2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

Leveraging Spatio-Temporal Evidence and Independent Vision Channel to Improve Multi-Sensor Fusion for Vehicle Environmental Perception

Juwang Shi, Wenxiu Wang, Xiao Wang, Hongbin Sun*, Xuguang Lan, Jingmin Xin and Nanning Zheng

Abstract— **For intelligent vehicles, multi-sensor fusion is of great importance to perceive traffic environment with high accuracy and robustness. In this paper, we propose two effective methods, i.e. spatio-temporal evidence generating and independent vision channel, to improve multi-sensor track-level fusion for vehicle environmental perception. The spatio-temporal evidence includes instantaneous evidence, tracking evidence and tracks matching evidence to improve existence fusion. Independent vision channel leverages the specific advantage of vision processing on object recognition to improve classification fusion. The proposed methods are evaluated by using the multi-sensor dataset collected from real traffic environment. Experimental results demonstrate that the proposed methods can significantly improve the multi-sensor track-level fusion in terms of both detection accuracy and classification accuracy.**

I. Introduction

Robust object perception in traffic environment is significant for both autonomous driving and advanced driver assistant system (ADAS). Object perception can be achieved by different kinds of sensors, such as lidar, radar, and camera. Lidar and radar are active sensors, which can measure accurate distance of objects, but have poor capability to recognize the classification of objects. On the contrary, cameras show excellent performance in object recognition but have poor capability to measure the distance of objects. Radar is more sensitive to measure the velocity of objects, and lidar is more suitable for perceiving the shape of objects. In summary, each kind of these sensors have its advantages and disadvantages [1]. Thus, it is promising to fuse different sensors for high accuracy and robustness of object perception.

Sensor fusion systems have been well studied over decades. In general, sensor fusion methods can be divided into three categories: low-level, feature-level and high-level fusion methods. Low-level fusion architecture is with no pre-processing of raw data taking place at the sensor-level. Although low-level fusion [2] has a strong description capability of objects, it requires high data bandwidth and can be complex to implement in practice. Feature-level fusion [3] attempts to extract certain features from raw data through a pre-processing step, before carrying out the data fusion. Feature-level fusion can reduce complexity but still hard to implement in practice. In high-level fusion [4], each sensor independently carries out a tracking algorithm and generates an object list. High-level fusion can produce the best perception performance because of its modularity, practicality and scalability. Hence in this paper, we focus on high-level fusion, which is also called as track-level fusion. One potential drawback of track-level fusion is its poor description capability of objects. Therefore, in track-level fusion designs, we should pay more attention to the complete description of objects, such as classification, shape and so forth.

Many track-level fusion methods [5]–[7] have been proposed and made great contributions to object perception. While tracking an object, we mainly focus on three kinds of information: 1) the existence probability of an object; 2) the accuracy of the state of an object, including position, velocity, and orientation, which is reflected on its global track; 3) the completeness of its other description information, such as classification, shape and so forth. Therefore, track-level fusion mainly includes existence fusion, track-to-track fusion and classification fusion. Reference [5] and [6] employ an existence fusion method based on Dempster-Shafer evidential theory (DST) to estimate the existence probability of an object. Nevertheless, their existence evidence only takes into account instantaneous spatial evidence, while ignoring the temporal evidence, hence it produces relatively high false positive rate and false negative rate of detection, especially under highly dynamic traffic environment. Reference [5] and [7] proposed track-to-track fusion methods to estimate the state of an object using information matrix fusion (IMF). However, they fuse the image objects the same as other active sensors, while ignoring the inaccurate position of the image objects, which result in relatively high false negative rate, false positive of detection and relatively low correct classification rate of recognition.

