Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Review:

A survey on multi-sensor fusion based obstacle detection for intelligent ground vehicles in off-road environments*

Jin-wen HU†, Bo-yin ZHENG†‡, Ce WANG, Chun-hui ZHAO, Xiao-lei HOU, Quan PAN, Zhao XU

Key Laboratory of Information Fusion Technology, Northwestern Polytechnical University, Xi'an 710072, China

†E-mail: hujinwen@nwpu.edu.cn; zhengboyin@mail.nwpu.edu.cn

Received Sept. 24, 2019; Revision accepted Dec. 2, 2019; Crosschecked Apr. 1, 2020

Abstract: With the development of sensor fusion technologies, there has been a lot of research on intelligent ground vehicles, where obstacle detection is one of the key aspects of vehicle driving. Obstacle detection is a complicated task, which involves the diversity of obstacles, sensor characteristics, and environmental conditions. While the on-road driver assistance system or autonomous driving system has been well researched, the methods developed for the structured road of city scenes may fail in an off-road environment because of its uncertainty and diversity. A single type of sensor finds it hard to satisfy the needs of obstacle detection because of the sensing limitations in range, signal features, and working conditions of detection, and this motivates researchers and engineers to develop multi-sensor fusion and system integration methodology. This survey aims at summarizing the main considerations for the onboard multi-sensor configuration of intelligent ground vehicles in the off-road environments and providing users with a guideline for selecting sensors based on their performance requirements and application environments. State-of-the-art multi-sensor fusion methods and system prototypes are reviewed and associated to the corresponding heterogeneous sensor configurations. Finally, emerging technologies and challenges are discussed for future study.

Key words: Multi-sensor fusion; Obstacle detection; Off-road environment; Intelligent vehicle; Unmanned ground vehicle

https://doi.org/10.1631/FITEE.1900518 **CLC number:** TP242.6

1 Introduction

In recent years, multi-sensor fusion based methods have been widely applied in intelligent ground vehicles for both civilian and military purposes, such as logistics, environmental exploration, search and rescue, and surveillance. The multi-mode sensor suite onboard enables the vehicle to function in more varied environments than the single-mode type. Many organizations and institutions thus have been motivated to develop advanced driver assistance systems (ADASs) and autonomous driving vehicles (ADVs) (https://medium.com/waymo/where-the-next-10-million-miles-will-take-us-de51bebb67d3; https://www.tesla.com/model3; Aufrère et al., 2003; Li THS et al., 2010; Son et al., 2015). Unlike ADVs and ADASs, unmanned ground vehicles (UGVs) do not transport human beings. The first initiatives were for military use and space applications, such as the famous Mars Exploration Rover Project. Then, they were widely adopted for civilian and commercial applications (such as agriculture, manufacturing, mining, and industry). Thus, they can be used usually in non-structured and off-road situations, e.g., Black Knight (https://www.nrec.ri.cmu.edu/solutions/defense/

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61603303, 61803309, and 61703343), the Natural Science Foundation of Shaanxi Province, China (No. 2018JQ6070), the China Postdoctoral Science Foundation (No. 2018M633574), and the Fundamental Research Funds for the Central Universities, China (Nos. 3102019ZDHKY02 and 3102018JCC003)

ORCID: Jin-wen HU, https://orcid.org/0000-0003-2771-2393; Bo-yin ZHENG, https://orcid.org/0000-0003-3908-9332

other-projects/black-knight.html), Crusher (https://www.nrec.ri.cmu.edu/solutions/defense/other-projects/crusher.html), Dragon Runner (https://qinetiq-na.com/products/unmannedsystems/dragon-runner/dragon-runner-10/), and Terramax (https://oshkoshdefense.com/advanced-technologies/terramax-unmanned-ground-vehicle-technology/). Compared with roads, it is more difficult to identify a safe moving space for vehicles when more complicated obstacles exist, such as trees, bushes, holes, hills, and rocks. Therefore, UGVs always require more powerful sensor configurations to obtain full knowledge of the environment. Several typical ADASs/ADVs and UGVs are listed in Table 1, along with their corresponding sensor configurations where lidar and camera are the two main sensors for them. The use of other sensors varies with the applications. For example, the Tesla Model 3 (https://www.tesla.com/model3) is for commercial use, so it is equipped with sonars for near-range detection such as parking assistance. In addition, most UGVs are developed to work at night, so infrared cameras are usually employed for night vision detection. Therefore, a good choice of sensor suite depends on not only the sensor abilities, but also its application.

There are several review papers related to autonomous driving, but many of them focus on only the structured roads. Zhu et al. (2017) reviewed mainly the methods related to four practical abilities that a self-driving car should have, i.e., lane detection, traffic light recognition, vehicle tracking, and scene understanding. van Brummelen et al. (2018) contributed a comprehensive review of technologies of ADV perception, while the complexity of obstacle types for off-road was not considered. Arnold et al. (2019) reviewed monocular-based, point-cloud-based, and fusion-based methods for three-dimensional (3D) object detection, but they focused on only self-driving cars in urban scenes and discussed only lidar-camera fusion. Rosique et al. (2019) not only gave a detailed comparison of different sensors used for both obstacle detection and self-positioning, but also discussed the existing simulators and regulations in different countries for autonomous vehicles (AVs).

In this review, we aim at summarizing the main considerations for the onboard multi-sensor configuration of intelligent ground vehicles in a complex environment and providing users with a guideline to select sensors based on their performance requirements and application environments, such as weather, lighting condition, and obstacle type.

2 Sensor characteristics and analysis

In this section, we give a detailed analysis of different sensors with their pros and cons for obstacle detection. The final objective of obstacle detection is to determine the places where the vehicle cannot move across because of the specific terrain or objects. Such terrain or objects are considered as obstacles. The active or passive sensors that are useful in determining the geometric properties as well as the positions of the obstacles are all under consideration, and are listed in Table 2. Selecting the best sensors from a variety of products and manufacturers is a difficult and crucial task. Trade-offs need to be considered carefully among the following characteristics: active or passive; limited capability or all-weather day-or-night operation; range, direction, or color measurements; their prices. These characteristics have direct impacts on the perception performance of the system. Range measurement is a key characteristic in determining the obstacle location. A two-dimensional (2D) image can offer color, texture, and direction information, but lacks the

Table 1 Surround sensors adopted by typical ADASs/ADVs and UGVs

Vehicle	Product	Color camera	Stereo camera	Infrared camera	Lidar	Radar	Sonar
ADASs/ADVs	Navlab11 (Aufrère et al., 2003)	✓			✓		
	Waymo Firefly (Krafcik, 2018)	✓			✓	✓	
	Tesla Model 3	✓				✓	✓
UGVs	Terramax	✓		✓	✓	✓	
	Crusher	✓	✓		✓		
	Black Knight		✓	✓	✓		
	Dragon Runner	✓		✓			

Table 2 Sensors for obstacle detection

Sensor	Type	Measurement	Resolution	Sensing range	Fog, rain, and snow robustness	Night time robustness	Cost
Sonar	Active	Range	Sparse	+	+++	+++	++
Lidar	Active	Angle, range, and size	Sparse	++	+	+++	+++
Radar	Active	Angle, range, size, and velocity	Highly sparse	+++	+++	+++	++
RGB-D (SL)	Active	Range, color, and direction	Dense	+	+	++	++
RGB-D (ToF)	Active	Range, color, and direction	Dense	+	+	++	++
RGB-D (stereo)	Passive	Range, color, and direction	Dense	+	++	+	+
Monocular camera	Passive	Color and direction	Dense	−	++	+	+
Thermal camera	Passive	Relative differences of radiation and direction	Dense	−	++	+++	++
Polarized camera	Passive	Polarization of reflected light and direction	Dense	−	++	+	+

depth information of the obstacles. For both range and image measurements, 3D detection algorithms have been researched extensively in recent years. Depth information can be obtained by visual odometry, lidar, sonar, and so on (Arnold et al., 2019). We present a detailed analysis of the sensors in the following subsections by dividing them into three categories: range-based, image-based, and hybrid sensors (Table 3).

Table 3 Category of sensors used for intelligent vehicles

Sensor	Vehicle
Range-based	Lidar; radar; sonar
Image-based	RGB monocular; infrared thermal camera; polarization camera
Hybrid (RGB-D)	RGB stereo; structured light; time-of-fight (ToF)

2.1 Range-based sensors

Lidar (light detection and ranging), sonar (sound navigation and ranging), and radar (radio detection and ranging) are the most common range sensors used for intelligent vehicles. Lidar measures the distance in its field-of-view (FoV) by calculating the time taken by a pulse of light that travels to an object and backs to the sensor. Table 4 illustrates several widely used lidars and classifies them into two types: traditional mechanical lidars and solid-state lidars (SSLs). They are compared in terms of factors, including detection range, FoV, weight, power consumed, and cost.

For traditional mechanical lidars, the Hokuyo UTM-30LX (https://www.hokuyo-aut.jp/search/single.php?serial=170) and velodyne lidar series (VLP16, https://velodynelidar.com/vlp-16.html; HDL32, https://velodynelidar.com/hdl-32e.html; HDL64, https://velodynelidar.com/hdl-64e.html) represent individually typical 2D and 3D lidars. 3D lidars offer us richer 3D point cloud data than single beam lidars. This characteristic makes it easier to extract the features from the 3D point cloud data, so many intelligent vehicles or ADASs have adopted such sensors. However, these 3D lidars are more expensive. Note that the rotation rate and horizontal resolution of 2D lidars are more competitive than those of the 3D ones. When considering these factors, people tend to choose 2D lidars. For example, 2D lidars have been used on the "Stanley" by a Stanford team, the winner of the 2005 DARPA Grand Challenge (Thrun et al., 2006).

