnar

structure extraction method in detail.

**RANSAC-based Columnar Structure Extraction.** Moving object has a complete geometry structure in the event point cloud. Typical methods for extracting regular structure include Hough transformation [20], RANSAC [15], etc. The essence of RANSAC lies in random sampling consensus and iterative optimization, which maximizes the extraction of points covered by the moving object structure in the event point cloud. Given a compensated event stream, we transform the events to the point cloud  $D$  along the timestamp. We fit the columnar structure in the event point cloud as follows:

*Step 1:* We first randomly sample a set 3D points  $P$  from the event point cloud  $D$ . Since the moving object is a regular columnar structure, we set the points  $P$  as  $\{p_1, p_2, p_3, p_4\}$ .

*Step 2:* Columnar structure could be modeled by axis and radius characterization. We compute the normal vector of the columnar structure as its axis  $\vec{n}_z$ :

$$\vec{n}_1 = \frac{p_2 - p_1}{\|p_2 - p_1\|_2}, \vec{n}_2 = \frac{p_3 - p_1}{\|p_3 - p_1\|_2}, \quad (8)$$

$$\vec{n}_z = \vec{n}_1 \times \vec{n}_2. \quad (9)$$

*Step 3:* We calculate the distance between the point  $p_4$  and the normal vector  $\vec{n}_z$  as the radius of the columnar structure:

$$\vec{n}_3 = \frac{p_4 - p_1}{\|p_4 - p_1\|_2}, r_z = \frac{\|\vec{n}_z \times \vec{n}_3\|_2}{\|\vec{n}_3\|_2}. \quad (10)$$

*Step 4:* We further compute the distance between the other point  $p_i$  and the constructed model:

$$\vec{n}_i = \frac{p_i - p_4}{\|p_i - p_4\|_2}, r(p_i) = \frac{\|\vec{n}_z \times \vec{n}_i\|_2}{\|\vec{n}_i\|_2}. \quad (11)$$

*Step 5:* During one iteration, the whole event point cloud data is fed into the objective function that can determine the number of inliers:

$$\max |\{p_i | |r(p_i) - r_z| \leq \theta, p_i \in D\}|. \quad (12)$$

where  $\theta$  is the distance threshold. We iteratively optimize the object function Eq. 12 from *step 1* to *step 5* to obtain the optimal number of inliers, and the corresponding inliers are the event points from the moving object structure. Thus, RANSAC could fit the complete regular structure of the moving object, and benefit to relieve the limitation of the confidence-based spatial reasoning alone.

**Moving Object Fusion Localization.** The spatial reasoning can provide the position of the independent moving object, and the key is how to use the result of the temporal reasoning to ensure the structure integrity of the moving object. In this part, we design a contour-based fusion localization strategy. First, we back-project the extracted moving object point cloud into the image coordinate to obtain a clear contour. Next, we take this contour as the maximum boundary of the connected component, and expand the corresponding position of the moving object segmented by the spatial reasoning to the surrounding via connected component searching strategy [21]. If the boundary is exceeded during searching process, we take the region within the boundary of the connected component as the final moving object region. If the boundary is not exceeded, we choose the updated connected component as the final moving object region.