To address the abovementioned two problems, we propose two methods, i.e. spatio-temporal evidence generating (STEG) and independent vision channel (IVC), to improve multi-sensor track-level fusion. STEG method increases the accuracy of existence estimation, and hence improves the detection accuracy of track-level fusion. IVC method not only improve the detection rate of the track-level fusion, but also increases the correct classification rate of the track-level fusion. The proposed methods are evaluated using the multi-sensor dataset collected from real traffic environment. Experimental results clearly demonstrate the effectiveness of the proposed methods on track level fusion. The proposed STEG method reduces the false negative rate of object detection by 0.06 in the case of similar false positive rate compared with the method without STEG, and reduces the

This work was supported partly by National Key R&D Program of China (No. 2017YFC0803907), National Natural Science Foundation of China (No. 61790563, 91748208), and Joint Foundation of Ministry of Education of China (No. 6141A02033303)

The authors are with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049 P.R. China

*Corresponding author: Hongbin Sun, hsun@mail.xjtu.edu.cn

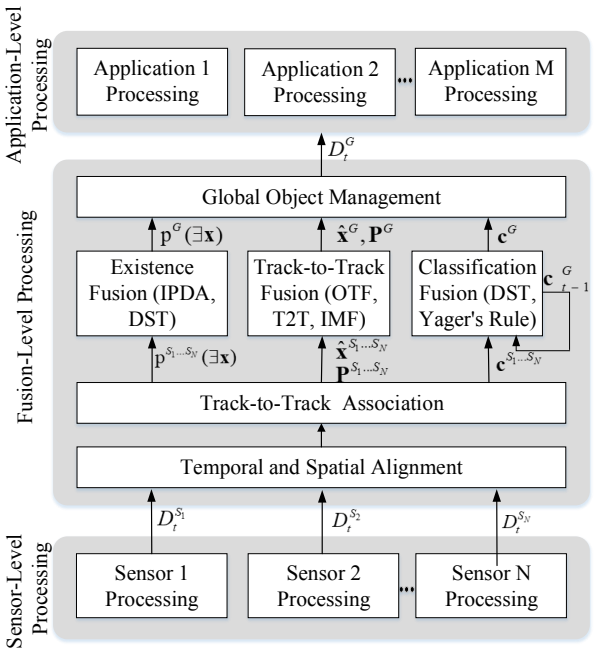


Fig. 1: Framework of state-of-the-art track-level fusion

false positive rate by 0.08 in the case of similar false negative rate compared with the method without STEG. Meanwhile, the proposed IVC method reduces the false negative rate of object detection by 0.01, and increases the correct classification rate of object recognition by 0.19.

II. Background of Track-Level Fusion

In order to correctly detect objects in outdoor environments and mostly not to lose object information by entering and exiting the field of view, the multi-sensor data should be fused at track-level. As shown in Fig. 1, the framework of the track-level fusion consists of sensor-level processing, fusion-level processing, and application-level processing. In sensor-level processing, we get sensor local object lists from each sensor independently

$$D_t^{S^i} = \{O_1, O_2, ..., O_N\} \tag{1}$$

where t is the timestamp, S_i represents the sensor i, and O_j is the description of detected object, consisting of the state $\hat{\mathbf{x}}$, state covariance $\hat{\mathbf{P}}$, the existence probability p, and the classification $\mathbf{c}$. In fusion-level processing, the object lists from different sensors firstly are aligned spatially and temporally to a common coordinate system. After track-to-track association, the sensor-level object lists $D_t^{S^1 \cdots S^N}$ are fused together to form global object lists D_t^G by existence fusion, track-to-track fusion, and classification fusion. In application-level processing, the global object lists D_t^G are in conjunction with other data sources for specific applications.

Existence fusion is to fuse object existence probability from independent sensor-level estimations to generate the global object existence probability estimation. It is very important to the robustness improvement of object tracking. For single sensor object tracking, existence probability is mostly estimated by normalized innovation squared (NIS). Recently several more advanced methods are also proposed. For example, [8] presents an approach using several cues from stereo vision and the tracking process to estimate the existence probability of objects. For multi-sensor object tracking, object existence probability estimation was developed in the integrated probabilistic data association (IPDA) framework as a quality measure of detected objects [9]. Reference [6] proposes a DST based fusion method for object existence probability estimation, in which object existence estimated by each sensor is combined.

