Moving Object Detection and Tracking Method Based on Composite Dynamic Models and Evidence Fusion Architecture: Postprint

Authors: Cheng Wei, Wu Haibin, Zheng Hongqing

Date: 2018-08-13T00:00:00+00:00

Abstract

In the environmental perception function of Advanced Driver Assistance Systems (ADAS), the accuracy of moving object detection and tracking is of paramount importance. To address the limitation of low precision in existing methods, this paper proposes a moving object description and perception method based on multi-sensor detection and classification: a composite model incorporating core object dynamic features and classification descriptions is established, upon which an information perception and fusion method based on an evidence framework is designed to achieve moving object detection and tracking by integrating dynamic models and uncertainty characteristics. To validate the effectiveness of the proposed method, relevant experiments were conducted on a demonstration vehicle equipped with radar, LiDAR, and cameras, detecting and tracking three types of moving objects—pedestrians, trucks, and cars—under various driving scenarios. The experimental results demonstrate that the proposed method achieves very high accuracy.

Full Text

Preamble

Moving Object Detection and Tracking Based on Composite Dynamic Model and Evidential Fusion Framework

Cheng Wei[1], Wu Haibin[2], Zheng Hongqing[1]
(1. School of Electronic and Electrical Engineering, Minnan Polytechnic Institute, Shishi, Fujian 362700, China;
2. Institute of Machinery & Automation, Fuzhou University, Fuzhou 350116, China)

Abstract: Accurate detection and tracking of moving objects is a critical aspect of environmental perception in advanced driver assistance systems. To overcome the low precision limitations of existing methods, this paper proposes a novel description and perception approach based on multi-sensor detection and classification. A composite model containing core object dynamic features and classification descriptions is established, upon which an evidential framework-based information perception and fusion method is designed to achieve moving object detection and tracking by integrating dynamic models and uncertainty characteristics. To validate the proposed method, experiments were conducted on a demonstration vehicle equipped with radar, lidar, and camera. Detection and tracking tests were performed on three types of moving objects—pedestrians, trucks, and cars—under various driving scenarios. Experimental results demonstrate that the proposed method achieves very high accuracy.

Keywords: moving object detection and tracking; multi-sensor system; classification algorithms; composite dynamic model; evidential fusion framework

Introduction

With continuous advances in intelligent control, information communication, and industrial manufacturing technologies, intelligent vehicles have attracted extensive research and commercial development. The most compelling feature of intelligent vehicle systems is their ability to operate reliably in uncertain, dynamic, and unstructured environments. Advanced driver assistance systems in intelligent vehicles play a crucial role in helping drivers perform complex driving tasks to avoid hazardous situations. Specifically, through environmental perception methods, these systems issue warning messages before collisions or scrapes occur, trigger safety devices, automatically brake to avoid obstacles, and alert drivers to maintain focus to prevent imminent collisions.

The mechanism of environmental perception involves obtaining detailed descriptions of the external environment and accurate identification of objects of interest through rational sensor selection. Vehicle environmental perception primarily includes two tasks: simultaneous localization and mapping (SLM) and moving object detection and tracking (MODT). SLM generates real-time maps of the vehicle and its surroundings based on given sensor parameters, while MODT detects and tracks moving objects around the vehicle and predicts their future motion.

Managing and perceiving incomplete information represents an important functional requirement in environmental perception systems. Such incomplete information may arise from sensor calibration issues, hardware failures, uncertain detections, asynchronous scanning, or environmental disturbances like occlusion, weather conditions, and object motion. In most practical outdoor scenarios, it is necessary not only to detect non-moving objects in noisy environments but also to track moving objects and perform data association. Correct detection of moving objects is a critical aspect of motion target tracking systems and often requires assistance from numerous sensors.

Classifying moving objects in the surrounding environment enables intelligent vehicles to achieve better understanding of current driving conditions. Object classification serves as either an independent task within MODT or as part of the output information set of environmental perception. Classification enriches detection results by incorporating environmental information from different sensors, including lidar point clouds and image patches from cameras. Moving object classification can distinguish and confirm targets and provide clues for data association. Furthermore, classification of moving targets helps understand motion models for tracking purposes. In the early recognition stage, classifying objects of interest through multi-sensor data fusion can improve detection and tracking performance while reducing false positive rates and misclassification rates.