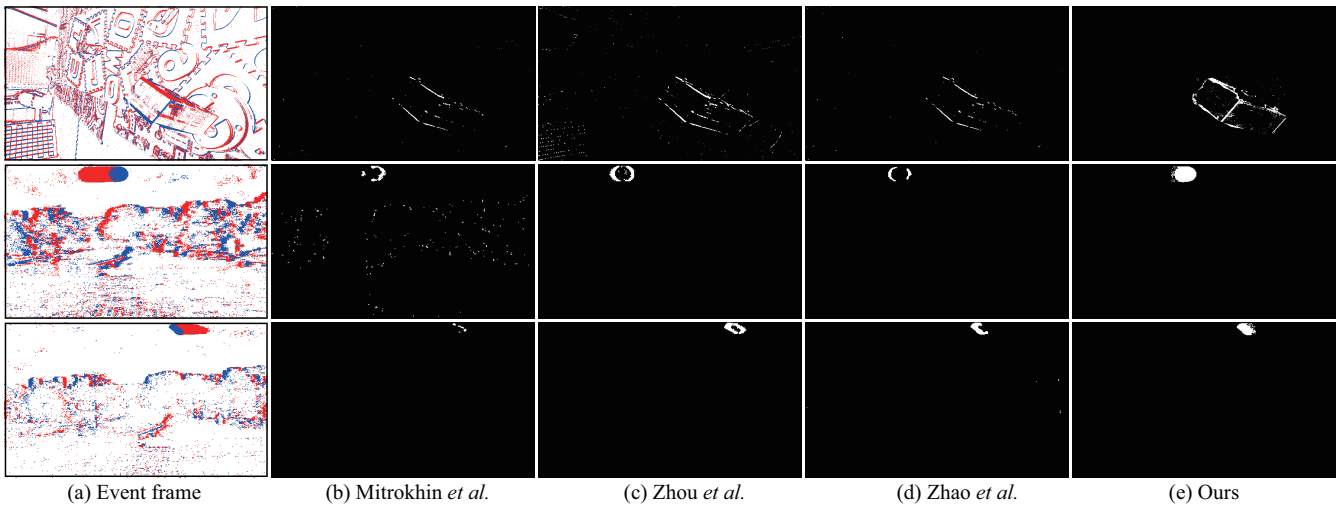


Fig. 4: Visual comparison of moving object detection on public (first row) and self-collected (last two rows) datasets.

TABLE I: Quantitative results on different datasets.

Method	EVIMO2 [22]		VIE	
	IoU	Accuracy	IoU	Accuracy
Mitrokhin <i>et al.</i> [8]	0.61	75.00%	0.54	60.41%
Zhou <i>et al.</i> [23]	0.63	71.42%	0.57	64.92%
Zhao <i>et al.</i> [12]	0.62	75.00%	0.62	66.70%
Ours	<b>0.64</b>	<b>78.57%</b>	<b>0.71</b>	<b>79.95%</b>

## V. EXPERIMENTS

### A. Experiment Setup

**Datasets.** We conduct extensive experiments on the public dataset EVIMO2 [22] and self-collected various-illumination event (VIE) dataset. Different from the public dataset EVIMO2, the proposed VIE dataset contains various ego-motion patterns (*i.e.*, rotation and translation) and fast moving objects under various illumination conditions (*i.e.*, ideal light and low light conditions), where we utilize DAVIS346 camera to collect the paired RGB-event data with manually handcrafted moving object bounding box labels.

**Comparison Methods.** We choose the optimization-based methods (Mitrokhin *et al.* [8] and Zhou *et al.* [23]) and the IMU-based method (Zhao *et al.* [12]) for fair comparison. The optimization-based methods use the motion estimated from the events for compensation and moving object detection, while the IMU-based method uses the readout data from the hardware IMU. In addition, we choose the Intersection over Union (IoU) and the accuracy of the moving object detection as the quantitative metric.

### B. Comparison Experiments

In Fig. 4 and Table I, we show the qualitative and quantitative comparison on public and self-collected datasets. We have two conclusions. First, our method and the IMU-based method Zhao *et al.* [12] significantly outperform the optimization-based methods Mitrokhin *et al.* [8] and Zhou

TABLE II: Ablation study on spatial and temporal reasoning.

Methods	IoU	Accuracy
Frame difference	0.21	0%
Confidence-based spatial reasoning	0.69	76.16%
Structure-based temporal reasoning	0.59	71.15%
Joint spatio-temporal reasoning (ours)	<b>0.71</b>	<b>79.95%</b>

*et al.* [23] by a large margin. The main reason is that, optimization methods require the ideal handcraft features, while there exist varying degrees of noise in dynamic event stream may weaken the discriminative features. On the contrary, IMU that is robust to the environment is able to provide accurate measure values for motion compensation. Second, the proposed method performs better than the IMU-based method Zhao *et al.* [12]. This is because Zhao *et al.* [12] only extracts the moving object via the spatial motion confidence segmentation, limiting the structure integrity of the moving object. In contrast, the proposed method could further improve moving object detection by extracting object structure from the event point cloud.

### C. Ablation Study and Discussion

#### Effectiveness of Spatio-Temporal Reasoning Framework.

In Table II, we demonstrate the effectiveness of the joint spatio-temporal reasoning framework. Without any reasoning approach, we only use the frame difference method [24] to detect the moving object, which hardly performs well. With only confidence-based spatial reasoning, the object performance improves significantly. With only structure-based temporal reasoning, the object performance is slightly weaker than only spatial reasoning. The reason is that, although the object extracted from event point cloud has a complete structure, the moving object region is hollow, limiting the performance. Instead, the proposed joint spatio-temporal reasoning can fuse the advantages of the two reasoning strategies for improving the moving object detection.

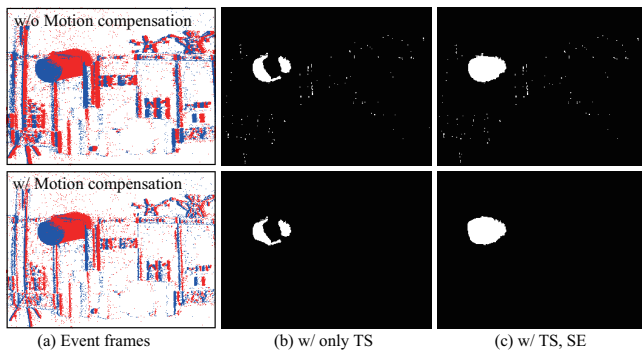


Fig. 5: Effectiveness of different modules. (a) Event frames with and without motion compensation. (b) Detection results with only threshold segmentation (TS). (c) Detection results with threshold segmentation and structure extraction (SE).