Track-to-track fusion fuse object state and its covariance from independent sensor-level estimations to generate the global object state estimation. Reference [10] presents object track fusion (OTF) method, in which sensor-level tracks are seen as measurements to global object tracks, ignoring the correlation and information redundancy. Reference [11] introduces track to track fusion (T2T) method calculating the cross correlation by approximation techniques, which are subject to exibility drawbacks. IMF is used in [5] to fuse multi-sensor respective object tracks into global tracks and showed an excellent performance. IMF based approach is presented in [7], which uses IMF to handle temporal correlation, achieved centralized architecture comparable performance. The accuracy comparison among existed track-to-track fusion methods for object state estimation, has been done in [12]. It shows that IMF is the most robust to process noise and is most accurate and consistent during process model deviation.

Classification fusion aims to improve classification estimation of global objects. Most of classification fusion methods are based on evidential theory, which is a useful tool to make decision based on incomplete and uncertain information. For global object classification fusion, it is critical to find appropriate evidence. Reference [13]–[15] use DST based method to estimate the classification of the global object track. In [13], the proposed fusion method based on DST relies on two main pieces of evidence: the instantaneous fusion evidence, obtained from the combination of evidence provided by individual sensor per object at current time; and the dynamic fusion evidence, which combines evidence from previous result with the instantaneous fusion result. In [14], the authors presented an evidential theory based fusion of object grid maps to decide a grid occupied or not. A DST based classification approach [15] is proposed to fuse instantaneous classification result and the previous classification result of global tracks. Reference [16] and [17] use Yager's rule to combine an object classification evidence from different sensors, is further used to associate objects from different sensors.

III. Proposed Track-Level Fusion

A. Overall Framework

Fig. 2 illustrates the framework of the proposed track-level fusion. The overall framework is very similar to

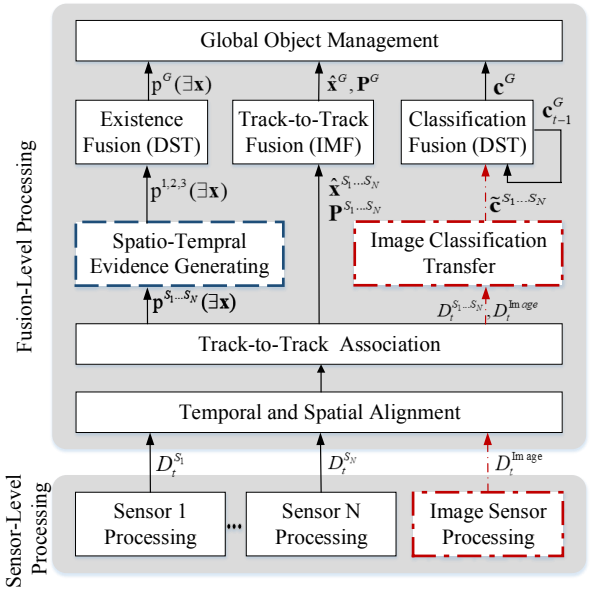


Fig. 2: Framework of proposed track-level fusion

the framework of the state-of-the-art track-level fusion. In particular, for temporal and spatial alignment, we employ the same method introduced in our prior work [18] to synchronize and calibrate the independent sensor-level object lists to a common coordinate system. Then we use Hungarian algorithm to associate the temporal and spatial aligned object lists from multiple sensors. We employ DST for existence fusion and classification fusion, and IMF for track-to-track fusion. This is mainly because DST and IMF are the most effective methods respectively.

The major differences between the proposed and the state-of-the-art frameworks can be described as follows.

- We propose a spatio-temporal evidence generating method for existence fusion based on DST. We simultaneously consider the spatial instantaneous evidence, the temporal tracking evidence, and tracks matching evidence to improve the existence fusion.
- We employ an independent vision channel to fuse the image object information to the global tracks. The use of IVC can not only improve the classification fusion, but also avoid the negative influence on track-to-track fusion.

