2022 7th International Conference on Communication, Image and Signal Processing (CCISP) | 978-1-6654-5959-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/CCISP55629.2022.9974277

Environment Perception Technology for Intelligent Robots in Complex Environments: A Review

Jiajun Wu
[1]*State Key Laboratory of Precision Blasting*
[2]*School of Smart Manufacturing*
Jianghan University
Wuhan, China
297788851@qq.com

Jun Gao*
[1]*State Key Laboratory of Precision Blasting*
[2]*School of Smart Manufacturing*
Jianghan University
Wuhan, China
gaojun407104739@163.com

Jiangang Yi*
[1]*State Key Laboratory of Precision Blasting*
[2]*School of Smart Manufacturing*
Jianghan University
Wuhan, China
yjg_wh@yeah.net

Peng Liu
[1]*State Key Laboratory of Precision Blasting*
[2]*School of Smart Manufacturing*
Jianghan University
Wuhan, China

Changsong Xu
[1]*State Key Laboratory of Precision Blasting*
[2]*School of Smart Manufacturing*
Jianghan University
Wuhan, China

***Abstract*—Environmental perception is a necessary prerequisite for intelligent robots to perform specified tasks, and is the basis for subsequent control and decision-making. In recent years, with the rapid development of deep learning technology and the dramatic improvement of hardware performance, vision-based environmental perception technologies, such as target recognition and target detection, have made significant progress. However, most vision algorithms are developed based on images with stable lighting conditions and no significant disturbances. In fact, robots often need to operate in unstructured, complex conditions or visually degraded environments. Visual perception alone cannot meet the job requirements and it lacks the ability to adapt to the environment. Therefore, the environment perception technology based on multi-sensor fusion has become a popular research direction. In this paper, we first analyze the characteristics of sensors required for perception, and briefly review the uni-modal sensor application status in complex environments such as mines, railways, highways, tunnels, *etc*. Secondly, we introduce the datasets and sensor fusion methods for robotics perception. Thirdly, we provide an overview of the multi-modal perception technology applied on intelligent robot. Finally, we summarize the challenges and future development trends in this direction.**

Keywords—environmental perception, multi-sensor fusion, LiDAR, millimeter wave radar

I. INTRODUCTION

From the initial simple remote-controlled industrial robot to the advanced automation equipment integrating automatic control, mechanical combination, multi-sensor fusion, artificial intelligence and other high-tech cutting-edge technologies [1], the technology and application of robots are constantly improving and perfecting. An intelligent robot is a comprehensive system equipment integrating multiple functions such as environmental perception, intelligent decision-making, and autonomous execution. The rapid development of artificial intelligence and information technology in recent years has laid a solid foundation for the intelligence of robots. Intelligent robots are not only widely used in industries such as industry, agriculture, medical care, and services, but also gradually applied in harmful and dangerous environments such as safety protection, national defense, and environmental exploration [2]. In industry, tunnels, mines and other related areas are all complex environments. Due to rock blasting, tunnels or mines frequently generate a significant amount of rockfall, dust, and smoke during mining and drilling. Therefore, in the harsh and complex environmental phenomena such as the humidity of the layer, the work efficiency and work safety of the entire project face serious challenges [3]. In addition, when the operating environment changes or the complexity of the operation increases, compared with the ordinary manual mechanical excavation method, the use of intelligent robots to complete the above specified tasks has significant advantages of avoiding casualties, improving work efficiency, and maintaining construction effects. Environmental perception is the premise for intelligent robots to complete target tasks in complex environments such as tunnels and mines. Visual perception such as object detection, object tracking, and pose estimation has been widely used for robotic perception system [4]. However, vision-based environment perception has high requirements on the lighting conditions of the environment, and there are still great limitations in using such a single modality for perception. Therefore, the use of different sensors for functional complementary fusion becomes the main research direction in recent years [5]. A typical robotic perception system is shown in Fig. 1.

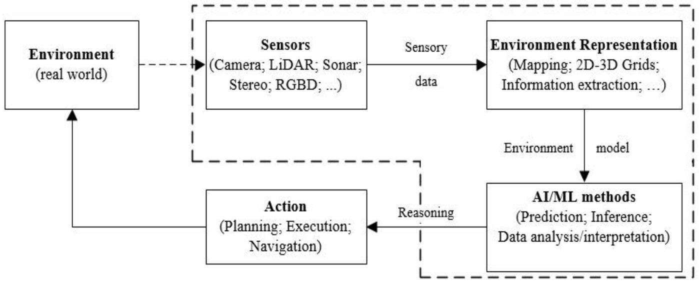
Environment
(real world)
Sensors
(Camera; LiDAR; Sonar; Stereo; RGBD; ...)
Sensory
data
Environment Representation
(Mapping; 2D-3D Grids; Information extraction; ...)
Environment
model
AI/ML methods
(Prediction; Inference; Data analysis/interpretation)
Reasoning
Action
(Planning; Execution; Navigation)

Fig. 1. Key modules of a typical robotic perception system

This research is supported by State Key Laboratory of Precision Blasting, Jianghan University (No. PBSKL2022302).