TABLE III: Discussion on the impact of various ego motion patterns of the camera on moving object detection.

Rotation	Translation	Zhao <i>et al.</i> [12]	Ours
×	×	100%	100%
√	×	80.23%	<b>87.36%</b>
×	√	62.51%	<b>81.82%</b>
√	√	54.65%	<b>80.00%</b>

**How do Various Modules Improve Detection?** We conduct an ablation study on the impact of different modules on the final moving object detection results in Fig. 5. Motion compensation can align the background events so that background boundaries are sharp. Confidence threshold segmentation can extract the location of the moving object according to the motion confidence, while the object structure is incomplete. Structure extraction from the event point cloud could further refine the contour of the extracted object, improving moving object detection.

**Adaptability on Various Ego Motion Patterns.** In Table III, we analyze the adaptability of the IMU-based method Zhao *et al.* [12] and our method to different ego motion patterns of camera itself. With only the ego motion of rotation, both the Zhao *et al.* [12] and our method can perform well. With only the ego motion of translation, the performance of the IMU-based method declines seriously, while the proposed method can still maintain a reliable performance. Under the complex ego motion including rotation and translation, our performance only declines slightly. This is because IMU only provides globally identical translation motion, and ego motion of translation makes it difficult to align the background events at different depths. On the contrary, the proposed method can exploit the regular structure of the moving object from the event point cloud to improve detection performance.

**Generalization for Different Conditions.** In Fig. 6, we discuss the generalization of the proposed method for different illumination conditions. In Fig. 6 (a), we set five illumination conditions, including three low light conditions and two ideal light conditions. As shown in Fig. 6 (a)-(c), we can conclude

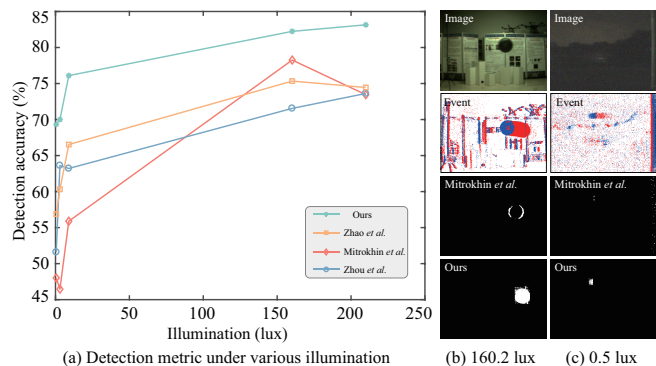


Fig. 6: Generalization for various illumination conditions. (a) Comparison curve of different methods under various illumination. (b) Visualization under ideal light (160.2 lux). (c) Visualization under low light (0.5 lux).

that, the optimization-based methods perform well under ideal illumination, while suffering degradation caused by the noise events under low light conditions, thus interfering with the segmentation results. On the contrary, with the help of IMU, the proposed method shows a good performance in both ideal and low light conditions.

**Limitation.** The proposed method could extract the independent moving object with regular geometric shapes. However, in autonomous driving, though our method can effectively locate the object, it is ineffective in ensuring the structure integrity of irregular objects (*e.g.*, pedestrians). The reason is that irregular objects do not follow the regular structure in the event point cloud, resulting in the ineffectiveness of the structure-based temporal reasoning. In the future, we will introduce deep models to non-linearly learn the temporal structure of moving objects to improve the robustness of the proposed method in various practical scenes.

## VI. CONCLUSIONS

In this work, we propose a novel joint spatio-temporal reasoning framework for moving object detection from motion confidence and geometry structure. To the best of our knowledge, we are the first to discover the columnar structure of the moving object in the event point cloud and investigate the RANSAC-based regular structure extraction for moving object detection. The proposed method can effectively extract the moving object with complete contour from dynamic event stream. We demonstrate that the proposed method significantly outperforms the state-of-the-art methods on both the public and self-collected datasets. We believe that our work could not only facilitate the development of moving object detection but also enlighten the researchers of the broader field, *i.e.*, scene understanding under adverse conditions.

## ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 62371203, in part by the National Natural Science Foundation of China under Grant 62176100, and in part by Xiaomi Young Talents Program.