Perception systems include different fusion levels. Low-level fusion occurs in SLM, while detection-level and tracking-level fusion take place in MODT. During the detection stage, fusion is implemented based on moving target lists provided by each sensor. During the tracking stage, tracking lists from individual sensor modules are used to generate a final tracking list. SLM has been extensively studied in the literature with novel and practical results. Therefore, this paper focuses on MODT, investigating evidence from different class sets based on fused data from three sensor devices: lidar, camera, and radar.

Multi-sensor tracking fusion requires obtaining tracking lists from a series of sensors and fusing them into a combined list. Previous work has addressed this by focusing on relationships in different tracking fissions and randomly combining related targets. During the tracking stage, effective fusion strategies can significantly reduce erroneous tracking, improve tracking accuracy, and supplement classification information for final output. The detection stage aims to collect and combine initial data from sensors. Some researchers recommend performing fusion at this level to reduce error detection that may lead to tracking failure. Others focus on resolving data redundancy from active and passive sensors and improving target detection accuracy according to physical conditions or constraints. However, these studies do not fully utilize all available motion state and appearance information. Moreover, appearance information from sensor measurements is considered less important than motion state information for distinguishing moving objects from static ones.

When classification is treated as an independent module within the perception scheme, it is typically used to implement single-class classification (e.g., pedestrians only) or single-sensor-based classification processes. This approach can exclude data differences among sensors. Research shows that if classification information is effectively managed in the early perception stage to enhance data association, object tracking accuracy can be directly improved. Among multi-sensor fusion methods, probability-based approaches are most common. However, evidential framework-based methods can not only fuse multi-sensor data but also perceive vehicles. Compared with probability-based methods, evidential framework methods fully utilize incomplete and inaccurate information, thereby improving detection accuracy.

Building upon previous research, this paper proposes a novel method addressing sensor data association and target detection and tracking based on sensor fusion. First, a composite model containing core object dynamic features and classification descriptions is established. Based on this model, an evidential framework-based information perception and fusion method is designed to achieve moving object detection and tracking by integrating dynamic models and uncertainty characteristics. Additionally, knowledge from different sensors enhances object description accuracy: radar information is used for preliminary moving object detection; lidar data estimates target distance and size; and camera-based classification information generates hypotheses for detected targets. To validate the proposed method, experiments were conducted on a demonstration vehicle, detecting and tracking three moving objects—pedestrians, trucks, and cars—with results proving high accuracy.

1. Introduction to Evidential Framework and SLM

1.1 Overview of Evidential Framework

The evidential framework generalizes the Bayesian framework of subjective probability, deriving confidence levels for related problems based on available evidence. It represents the entire real world using a set of mutually exclusive propositions called the frame of discernment Ω. The confidence function is formulated as $m : 2^{\Omega} \to [0, 1]$. In practical applications, the basic belief function shown in Equation (1) is commonly used:

$$m(\phi) = 0, \quad \sum_{A \subseteq \Omega} m(A) = 1$$

where A represents the dataset to be identified.

In practice, to avoid anomalous results that may significantly conflict with actual situations, the belief function shown in Equation (1) requires regularization as shown in Equation (2), distributing conflict values across all elements of the frame of discernment rather than just those with intersecting combined masses:

$$(m_1 \oplus m_2)(A) = \begin{cases} 0, & A = \phi \\ \frac{\sum_{X_i \cap Y_j = A} m_1(X_i) m_2(Y_j)}{1 - \sum_{X_i \cap Y_j = \phi} m_1(X_i) m_2(Y_j)}, & A \neq \phi \end{cases}$$

The evidential framework can express incomplete evidence and lacks prior probability objectives, enabling the use of implicit information when defining the frame of discernment. Additionally, discount coefficients are important mechanisms for integrating multi-source evidence, allowing coefficients to be set according to sensor performance, while combination rules serve as effective tools for integrating evidence from different sources. However, when the quantity is too large, the computational complexity of the evidential framework increases significantly due to the need to calculate confidence for all hypotheses. Therefore, in practical applications, Ω can be transformed into a streamlined version.