B. Spatio-Tempral Evidence Generating

1) Instantaneous Evidence Generating: When an object appears in the range of perception, we need evidence to instantaneously support the existence fusion. The robustness of single sensor existence estimation is actually very low. Therefore, the direct use of single sensor existence estimation [6] for existence fusion is is inappropriate. To improve the robustness of object detection, we use the matching information of instantaneous measurements as evidence to support the object existence estimation. We use Hungarian algorithm to obtain the instantaneous matching information of different object lists from different sensors, in which we calculate Euclidean distance between objects from different lists as weight matrix.

We define the critical distance as d^1_{max}, and define the distance of two objects from independent sensors as $d_{i,j}$. If $d_{i,j}$ is less than d^1_{max} and object i is matched with object j, we define instantaneous evidence mass which uses Sigmod function to map $d_{i,j}$ to the mass set $[a, b]$. The Sigmod function is defined as

$$\begin{cases} m^1(A) = & \dfrac{1}{1+\mathrm{e}^{-\hat{d}_{i,j}}} \\ \hat{d}_{i,j} = & \dfrac{d^1_{\max}}{\varepsilon + d_{i,j}} \end{cases} \tag{2}$$

where ε is a positive constant closed to zero.

2) Tracking Evidence Generating: When an object is detected only by one sensor during a period, we cannot obtain evidence from multiple sensors. As the instantaneous measurements of objects from single sensor is not robust, we take into account object tracking history as evidence for existence fusion.

We calculate the average Euclidean distance between the objects of current frame and previous frame on the same track during a period to obtain the mass to support the existence estimation. The mass of the evidence is defined as

$$\begin{cases} m^2(A) = & \dfrac{1}{1+\mathrm{e}^{-\hat{d}_{avg}}} \\ \hat{d}_{avg} = & \dfrac{d^2_{\max}}{\varepsilon + d_{avg}} \end{cases} \tag{3}$$

where ε is a constant as mentioned in subsection B1, and d^2_{max} is the critical distance that is associated with the velocity of an object. d_{avg} is estimated by the length and smoothness of the track, which is defined as

$$d_{avg} = \frac{\sqrt{\sum_{i=k-n}^{k}(x_{i-1}-x_i)^2 + (y_{i-1}-y_i)^2}}{n-1} \tag{4}$$

where k is the current timestamp, and n-1 is the selected length of the track.

3) Tracks Matching Evidence Generating: If an object is simultaneously observed by several sensors during a period in their common field of view, it means that the object almost certainly exists. This is strong evidence for object existence fusion.

In this paper we calculate the average Euclidean distance of the two object tracks from different sensors to obtain the mass to support existence estimation. The mass of this strong evidence is defined as

$$\begin{cases} m^3(A) = & \dfrac{1}{1+\mathrm{e}^{-\hat{d}_{t2t}}} \\ \hat{d}_{t2t} = & \dfrac{d^3_{\max}}{\varepsilon + d_{t2t}} \end{cases} \tag{5}$$

where ε is a positive constant closed to zero, and d_{t2t} is the average distance between two tracks. d_{t2t} is defined as

$$d_{t2t} = \frac{\sqrt{\sum_{ki=k-n}^{k} (x_{ki}^{i} - x_{ki}^{j})^2 + (y_{ki}^{i} - y_{ki}^{j})^2}}{n} \quad (6)$$

where ki is the timestamp, and i, j represents different sensors, and n is track length.

4) DST based Existence Fusion : For object existence probability fusion, we obtain the abovementioned three kinds of evidence, namely instantaneous evidence, tracking evidence and tracks matching evidence.