## REFERENCES

- [1] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(1):154–180, 2020.
- [2] Mohsi Jawaid, Ethan Elms, Yasir Latif, and Tat-Jun Chin. Towards bridging the space domain gap for satellite pose estimation using event sensing. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11866–11873. IEEE, 2023.
- [3] Gregory Cohen, Saeed Afshar, Brittany Morreale, Travis Bessell, Andrew Wabnitz, Mark Rutten, and André van Schaik. Event-based sensing for space situational awareness. *The Journal of the Astronautical Sciences*, 66:125–141, 2019.
- [4] Saeed Afshar, Andrew Peter Nicholson, Andre Van Schaik, and Gregory Cohen. Event-based object detection and tracking for space situational awareness. *IEEE Sensors Journal*, 20(24):15117–15132, 2020.
- [5] Davide Falanga, Kevin Kleber, and Davide Scaramuzza. Dynamic obstacle avoidance for quadrotors with event cameras. *Science Robotics*, 5(40):eaz9712, 2020.
- [6] Chethan M Parameshwara, Simin Li, Cornelia Fermüller, Nitin J Sanket, Matthew S Evanusa, and Yiannis Aloimonos. Spikems: Deep spiking neural network for motion segmentation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3414–3420. IEEE, 2021.
- [7] Timo Stoffregen, Guillermo Gallego, Tom Drummond, Lindsay Kleeman, and Davide Scaramuzza. Event-based motion segmentation by motion compensation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7244–7253, 2019.
- [8] Anton Mitrokhin, Cornelia Fermüller, Chethan Parameshwara, and Yiannis Aloimonos. Event-based moving object detection and tracking. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–9. IEEE, 2018.
- [9] Xiuyuan Lu, Yi Zhou, and Shaojie Shen. Event-based motion segmentation by cascaded two-level multi-model fitting. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4445–4452. IEEE, 2021.
- [10] Nitin J Sanket, Chethan M Parameshwara, Chahat Deep Singh, Ashwin V Kuruttukulam, C Fermuller, Davide Scaramuzza, and Yiannis Aloimonos. Evdodge: Embodied ai for high-speed dodging on a quadrotor using event cameras. *arXiv preprint arXiv:1906.02919*, pages 31–45, 2019.
- [11] Benedek Forrai, Takahiro Miki, Daniel Gehrig, Marco Hutter, and Davide Scaramuzza. Event-based agile object catching with a quadrupedal robot. pages 12177–12183, 2023.
- [12] Chunhui Zhao, Yakun Li, and Yang Lyu. Event-based real-time moving object detection based on imu ego-motion compensation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 690–696. IEEE, 2023.
- [13] Johan Ernest Mebius. Derivation of the euler-rodrigues formula for three-dimensional rotations from the general formula for four-dimensional rotations. *arXiv preprint math/0701759*, 2007.
- [14] Manish Nagaraj, Chamika Mihiranga Liyanagedera, and Kaushik Roy. Dotie-detecting objects through temporal isolation of events using a spiking architecture. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4858–4864. IEEE, 2023.
- [15] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [16] Wuyong Tao, Shaoping Xu, Wei Huang, Shufan Hu, and Meng Pang. A distinctive binary descriptor and 2-point ransacwc for point cloud registration. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.
- [17] Yong-Jin Liu, Jun-Bin Zhang, Ji-Chun Hou, Ji-Cheng Ren, and Wei-Qing Tang. Cylinder detection in large-scale point cloud of pipeline plant. *IEEE transactions on visualization and computer graphics*, 19(10):1700–1707, 2013.
- [18] Feida Zhang, Xin Li, Taoyang Wang, Guo Zhang, Jianzhi Hong, Qian Cheng, and Tiancheng Dong. High-precision satellite video stabilization method based on ed-ransac operator. *Remote Sensing*, 15(12):3036, 2023.
- [19] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G Lowe. Unsupervised learning of depth and ego-motion from video. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1851–1858, 2017.
- [20] Richard O Duda and Peter E Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.
- [21] Lifeng He, Yuyan Chao, Kenji Suzuki, and Kesheng Wu. Fast connected-component labeling. *Pattern recognition*, 42(9):1977–1987, 2009.
- [22] Anton Mitrokhin, Chengxi Ye, Cornelia Fermüller, Yiannis Aloimonos, and Tobi Delbruck. Ev-imo: Motion segmentation dataset and learning pipeline for event cameras. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6105–6112. IEEE, 2019.
- [23] Yi Zhou, Guillermo Gallego, Xiuyuan Lu, Siqi Liu, and Shaojie Shen. Event-based motion segmentation with spatio-temporal graph cuts. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [24] Nishu Singla. Motion detection based on frame difference method. *International Journal of Information & Computation Technology*, 4(15):1559–1565, 2014.