1.2 Overview of SLM

Although this paper focuses on the MODT component, SLM is needed to obtain maps and vehicle poses. Specifically, lidar data ($z_{1:t}$) is used to populate a two-dimensional Bayesian occupancy grid map M, where each cell is associated with the probability of being occluded by obstacles. Vehicle position is obtained using maximum likelihood estimation based on shape models $P(\omega)$ and likelihood models $P(z_{1:t}|\omega)$. Since objects to be detected have different dynamic definitions, interacting multiple models based on constant velocity, constant acceleration, and turn rate models are used to track each target. Data association between targets is obtained through heterogeneous multiple hypothesis tracking methods.

2. Moving Target Detection and Tracking

The first stage of moving target tracking utilizes data from different sensors (including lidar, camera, and radar) to detect interesting moving objects around the intelligent vehicle.

2.1 Target Detection

Lidar scanners based on light detection and ranging serve as the primary sensor for moving target detection due to their high resolution and obstacle detection accuracy. The main detection objective of lidar sensors is to obtain precise measurement data of the shape of moving obstacles ahead of the vehicle. The detection principle based on lidar sensors identifies contradictions between free space and occupied areas in grid map M: if a previously free area is detected as occupied, it should belong to a moving object; if an occupied cell is detected as freed, it likely belongs to a static object. Using distance clustering methods, grid clouds belonging to moving objects can be clearly identified while providing estimates of shape, size, and distance of potential moving targets.

To obtain appearance information of moving objects from camera images, distinctive visual features must be extracted. The Histogram of Oriented Gradients (HOG) method can effectively detect moving vehicles and pedestrians, making it suitable as the core visual descriptor for vehicles and pedestrians. HOG generates visual descriptions of local image information for determining whether these regions contain the moving targets to be detected. To improve detection accuracy, a sparse HOG focusing on specific regions in images is proposed, effectively reducing computational complexity for high-dimensional descriptors. Additionally, to accelerate feature computation, the proposed method follows relevant rules for integrated image processing schemes.

After obtaining regions of interest through radar detection images, visual features are extracted for each region, and classifiers determine whether targets are within the regions of interest. Therefore, classifier selection critically impacts final processing speed and results. A classifier based on discrete boosting learning is selected, which combines multiple weak classifiers to form a strong classifier, where each weak classifier only needs to perform slightly better than random guessing. For each moving target to be detected (including pedestrians, trucks, and cars), a corresponding binary classifier must be trained. During the training phase, information from different perspectives (e.g., front, back, and profile views) of target categories can be obtained from public databases and manually annotated images containing classified targets. Next, the likelihood of each target classification must be estimated. This estimation follows the principle that the more positive regions containing the target and the higher their confidence, the more certain the target belongs to a specific class.

Radar sensors use built-in mechanisms to detect moving obstacles or targets and provide a list of n targets as input to the perception algorithm. Each element in the list includes the distance, bearing, and relative speed of the detected target. Radar sensors generate a target point for each moving object with a specific cross-section. However, it should be noted that target points may correspond to either static objects or other moving obstacles, leading to false detections. For example, weak objects like pedestrians are highly likely to be undetected.

2.2 Target Classification

As analyzed in Section 3.1, fusing classification information during the detection stage enables the proposed scheme to better describe moving target kinematic characteristics than general methods. This approach not only improves detection accuracy and enables precise estimation of moving target motion but also reduces tracking error rates. However, there is insufficient information during the detection stage to determine target classification. Therefore, a composite dynamic model consisting of kinematic information and appearance information must be established for moving objects to be detected. Kinematic information includes two-dimensional spatial position and shape information inferred from moving target detection. Appearance information includes evidence distributions for all possible hypothesis classes, where $\Omega = \{p, b, c\}$ represents the set of categories to be detected, with p, b, and c representing pedestrians, trucks, and cars, respectively.