The framework of discernment of existence probability is defined as

$$\Omega = \{\exists, \nexists\} \quad (7)$$

where $\exists$ represents existence. Actually, we calculate the mass $m(\exists)$, and the mass $m(\Theta)$, where Θ is the unknown proposition. In [6], if an object is in the range of a sensor and the sensor fails to detect the object, they define a mass for $m(\exists)$. This is inappropriate because of the occlusion issue and the unreliability of the sensor. Therefore, assuming A represent existence or unknown proposition, we have three mass values to support the proposition A, $m_1(A)$, $m_2(A)$ and $m_3(A)$ as calculated above. We calculate the fusion existence probability using the combination and discriminant rule presented in [6], which is defined as

$$m^i\{\Theta\} = 1 - m^i\{\exists\} \quad (8)$$

$$m(A) = \frac{\sum_{B\cap C\cap D=A} m^1(B)m^2(C)m^3(D)}{1-K} \quad (9)$$

where K is defined as

$$K = \sum_{B\cap C\cap D=\emptyset} m^1(B)m^2(C)m^3(D) \quad (10)$$

C. Indenpendent Vision Channel

Among all sensors, the recognition capability of camera sensors is the most excellent. Therefore image information must be fused for comprehensive object perception. However, the real-wrold position of object detected by image sensor is not accurate as other active sensors, because camera calibration parameters cannot fit all conditions of the outdoor traffic environments. For example, the camera mounted on vehicle may shake with the vehicle moving on uneven roads. If IMF method uses the image object lists the same as other sensors, it will result in relatively high false negative and false positive rate of object detection, and consequently bring relatively low correct classification rate of object recognition. Reference [19] points out the inaccurate position of image object and hence proposes a method to fuse image information. Nevertheless, the method demands other sensors must obtain geometrical information to match the shape of image objects, which is not generally available for sensors without capability to perceive shape information.

To address the abovementioned problem, we propose an independent vision channel method to fuse image object information more appropriately. The proposed independent vision channel processes the image object lists independently, avoiding the negative influence on track-to-track fusion. In addition, the independent vision channel transfers the image information to objects of active sensors, hence can improve the classification estimation of the track-level fusion as well. We firstly match the tracks of image object with the tracks of other sensors tracks using the tracks matching algorithm mentioned in subsection A3. When d_{t2t} of the current timestamp or the previous timestamp is less than d^3max, we transfer the image object information such as classification to objects of another active sensor. After that, we firstly fuse the transfered classification $\tilde{c}^i$ and the self-contained classification c^{S_j} to be $\tilde{c}^{S_j}$ using DST as equation (12). Then, we fuse the current classification and the previous classification using DST as equation (13).

$$m_k^{\tilde{c}^{S_j}}(C) = \frac{\sum_{A\cap B=C} m_k^{\tilde{c}^i}(A)m_k^{c^{S_j}}(B)}{1 - \sum_{A\cap B=\emptyset} m_k^{\tilde{c}^i}(A)m_k^{c^{S_j}}(B)} \quad (11)$$

$$m_k^{G}(C) = \frac{\sum_{A\cap B=C} m_k^{\tilde{c}^{S_j}}(A)m_{k-1}^{G}(B)}{1 - \sum_{A\cap B=\emptyset} m_k^{\tilde{c}^{S_j}}(A)m_{k-1}^{G}(B)} \quad (12)$$

IV. Experimental Results

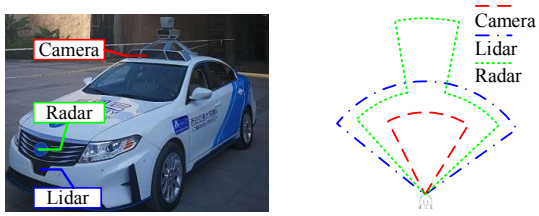
A. Experimental Setup

All the object perception experiments are conducted on "Discovery" autonomous vehicle research platform developed by Xian Jiaotong University (XJTU). The platform is designed to meet the requirements of general autonomous driving research, while great efforts have been made to address the challenge of environmental perception."Discovery" won the China Intelligent Vehicle Future Challenge (IVFC) in 2017. "Discovery" is mounted with one Delphi ESR MMW radar, one monocular PointGrey camera and one ibeo LUX-8L as shown in Fig. 3 (a). Fig. 3 (b) shows specific perception ranges of the three sensors. Table I lists the types, field of views (FOVs), ranges, and update rates of these three sensors.