978-1-6654-5959-4/22/$31.00 ©2022 IEEE

TABLE I. COMPARISON OF ADVANTAGES AND DISADVANTAGES OF DIFFERENT SENSORS

Sensor	*Advantage*	*Disadvantage*	*Cost*	*Application*
LiDAR	High measurement accuracy Less affected by light Quick response	High cost Large rain fog dust environment signal attenuation	High	Obstacle ranging Robot positioning
Millimeter wave radar	Small size and light weight Strong ability to penetrate rain fog dust	Compared with other radars, the accuracy is lower [6] Difficulty recognizing static objects [7]	Medium	Obstacle detection Robot positioning
Camera	High-resolution easy data processing	High requirements for light conditions susceptible to fog	Low	Target detection and recognition

II. SENSORS FOR ENVIRONMENT PERCEPTION

When encountering a complex operating environment, intelligent robots need a large number of sensors to collect surrounding environment information for the system to comprehensively analyze and make action decisions. In different environment, different sensors have their own unique advantages. Table I summarizes the advantages and disadvantages of different sensors. At present, the most widely used at domestic and foreign are generally classified into three categories, *i.e.*, LiDAR, Millimeter wave radar and Visual camera.

The position estimation method based on Laser Detection and Ranging (LiDAR) system completely relies on one or more LiDAR sensors. LiDAR has the advantages of accurate measurement, high accuracy of distance measurement, and strong ability to obtain three-dimensional information. However, although manufacturers in the industry have been working to reduce the cost of LiDAR production in recent years, they still have higher costs compared to millimeter-wave radars and cameras. In a typical position estimation method based on the combination of LiDAR and camera, the primary application of LiDAR is the creation of a map of the surroundings of robot, while the camera sensor is used to estimate the position of robot in the map, thereby reducing costs. However, due to blasting during mining or during tunnel rock drilling, most working environments are dark and accompanied by a lot of dust and smoke, location estimation based solely on cameras is usually inaccurate and unreliable [8].

A. Applications of LiDAR

The working principle of LiDAR is to release multiple laser beams, receive the reflected signal of the object, calculate the distance between the sensor transmitter and the target, and obtain and analyze the reflected energy, spectral amplitude, spectral frequency and other information on the surface of the object. LiDAR is mostly used for autonomous navigation and positioning of intelligent unmanned platforms, and the most widely used is LiDAR-based Simultaneous Localization and Mapping (SLAM) [9]. Among the localization methods with LiDAR as the main detection sensor, the method of fusion with other auxiliary sensors such as inertial measurement unit (IMU) has the highest accuracy [10]. Based on the advantages of less data volume and high ranging accuracy, 2D LiDAR is the preferred radar for indoor environment navigation and positioning related research. However, with the increasing demand for complex environmental scenarios, the application scenarios of intelligent robots are constantly expanding. Thanks to the DARPA Underground Challenge in the United States, 3D LiDAR technology has developed rapidly and can be widely applied to outdoor scenes. 3D LiDAR can provide point clouds that are stronger than 2D LiDAR with richer 3D data and better inter-frame matching methods, which has the characteristics of high ranging accuracy, little influence by light, and anti-electromagnetic interference. For example, He et al. [11] fused LiDAR with visual inertial odometry to overcome the limitations of a single sensor and to enhance the precision of the positioning. 3D LiDAR has become core sensor in the field of environmental perception, but the drawbacks of high production cost and large signal attenuation in dense smoke and fog environment are the main factors restricting its large-scale deployment in practical applications.

