EVENT-BASED VISION ENHANCED: A JOINT DETECTION FRAMEWORK IN AUTONOMOUS DRIVING

Jianing Li[1], *Siwei Dong*[1], *Zhaofei Yu*[1,3], *Yonghong Tian*[1,2,3*] , *Tiejun Huang*[1,2,3]

[1]National Engineering Laboratory for Video Technology, School of EE&CS, Peking University, Beijing, China
[2] School of ECE, Shenzhen Graduate School, Peking University, Shenzhen, China
[3]Pengcheng Laboratory, Shenzhen, China

ABSTRACT

Due to the high-speed motion blur and low dynamic range, conventional frame-based cameras have encountered an important challenge in object detection, especially in autonomous driving. Event-based cameras, by taking the advantages of high temporal resolution and high dynamic range, have brought a new perspective to address the challenge. Motivated by this fact, this paper proposes a joint framework combining event-based and frame-based vision for vehicle detection. Specially, two separate event-based and frame-based streams are incorporated into a convolutional neural network (CNN). Besides, to accommodate the asynchronous events from event-based cameras, a convolutional spiking neural network (SNN) is utilized to generate visual attention maps so that two streams can be synchronized. Moreover, Dempster-Shafer theory is introduced to merge two outputs from CNN in a joint decision model. The experimental results show that the proposed approach outperforms the state-of-the-art methods only using frame-based information, especially in fast motion and challenging illumination conditions.

Index Terms— Event-based Vision, Neuromorphic Cameras, Convolutional Neural Networks, Spiking Neural Networks, Dempster-Shafer Theory

1. INTRODUCTION

Autonomous driving systems have been widely researched in recent years, and it will be increasingly adopted by the general public in future [1, 2]. At present, vision cameras, along with radar, LiDAR, ultrasound, form the backbone of autonomous driving systems [3], and can obtain high spatial resolution and adequate videos for machine vision models [4, 5]. In fact, vision sensors have played a key role to understand the real driving scenes, and accurately and promptly detecting dangerous vehicles in vision-based intelligent systems is extremely important for preventing traffic accidents.

Previously, much research has been done focusing on vehicle detection [4, 6] using frame-based cameras, namely active-pixel sensors (APS). The earliest methods to achieve real-time detection are mainly cascade detection based on local features. After that, object detection systems have been enhanced significantly by deep neural networks such as Fast R-CNN [7] and Faster R-CNN [8]. Furthermore, end-to-end object detection models which include SSD [9] and YOLOs [10, 11] have appeared. However, those frame-based methods can achieve a satisfactory performance only under special conditions, including slow motion and proper illumination. Actually, vehicle detection is still challenging for frame-based cameras due to the complicated road conditions with large illumination variations, especially in over-exposed and insufficient light scenes. In addition, frames suffer from motion blur in high-speed movement so that subsequent algorithms have failed to capture object.

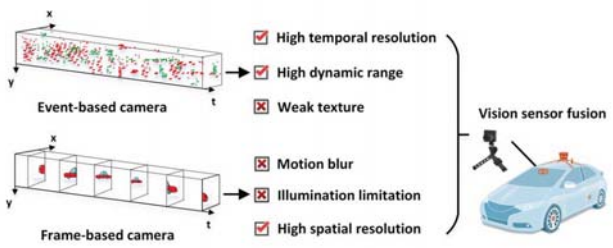


Fig. 1. Combining event-based and frame-based vision for vehicle detection.