For larger targets detected by lidar, a rectangular box described by set $\{x, y, w, l\}$ can be used, where x, y are the box center coordinates and w, l represent target width and length. For smaller targets, a point model described by set $\{x, y\}$ can be used, where x, y represent the target center point. Target position and size can be obtained by measuring the number of occupied two-dimensional grids, while target classification is inferred from visual size and subsequent fixed fitting model methods. However, due to short visual duration of moving targets, precise classification cannot be made. For example, if a detected object' s width is smaller than a set threshold ω_{small}, the detection system might mistakenly identify it as a pedestrian or truck but cannot determine the object' s actual size. Therefore, typical sizes of detected classes must be defined based on prior knowledge of physical size distributions. To distinguish from general methods, this paper defines the basic belief assignment $m_1(A)$ for each $A \in \Omega$ as shown in Equation (4), rather than simply deciding that a target belongs to only one class. The basic belief assignment describes the evidence distribution for classifying moving targets detected by lidar. The formula uses classification-related factors α_p, α_b, α_c to describe lidar performance in detecting pedestrians, trucks, and cars. Additionally, discount factors γ_b and γ_c represent uncertainty in lidar detection conclusions for trucks or cars.

$$m_1(A) = \begin{cases} \alpha_p, & A = \{p\} \\ \alpha_b(1-\gamma_b), & A = \{b\} \\ \alpha_c(1-\gamma_c), & A = \{c\} \\ \alpha_b\gamma_b, & A = \{b, c\} \\ \alpha_c\gamma_c, & A = \{b, c\} \\ 1 - \sum_{A \subset \Omega} m_1(A), & A = \Omega \end{cases}$$

Camera images of moving targets are processed to obtain appearance-based evidence distributions of target classes, which are then used with offline classifiers. Based on the possible relationships between c_{am} and c_{bm}, define: 1 means c_{am} and c_{bm} are the same object; 0 means they are different objects; Φ represents an unknown relationship between c_{am} and c_{bm}. Let $\Lambda = \{1, 0, \Phi\}$ be the frame of discernment representing these propositions. $m^{\theta}_{a,b}$ and $m^{\psi}_{a,b}$ quantify evidence supporting propositions θ and ψ, respectively, while $m^{\Lambda}_{a,b}$ represents unknown evidence that cannot support other propositions. These propositions are described by finding similarity metrics between detected targets in c_{am} and c_{bm}.

Sensors S_1 and S_2 can provide different types of information based on position, shape, or appearance. Similarity measures encoding all this detection information can be defined. Camera-based classification requires generating multiple sub-regions in each area to be tested, covering as many target structures of different sizes as possible. After obtaining object classifications for each region, the basic belief assignment m_2 shown in Equation (5) represents the evidence distribution for classification hypotheses of each target Ω detected by the image processing procedure.

$$m_2(A) = \begin{cases} \delta_p, & A = \{p\} \\ \delta_b, & A = \{b\} \\ \delta_c, & A = \{c\} \\ 1 - \delta_p - \delta_b - \delta_c, & A = \Omega \end{cases}$$

where δ_p, δ_b, δ_c are confidence factors and σ is camera accuracy.

Radar target detection provides preliminary moving target detection. To obtain target class, the relative speed detected by radar is used to calculate the basic belief assignment m_3 for the radar sensor based on Equation (6):

$$m_3(A) = \begin{cases} \zeta, & A = \{b\}, v_t < v_s \\ \tau, & A = \{c\}, v_t > v_s \\ 1 - \zeta - \tau, & A = \Omega \end{cases}$$

where v_s is a speed threshold obtained through statistical estimation of the slowest urban vehicle speed, v_t is target speed, and ζ, τ are class-specific confidence factors.

2.3 Target Tracking

After detecting moving targets from each sensor input and defining composite object representations, fused target detection and tracking can be performed. This paper proposes a detection-level multi-sensor fusion framework that defines additional detection modules to obtain more evidence sources.

When using evidence from multiple signal sources, it is necessary to determine which detection target each sensor’ s (evidence source’ s) detection list is associated with. Since information combination at the detection level can reduce adverse effects from sensor measurement errors, classification module uncertainties, incomplete information, and information conflicts, multi-source detection can effectively improve detection result reliability.

1) Position Information Similarity Extraction

Consider two evidence sources S_1 and S_2, each providing a detection list: $A = \{a_1, a_2, ..., a_m\}$ and $B = \{b_1, b_2, ..., b_n\}$, where each element represents association evidence between S_1 and S_2. To combine information from these different sources, relationships between detection data in lists A and B must be found. Based on position information from detections a_i and b_j, a distance d_{a_i,b_j} satisfying pseudo-distance metric properties is defined. Since Mahalanobis distance can represent correlation between distances, it is used as the distance metric.