B. Applications of millimeter wave radar

Radars operating in the millimeter wave band are called millimeter wave radars and its principle is similar to LiDAR. But unlike LiDAR, millimeter-wave radar uses much longer radiation wavelengths than LiDAR or cameras, Therefore, the ability of millimeter wave radar to penetrate smoke, fog and dust is stronger than other radars, and it has the characteristics of all-weather, anti-jamming and anti-stealth. In addition, the millimeter-wave radar antenna is small in size and light in weight, which is easy to carry on intelligent unmanned platforms. Therefore millimeter-wave radars are the first choice in complex environments [12]. In a smoke-filled environment, Muhammad et al. [13] conducted a comparative experiment on the perception of surrounding objects between a robot equipped with millimeter-wave radar and a robot equipped with optical sensors. The experimental results confirm that the millimeter-wave radar can detect the surrounding environment targets, while the optical sensors do not. Millimeter-wave radar can scan the precise distance of multiple targets, and can be fused with IMU sensors to obtain more accurate data, which is widely used in target detection [14] and mobile robot obstacle avoidance [15]. With the development of millimeter-wave radar in recent years, there are three main frequency bands in different countries - 24 GHz, 60 GHz and 77 GHz. Europe and the United States chose to focus on 77 GHz research, while Japan chose the 60 GHz frequency band. The wavelength of the 77GHz radar is less than one-third of the 24GHz radar, so the area of the transceiver antenna is greatly reduced, and the size of the entire radar is effectively reduced, which is very beneficial for the pursuit of miniaturization. Recently, 4D millimeter-wave imaging radar has caught the attention of researchers. With the advantages of height detection, ultra-high sensitivity and high resolution, 4D millimeter-wave imaging radar may replace LiDAR as the core sensor for environmental perception in the future.

C. Applications of camera

Vision sensors are the direct source of information for the entire machine vision system. The current mainstream vision systems include monocular vision, binocular vision, multi vision, and panoramic vision. With low cost, cameras are the basic sensors of robot vision systems, and are currently the main choice in industry and academia. Cameras can capture the color, texture, shape and other information of targets in the surrounding environment, and can identify different targets in non-extreme environments, so they are mostly used for tasks such as target detection, tracking, and identification. Compared with millimeter wave radar and LiDAR, Cameras have the advantages of dense data and high resolution. However, its shortcomings are also obvious: camera is easily affected by complex environments, the perception performance is greatly reduced in dust, smoke and other environments, and it is also very sensitive to scenes with sudden changes in light, such as the dust environment and shadows caused by rock blasting. In complex environments with poor light source conditions or heavy dust and smoke, autonomous mobile robots usually use sensors such as millimeter-wave radar, visual cameras, and inertial sensors for environmental perception [16]. But in some cases, the environment in the tunnel will change after manual intervention, such as actively providing lighting to achieve good lighting conditions. At this time, the fusion of vision sensors and radar will have a more accurate effect.

The use of multiple sensors in combination can complement the shortcomings of each sensor, overcome the functional limitations of a single type of sensor, improve redundancy in detection accuracy, and avoid detection failure caused by false detection of a sensor or even work stoppage. The robustness and detection accuracy of the perception system can be enhanced using this method [17].

III. Intelligent Robotic Perception

In perception models of intelligent robot, visual perception is the most widely used technology. For example, Ryoma et al. [18] combined a deep Q-network (DQN) with a curriculum learning framework to allow robots to avoid walls, obstacles, and other robots through images captured by cameras.

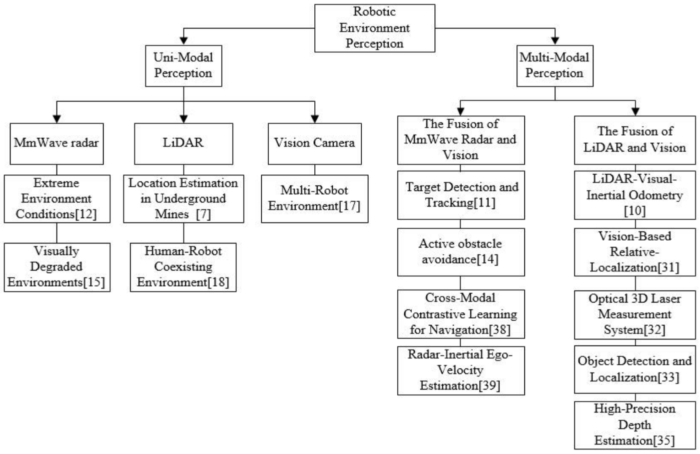
Robotic Environment Perception
Uni-Modal Perception
Multi-Modal Perception
MmWave radar
LiDAR
Vision Camera
The Fusion of MmWave Radar and Vision
The Fusion of LiDAR and Vision
Extreme Environment Conditions[12]
Location Estimation in Underground Mines [7]
Multi-Robot Environment[17]
Target Detection and Tracking[11]
LiDAR-Visual-Inertial Odometry [10]
Visually Degraded Environments[15]
Human-Robot Coexisting Environment[18]
Active obstacle avoidance[14]
Vision-Based Relative-Localization[31]
Cross-Modal Contrastive Learning for Navigation[38]
Optical 3D Laser Measurement System[32]
Radar-Inertial Ego-Velocity Estimation[39]
Object Detection and Localization[33]
High-Precision Depth Estimation[35]

Fig. 2. Robotic environment perception

In the human-robot coexistence environment, Pritam et al. [19] proposed a new variant of the CFAST matching algorithm, which obtains the position of target person by employing a visual sensor to capture the position of human shoes in consecutive frames of the camera in real time. However, in a complex and changeable environment, environmental perception based on a single sensor will fail to meet the requirements because the method is too simple [20]. Therefore, multi-sensor fusion is a necessary condition to maintain robustness in robotic environment perception tasks. Fig. 2 illustrates an overview of the perception methods for intelligent robot.