In order to appreciate how biological approaches and neuromorphic engineering techniques could be beneficial for advancing artificial vision [12], it is inspiring to look at some shortcomings of frame-based cameras. Recently, event-based cameras, namely neuromorphic cameras, such as dynamic vision sensor (DVS) [13], ATIS [14] and DAVIS [15], are bio-inspired vision sensors that, in contrast to frame-based cameras, work in a completely different way: acquiring a stream of asynchronous events for independent pixels, instead of providing a sequence of frame-based images at a fixed rate, as shown in Fig.1. Event-based cameras have some key advantages over frame-based cameras: high temporal resolution (μs), high dynamic range (HDR) and low power consumption due to convey sparse events with little redundancy. In addition, events are outputted only when intensity changes so that event-based cameras are natural object moving detectors and have a flaw with weak texture in spatial structure. Indeed, event-based cameras are gradually applied to computer vision tasks [16, 17, 18] related to motion estimation.

Aiming at the shortages of frame-based cameras, some researchers have focused on object detection [17, 19, 20, 21]

* Corresponding author (Y. Tian, email: yhtian@pku.edu.cn).

978-1-5386-9552-4/19/$31.00 ©2019 IEEE
DOI 10.1109/ICME.2019.00242

based on event-based cameras. Mesa *et al.* [17] proposed an event-driven stereo object detection and tracking algorithm which can solve high-speed movement object occlusion. Li *et al.* [19] introduced a recursive adaptive temporal pooling method to extract motion invariant features for object detection. Anton *et al.* [20] presented a multiple moving object detection approach in challenging conditions with fast motion or lighting variations. Moreover, Chen [21] used pseudo-labels for supervised learning on DVS data to object detection under ego-motion. However, those methods utilized only event-based streams without using frame-based cameras. In fact, event-based cameras can be the principal information to auxiliary frame-based object detection [22, 23]. In other words, combining event-based and frame-based vision can further improve detection performance.

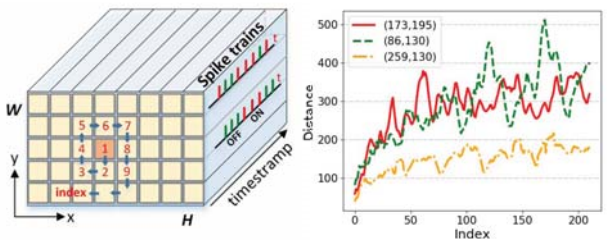
Inspired by the pros and cons of previous works, this paper proposes a joint detection framework (JDF) based on DAVIS, which outputs both conventional frames and event streams. As a matter of fact, the goal of this work is not to develop a state-of-the-art detector [8, 9, 10, 11]. In contrast, we aim at overcoming the following challenges: 1) How to take advantages of event-based and frame-based streams; 2) Sparse and asynchronous events are well applied to computer vision algorithms, especially exploiting effective spatial-temporal features from event streams. As a result, two separate streams are integrated into convolutional neural network (CNN). Besides, to accommodate the asynchronous events from event-based cameras, a convolutional spiking neural network (SNN) is utilized to generate visual attention maps so that two streams can be synchronized. Moreover, Dempster-Shafer mechanism [24] is introduced in joint decision model, and it achieves impressive performance on DDD17 [25] dataset, especially in fast motion and challenging illumination conditions.

The main contributions are summarized as follows: 1) We introduce Dempster-Shafer theory to the proposed JDF by combining event-based and frame-based vision; 2) We present a convolutional SNN to generate visual attention maps so that it builds a bridge linking asynchronous events to deep learning algorithms. Furthermore, we show that it is possible to utilize transfer learning from pretrained models on detection tasks; 3) We provide a labeled and synchronized dataset [†] including frames and event streams, and the experiments have validated the effectiveness of the proposed framework, in which JDF outperforms the state-of-the-art methods on the basis of frame-based cameras.

2. OUR APPROACH

In this section, we first explain spatial-temporal events from event-based cameras and describe the basic concepts in the convolutional SNN. Then, we present the components in the proposed JDF. Finally, we introduce Dempster-Shafer theory applied in joint decision model.

[†]https://pkuml.org/resources/pku-ddd17-car.html



(a) Event stream (b) The distance of spike trains

Fig. 2. Measuring spike train distance from event streams.