In recent years, SSLs have been well studied because they supply the automotive market a low-cost automotive-grade laser light based detecting and sensing component without moving parts. Traditional mechanical lidars are expensive and large, and have spinning mechanical pieces that could easily break during operations. Thus, SSLs are more robust to vibrations and have a longer life span. Electromechanical lidar can run 1000–2000 h before a failure, while SSLs can run up to 100x longer. SSLs are principally based on three technologies: flash lidar, optical-phased array (OPA), and micro-electromechanical system (MEMS). Flash lidars are used especially for short-range measurement. In a flash lidar, the transmitter illuminates the whole scene and an array of detectors measure the distance of each pixel on the image. LeddarTech (Quebec

Table 4 Classification of lidars and their characteristics

Lidar	Model	Range	Horizontal FoV	Vertical FoV	Cost	Weight	Power
Mechanism	2D lidars (UTM-30LX)	Short	Wide	–	Medium	Light	Low
	3D lidars (VLP16, HDL32, and HDL64)	Long	Wide	Wide	High	Varies	Varies
Solid-state	Flash lidars (LeddarTech)	Short	Small	Small	Low	Light	Low
	OPA-based lidars (Quanergy)*	–	–	–	–	–	–
	MEMS-based lidars (Innoviz)	Medium	Small	Medium	Low	Medium	Medium

OPA: optical-phased array; MEMS: microelectromechanical system. * represents that there are no successful commercial OPA-based lidars on the market

City, Canada, https://leddartech.com/lidar/m16-multi-segment-sensor-module/) provides affordable flash lidar sensors to the market. Although the costs of flash lidars are low, their detection distance is relatively short. The OPA technology enables electronic beam to steer. It uses an OPA as a transmitter which steers laser pulses by shifting the phase. Since there is no moving part, OPA permits a high scanning speed of over 100 kHz over large angles. Quanergy (Sunnyvale, CA, USA, https://quanergy.com) is one of the companies focusing on the development of such lidars. Quanergy company has advertised its model S3-8 as the first affordable SSL sensor in the world, having a 120° horizontal FoV with a 0.05° resolution. The last one, which is also the most promising one in intelligent vehicles, is MEMS-based lidars. Innoviz (Kefar Sava, Israel, https://innoviz.tech/innovizpro/) released an MEMS-based solid state lidar (InnovizPro), which allows ranging up to 150 m with a 73° × 20° FoV with a 0.15° horizontal resolution. Yoo et al. (2018) investigated the current development and research on MEMS-based lidars.

For radars, there are commonly two operating modes. One is based on the time-of-fight (ToF) method similar to lidar. The difference is that they use radio pulses instead of light for the ranging process. The resolution of the sensors can be adjusted by changing the pulse width and time length you listen for a response (a ping back). These sensors often have fixed antennas leading to a smaller operating FoV (compared with lidar). The other type of radar relies on frequency-modulated continuous waves (FMCWs). FMCW radars use the frequency difference between the reflected and transmitted signals to determine a frequency shift. That frequency shift can be used to determine the range to the object that reflected it. Meanwhile, FMCW radars can accurately determine the relative traffic speed or the velocity of a moving object using the Doppler frequency shift. At present, there are two specifications of well-used microwave radars: 24–29 GHz for short range and 76–77 GHz for long range. The 24 GH radars have a length of around 1.25 cm, and are used mainly for short-distance sensing, surrounding environment perception (such as pedestrians and vehicles), parking assistance, lane change assistance, and other functions. The wavelength of the 77 GHz radars is normally shorter than 4 mm, which have better performance for mid/long-range measurement and better resolution.

Ultrasonic sensors work on the principle of reflected sound waves and are used to measure distance. Sound waves are emitted by the ultrasonic sensor and they are reflected back if there is an object in front of the sensor. The distance measurement is ToF-based. Since it is based on sound, it is insensitive to light, dust, vapor, and smoke hindering factors in the environment. In addition, since ultrasonic waves can reflect from a glass or liquid surface and return to the sensor head, transparent targets can be even detected.

Table 2 shows a comparison of different range sensors with respect to different environmental conditions and applications. The lidar and millimeter-wave radar are complementary, and they can work together to make up for individual shortcomings. Ultrasonic sensing is usually used for short-distance applications at low speeds, such as parking assistance, self-parking, and blind-spot detection because of its limited range.

2.2 Image-based sensors

RGB monocular cameras and infrared thermal cameras are well-adopted passive imaging sensors in obstacle avoidance tasks. RGB cameras produce visible image data, and infrared thermal cameras can be used to detect objects with different body surface temperatures based on thermal characteristics. For field scenes, water hazards can be detected with higher robustness using the polarization characteristic of the reflected light. The polarized sensors can detect and filter angles of polarization from reflected and refracted light. In most cases, they are made by adding polarisers in front of a monocular or a stereo camera. Detailed discussion will be given in Section 4, where water hazards detection is discussed.

RGB cameras use visible light as their work principle to obtain rich image information. They project the 3D world onto a 2D image plane. Although during this process depth information is lost, rich texture information can be conserved on the image plane. With the development of deep learning methods and improvements in computational ability, many algorithms have brought a faster and easier way to solve the detection and classification problem, and become more practical for implementation in an embedded system (such as intelligent vehicles or drones). However, RGB cameras suffer from illumination and weather degradation, loss of depth information, and the difficulties in computation for large outdoor scenes. In addition, less texture environment will cause cameras to be less efficient. To compensate for the loss of depth information and to obtain the positions of surrounding obstacles from a monocular, the technologies of visual simultaneous localization and mapping (SLAM) and structure from motion (SFM) have been developed (Cadena et al., 2016). The core of SLAM and SFM is using multi-view geometry to estimate the motions (rotation and translation) and construct the unknown surrounding environment. Even though these methods can bring localization information of the surroundings, they are robust to large scenes or quick movements.

In March 2018, a self-driving Uber car was involved in a fatal accident, and in May 2018, a Tesla car hit a police car while driving with the autopilot system. Both cars were equipped with fused sensors (the Uber car had visible light cameras, lidars, and radars). Since then, infrared thermal cameras have been discussed a lot because they have better vision in darkness. Furthermore, they perform equally well in daytime, offering redundancy for a RGB camera. In addition, the long-wavelength infrared (LWIR) camera can feel heat instead of seeing light, so it can reduce the impact of occlusion on the classification of obstacles from a cluttered background. The 2016 AWARE Vision Project (AWARE means all-weather all-roads enhanced) tested four different bands on the electromagnetic spectrum (visible RGB, near-infrared, short-infrared, and LWIR), and evaluated their detection performances in challenging visibility conditions (Pinchon et al., 2018). The results showed that LWIR cameras can detect pedestrians in full darkness and penetrate better than any other sensors.

2.3 Hybrid sensors

RGB-D sensors stand out for their capacity of measuring both color and depth information. Three RGB-D sensors (RGB stereo, structured-light (SL) based, and ToF-based) have been analyzed, and are listed in Table 5.

Like the monocular RGB camera, the RGB stereo camera (ZED, developed by Stereolabs, https://www.stereolabs.com/zed/) is a typical passive sensor which catches the reflected light in the environment. The RGB stereo camera works well only under conditions where the light is neither too strong nor too weak and the scenes under detection have rich textures. Stereo cameras work in the same way as the human vision system in inspecting the depth

Table 5 Comparisons of different types of three-dimensional RGB-D sensors

Sensor	Type	Extrinsic calibraion	Illumination sensibility	Darkness performance
Stereo vision	Passive	Yes	High	No
Structured-light-based	Active	Yes	Low	Yes
ToF-based	Active	No	Low	Yes

Sensor	Outdoor scene	Depth accuracy	Image resolution	Cost
Stereo vision	Yes	Middle (mm–cm)	Camera-dependent	Low
Structured-light-based	No	High (μm–cm)	Camera-dependent	Middle
ToF-based	Yes	Middle (mm–cm)	Low	High

information in the scene. The estimate of depth is strongly constrained by the length of the baseline in a stereo camera system. The short distance between two cameras results in a limited depth accuracy, while wide-baseline cameras suffer from more frequent occlusions and partial loss of spatial data (Seitz et al., 2006). In addition, stereo-based depth estimation is computationally difficult and needs to be implemented along with graphic processing unit (GPU) processors. The SL- and ToF-based cameras are active sensors which emit their own light, so they are more robust in a low-light environment. The purpose of these cameras is to solve the complexity and robustness problem of the matching algorithm in stereo cameras. However, the laser speckle may be submerged under strong light, and thus they are not suitable for outdoor scenes. Finally, the ToF depth cameras emit infrared light and measure the ToF to the observed object for depth measurements. They offer relatively accurate depth information with registered images at rates suitable for real-time applications. However, ToF cameras have low resolution and significant acquisition noises. Kinect I and Kinect II from Microsoft represent the SL- and ToF-based RGB-D cameras, respectively. Kinect I combines a color camera with a depth sensor projecting invisible structured light, while Kinect II couples a color camera with a ToF camera.

3 Multisensory fusion methods and comparisons

As described in Section 2, there is no single sensor capable of sensing all the measurements, and an intelligent vehicle needs to precisely perceive in a complex environment. Therefore, people tend to introduce many sensors to complement or verify the information provided.