All propositions belonging to frame of discernment Λ are expressed as $m^1_{a_i,b_j}$, $m^0_{a_i,b_j}$, and $m^{\Phi}_{a_i,b_j}$. $m^1_{a_i,b_j}$ represents evidence supporting the proposition that

a_i and b_j are the same target, while larger values indicate they are different targets.

2) Classification Information Similarity Extraction

Classification does not provide direct evidence supporting the proposition $P_{a_i,b_j} = 1$. Even if two detections are identified as the same category, they cannot be concluded to be the same object, as two different objects in the same driving scenario may belong to the same category. However, if two detections belong to different classes, they likely belong to different objects. Therefore, classification information is used as difference evidence $m^c_{a_i,b_j}$ between detections. The method for transforming evidence from Ω to Λ is:

$$m^c_{a_i,b_j}(A) = \begin{cases} 0, & A = 1 \\ \sum_{c_a \cap c_b = \emptyset} m_{ic_a} m_{jc_b}, & A = 0 \\ 1 - \sum_{c_a \cap c_b = \emptyset} m_{ic_a} m_{jc_b}, & A = \Phi \end{cases}$$

where c_{am} and c_{bm} represent classification hypotheses in detection lists A and B, respectively. Equation (10) shows that substantial information about classification differences between detection lists A and B can improve detection reliability.

Based on the combined target detection list generated by the fusion algorithm, an optimal trajectory estimation method within a sliding time window is proposed based on composite representation:

$$\omega^* = \arg\max_{\omega \in \Omega} P(\omega | z_{1:t}) \propto P(z_{1:t} | \omega) P(\omega)$$

By inputting generated targets into a module considering all object dynamic models and sensor models, and utilizing classification evidence distributions for each target detection, precise retrieval of all possible neighboring hypotheses can be achieved within a search space considering only significant evidence. If two targets have the same category attributes, their classifications are similar.

3. Experimental Analysis and Comparison

To validate the proposed method, datasets were collected from multiple driving scenarios using a demonstration vehicle equipped with radar, lidar, and camera. To enable continuous detection, the vehicle was equipped with processing units, driver interaction components, and forward-facing sensors. The camera captures black-and-white images with a field of view of $15 \pm 5°$; the medium-range radar provides moving target information with a maximum detection range of 100 m, maximum measurable speed of 200 km/h, field of view of $84 \pm 5°$, and angular measurement accuracy of 0.3°; the lidar provides a two-dimensional list of feature points with a maximum measurement range of 160 m, angular accuracy of

0.15°, distance accuracy of 6 cm, and field of view of 90°. [Figure 1: see original paper] shows the multi-sensor perception system used in the experiments. Kinematic and appearance information of moving targets can be obtained from lidar and radar sensors, while appearance information is obtained from the camera.

Three datasets were collected from real-world scenarios using the three configured sensors: two from urban areas and one from highways. To obtain clear references, all datasets were manually annotated. To verify the fusion performance at detection and tracking levels, the proposed method was compared with the method in reference [20].

In the proposed MODT solution, sensor data fusion operates at the detection level. First, the SLM method from Section 1 is applied to lidar measurements to detect potential moving objects. Based on the detected two-dimensional position states, the frame of discernment $\Omega = \{p, b, c\}$ is defined for each evidence class distribution, with 2^{Ω} possible classification hypotheses in each detection. Second, according to the methods proposed in Sections 2 and 3, data from lidar, radar, and camera detections are extracted to obtain representations of moving targets to be detected. Finally, after obtaining target representations, they are fused at the detection level using the method from Section 4, and targets are tracked based on the fused list.

[Figure 2: see original paper] and [Figure 3: see original paper] show two groups of detection results in urban and highway scenarios using the proposed method. Both scenarios feature high traffic flow with numerous moving objects around the vehicle. Subfigure (a) shows camera images with identified moving objects; subfigure (b) shows top-down views where rectangular boxes represent moving objects, points represent lidar detections, and circles represent radar detections.