2.1. Spatial-Temporal Events

Given an event-based camera with a resolution of $W \times H$, as is shown in Fig.2(a), a stream of events $\boldsymbol{\varepsilon}$ can be mathematically defined as:

$$\boldsymbol{\varepsilon} = \{e_i\}_{i=1}^{I}, with \quad e_i = [x_i, y_i, t_i, p_i]^T \tag{1}$$

where e_i is the ith event, namely spike, and it includes timestamp (t_i), event location ($[x_i,y_i] \in W \times H$), polarity (p_i), with $p_i \in \{1, -1\}$ representing ON and OFF events respectively.

In fact, there is merit to exploring spatial-temporal characteristics from event-based cameras to serve motion estimation. Since event streams are asynchronous and sparse point process, it imposes important challenges to signal processing methods. For two spike trains $s^m, s^n \in s(\Gamma)$, the inner product [26] is introduced to measure distance in the Hilbert space as follows:

$$F(s^m, s^n) = \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} \kappa\left(t_i^m, t_j^n\right) = \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} e^{-\frac{\left(t_i^m - t_j^n\right)^2}{2\sigma^2}} \tag{2}$$

where kernel κ is the autocorrelation of Gaussian smoothing function $h(t) = exp(-t^2/2\sigma^2)$, with events firing times $t_i^m, t_j^n \in \Gamma$.

In this work, one pixel can be chosen as the ordinate origin from event streams, and the surrounding pixels make up spike trains in Fig.2(a). The distance of spike trains for three selected pixels are shown in Fig.2(b), it illustrates that spike trains from neighboring pixels are more relevant.

2.2. Convolutional Spiking Neural Network

Many works [18, 19, 20, 21] have shown that event streams can be converted to images [18, 19, 21] or time surfaces [20] in constant interval, so that it can be applied to computer vision algorithms. However, those methods using rate-based strategy have not yet exploit spatial-temporal characteristics. Inspired by biologically visual receptive fields, a convolutional SNN is utilized to generate visual attention maps based on firing rates of output neurons.

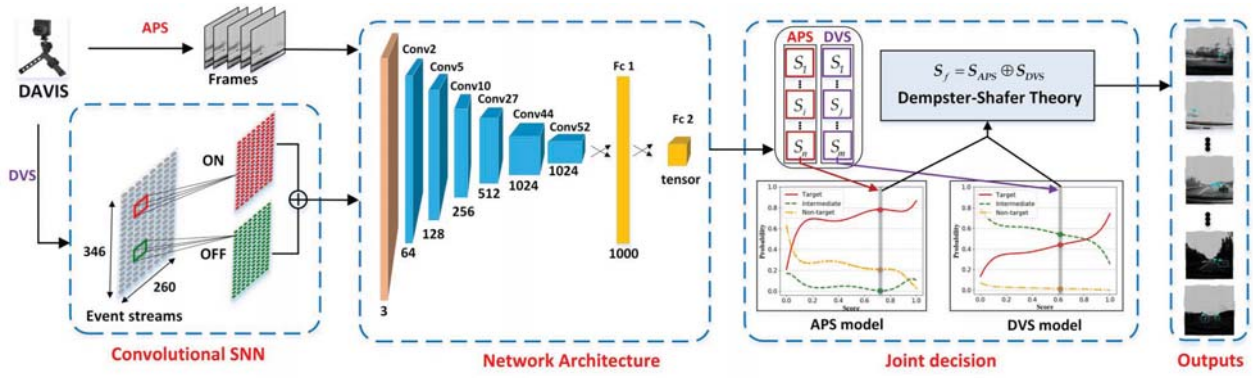


Fig. 3. The joint detection framework combining event-based and frame-based vision. Specially, convolutional SNN is used to generate visual attention maps so that two streams can be synchronized. Then, two separate event-based and frame-based streams are incorporated into network architecture. Finally, Dempster-Shafer theory is introduced in joint decision model.