TABLE I: Sensor Specifications

Sensor	Camera	Lidar	Radar
Type	PointGray GS3	Ibeo LUX-8L	Delphi ESR
FOV	$53°$	$110°$	$90°, 20°$
Range	$40m$	$55m$	$175m$
Update Rate	$30fps$	$6.25fps$	$20fps$

The dataset used in the experiment is collected by multiple sensors mounted on "Discovery" from urban roads in Xi'an City. The capturing rates of camera and radar are $10fps$, and that of lidar is $6.25fps$. We have totally captured a synchronized dataset of 45287 frames simultaneously including camera, lidar, and radar data. From the dataset, we select 4 sessions to test the proposed methods as shown in Table



(a) Platform (b) Perception ranges

Fig. 3: Sensor configurations of XJTU autonomous vehicle research platform "Discovery"

II. We have pre-processed the 4 selected sessions. MMW radar objects are read from the radar sensor directly, image objects are detected by single shot multibox detector (SSD) model [20], and lidar data is processed by density based spatial clustering of applications with noise (DBSCAN) [21]. The objects of radar and lidar are tracking by Kalman Filter respectively.

TABLE II: Selected Sessions from Dataset

Datasets	Seconds	Objects	Vehicles	Pedestrians	Bicycles
Session 1	100	1899	1510	389	0
Session 2	100	3869	2751	168	950
Session 3	100	2980	2980	0	0
Session 4	140	2856	2856	0	0

B. Results of Spatio-Temporal Evidence Generating

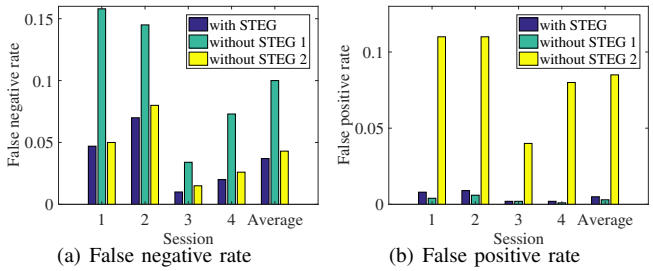
We evaluate the proposed spatio-temporal evidence generating method by comparing with the previous method presented in [6] in terms of the false negative rate and false positive rate of object detection. If an object exists, we fuse it to the global track. Therefore, to evaluate existence fusion performance, we calculate the global false negative rate and false positive rate of object detection respectively.

In our work, we select one set of parameters for the STEG method, $d^1_{max} = d^2_{max} = d^3_{max} = 2.2m, \varepsilon = 0.0001$. The method without STEG can adjust false negative rate and false positive rate of object detection by changing its parameters. Therefore, we select two sets of parameters of the method, i.e. without STEG 1, which is with similar false positive rate compared with the STEG method, and without STEG 2, which is with similar false negative rate compared with the STEG method. Table III lists the two sets of parameters of the previous methods. R^{lidar}_{trust} is the perception range of lidar sensor. In the perception range, if radar sensor detects an object and the lidar sensor fails, the result is the inexistence of the object.

As shown in Fig. 4, the proposed STEG method reduces the false negative rate of object detection by 0.06 in the case of similar false positive rate compared with the method without STEG, and reduces the false positive rate by 0.08 in the case of similar false negative rate compared with the method without STEG.

TABLE III: Previous Parameters Sets

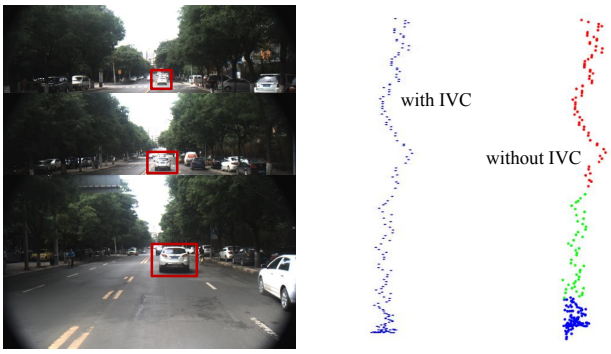
Experiments	p^{lidar}_{trust}	p^{radar}_{trust}	R^{lidar}_{trust}
without STEG 1	0.99	0.9	$50m$
without STEG 2	0.99	0.9	$20m$