In both scenarios, all oncoming vehicles can be well detected, tracked, and classified, including several cars and trucks in the highway scenario, and cars, trucks, and pedestrians in the urban scenario. Additionally, static objects like obstacles are quickly detected and correctly identified. Since composite object representation provides speed and direction information, moving object speeds are estimated through model-based tracking modules. In the early fusion stage, when moving direction is known, radar Doppler velocity information can effectively improve lidar’ s accuracy in estimating target speed. Furthermore, because three different sensors provide different classification hypotheses as classification distributions, generating fused information, classification accuracy can be substantially improved.

and compare the proposed detection-level fusion algorithm with the algorithm in reference [20]. Experiments used three datasets: two from highway scenarios and one from urban scenario. Results show that for highway scenarios, detection-level fusion does not significantly improve tracking-level fusion. However, at high speeds, determining moving vehicles is crucial. Therefore, the proposed algorithm is essential for practical applications in continuous support systems. Urban road conditions pose challenges for accurate prediction in modern vehicles, and detection-level fusion shows significant improvement over tracking-level fusion in urban environments. This is because detection-level fusion can detect moving vehicles earlier through rich sensor monitoring and data association.

For pedestrian classification, the proposed method achieves similar performance improvements as vehicle detection. However, small moving targets detected by lidar cannot be identified as pedestrians, a problem that can be solved by early fusion of radar and image data. Additionally, compared with reference [20], the proposed method requires fewer average sensor scans for moving target classification. This is because classification information for detected targets in early knowledge integration is concentrated in c_{am} and c_{bm}, which are directly related to simplified shape and motion model spaces.

To further validate effectiveness, comparative experiments between the proposed method and reference [20] were conducted in three road conditions: urban, suburban, and racetrack. shows the numbers of targets to be detected in the three scenarios. and show experimental results.

Correct classification indicates moving objects correctly classified, while **incorrect classification** indicates misclassified moving objects. For clarity, correct and incorrect classification numbers are expressed as percentages. Three moving target types are considered: pedestrians (p), trucks (b), and cars (c). According to -and related data, the proposed algorithm has an average computation time of 30 ms, fully meeting real-time platform processing requirements (75 ms). Thus, the dynamic response speed fully satisfies real-time processing demands. Moreover, the proposed algorithm' s accuracy is significantly higher than reference [20].

In urban scenarios, despite numerous moving targets and environmental noise interference, vehicle detection and classification accuracy remain high, though some false detection rates persist. This is primarily because sensor fields of view are forced to narrow in highly dynamic, high-traffic environments. Additionally, higher pedestrian misclassification rates occur because the proposed method misidentifies traffic posters as pedestrians. These errors indicate the system needs a more robust visual classification system or more descriptive visual features. Undeniably, to effectively improve algorithm robustness, not only must experimental environments be selected and planned (i.e., the current algorithm has lower misclassification rates in relatively simple environments), but image acquisition system sensitivity and accuracy must also be enhanced (i.e., improving input image accuracy at the algorithm' s input stage).

In suburban scenarios, the proposed method's processing time is less than 20 ms, and classification accuracy is relatively high. This is because suburban scenarios have fewer moving targets and less environmental interference, allowing the algorithm to fully leverage its classification capabilities.

In racetrack scenarios, with only a few pedestrians and vehicles present, pedestrian and vehicle detection and recognition rates are nearly perfect (98%-100%).

This scenario does not include common driving situations such as multiple moving objects and high dynamic traffic flow.

4. Conclusion

For moving target detection and classification in intelligent vehicles, this paper leverages classification information as a key component of composite object mathematical models, enabling full fusion of motion and appearance information during moving target detection, classification, and tracking. The paper further analyzes composite object descriptions generated by detection-level multi-sensor fusion and defines, studies, tests, and evaluates the proposed fusion algorithm based on three primary sensors: lidar, radar, and camera. Finally, the complete perception method was experimentally analyzed on a test vehicle. Results demonstrate that integrating classification information at the detection level enables the fusion process to fully consider evidence distributions of detected targets, substantially improving detection accuracy.

Currently, research on multi-sensor fusion methods is gradually deepening. Through reasonable expansion of the proposed method, a system classification method integrating sensor motion perception can be designed, enabling the device' s computing core to possess learning representation and motion control functions when interacting with the environment.

References

[1] Wu Libing, Fan Jing, Nie Lei, et al. An urban vehicle cooperative routing method under the environment of vehicle interconnection [J]. Acta Computer Science, 2017, 40 (7): 1600-1613.

[2] Han Feilong, Ying Jie, Zhu Dandan. A new vehicle aided driving dynamic obstacle detection and classification method [J]. Application Research of Computers, 2017, 34 (6): 1909-1912.