In network topology, two layers networks are connected using a 3×3 convolutional kernels, as is shown in Fig.3. The first layer is the input of event streams, and which has $2{\times}W{\times}H$ neurons. In other words, each pixel includes two neurons representing ON and OFF ganglion cells in biological retina. In addition, the second layers are outputs of visual attention maps, and which has $W{\times}H$ neurons to obtain firing rates of ON or OFF layer, respectively.

In this study, we use the leaky-integrated-and-fire (LIF) neuron [27] to emulate neuronal dynamics in SNN as follows:

$$\tau_m \frac{dV}{dt} = -V + w * \delta\left(t - t_i\right) \tag{3}$$

where V is the neuronal membrane potential, w is the synaptic weight, and τ_m is the time constant. Between two spikes, the membrane potential of LIF is presented as:

$$V_i = V_{i-1} e^{-\frac{t_i - t_{i-1}}{\tau_m}} + w \tag{4}$$

when V_i reaches the threshold, it fires an spike. Then, V_i is reseted to zero without again until over the refractory period.

2.3. Joint Detection Framework

The proposed framework is shown in Fig.3, there are two branches which aim at processing frames and event streams. Indeed, the one core of this framework is that event streams are generated visual attention maps based on convolutional SNN, the details are presented in Section 2.2. Moreover, the other is that two separate streams are fed into CNN which leverages transfer learning from pretrained detection models.

It is important that we make great effort to build a bridge linking event streams to existing detectors [7, 8, 9, 10, 11]. In this sense, we choose YOLOv3 architecture [11] as a fair benchmark considering the balance of accuracy and complexity. As it has $G{\times}G$ grid, B predicting bounding boxes for each grid cell, and C class predictions. In this work, the last fully-connected (FC) layer is adjusted as an $G{\times}G{\times}B{\times}(C+5)$ tensor.

Finally, two outputs of the last CNN layer are integrated as detection results based on Dempster-Shafer theory [24], which obtains synthetic judgment by combing evidences from probabilities of related hypotheses.

For object detection, the universal set represents various possible states for bounding boxes, and it is defined as:

$$\Omega = \{T, \neg T, \{T, \neg T\}\} \tag{5}$$

where T is target hypothesis, $\neg T$ is non-target, and $\{T, \neg T\}$ is intermediate state.

To obtain dynamic probability assignment, the trained precision-recall model is utilized to represent prior information of detectors. Meanwhile, we introduce a theoretical best possible detector, as is shown in Fig.4, and it is modeled as:

$$\hat{p_b} = 1 - r^k \tag{6}$$

where $\hat{p_b}$ is a theoretical limit for best possible detector in recall r and performance parameters k.

To compute joint probabilities of the hypotheses for output belief scores S_1 and S_2, the combination rule based on Dempster-Shafer theory is presented as:

$$S_f\left(A\right) = S_1 \oplus S_2 = \frac{1}{K} \sum_{B_1 \cap B_2 = A} S_1\left(B_1\right) S_2\left(B_2\right) \tag{7}$$

where B_1 and B_2 are subsets of Ω, and K is normalization constant measuring the amount of conflict between universal sets, and it is given by:

$$K = 1 - \sum_{B_1 \cap B_2 \neq \emptyset} S_1\left(B_1\right) S_2\left(B_2\right) \tag{8}$$

After dynamic probability assignment, the issue becomes how to combine independent sources, in other words, how to combine bounding boxes and probabilities from two streams, respectively. The joint probabilities are merged based on Dempster-Shafer theory, then non-maximum suppression (NMS) [6] is adopted to integrate bounding boxes.

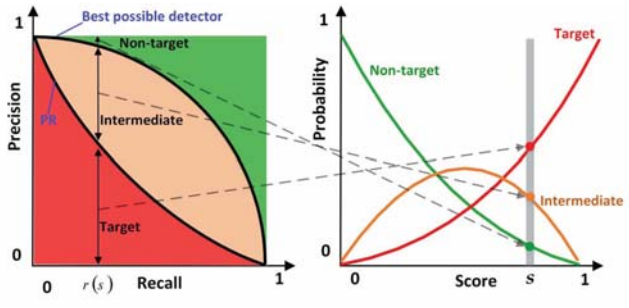


Fig. 4. Dynamic probability assignment.