Khaleghi et al. (2013) discussed in depth the data-related problems and multiple fusion methods. They proposed a taxonomy of data fusion methodologies. A more global and comprehensive review has been made by Lahat et al. (2015), where they emphasized complementarity as the principal reason to use multi-mode sensors and gave guidelines on how to approach a data fusion problem. Fourati (2015) collected the latest data fusion concepts and algorithms as well as many applications, such as intelligent transportation, medical diagnosis, and robotics. According to these works, there are mainly two purposes of fusion-based methods: redundancy and complementarity. Redundancy uses the same physical measurands from either the heterogenous sensor (range information from radar and lidar) or the same sensor in time series. Redundancy fusion aims at improving the accuracy of the measurements and managing the uncertainty of data by identifying and associating the data to update the confidence level of fused information. Complementary information has two types: one is using different physical measurements (range, color, temperature, etc.) for the same FoV and the other is complementarity of the FoV. Using complementary information can help enrich the representation of the surrounding environment.

Therefore, we review and compare the existing sensor fusion methods in this section. They are divided into three categories: probability-, classification-, and inference-based methods. Table 6 shows a comparison and applications of the existing fusion methods related to different modalities' fusion.

3.1 Probability-based methods

Probability-based methods are built upon the Bayesian recursive rule, which provides a method for computing the posterior (conditional) probability density of the hypothetical state x_k at time k given the set of measurement $Z^k = \{z_1, z_2, ..., z_k\}$ and the prior distribution $p(x_k|Z^k) = \dfrac{p(z_k|x_k)p(x_k|Z^{k-1})}{p(z_k|Z^{k-1})}$. $p(z_k|x_k)$ is called the likelihood function and is based on the given sensor model. $p(x_k|Z^{k-1})$ is called the prior distribution and stores all the past information. The denominator is a normalizing term to ensure that the probability density function integrates to one.

3.1.1 Probabilistic occupancy grids

Probabilistic occupancy grids (POGs) is the simplest approach for implementing Bayesian data fusion. Although simple, the occupancy grid is useful for vehicles' perception tasks: mapping (Elfes, 1989; Vu, 2009), moving object detection (Baig et al., 2011), and sensor fusion (Grabe et al., 2009). The occupancy grids discretize the environment of interest into a grid of equal-size spatial cells. Each cell carries the probability value that represents the occupancy

Table 6 Fusion method applications and analysis

Type	Method	Reference	Fusion level	Main contributions (+) and limitations (−)
Probability-based	Grid map	Sock et al., 2016	Decision	+ Easy implementation for traversable map generation + Ability for non-linear systems − Computationally intractable in high dimensions
	Kalman filter	Cho et al., 2014; Asvadi et al., 2016	Decision	+ Widely adopted and easily implemented for moving object tracking + Offering an analytical solution to the dynamic system − Limited in handling non-linearities
	Particle filter	Liu and Sun, 2012	Decision	+ Widely adopted and easily implemented for moving object tracking + Being able to deal with non-linearities − More computation burdens than those of KF-based methods
Classification-based	SVM	Li QQ et al., 2014	Feature	+ Widely used for road detection + Being able to deal with non-linearities − Hand-selected features − High complexity in large-scale tasks
	CRF/MRF	Häselich et al., 2011; Xiao et al., 2018	Decision/Feature	+ Widely used for image labeling and point cloud labeling + Applications for road detection and terrain classification − Complex and time-consuming for energy minimization functions
	Deep learning	Chen XZ et al., 2017; Ku et al., 2018	Feature	+ Widely used for 3D object detection in urban scenes + No hand-selected features before the classification step + An end-to-end detection process − High storage and computation requirements
Inference-based	Fuzzy logic	Zhao et al., 2014; Wei et al., 2018	Decision	+ Adopted mainly for classification confidence fusion + Intuitive approach to deal with vague data − Choice of good membership functions and fuzzy rules strongly influencing fusion reliability
	Evidence theory	Chavez-Garcia, 2014; Starr and Lattimer, 2017	Decision	+ Adopted mainly for classification confidence fusion + Being enable to fuse uncertain and ambiguous data − Inability for fusion of highly conflicting data and incapacity to deal with data imprecision

state of that cell. Usually, a cell can be empty or occupied. To construct a map of the robot world, estimates of the cell state are obtained by interpreting the incoming range readings using probabilistic sensor models. The Bayesian estimator allows the incremental updating of the occupancy grid. The tendency now is to construct a globally consistent map from time series or multiple robots (Yue et al., 2017, 2018). Computational cost of the grid-based fusion method is strongly related to the resolution and the size of the environment, and thus it is more adaptive to situations where dimensions are modest or with certain assumptions to reduce the size of the grid. Several improvements have been made by the occupancy grid method, such as hierarchical (quadtree) grids (Bosch et al., 2007) , irregular (triangular and pentagonal) grids (Azim and Aycard, 2012), and working on a local map (Vu, 2009) to avoid the global map updating process.

3.1.2 Kalman filter

The prior distribution term and normalizing term inside the Bayesian estimator contain integrals that cannot be evaluated analytically. Thus, the Kalman filter (KF) arose as an exact analytical solution under constraints of system dynamics being linear Gaussian. In addition, it is easy to implement a KF because of its simplicity and optimality in a mean-squared error sense. One of the extensions of the KF is the extended KF (EKF) that can be employed when at least the state model or the observation model is non-linear. Another extension version of the KF is the unscented KF (UKF). In the UKF, the probability density is estimated by a deterministic sampling of points which represent the underlying distribution as a Gaussian. The non-linear transformation of these points is intended to approximate the posterior distribution. In the fusion frame, KF-based methods are often used for dynamic object tracking or SLAM by estimating the motions on time series.

3.1.3 Monte-Carlo methods

Monte-Carlo (MC) simulation based methods describe probability density as a set of weighted samples of an underlying state space. They are flexible as they do not make any assumptions regarding the probability densities to be approximated. Sample-based methods can represent general probability densities, so that MC methods are suited for non-linear problems. The Markov chain Monte-Carlo (MCMC) methods and the sequential Monte-Carlo (SMC) methods (also known as particle filters) are two types of MC methods for obtaining the estimate of the optimal state with non-linear non-Gaussian state-space models.

For large data sets, MCMC methods require complete "browsing" of the observations from the initial time to the current time, while SMC methods reduce computational complexity by iteratively involving only the latest observation at the current time. They depend on the sampling of the posterior distribution. Thus, the computation load of SMC methods is at least $O(N)$, where N is the number of samples used to approximate the posterior distribution (del Moral et al., 2012), while the computation load of the MCMC methods is at least $O(tN)$, where t is the discrete time index.

Compared with the EKF, SMC methods are more computationally expensive because they may require a large number of random samples. In other words, the number of samples selected randomly will increase exponentially with the dimension of the state spaces. The disadvantage of this method is that it is difficult to determine the optimal number of particles (Gning et al., 2007).

3.2 Classification-based methods

The most commonly used classification method is the support vector machine (SVM), where the features from different sensors are concatenated to a combined feature and then an SVM classifier is trained for a certain object detection. This method is capable of handling both non-linear and non-monotonic data, producing an intuitive and unique result (Burges, 1998). However, the internal working process of this method is not clear enough and lacks the transparency of results. Also, the choice of the kernel function is a problem. In large-scale tasks, it has high complexity and large storage requirements (Avidan, 2004).

The conditional random field method and Markov random field method were initially used for image labeling and lidar point labeling. They can be taken as a process of optimizing a probabilistic graphical model, defined as $\varsigma = (\nu, \varepsilon)$, where $\nu = \{X_1, X_2, ..., X_N\}$ is the discrete random variable to be predicted and ε defines the neighboring or connectivity bewteen the random variables. According to the Hammersley-Clifford theorem, the posterior distribution probability $\Pr(x|Y)$ over the labelings of the conditional random field (CRF) is a Gibbs distribution. Then, the maximization for the posterior ditribution is turned into minimizing the Gibbs energy function. This is a global optimization and can result in a smooth and accurate prediction.

3.3 Inference-based methods

For inference-based methods, the input information is obtained from each individual sensor and has a symbolic representation. The purpose of the fusion is to combine the symbolic representations with associated uncertainty metrics (confidence scores) into a comprehensive decision. For example, a lidar and a camera detect the object in the same position as a pedestrian with different confidence scores, and then two confidence scores are fused to be more reliable by inference-based methods.

The Dempster-Shafer evidence theory is one of these methods. It can clearly model the uncertainty that the Bayesian method cannot properly represent. This method models not only the underlying assumptions but also the composite hypotheses. It is more flexible than the method of probability theory. However, the disadvantage of this method is that as the number of components of the identification framework increases, the computational complexity increases as well. There are still conflicting problems in this method, so the assumption of independent evidence is not necessarily credible, and some scholars have proposed improvements for conflicting evidence (Denoeux, 2000; Zhan et al., 2009).

Fuzzy reasoning is another inference-based method and can handle problems with vague data. It deals with reasoning that is approximate (a value between 0 and 1) rather than exact. It is concerned with degrees of membership in vaguely defined sets. This degree of set membership indicates the proposition degree of truth. Fuzzy reasoning is widely used in control applications. However, in the fusion area, it is limited with some drawbacks: it is not straightforward to settle good membership functions and fuzzy rules. In addition, fuzzy rules need extensive tests for the verification and validation of the system. This is very important in a vehicle safety system (Subramanian et al., 2009; Majumder and Pratihar, 2018).