(a) False negative rate (b) False positive rate

Fig. 4: STEG detection accuracy comparisons

C. Results of Independent Vision Channel

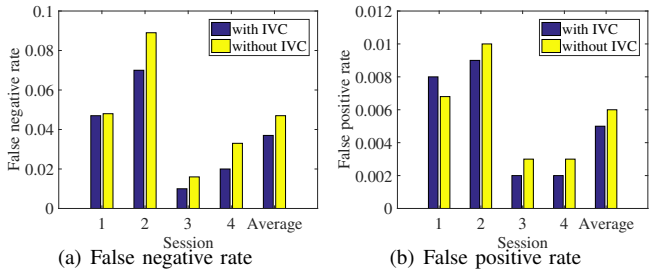


(a) Traffic scene (b) Tracks of with and without IVC

Fig. 5: An example of object ID change comparison

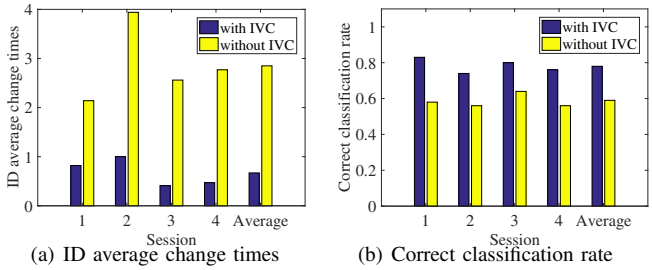
The larger false negative rate of detection results in more identification number (ID) change of the global object track. IVC method can reduce false negative rate and ID change times . Fig. 5 (a) shows a traffic scene, in which a frontal car (marked by the red bounding box) is moving. Fig. 5 (b) illustrates the ID change comparison of the car tracks provided by the method with IVC and the method without IVC respectively, in which each kind of color represents one ID. The ID of the track does not change during 20 seconds provided by the proposed IVC method, but it changes two times provided by the method without IVC. Fig. 6 illustrates the comparisons of the false negative rate and false positive rate of object detection. The comparisons demostrate that the method with IVC can reduce the false negative rate by 0.01 compared with the method without IVC. Although the improvement of the false negative rate of object detection is relatively little, it can reduce the ID average change times and hence improves the clssification fusion effectively.

In this paper, we use the DST based classification fusion method, in which the classification of an object is decided



(a) False negative rate (b) False positive rate

Fig. 6: IVC detection accuracy comparison



(a) ID average change times (b) Correct classification rate

Fig. 7: IVC classification accuracy comparison

by current sensor evidence and previous global classification evidence. Therefore, the previous classification evidence is lost when an object ID changes, which results in low correct classification rate. Fig. 7 illustrates the average change times of object IDs and correct classification rate. It shows that the average change times of the method with IVC is 0.67, which is more than 3 times lower than that of the method without IVC. The correct classification rate of the method with IVC is 0.78, which is 0.19 higher than that of the method without IVC.

V. Conclusion

In this paper, we propose two methods to improve the track-level fusion for object perception. Firstly, we propose a spatio-temporal evidence generating method for object existence probability fusion to lower false negative rate and false positive rate of detection. Secondly, we propose an independent vision channel method to improve the track-to-track fusion and classification fusion. Finally, the proposed methods are evaluated by the multi-sensor dataset collected from real traffic environment. Experimental results demonstrate that the proposed methods can significantly improve the multi-sensor track-level fusion in terms of both detection accuracy and classification accuracy. In the future work, we will consider to use road scene understanding from camera to help object fusion perception.

References

[1] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1773–1795, 2013.

[2] S. Pietzsch, T. D. Vu, O. Aycard, T. Hackbarth, N. Appenrodt, J. Dickmann, and B. Radig, "Results of a precrash application based on laser scanner and short-range radars," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 584–593, 2009.