3. EXPERIMENTS

In this section, detailed experimental settings, performance scores and representative results can be found as follows.

3.1. Experimental Settings

To verify the effectiveness of our methods, we conduct experiments on DDD17 [25] dataset including frames and event streams, and which has over 400GB and 12 hours of a 346×260 pixel DAVIS sensor recording driving scenes. In order to obtain accurate labels for vehicle detection, we provide a hand-labeled dataset including synchronized frames and event streams, as shown in Table 1. Moreover, we fine-tune the pretrained YOLOv3 model on the labeled DDD17 dataset, set 0.5 overlap threshold and 0.5 scores throughout the experiments. All timing information is on a Tesla K80.

Table 1. Details of the labeled DDD17 dataset.

File(.hdf5)	Condition	T(s)	Label	Type
1487339175	day	347	27	test
1487417411	day	2096	419	test
1487419513	day	1976	204	train
1487424147	day	3040	388	train
1487430438	day	3135	343	train
1487433587	night-fall	2335	145	train
1487593224	day	524	40	test
1487594667	day	2985	196	train
1487597945	night-fall	50	16	test
1487598202	day	1882	618	train
1487600962	day	2143	218	test
1487608147	night-fall	1208	348	train
1487609463	night-fall	101	183	test
1487781509	night-fall	127	10	test

To conduct comprehensive evaluation of the proposed joint detection framework (JDF), we compare JDF with the state-of-the-art models and two baselines, including:

(1) APS [11]: A frame-based detection method that only takes APS frames as input.

(2) Rate-based DVS (R-DVS) [21]: The approach that adopts rate-based methodology to convert event streams into frames in 10ms interval as input.

Table 2. Effectiveness test on day condition.

Methods	Precision	Recall	AP	FPS
R-DVS [21]	0.874	0.212	0.357	9
S-DVS	0.889	0.256	0.414	9
APS [11]	0.889	0.755	0.867	9
S-DVS + APS	0.898	0.783	0.897	9
JDF	**0.941**	**0.806**	**0.908**	**9**

Table 3. Effectiveness test on night-fall condition.

Methods	Precision	Recall	AP	FPS
R-DVS [21]	0.823	0.267	0.382	9
S-DVS	0.868	0.293	0.437	9
APS [11]	0.852	0.574	0.744	9
S-DVS + APS	0.872	0.621	0.761	9
JDF	**0.926**	**0.672**	**0.833**	**9**

(3) Spike-based DVS (S-DVS): Different from R-DVS, S-DVS that uses spike-based methodology by Convolutional SNN to generate visual attention maps.

(4) APS + S-DVS: Two streams of APS and S-DVS are merged into frames before input the detector.

To compare different approaches, precision, recall and average precision (AP), frames per second (FPS) are adopted as evaluation metrics, which are the most widely used metrics in object detection.

3.2. Effectiveness Test

In this section, we will test JDF on DDD17 dataset and report the results. We will also explore several experiments to see why and how JDF works.

Evaluation on day condition. Performances of all approaches on day condition can found in Table 2. From Table 2, we can see that the proposed JDF achieves impressive performance on day condition in DDD17 dataset. In particular, JDF outperforms APS, R-DVS, or S-DVS that only using either frames or event streams, respectively. Meanwhile, we can find that the remarkable performance enhancement from S-DVS + APS to JDF after joint decision model based on Dempster-Shafer theory. In addition, S-DVS obtains better performance than R-DVS due to spike-based methodology. This may be caused by the fact that convolutional SNN towards better representations for event streams than rate-based methodology.