The above classification of the fusion methods is based on their intrinsic working principles. There is another popular way to categorize fusion methods (according to different levels of fusions): raw data level, feature level, and decision level. Raw data can be directly combined if the sensor data is commensurate (i.e., the sensors measure the same physical phenomenon such as two images or two point clouds). In feature level fusion, features from two sensors are combined into one vector for posterior classification (Kaempchen et al., 2005; Douillard et al., 2009; Schneider et al., 2010; Häselich et al., 2011). Raw data level fusion and feature level fusion can be included in the low level fusion category. In this category, fused data should be inside the FoV of all the sensors, thus leading to the limitation that the FoV of the system relies only on the sensor who has the narrowest FoV. Decision level fusion can adjust to this issue. When the sensor data is not commensurate, feature level fusion or decision level fusion will be adopted. In feature level fusion, features are extracted from different sources and combined into a single feature vector with concatenation operations (Su et al., 2015; Cai et al., 2016; Hosang et al., 2016). Decision level fusion combines a series of sensor preliminary determined information (entity's location, label, or other attributes). Typical methods in this level contain weighted decision methods, fuzzy logic algorithms, the Bayesian inference, and the Dempster-Shafer theory (Seraji, 2003; Shirkhodaie et al., 2005; Wei et al., 2018). One of the most recent studies on beacon detection in an industrial environment combines the detections from the camera and a sparse lidar (eight beams) to realize the fusion (Wei et al., 2018). First, Wei et al. (2018) separately used the two sensors to realize beacon detection. Then, they used a neural network to transform the bounding box from camera detection to the range and angle information in the lidar coordinate. Next, they fused the confidence scores from each sensor in one frame by fuzzy logic theory. Their method has the advantage of reducing the false positive detections. Another well-adopted method to detect the obstacle in an outdoor environment is the Bayes inference. Sock et al. (2016) used a lidar and a camera to generate an individual drivability map and then fused it onto a 2D probability grid map using Bayes' rules.

4 Prototype systems and applications for obstacle detection in diverse environments

In this section, an obstacle detection system architecture is presented, and then the prototype systems are reviewed with different obstacle types and environmental conditions.

4.1 System architecture

The overall architecture of the obstacle detection system based on multi-sensor fusion methods is illustrated in Fig. 1. The overall system is composed of four independent units as follows:

1. Data acquisition unit. This unit is responsible for acquiring and storing raw sensor information.

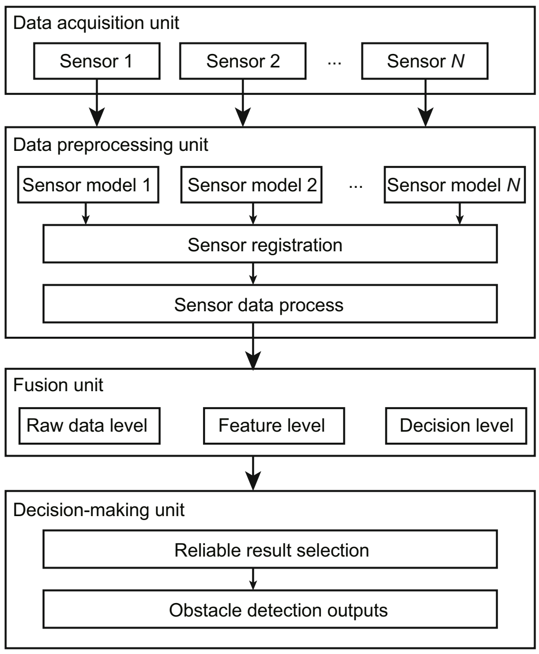

Fig. 1 Architecture of a multi-sensor fusion obstacle detection system

It has multiple sensors as its inputs. These sensors acquire and store their raw sensor data via an interface. This unit runs independently of any other unit.

2. Data preprocessing unit. This unit receives the sensor raw data from the data acquisition unit. Extrinsic calibration is obtained to do sensor registration and data association.

3. Fusion unit. This unit performs raw data level fusion, feature level fusion, and decision level fusion on the appropriate pre-processed sensor data at different levels as needed.

4. Decision-making unit. This unit selects and determines good fusion results for obstacle detection tasks based on the results given by the multi-sensor fusion output, and then passes the obstacle information (such as locality, label, size, and velocity) to other modules in an automated system (such as the path planning or control modules).

4.2 Applications with different types of obstacles

For off-road environments, UGVs will encounter not only vehicle-human obstacles, but also many irregular ones. Vehicle-human detection has been widely researched for on-road scenes, and therefore the related algorithms can be easily transferred for off-road vehicle-human detection. To give a comprehensive coverage of all obstacles, in this subsection, we briefly review several on-road detection algorithms, point out their limitations, and analyze off-road detection algorithms according to the categorized types of obstacles.

For vehicle-human detection, the system tends to require high sensing abilities because a single failure of detection or misclassification can cause dramatic accidents. A wide horizontal FoV is necessary because incidents can come from any direction in complex and cluttered urban areas. It can be seen from Cho et al. (2014) that the horizontal FoV has been well covered with the multi-mode sensor configuration. However, for irregular obstacles in off-road scenes, it is difficult to figure out the semantic information of each obstacle because of their variety. The main concern is to obtain more accurate estimates of location and size for obstacle avoidance, path planning, and navigation. Comparisons considering obstacle types can been found in Table 7.

4.2.1 On-road obstacle detection

Perceiving the city-scene environment usually consists of two parts: One is to realize road detection and the other is to detect, recognize, and track the vehicles and pedestrians. For road detection, multi-sensor fusion methods have been widely used (Hu et al., 2014; Li QQ et al., 2014; Xiao et al., 2018). Hu et al. (2014) proposed lidar-camera-based road detection and segmentation, but they used the two sensors separately. They analyzed and processed the lidar data for ground seed extraction, while road detection and segmentation were dominated by a camera. Li QQ et al. (2014) combined a 2D lidar and a camera and proposed robust road detection by feature level fusion. Another state-of-the-art method has been proposed by Xiao et al. (2018). They suggested a novel hybrid-CRF-based fusion method to improve the performance of road detection.

For on-road obstacle detection, tremendous work has been done for single sensor detections. Lidar-based methods have been well developed (Montemerlo and Thrun, 2006; Hadsell et al., 2010; Bogoslavskyi and Stachniss, 2017; Siritanawan et al., 2017; Zhou and Tuzel, 2018). The Voxelnet Network (Zhou and Tuzel, 2018) developed by the Apple team is one of the latest methods for potential dynamic obstacle detection (human and vehicles) in both the city scene and field area. Montemerlo and Thrun (2006) introduced a multi-grid representation method using several maps with different resolutions to improve the performance of obstacle detection

Table 7 Comparisons of detection requirements for on-road and off-road scenes

Scene	Ability required by perception	Potential dynamic obstacle	Potential static obstacle
On-road	Road/Lane detection; obstacle detection in all views	Vehicles; pedestrians	Infrastructure; traffic signs
Off-road	Unknown obstacle detection; balance of range; FoV	Vehicles; pedestrians	Convex (rock, tree, and slope); concave (ditch, hole, slope, and pit); water hazards

for urban structures. To accurately estimate the terrain, Hadsell et al. (2010) proposed an algorithm to reconstruct 3D surfaces. The main disadvantage of these methods is that their high computational burden makes them hard to satisfy real-time requirements. Bogoslavskyi and Stachniss (2017) proposed an efficient point cloud segmentation method based on the range image. The point cloud is segmented directly by processing raw lidar measurements, so the speed is fast and not limited to different types of obstacle. Siritanawan et al. (2017) proposed a real-time 3D clustering algorithm for non-uniform and sparse lidar point cloud. In particular, their work can bring more robustness to key point extraction for SLAM applications.

Visible-light cameras were adopted in positive obstacle detection (Labayrade et al., 2002; Pfeiffer and Franke, 2011). Labayrade et al. (2002) proposed the concept of "V-disparity" to simplify obstacle detection on the road surface. In the "V-parallax" space, road planes and obstacles are described as piecewise linear curves and vertical lines, respectively, which become easy to distinguish. Pfeiffer and Franke (2011) proposed the concept of "Stixel," where obstacles are represented as columnar pixels. This reduces the influence of noise and improves the stability of the detection system. As for lidar-camera fusion, extrinsic calibration is a fundamental step and has always drawn a lot of attention. Most works correspond to the points viewed by the two sensors and use various checkboard patterns (García-Moreno et al., 2013; Gong et al., 2013a, 2013b; Levinson and Thrun, 2013; Napier et al., 2013; Park et al., 2014; Castorena et al., 2016). These methods require the lidar data have a certain density to ensure the detection of certain points. After extrinsic detection, a transformation matrix will be gained from a lidar space to a camera space.

For most detection methods for vehicles and pedestrians, a common fusion process of camera and lidar is like this: lidar data will generate the region of interests (RoIs) on the image by extrinsic calibration and then an image-based segmentation and detection method is used, and the vehicles and pedestrians will be detected (Premebida and Nunes, 2013; García et al., 2014; Zhao et al., 2014; Rubaiyat et al., 2018).

The above methods work well in the urban environment because the features in cities are human designed, so it is easier for sensors to detect the corners or other types of features. However, they are not robust enough against varying and complicated obstacles in an unknown environment such as a field. In the following, we will discuss off-road obstacle detection by classifying the obstacles into three types: convex obstacles (tree, slope, and hill), concave obstacles (pit, holes, and ditches), and water hazards.