[3] N. Kaempchen, M. Buhler, and K. Dietmayer, "Feature-level fusion for free-form object tracking using laserscanner and video," in *IEEE Intelligent Vehicles Symposium*, 2005, pp. 453–458.

[4] M. Aeberhard and N. Kaempchen, "High-level sensor data fusion architecture for vehicle surround environment perception," in *International Workshop on Intelligent Transportation*, 2011.

[5] M. Aeberhard, S. Schlichtharle, N. Kaempchen, and T. Bertram, "Track-to-track fusion with asynchronous sensors using information matrix fusion for surround environment perception," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1717–1726, 2012.

[6] M. Aeberhard, S. Paul, N. Kaempchen, and T. Bertram, "Object existence probability fusion using dempster-shafer theory in a high-level sensor data fusion architecture," in *IEEE Intelligent Vehicles Symposium*, 2011, pp. 770–775.

[7] H. Li, F. Nashashibi, B. Lefaudeux, and E. Pollard, "Track-to-track fusion using split covariance intersection filter-information matrix filter (SCIF-IMF) for vehicle surrounding environment perception," in *IEEE International Conference on Intelligent Transportation Systems*, 2013, pp. 1430–1435.

[8] S. Gehrig, A. Barth, N. Schneider, and J. Siegemund, "A multi-cue approach for stereo-based object confidence estimation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 3055–3060.

[9] D. Musicki, R. Evans, and S. Stankovic, "Integrated probabilistic data association," *IEEE Transactions on Automatic Control*, vol. 39, no. 6, pp. 1237–1241, 1994.

[10] Y. Bar-Shalom and X. Li, *Multitarget-multisensor tracking: principles and techniques*. Storrs, CT: YBS Publishing, 1995.

[11] H. Chen, T. Kirubarajan, and Y. Bar-Shalom, "Performance limits of track-to-track fusion versus centralized estimation: theory and applica- tion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 2, pp. 386–400, 2003.

[12] S. Nilsson and A. Klekamp, "A comparison of architectures for track fusion," in *IEEE International Conference on Intelligent Transportation Systems*, 2015, pp. 517–522.

[13] T. D. Vu, O. Aycard, and F. Tango, "Object perception for intelligent vehicle applications: A multi-sensor fusion approach," in *IEEE Intelligent Vehicles Symposium*, 2014, pp. 774–780.

[14] D. Nuss, M. Thom, A. Danzer, and K. Dietmayer, "Fusion of laser and monocular camera data in object grid maps for vehicle environment perception," in *IEEE International Conference on Information Fusion*, 2014, pp. 1–8.

[15] M. Aeberhard and T. Bertram, "Object classification in a high-level sensor data fusion architecture for advanced driver assistance systems," in *IEEE International Conference on Intelligent Transportation Systems*, 2015, pp. 416–422.

[16] R. O. Chavez-Garcia, T. D. Vu, and O. Aycard, "Fusion at detection level for frontal object perception," in *IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 1225–1230.

[17] R. O. Chavez-Garcia and O. Aycard, "Multiple sensor fusion and classification for moving object detection and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 525–534, 2016.

[18] X. Wang, L. Xu, H. Sun, J. Xin, and N. Zheng, "On-road vehicle detection and tracking using mmw radar and monovision fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 2075–2084, 2016.

[19] H. Cho, Y. W. Seo, B. V. K. V. Kumar, and R. R. Rajkumar, "A multi-sensor fusion system for moving object detection and tracking in urban driving environments," in *IEEE International Conference on Robotics and Automation*, 2014, pp. 1836–1843.

[20] L. Wei, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: single shot multibox detector," in *European Conference on Computer Vision*, 2016, pp. 21–37.

[21] X. Zhang, W. Xu, C. Dong, and J. M. Dolan, "Efcient L-shape tting for vehicle detection using laser scanners," in *IEEE Intelligent Vehicles Symposium*, 2017, pp. 54–59.