Evaluation on night-fall condition. By comparing Table 3 and Table 2, we can see that even with night-fall condition the performance of both S-DVS and R-DVS still maintain stable. However, APS drops sharply in such a challenging setting. In other words, after incorporating the auxiliary event streams, the performance of JDF can significantly improve over APS. Actually, event-based cameras, by taking the advantages of high temporal and HDR, have brought a new perspective to overcome the shortages of frame-based cameras.

Benefit from JDF. Some representative detection results are illustrated in Fig.5. As is shown in Fig.5, JDF have bet-

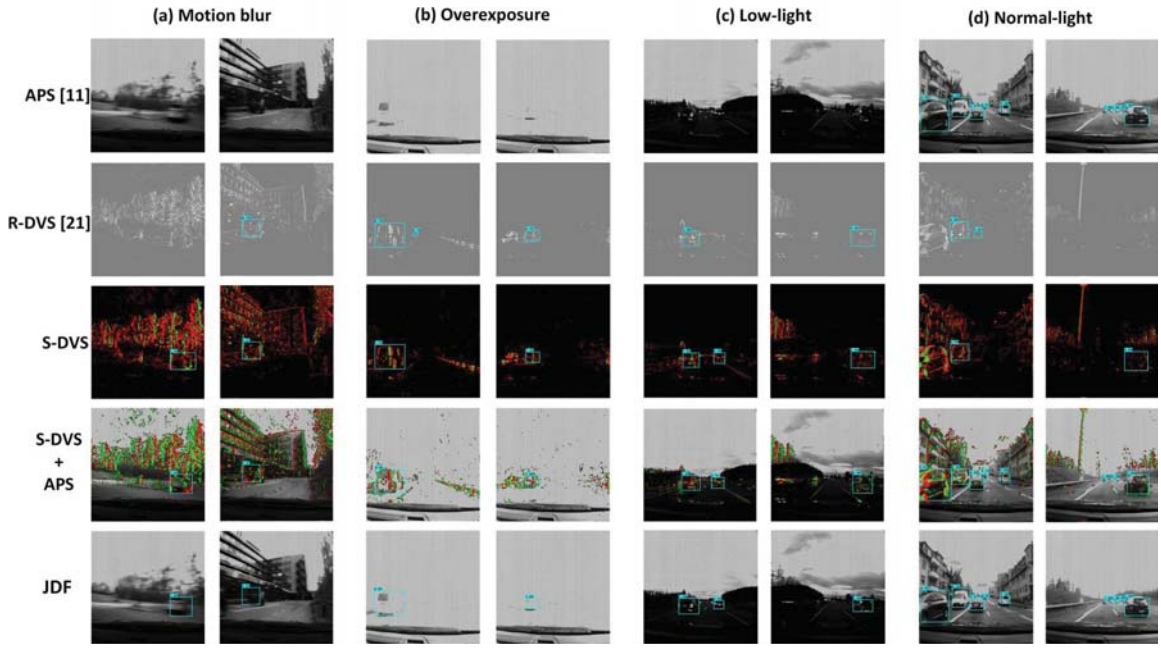
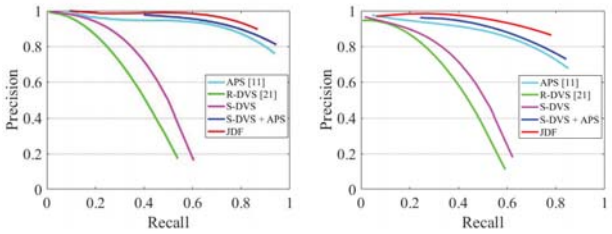


Fig. 5. Representative results on DDD17 dataset. (a) Motion blur; (b) Overexposure; (c) Low-light; (d) Normal-light. It is clear that JDF achieves better performance than other methods on various conditions, especially in fast motion and challenging illumination scenes.

Table 4. Scalability test for the state-of-the-art methods on DDD17 dataset.