4.2.2 Convex obstacle detection

Convex obstacle detection in the field differs from the city scene, where there are no well-structured roads but different types of terrain and where the shapes and sizes of obstacles have no common characteristics. In this case, terrain classification draws attention because it is adaptive to the field scenes and the prediction will be a traversibility map. This is enough for obstacle avoidance and path planing in the field. Like the city scene, researchers first developed single-sensor-based methods. For example, Matthies et al. (1995) proposed a stereo vision based approach for UGV navigation. Radars have a strong penetrating ability and can work well in bad weather. Jing et al. (2013) used the relationship between Doppler shift and obstacle height to analyze the height image characteristics of flat ground, convex obstacles, and pits. Based on this, a Doppler-feature-based method was proposed to estimate the height of the obstacle and to classify the obstacle. Although this method can work all day, the detection distance is not long and the accuracy is not high. Therefore, it is hard for millimetre wave radar detection alone to replace the accurate lidar detection methods, so it is common to fuse the millimeter wave radar to improve the adaptability to weather. Sensor-fusion-based methods are mainly for traversibility map generation. There are usually several vague classes for classification, such as road, rough road, obstacles, and unknowns. Shinzato et al. (2014) used a lidar-camera-based approach, in which the lidar points are projected onto the image plane, but the feature used for obstacle classification is derived from the height information of lidar points regardless of the pixel information. Sock et al. (2016) used a lidar and a camera to generate a traversibility map separately and then fused them by Bayesian inference. Häselich et al. (2011) realized feature level fusion based on a 2D grid map.

4.2.3 Concave obstacle detection

Concave obstacles are difficult to detect in general because of their negative nature. Infrared thermal cameras can be used to detect concave obstacles by measuring their temperature difference in the environment. Matthies and Rankin (2003) used a thermal infrared camera to detect concave obstacles. However, this method is applicable only when the temperature difference is large at night because the surface temperatures of the ground and the pit are almost the same during the day. For lidar sensors, Larson and Trivedi (2011) proposed a lidar-based method for concave obstacle detection, which uses mainly geometric features of the contour of concave obstacles. Then, these features are sent to an SVM classifier to detect the obstacles. Morton and Olson (2011) distinguished the negative obstacles from the positive ones by defining the height-length-density (HDL) measure. Fernández et al. (2012) used differences of height on adjacent points to identify the negative slope of speed humps. Chen L et al. (2017) proposed a lidar histogram to combine road detection, convex obstacle detection, and concave obstacle detection into one framework. On this histogram, 3D travelable segments can be projected into a straight line. Two types of obstacles are separately projected above and under the line, thereby transforming the road and obstacle detection problem into a linear classification problem in a 2D space.

However, since the laser can be reflected repeatedly in the pit, the observation information may be lost. Thus, the research tendency is to combine lidars with other sensors for better concave obstacle detection. Bajracharya et al. (2013) used a grid-map-based method to represent the negative obstacles from stereo data. They considered the difference between the maximum and minimum elevations in adjacent grid cells. Karunasekera et al. (2017) proposed a new method to detect the negative obstacles and saliency on the ground plane using a stereo camera.

4.2.4 Water hazard detection

In the field, vehicles often face water hazards. Water can cause various destructive effects such as a short circuit of the vehicle. At the same time, various field roads are often on soil, and it is possible that the vehicle is trapped in a muddy area and cannot be pulled out by itself. Therefore, water hazard is an important issue that must be considered for UGVs. A color-imagery-based scheme was proposed by Matthies et al. (2003) who have suggested, for still water, a mixture of Gaussian models to train a classifier on water regions in an RGB color space. However, this method is easily influenced by light and shadow. It works well only when sky is the main component reflected by the water.

Polarization cameras can be easily made by combining a normal monocular camera and a linear polarizer filter. They use the polarization of the reflected light in the outdoor water environment to realize the detection of water hazards. The visualization of water detection by an infrared polarization camera is shown in Fig. 2. Water hazards can be detected by a comparison of polarization degree and similarity of the polarization phases. It is good at detection (Sarwal et al., 2004) but requires three cameras with a good calibration, and the computational cost is high. Yamada et al. (2005) proposed the polarized light to distinguish the wet and dry roads and to give more detailed data, but the experimental data is based only on the structured roads. Iqbal et al. (2009) surveyed several detection of water hazards based on different sensors, including a visible-light camera, a short-wave infrared camera, a thermal infrared camera, lidars, and polarization cameras, and compared sensor characteristics for water detection.

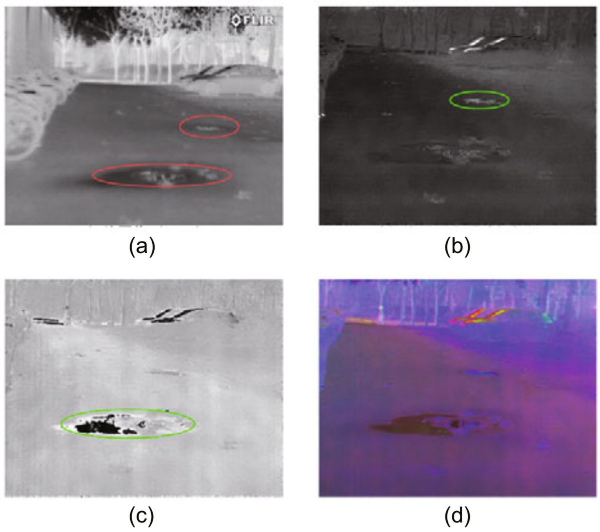

Fig. 2 Water detected by a polarization infrared camera: (a) raw image from the infrared camera; (b) polarization phase image; (c) degree of polarization; (d) pseudo color image

4.3 Applications for adverse environments

4.3.1 Low visibility detection

Obstacle detection in low visibility environments has been investigated for years. Low visibility arises in scenes without enough light or with too much smoke. Sensing ability decreases rapidly as the visibility gets lower for the RGB cameras. Thermal cameras are considered as a replacement in such cases and have been combined with other sensors to improve the detection results in applications. Liu and Sun (2012) fused color images and thermal images for object tracking. Color images are more effective than thermal images when the objects are at "thermal crossover," while thermal cameras are more adaptive to poor lighting conditions. A survey about these two sensor fusion methods and applications has been performed by Ma et al. (2019). It gives a very comprehensive view of image-based fusion methods and a well-constructed series of evaluation standards for fusion results. Thermal-visible images were used to detect pedestrians (González et al., 2016). González et al. (2016) created their own thermal-visible image dataset (CVC-14) to evaluate pedestrian detectors with a number of combinations of three factors: visible/far-infrared (FIR) modalities, pedestrian models, and lighting conditions. Then, they used the Korea Advanced Institute of Science and Technology (KAIST) multispectrum pedestrian dataset (Hwang et al., 2015) to compare single-sensor results with the sensor fusion results at daytime and at night. For solving smoke scene detection problems, several fusion systems were proposed especially for the fire environment (Kim JH et al., 2015; Starr and Lattimer, 2017). Kim JH et al. (2015) proposed a fusion system of the stereo infrared camera and a radar sensor. The stereo infrared camera is to obtain the 3D obstacle information using a disparity map and an FMCW radar obtained in parallel multi-object detection. Then, these two results are combined by calculating the correspondence of obstacles detected by each sensor. When the obstacles are matched, the distance information will be extracted from the radar which is more accurate. Starr and Lattimer (2017) combined a thermal camera with a 3D lidar. In their work, two sensors detect objects separately and then adopt the evidential fusion method (Dempster-Shafer theory) to obtain a good result. Zhang et al. (2018) first proposed a two-step method to realize extrinsic calibration between a 3D lidar and a thermal camera. With the extrinsic calibration parameters obtained from this work, the fusion algorithms between these two sensors are not limited to the decision level and can be extended for feature level fusion.

4.3.2 Adverse weather detection

Rain, snow, and fog can decrease performance for both cameras and lidars. More and more researchers tend to combine other sensors to ensure system security in bad weather using sensors such as radars. The extrinsic calibration of radar and camera was proposed by Kim J et al. (2018). Since radars have good performance for rain and snow conditions, Radecki et al. (2016) fused radars, cameras, and lidars, and proposed a perception algorithm for tracking and classification under different weather conditions. They evaluated the performances of different combinations of these three sensors (camera+lidar, camera+radar, lidar+radar, and camera+lidar+radar) under different weather conditions, and the experimental results proved that the full set of the three sensors can bring the best detection and classification results. Wang JG et al. (2018) took radars as a primary sensor and vision as a second sensor to help correct the false positives from radars and obtain classification information. Their method has shown good results for vehicle detection during heavy rain.

5 Emerging technologies and challenges

The intelligent control of unmanned vehicles is greatly constrained by the sensing and communication technologies. More and more efforts have been devoted to the emerging technologies, such as Internet of Things (IoT), vehicle networks, low-cost high-precision sensors, and cooperative detection systems, to enhance the capability of environment perception. Most of them are still under development with a lot of technological barriers and challenges to overcome.

5.1 Internet of Things

IoT is involved in a wide range of applications. Within the scope of obstacle detection, it is worth mentioning that the vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) technologies improve transportation safety in the urban area surrounded by infrastructures and dense traffic. V2I communication can provide a network for on-road information collection such as traffic lights, speed limits, and road signs. Similarly, V2V communication allows vehicles to share data, such as vehicle localization, intention to change lanes, and data acquired by some onboard sensors with other vehicles. Müller et al. (2016) proposed a cooperative localization fusion method based on onboard radar and V2V communication. A review on ranging sensors and cooperative techniques for relative positioning of vehicles was given by Müller (2017). In the following year, Tian et al. (2018) first used connected vehicle identity information to improve the accuracy of data association and to evaluate their approach on real-world collected data. Besides vehicle tracking, Fukatsu and Sakaguchi (2019) presented a protocol aiming at improving the safety of vehicle overtaking. Simulation results showed that with cooperative driving technology, it is possible to safely realize overtaking with a higher velocity. Also, their work gives an analysis of suitable bandwidth for this cooperation in real time.