[3] Chen Cuntong, Zhao Junqiao, Ye Chen, et al. Asynchronous transmission mechanism of inter-process messages for intelligent unmanned vehicles based on shared memory [J]. Journal of Software, 2017, 28 (5): 1315-1325.

[4] Zhu Hao, Yuen K V, Mihaylova L, et al. Overview of environment perception for intelligent vehicles [J]. IEEE Trans on Intelligent Transportation Systems, 2017, 32 (9): 1-18.

[5] Ou Sheng, Kang Hengbin. Object detection and classification by decision-level fusion for intelligent vehicle systems [J]. Sensors, 2017, 17 (1): 207-218.

[6] Li Yan, He Lin, Shuai Changgeng, et al. Active control of low-frequency sinusoidal vibration transmission of ship machinery [J]. Journal of Ship Mechanics, 2015, 19 (12): 1549-1563.

[7] Wang Xing, Zhang Chun, Liu Feng, et al. Exponentially weighted particle

filter for simultaneous localization and mapping based on magnetic field measurements [J]. IEEE Trans on Instrumentation & Measurement, 2017, 66 (7): 1658-1667.

[8] Yi Qun, Li Caili. Research status and development trend of active control technology of mechanical vibration [J]. Mechanical Engineering & Automation, 2016, 15 (3): 220-221.

[9] Varanis M, Silva A L, Mereles A. On mechanical vibration analysis of a multi degree of freedom system based on arduino and MEMS accelerometers [J]. Revista Brasileira De Ensino De Física, 2017, 40 (1): 1-10.

[10] Wu Yong, Guan Shengxiao. Improved SLAM problem solving method based on unscented Kalman filter [J]. Computer System Application, 2017, 26 (3): 30-36.

[11] Li Jianpo, Mu Baochun. Based on collaborative prediction algorithm for all mobile nodes localization in wireless sensor networks [J]. Application Research of Computers, 2017, 34 (1): 186-188.

[12] Liu Yang, Yang Chao. Multi-sensor based on dynamic spatial and temporal characteristics of soft and hard collaborative decision making target detection research [J]. Application Research of Computers, 2016, 33 (2): 447-449.

[13] Martynowicz P. Study of vibration control using laboratory test rig of wind turbine tower-nacelle system with MR damper based tuned vibration absorber [J]. Bulletin of the Polish Academy of Sciences Technical Sciences, 2016, 64 (2): 347-359.

[14] Zhang Shengqi, Yue Xihong, Wang Zhengxue, et al. Active vibration control of a piezoelectric bonded smart cantilever beam using fuzzy logic [J]. Zhendong Gongcheng Xuebao//Journal of Vibration Engineering, 2017, 30 (1): 110-117.

[15] Garcia F, Martin D, Escalera A D L, et al. Sensor fusion methodology for vehicle detection [J]. IEEE Intelligent Transportation Systems Magazine, 2017, 9 (1): 123-133.

[16] Zhao Wei. Automatic motion classification method based on DWT and random forest [J]. Journal of Xiangtan University Natural Science, 2018, 40 (1): 107-110.

[17] Zhao Li, He Zhen, Cao Wei, et al. Real-time moving object segmentation and classification from HEVC compressed surveillance video [J]. IEEE Trans on Circuits & Systems for Video Technology, 2017, 32 (9): 14-21.

[18] Ke Hui, Wang Xiaohui, Zhang Jianjun, et al. Active suppression for terminal residual vibration of a 3-DOF parallel robot with limbs of structures [J]. China Mechanical Engineering, 2016, 27 (12): 1648-1655.

[19] Yi Wen, Jiang Min, Hoseinnezhad R, et al. Distributed multi-sensor fusion using generalised multi-Bernoulli densities [J]. IET Radar Sonar & Navigation, 2017, 11 (3): 434-443.

[20] Xie Huan, Pu Jinlong. Research on information fusion method of multi-sensor based on dempster-shafer evidence theory [J]. Electro-Optic Technology Application, 2017, 26 (4): 562-570.

[21] Proença H, Neves J C. Deep-PRWIS: periocular recognition without the iris and sclera using deep learning frameworks [J]. IEEE Trans on Information Forensics & Security, 2017, 34 (9): 34-42.

Note: Figure translations are in progress. See original paper for figures.