Architectures	Precision		Recall		AP		FPS	
	APS	**JDF**	APS	**JDF**	APS	**JDF**	APS	**JDF**
Faster-RCNN[8]	0.749	**0.829**	0.863	**0.907**	0.802	**0.866**	3	**3**
SSD [9]	0.849	**0.886**	0.641	**0.664**	0.731	**0.759**	12	**12**
YOLOv2 [10]	0.714	**0.829**	0.691	**0.832**	0.702	**0.778**	15	**15**
YOLOv3 [11] (**benchmark**)	0.928	**0.939**	0.695	**0.762**	0.795	**0.841**	9	**9**



(a) Day condition (b) Night-fall condition

Fig. 6. Precision-recall curves on DDD17 dataset.

ter detection results than other methods on various conditions. Moreover, it is clear that APS fails to capture targets in challenging scenes, including motion blur, overexposure, night-fall. This is an interesting findings, implying that the usage of the auxiliary event stream can improve performance, especially in fast motion and challenging illumination conditions.

Validate the robustness. The curves of precision-recall models on DDD17 dataset are illustrated in Fig.6(a) and Fig.6(b), respectively. From these results, we find that the proposed JDF has not only better robustness than other methods on day condition but also fits for night-fall condition.

Evaluating time complexity. As shown in Table2 and Table 3, time complexity analysis among all approaches, in which the last column, FPS depicts the speed of detectors. The results agree with that the overall performance of JDF has significantly improved meanwhile the computational speed is almost comparable in contrast to other methods.

3.3. Scalability Test

Beyond effectiveness test, we also conduct several experiments to compare JDF and APS on several state-of-the-art methods [8, 9, 10, 11], as is shown in Table 4. For the fact that we select YOLOv3 as a fair benchmark considering the balance of the accuracy and time complexity. Actually, any detector can be an alternative owning to that our work has implemented a generic interface providing the input for the detector as well as joint decision model. Note that JDF can significantly improve performance for the state-of-the-art methods only using frame-based streams.

Moreover, the goal of this work is not to develop a powerful detector. On the contrary, we aim at the challenges including how to perform an effective joint detection framework combining two streams and how to build a bridge linking asynchronous events to algorithms.

4. CONCLUSION

In this paper, we propose a joint detection framework (JDF) combining event streams and frames, which aims to make two streams benefit from each other based on Dempster-Shafer theory. As demonstrated by the experimental results on the DDD17 dataset, our JDF can improve the two streams framework remarkably and outperforms the state-of-the-art methods only using frame-based cameras, especially in fast motion and challenging illumination conditions.

Acknowledgment. This work is partially supported by grants from the National Basic Research Program of China under grant 2015CB351806, the National Natural Science Foundation of China under contract No.U1611461, No. 61825101, No. 61806011 and No.61425025.

5. REFERENCES

[1] Sayanan Sivaraman and Mohan Manubhai Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Transactions on Intelligent Transportation Systems*, 2013.

[2] Jason Borenstein, Joseph Herkert, and Keith Miller, "Self-driving cars: Ethical responsibilities of design engineers," *IEEE Technology and Society Magazine*, 2017.

[3] Sujeet Milind Patole, Murat Torlak, Dan Wang, and Murtaza Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, 2017.

[4] Benjamin Ranft and Christoph Stiller, "The role of machine vision for intelligent vehicles," *IEEE Transactions on Intelligent Vehicles*, 2016.

[5] Nasim Arbabzadeh and Mohsen Jafari, "A data-driven approach for driving safety risk prediction using driver behavior and roadway information data," *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[6] Wenqing Chu, Yao Liu, Chen Shen, Deng Cai, and Xian-Sheng Hua, "Multi-task vehicle detection with region-of-interest voting," *IEEE Transactions on Image Processing*, 2018.

[7] Ross Girshick, "Fast r-cnn," in *CVPR*, 2015.

[8] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *NIPS*, 2015.