5.2 Reliable wireless communication

Wireless communication technologies assure the transference of data between multiple agents in a limited period. The adoption and deployment of various wireless communication technologies depend on the applications in the field of cooperative perception. For example, safe driving requires a short data communication latency. Therefore, dedicated short range communication is employed because it enables ADVs to communicate with their neighbor vehicles in a range up to 1000 m by periodically broadcasting cooperative awareness messages (CAMs) (Kenney, 2011). The use of other wireless communication technologies, such as WiFi, LTE, WiMAX, allows longer range communication at a lower price. Such technologies can be used on applications such as traffic data collection or forward collision warning, since these applications do not require a very short latency. Dey et al. (2016) compared and evaluated the performances of different wireless networks. However, V2V gives opportunities for attackers to pass false data (Raya and Hubaux, 2007; Lim and Tuladhar, 2019) when trusted authorities and infrastructures are not available.

5.3 Sensor technologies

Although sensor technologies have been widely developed for low-cost high-efficiency sensors, there are still difficulties in using them in the real world. For example, SSLs are good replacements for mechanical lidars, but at this level, they do not guarantee the same performance as the traditional ones because of low resolution and less data. For concave obstacle detection, the existing methods discussed in Section 4 are still limited to single-modality sensors. These methods are less efficient in dealing with long-range detection, and thus people turn to a new sensor type named ground-based interferometric synthetic aperture radar (InSAR) (Wang J et al., 2016). However, ground-based InSARs are cumbersome. According to the best of our knowledge, there is no commercial ground-based InSAR for ground vehicles.

5.4 Asynchronous information fusion

As presented in Section 4, multi-sensor-based methods have attracted a lot of attention for adverse environments, but the work is still not sufficient to deal with multiple heterogeneous sensors in real time, which are usually asynchronous. For ADVs, real-time detection is crucial for safe driving but many studies can achieve only fewer than 10 frames per second (Chen XZ et al., 2017; Ku et al., 2018; Zhou and Tuzel, 2018). Further improvements are required for quick and reliable detection algorithms. Temporal fusion is another useful and common way to increase continuity and reliability of the detection because it takes advantages of temporal information. Unfortunately, only a few studies have put emphasis on it (Zhao et al., 2014; Sock et al., 2016).

5.5 Heterogeneous platform cooperation

Unmanned aerial vehicles (UAVs) and UGVs often collaborate in certain high-risk missions (Li JQ et al., 2016; Arbanas et al., 2018; Nguyen et al., 2018). UAVs are often used to cover large areas but they cannot precisely localize ground features. UGVs can accurately locate the ground features but the mobility is limited to the terrain conditions and infrastructures. The prior information offered by UAVs can be used by UGVs for localization of the obstacles and path planning. Thus, this cooperation can bring higher visibility to the environment and is worthy of investigation.

6 Summary

In this paper, we reviewed obstacle detection methods and applications based on multi-sensor fusion for intelligent ground vehicles in off-road environments. Advantages and disadvantages of commonly used detection sensors, multi-sensor configurations, fusion methods, obstacles, and suitable environments were analyzed and compared. The architecture of the multi-sensor fusion obstacle detection system was presented, followed by a review of the existing prototype systems and their applications. Finally, the integration of emerging technologies for obstacle detection tasks was discussed and future challenges for obstacle detection were suggested.

Contributors

Jin-wen HU guided and designed the research. Bo-yin ZHENG and Ce WANG drafted the manuscript. Chun-hui ZHAO and Xiao-lei HOU collected the relevant materials. Quan PAN and Zhao XU helped organize the manuscript. Jin-wen HU revised and finalized the paper.

Compliance with ethics guidelines

Jin-wen HU, Bo-yin ZHENG, Ce WANG, Chun-hui ZHAO, Xiao-lei HOU, Quan PAN, and Zhao XU declare that they have no conflict of interest.

References

Arbanas B, Ivanovic A, Car M, et al., 2018. Decentralized planning and control for UAV-UGV cooperative teams. *Auton Robot*, 42(8):1601-1618. https://doi.org/10.1007/s10514-018-9712-y

Arnold E, Al-Jarrah OY, Dianati M, et al., 2019. A survey on 3D object detection methods for autonomous driving applications. *IEEE Trans Intell Trans Syst*, 20(10):3782-3795. https://doi.org/10.1109/TITS.2019.2892405

Asvadi A, Girão P, Peixoto P, et al., 2016. 3D object tracking using RGB and LIDAR data. IEEE 19th Int Conf on Intelligent Transportation Systems, p.1255-1260. https://doi.org/10.1109/ITSC.2016.7795718

Aufrère R, Gowdy J, Mertz C, et al., 2003. Perception for collision avoidance and autonomous driving. *Mechatronics*, 13(10):1149-1161. https://doi.org/10.1016/S0957-4158(03)00047-3

Avidan S, 2004. Support vector tracking. *IEEE Trans Patt Anal Mach Intell*, 26(8):1064-1072. https://doi.org/10.1109/TPAMI.2004.53

Azim A, Aycard O, 2012. Detection, classification and tracking of moving objects in a 3D environment. IEEE Intelligent Vehicles Symp, p.802-807. https://doi.org/10.1109/IVS.2012.6232303

Baig Q, Aycard O, Vu TD, et al., 2011. Fusion between laser and stereo vision data for moving objects tracking in intersection like scenario. IEEE Intelligent Vehicles Symp, p.362-367. https://doi.org/10.1109/IVS.2011.5940576

Bajracharya M, Ma J, Malchano M, et al., 2013. High fidelity day/night stereo mapping with vegetation and negative obstacle detection for vision-in-the-loop walking. IEEE/RSJ Int Conf on Intelligent Robots and Systems, p.3663-3670. https://doi.org/10.1109/IROS.2013.6696879

Bogoslavskyi I, Stachniss C, 2017. Efficient online segmentation for sparse 3D laser scans. *PFG J Photogramm Remote Sens Geoinform Sci*, 85(1):41-52. https://doi.org/10.1007/s41064-016-0003

Bosch A, Zisserman A, Munoz X, 2007. Representing shape with a spatial pyramid kernel. 6th ACM Int Conf on Image and Video Retrieval, p.401-408. https://doi.org/10.1145/1282280.1282340

Burges CJC, 1998. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov*, 2(2):121-167. https://doi.org/10.1023/A:1009715923555

Cadena C, Carlone L, Carrillo H, et al., 2016. Past, present, and future of simultaneous localization and mapping: toward the robust-perception age. *IEEE Trans Robot*, 32(6):1309-1332. https://doi.org/10.1109/TRO.2016.2624754

Cai ZW, Fan QF, Feris RS, et al., 2016. A unified multi-scale deep convolutional neural network for fast object detection. Proc 14th European Conf on Computer Vision, p.354-370. https://doi.org/10.1007/978-3-319-46493-0_22

Castorena J, Kamilov US, Boufounos PT, 2016. Autocalibration of lidar and optical cameras via edge alignment. Proc Int Conf on Acoustics, Speech and Signal Processing, p.2862-2866. https://doi.org/10.1109/ICASSP.2016.7472200

Chavez-Garcia RO, 2014. Multiple Sensor Fusion for Detection, Classification and Tracking of Moving Objects in Driving Environments. PhD Thesis, Université de Grenoble, Grenoble, France.

Chen L, Yang J, Kong H, 2017. Lidar-histogram for fast road and obstacle detection. Proc IEEE Int Conf on Robotics and Automation, p.1343-1348. https://doi.org/10.1109/ICRA.2017.7989159

Chen XZ, Ma HM, Wan J, et al., 2017. Multi-view 3D object detection network for autonomous driving. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.6526-6534. https://doi.org/10.1109/CVPR.2017.691

Cho H, Seo YW, Kumar BVKV, et al., 2014. A multi-sensor fusion system for moving object detection and tracking in urban driving environments. Proc IEEE Int Conf on Robotics and Automation, p.1836-1843. https://doi.org/10.1109/ICRA.2014.6907100

del Moral P, Doucet A, Jasra A, 2012. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Stat Comput*, 22(5):1009-1020. https://doi.org/10.1007/s11222-011-9271-y

Denoeux T, 2000. A neural network classifier based on Dempster-Shafer theory. *IEEE Trans Syst Man Cybern Part A*, 30(2):131-150. https://doi.org/10.1109/3468.833094

Dey KC, Rayamajhi A, Chowdhury M, et al., 2016. Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in a heterogeneous wireless network—performance evaluation. *Trans Res Part C*, 68(7):168-184. https://doi.org/10.1016/j.trc.2016.03.008

Douillard B, Brooks A, Ramos F, 2009. A 3D laser and vision based classifier. Int Conf on Intelligent Sensors, Sensor Networks and Information Processing, p.295-300. https://doi.org/10.1109/ISSNIP.2009.5416828

Elfes A, 1989. Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6):46-57. https://doi.org/10.1109/2.30720

Fernández C, Gavilán M, Llorca DF, et al., 2012. Free space and speed humps detection using lidar and vision for urban autonomous navigation. Proc IEEE Intelligent Vehicles Symp, p.698-703. https://doi.org/10.1109/IVS.2012.6232255