[9] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg, "Ssd: Single shot multibox detector," in *ECCV*, 2016.

[10] Joseph Redmon and Ali Farhadi, "Yolo9000: better, faster, stronger," *CVPR*, 2017.

[11] Joseph Redmon and Ali Farhadi, "Yolov3: An incremental improvement," *arXiv*, 2018.

[12] Christoph Posch, Teresa Serrano-Gotarredona, Bernabe Linares-Barranco, and Tobi Delbruck, "Retinomorphic event-based vision sensors: bioinspired cameras with spiking output," *Proceedings of the IEEE*, 2014.

[13] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck, "A 128×128 120 db 15μs latency asynchronous temporal contrast vision sensor," *IEEE journal of solid-state circuits*, 2008.

[14] Christoph Posch, Daniel Matolin, and Rainer Wohlgenannt, "A qvga 143 db dynamic range frame-free pwm image sensor with lossless pixel-level video compression and time-domain cds," *IEEE Journal of Solid-State Circuits*, 2011.

[15] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck, "A 240× 180 130 db 3 μs latency global shutter spatiotemporal vision sensor," *IEEE Journal of Solid-State Circuits*, 2014.

[16] Henri Rebecq, Guillermo Gallego, Elias Mueggler, and Davide Scaramuzza, "Emvs: Event-based multi-view stereo—3d reconstruction with an event camera in real-time," *International Journal of Computer Vision*, 2017.

[17] Luis A Camuñas-Mesa, Teresa Serrano-Gotarredona, Sio-Hoi Ieng, Ryad Benosman, and Bernabé Linares-Barranco, "Event-driven stereo visual tracking algorithm to solve object occlusion," *IEEE transactions on neural networks and learning systems*, 2018.

[18] Ana I Maqueda, Antonio Loquercio, Guillermo Gallego, Narciso Garcıa, and Davide Scaramuzza, "Event-based vision meets deep learning on steering prediction for self-driving cars," in *CVPR*, 2018.

[19] Jia Li, Feng Shi, Weiheng Liu, Dongqing Zou, Qiang Wang, Hyunku Lee, Paul-K.J Park, and Hyunsurk Eric Ryu, "Adaptive temporal pooling for object detection using dynamic vision sensor," in *BMVC*, 2017.

[20] Anton Mitrokhin, Cornelia Fermuller, Chethan Parameshwara, and Yiannis Aloimonos, "Event-based moving object detection and tracking," *IROS*, 2018.

[21] Nicholas FY Chen, "Pseudo-labels for supervised learning on dynamic vision sensor data, applied to object detection under ego-motion," in *CVPRW*, 2018.

[22] Christian Brandli, Lorenz Muller, and Tobi Delbruck, "Real-time, high-speed video decompression using a frame-and event-based davis sensor," in *ISCAS*, 2014.

[23] Hongjie Liu, Diederik Paul Moeys, Gautham Das, Daniel Neil, Shih-Chii Liu, and Tobi Delbrück, "Combined frame-and event-based detection and tracking," in *ISCAS*, 2016.

[24] Hyungtae Lee, Heesung Kwon, Ryan M Robinson, William D Nothwang, and Amar M Marathe, "Dynamic belief fusion for object detection," in *WACV*, 2016.

[25] Jonathan Binas, Daniel Neil, Shih-Chii Liu, and Tobi Delbruck, "DDD17: End-to-end davis driving dataset," in *ICML*, 2017.

[26] Il Memming Park, Sohan Seth, Antonio RC Paiva, Lin Li, and Jose C Principe, "Kernel methods on spike train space for neuroscience: a tutorial," *IEEE Signal Processing Magazine*, 2013.

[27] Chankyu Lee, Gopalakrishnan Srinivasan, Priyadarshini Panda, and Kaushik Roy, "Deep spiking convolutional neural network trained with unsupervised spike timing dependent plasticity," *IEEE Transactions on Cognitive and Developmental Systems*, 2018.