Fourati H, 2015. Multisensor Data Fusion: from Algorithms and Architectural Design to Applications. CRC Press, Boca Raton, USA. https://doi.org/10.1201/b18851

Fukatsu R, Sakaguchi K, 2019. Millimeter-wave V2V communications with cooperative perception for automated driving. IEEE 89th Vehicular Technology Conf, p.1-5. https://doi.org/10.1109/VTCSpring.2019.8746344

García F, García J, Ponz A, et al., 2014. Context aided pedestrian detection for danger estimation based on laser scanner and computer vision. *Expert Syst Appl*, 41(15): 6646-6661. https://doi.org/10.1016/j.eswa.2014.04.034

García-Moreno AI, Gonzalez-Barbosa JJ, Ornelas-Rodriguez FJ, et al., 2013. LIDAR and panoramic camera extrinsic calibration approach using a pattern plane. Proc 5th Mexican Conf on Pattern Recognition, p.104-113. https://doi.org/10.1007/978-3-642-38989-4_11

Gning A, Abdallah F, Bonnifait P, 2007. A new estimation method for multisensor fusion by using interval analysis and particle filtering. Proc IEEE Int Conf on Robotics and Automation, p.3844-3849. https://doi.org/10.1109/ROBOT.2007.364068

Gong XJ, Lin Y, Liu JL, 2013a. Extrinsic calibration of a 3D LIDAR and a camera using a trihedron. *Opt Laser Eng*, 51(4):394-401. https://doi.org/10.1016/j.optlaseng.2012.11.015

Gong XJ, Lin Y, Liu JL, 2013b. 3D LIDAR-camera extrinsic calibration using an arbitrary trihedron. *Sensors*, 13(2): 1902-1918. https://doi.org/10.3390/s130201902

González A, Fang Z, Socarras Y, et al., 2016. Pedestrian detection at day/night time with visible and FIR cameras: a comparison. *Sensors*, 16(6):820. https://doi.org/10.3390/s16060820

Grabe B, Ike T, Hoetter M, 2009. Evaluation method of grid based representation from sensor data. Proc IEEE Intelligent Vehicles Symp, p.1245-1250. https://doi.org/10.1109/IVS.2009.5164461

Hadsell R, Bagnell JA, Huber D, et al., 2010. Space-carving kernels for accurate rough terrain estimation. *Int J Robot Res*, 29(8):981-996. https://doi.org/10.1177/0278364910369996

Häselich M, Arends M, Lang D, et al., 2011. Terrain classification with Markov random fields on fused camera and 3D laser range data. Proc 5th European Conf on Mobile Robots, p.153-158.

Hosang J, Benenson R, Dollár P, et al., 2016. What makes for effective detection proposals? *IEEE Trans Patt Anal Mach Intell*, 38(4):814-830. https://doi.org/10.1109/TPAMI.2015.2465908

Hu X, Rodriguez FSA, Gepperth A, 2014. A multi-modal system for road detection and segmentation. Proc IEEE Intelligent Vehicles Symp, p.1365-1370. https://doi.org/10.1109/IVS.2014.6856466

Hwang S, Park J, Kim N, et al., 2015. Multispectral pedestrian detection: benchmark dataset and baseline. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.1037-1045. https://doi.org/10.1109/CVPR.2015.7298706

Iqbal M, Morel M, Meriaudeau F, 2009. A survey on outdoor water hazard detection. 5th Int Conf on Information and Communication Technology and Systems, p.33-39.

Jing X, Du ZC, Li F, 2013. Obstacle detection by Doppler frequency shift. *Electron Sci Technol*, 26(8):57-60 (in Chinese). https://doi.org/10.3969/j.issn.1007-7820.2013.08.018

Kaempchen N, Buehler M, Dietmayer K, 2005. Feature-level fusion for free-form object tracking using laserscanner and video. Proc IEEE Intelligent Vehicles Symp, p.453-458. https://doi.org/10.1109/IVS.2005.1505145

Karunasekera H, Zhang H, Xi T, et al., 2017. Stereo vision based negative obstacle detection. Proc 13th IEEE Int Conf on Control and Automation, p.834-838. https://doi.org/10.1109/ICCA.2017.8003168

Kenney JB, 2011. Dedicated short-range communications (DSRC) standards in the United States. *Proc IEEE*, 99(7):1162-1182. https://doi.org/10.1109/JPROC.2011.2132790

Khaleghi B, Khamis A, Karray FO, et al., 2013. Multisensor data fusion: a review of the state-of-the-art. *Inform Fus*, 14(1):28-44. https://doi.org/10.1016/j.inffus.2011.08.001

Kim J, Han DS, Senouci B, 2018. Radar and vision sensor fusion for object detection in autonomous vehicle surroundings. Proc IEEE 10th Int Conf on Ubiquitous and Future Networks, p.76-78. https://doi.org/10.1109/ICUFN.2018.8436959

Kim JH, Starr JW, Lattimer BY, 2015. Firefighting robot stereo infrared vision and radar sensor fusion for imaging through smoke. *Fire Technol*, 51(4):823-845. https://doi.org/10.1007/s10694-014-0413-6

Krafcik J, 2018. Where the Next 10 Million Miles Will Take Us at Waymo. https://santansun.com/2018/10/23/where-the-next-10-million-miles-will-take-us-at-waymo/ [Accessed on Jan. 15, 2020].

Ku J, Mozifian M, Lee J, et al., 2018. Joint 3D proposal generation and object detection from view aggregation. Proc IEEE/RSJ Int Conf on Intelligent Robots and Systems, p.1-8. https://doi.org/10.1109/IROS.2018.8594049

Labayrade R, Aubert D, Tarel JP, 2002. Real time obstacle detection in stereovision on non flat road geometry through "V-disparity" representation. Proc IEEE Intelligent Vehicle Symp, p.646-651. https://doi.org/10.1109/IVS.2002.1188024

Lahat D, Adali T, Jutten C, 2015. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proc IEEE*, 103(9):1449-1477. https://doi.org/10.1109/JPROC.2015.2460697

Larson J, Trivedi M, 2011. Lidar based off-road negative obstacle detection and analysis. Proc 14th Int IEEE Conf on Intelligent Transportation Systems, p.192-197. https://doi.org/10.1109/ITSC.2011.6083105

Levinson J, Thrun S, 2013. Automatic online calibration of cameras and lasers. Proc Robotics: Science and Systems, p.1-8. https://doi.org/10.15607/RSS.2013.IX.029

Li JQ, Deng GQ, Luo CW, et al., 2016. A hybrid path planning method in unmanned air/ground vehicle (UAV/UGV) cooperative systems. *IEEE Trans Veh Technol*, 65(12):9585-9596. https://doi.org/10.1109/TVT.2016.2623666

Li QQ, Chen L, Li M, et al., 2014. A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios. *IEEE Trans Veh Technol*, 63(2):540-555. https://doi.org/10.1109/TVT.2013.2281199

Li THS, Yeh YC, Wu JD, et al., 2010. Multifunctional intelligent autonomous parking controllers for carlike mobile robots. *IEEE Trans Ind Electron*, 57(5):1687-1700. https://doi.org/10.1109/TIE.2009.2033093

Lim K, Tuladhar KM, 2019. LIDAR: lidar information based dynamic V2V authentication for roadside infrastructure-less vehicular networks. Proc 16th IEEE Annual Consumer Communications and Networking Conf, p.1-6. https://doi.org/10.1109/CCNC.2019.8651684

Liu HP, Sun FC, 2012. Fusion tracking in color and infrared images using joint sparse representation. *Sci China Inform Sci*, 55(3):590-599. https://doi.org/10.1007/s11432-011-4536-9

Ma JY, Ma Y, Li C, 2019. Infrared and visible image fusion methods and applications: a survey. *Inform Fus*, 45(1):153-178. https://doi.org/10.1016/j.inffus.2018.02.004

Majumder S, Pratihar DK, 2018. Multi-sensors data fusion through fuzzy clustering and predictive tools. *Expert Syst Appl*, 107(04):165-172. https://doi.org/10.1016/j.eswa.2018.04.026

Matthies L, Rankin A, 2003. Negative obstacle detection by thermal signature. Proc IEEE/RSJ Int Conf on Intelligent Robots and Systems, p.906-913. https://doi.org/10.1109/IROS.2003.1250744

Matthies L, Kelly A, Litwin T, et al., 1995. Obstacle detection for unmanned ground vehicles: a progress report. Proc Intelligent Vehicles Symp, p.66-71. https://doi.org/10.1109/IVS.1995.528259

Matthies LH, Bellutta P, McHenry M, 2003. Detecting water hazards for autonomous off-road navigation. Unmanned Ground Vehicle Technology V, p.231-243. https://doi.org/10.1117/12.496942

Montemerlo M, Thrun S, 2006. Large-scale robotic 3-D mapping of urban structures. In: Ang MHJr, Khatib O (Eds.), Experimental Robotics IX. Springer, Berlin, p.141-150. https://doi.org/10.1007/11552246_14

Morton RD, Olson E, 2011. Positive and negative obstacle detection using the HLD classifier. Proc IEEE/RSJ Int Conf on Intelligent Robots and Systems, p.1579-1584. https://doi.org/10.1109/IROS.2011.6095142

Müller FDP, 2017. Survey on ranging sensors and cooperative techniques for relative positioning of vehicles. *Sensors*, 17(2):271. https://doi.org/10.3390/s17020271

Müller FDP, Diaz EM, Rashdan I, 2016. Cooperative positioning and radar sensor fusion for relative localization of vehicles. Proc IEEE Intelligent Vehicles Symp, p.1060-1065. https://doi.org/10.1109/IVS.2016.7535520

Napier A, Corke P, Newman P, 2013. Cross-calibration of push-broom 2D LIDARs and cameras in natural scenes. Proc IEEE Int Conf on Robotics and Automation, p.3679-3684. https://doi.org/10.1109/ICRA.2013.6631094

Nguyen H, Tran V, Nguyen T, et al., 2018. Apprenticeship bootstrapping via deep learning with a safety net for UAV-UGV interaction. https://arxiv.org/abs/1810.04344

Park Y, Yun S, Won CS, et al., 2014. Calibration between color camera and 3D LIDAR instruments with a polygonal planar board. *Sensors*, 14(3):5333-5353. https://doi.org/10.3390/s140305333

Pfeiffer D, Franke U, 2011. Modeling dynamic 3D environments by means of the stixel world. *IEEE Intell Trans Syst Mag*, 3(3):24-36. https://doi.org/10.1109/MITS.2011.942207

Pinchon N, Cassignol O, Nicolas A, et al., 2018. All-weather vision for automotive safety: which spectral band? In: Dubbert J, Müller B, Meyer G (Eds.), Advanced Microsystems for Automotive Applications 2018. Springer, Cham, p.3-15. https://doi.org/10.1007/978-3-319-99762-9_1

Premebida C, Nunes U, 2013. Fusing lidar, camera and semantic information: a context-based approach for pedestrian detection. *Int J Robot Res*, 32(3):371-384. https://doi.org/10.1177/0278364912470012

Radecki P, Campbell M, Matzen K, 2016. All weather perception: joint data association, tracking, and classification for autonomous ground vehicles. https://arxiv.org/abs/1605.02196

Raya M, Hubaux JP, 2007. Securing vehicular ad hoc networks. *J Comput Secur*, 15(1):39-68. https://doi.org/10.3233/JCS-2007-15103

Rosique F, Navarro PJ, Fernández C, et al., 2019. A systematic review of perception system and simulators for autonomous vehicles research. *Sensors*, 19(3):648. https://doi.org/10.3390/s19030648

Rubaiyat AHM, Fallah Y, Li X, et al., 2018. Multi-sensor data fusion for vehicle detection in autonomous vehicle applications. *Electron Image*, 2018(17):257-1-257-6. https://doi.org/10.2352/ISSN.2470-1173.2018.17.AVM-257

Sarwal A, Nett J, Simon D, 2004. Detection of Small Water-bodies. PercepTek Robotics 12395. Perceptek Inc., Littleton Co., MA, USA.

Schneider S, Himmelsbach M, Luettel T, et al., 2010. Fusing vision and LIDAR—synchronization, correction and occlusion reasoning. Proc IEEE Intelligent Vehicles Symp, p.388-393. https://doi.org/10.1109/IVS.2010.5548079

Seitz SM, Curless B, Diebel J, et al., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. IEEE Computer Society Conf on Computer Vision and Pattern Recognition, p.519-528. https://doi.org/10.1109/CVPR.2006.19

Seraji H, 2003. New traversability indices and traversability grid for integrated sensor/map-based navigation. *J Robot Syst*, 20(3):121-134. https://doi.org/10.1002/rob.10074

Shinzato PY, Wolf DF, Stiller C, 2014. Road terrain detection: avoiding common obstacle detection assumptions using sensor fusion. IEEE Intelligent Vehicles Symp, p.687-692. https://doi.org/10.1109/IVS.2014.6856454

Shirkhodaie A, Amrani R, Tunstel E, 2005. Soft computing for visual terrain perception and traversability assessment by planetary robotic systems. IEEE Int Conf on Systems, Man and Cybernetics, p.1848-1855. https://doi.org/10.1109/ICSMC.2005.1571416

Siritanawan P, Prasanjith MD, Wang DW, 2017. 3D feature points detection on sparse and non-uniform pointcloud for SLAM. 18th Int Conf on Advanced Robotics, p.112-117. https://doi.org/10.1109/ICAR.2017.8023504

Sock J, Kim J, Min JH, et al., 2016. Probabilistic traversability map generation using 3D-LIDAR and camera. Proc IEEE Int Conf on Robotics and Automation, p.5631-5637. https://doi.org/10.1109/ICRA.2016.7487782

Son J, Yoo H, Kim S, et al., 2015. Real-time illumination invariant lane detection for lane departure warning system. *Expert Syst Appl*, 42(4):1816-1824. https://doi.org/10.1016/j.eswa.2014.10.024

Starr JW, Lattimer B, 2017. Evidential sensor fusion of long-wavelength infrared stereo vision and 3D-LIDAR for rangefinding in fire environments. *Fire Technol*, 53(6):1961-1983. https://doi.org/10.1007/s10694-017-0666-y

Su H, Maji S, Kalogerakis E, et al., 2015. Multi-view convolutional neural networks for 3D shape recognition. Proc IEEE Int Conf on Computer Vision, p.945-953. https://doi.org/10.1109/ICCV.2015.114

Subramanian V, Burks TF, Dixon WE, 2009. Sensor fusion using fuzzy logic enhanced Kalman filter for autonomous vehicle guidance in citrus groves. *Trans ASABE*, 52(5):1411-1422. https://doi.org/10.13031/2013.29121

Thrun S, Montemerlo M, Dahlkamp H, et al., 2006. Stanley: the robot that won the DARPA grand challenge. *J Field Robot*, 23(9):661-692. https://doi.org/10.1002/rob.20147

Tian Z, Cai Y, Huang S, et al., 2018. Vehicle tracking system for intelligent and connected vehicle based on radar and V2V fusion. Proc Chinese Control and Decision Conf, p.6598-6603. https://doi.org/10.1109/CCDC.2018.8408291

van Brummelen J, O'Brien M, Gruyer D, et al., 2018. Autonomous vehicle perception: the technology of today and tomorrow. *Trans Res Part C*, 89(3):384-406. https://doi.org/10.1016/j.trc.2018.02.012

Vu TD, 2009. Vehicle Perception: Localization, Mapping with Detection, Classification and Tracking of Moving Objects. PhD Thesis, Institut National Polytechnique de Grenoble, Grenoble, France.

Wang J, Song Q, Jiang Z, et al., 2016. A novel InSAR based off-road positive and negative obstacle detection technique for unmanned ground vehicle. Proc IEEE Int Geoscience and Remote Sensing Symp, p.1174-1177. https://doi.org/10.1109/IGARSS.2016.7729297

Wang JG, Chen SJ, Zhou LB, et al., 2018. Vehicle detection and width estimation in rain by fusing radar and vision. 15th Int Conf on Control, Automation, Robotics and Vision, p.1063-1068. https://doi.org/10.1109/ICARCV.2018.8581246

Wei P, Cagle L, Reza T, et al., 2018. Lidar and camera detection fusion in a real-time industrial multi-sensor collision avoidance system. *Electronics*, 7(6):84. https://doi.org/10.3390/electronics7060084

Xiao L, Wang R, Dai B, et al., 2018. Hybrid conditional random field based camera-lidar fusion for road detection. *Inform Sci*, 432(03):543-558. https://doi.org/10.1016/j.ins.2017.04.048

Yamada M, Ueda K, Horiba I, et al., 2005. Detection of wet-road conditions from images captured by a vehicle-mounted camera. *J Robot Mech*, 17(3):269-276. https://doi.org/10.20965/jrm.2005.p0269

Yoo HW, Druml N, Brunner D, et al., 2018. MEMS-based lidar for autonomous driving. *Elektrotech Inftech*, 135(6):408-415. https://doi.org/10.1007/s00502-018-0635-2

Yue YF, Wang DW, Senarathne PGCN, et al., 2017. Robust submap-based probabilistic inconsistency detection for multi-robot mapping. Proc European Conf on Mobile Robots, p.1-6. https://doi.org/10.1109/ECMR.2017.8098660

Yue YF, Senarathne PGCN, Yang C, et al., 2018. Hierarchical probabilistic fusion framework for matching and merging of 3-D occupancy maps. *IEEE Sens J*, 18(21):8933-8949. https://doi.org/10.1109/JSEN.2018.2867854

Zhan YM, Leung H, Kwak KC, et al., 2009. Automated speaker recognition for home service robots using genetic algorithm and Dempster-Shafer fusion technique. *IEEE Trans Instrum Meas*, 58(9):3058-3068. https://doi.org/10.1109/TIM.2009.2016870

Zhang J, Siritanawan P, Yue Y, et al., 2018. A two-step method for extrinsic calibration between a sparse 3D LiDAR and a thermal camera. 15th Int Conf on Control, Automation, Robotics and Vision, p.1039-1044. https://doi.org/10.1109/ICARCV.2018.8581170

Zhao GQ, Xiao XH, Yuan JS, et al., 2014. Fusion of 3D-LIDAR and camera data for scene parsing. *J Visual Commun Imag Represent*, 25(1):165-183. https://doi.org/10.1016/j.jvcir.2013.06.008

Zhou Y, Tuzel O, 2018. VoxelNet: end-to-end learning for point cloud based 3D object detection. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.4490-4499. https://doi.org/10.1109/CVPR.2018.00472

Zhu H, Yuen KV, Mihaylova L, et al., 2017. Overview of environment perception for intelligent vehicles. *IEEE Trans Intell Trans Syst*, 18(10):2584-2601. https://doi.org/10.1109/TITS.2017.2658662