A. Datasets

In order to save costs and speed up the progress of experiments, researchers often use public domain datasets to train and verify the algorithms constructed in advance. High-quality datasets can train efficient detection algorithms, and some datasets do not target algorithms with specific needs, and provide a fair algorithm evaluation platform and benchmark, which is beneficial for researchers to compare and optimize their own models. Table II presents a brief description and comparison of several well-known datasets in different environments, such as Marulan [21], NCLT [22], ColoRadar [23], Chilean Underground Mine [24], and JRDB [25].

1) The Marulan multi-sensor perception dataset is a large, precisely calibrated and time-synchronized dataset collected from a multi-sensor unmanned ground vehicle (UGV), and the dataset contains image and radar point cloud data. The unmanned ground vehicle includes multiple laser scanners, a millimeter-wave radar scanner, a color camera, and an infrared camera, as well as a centimeter-accurate dGPS/INS system for localization. Environmental conditions of dust, smoke, and rain were artificially created during data collection, and some of the datasets were collected under night conditions.

2) The NCLT (The University of Michigan North Campus Long-Term Vision and LIDAR Dataset) is a large-scale, long-term autonomous robotics research dataset collected at the University of Michigan, Ann Arbor campus. The dataset includes omnidirectional images, 3D laser radar point cloud data, planar LiDAR point cloud data, GPS and related data of the Segway mobile robot proprioceptive sensor that collected the data. The data collection time span was long, a total of 15 months, approximately every two weeks. The same location, different weather, different time, and different seasons allow the dataset to capture a variety of challenging elements, including moving obstacles (such as pedestrians, bicycles, and cars), different light intensities, different perspectives, seasons and weather changes, and long-term structural changes brought on by construction projects. The NCLT dataset includes 147.4-kilometer robot trajectories and 34.9 hours of logs.

3) ColoRadar (Colorado mm-Wave Radar) dataset, a millimeter-wave radar robot perception dataset published by Kramer et al., including two different forms of dense, high-resolution radar data from two FMCW radar sensors, and a sparse radar point cloud produced by one of the radar sensors. Additionally, 3D LiDAR and inertial measurements are included in all datasets.

TABLE II. BRIEF DESCRIPTION AND COMPARISON OF DATASETS IN DIFFERENT ENVIRONMENTS

Datasets	*Marulan*	*NCLT*	*ColoRadar*	*Chilean Underground Mine*	*JRDB*
Scene	Outdoor	Campus area indoor, outdoor	Indoor, outdoor, creek path, mine	Underground mine	Human-machine coexistence
Weather scene	Day and night including dust, smoke or rain	Day, night, sunny day, four seasons	Sunny day, night	None	Day and sunny day
Sensors used	LiDAR Millimeter wave radar Color camera Infrared camera GPS	3D LiDAR Planar LiDAR Omnidirectional camera GPS	Millimeter wave radar 3D LiDAR IMU	Stereo camera 3D LiDAR 2D FM CW Radar	LiDAR 360° camera
Provided data	Images and radar point cloud data	Hundreds of gigabytes of radar data and tens of thousands of images	Dense, high-resolution radar data in two different forms	Images and radar point cloud data	RGB video at 15 frames per second laser point cloud
Application scenarios	2D/3D terrain representation Obstacle detection	Long-term SLAM Obstacle detection and tracking Repeat navigation	SLAM Mapping Self-positioning	SLAM	Autonomous robot navigation in human environments
Features	Dataset collected under human-controlled conditions (dust, smoke, rain)	The same environmental comparison data under long-term seasonal changes and visual comparison of different lighting conditions	Datasets collected in highly diverse 3D environments	Mine tunnel environment with variable lighting conditions	Extensive 2D/3D bounding box annotations

The datasets were collected in mines, built environments, and urban creek paths, and consisted of 52 sequences, including a total of over 145 minutes of 3D LiDAR, IMU, and 3D FMCW radar data. This dataset is helpful for the research of tasks such as automatic robot positioning and SLAM in complex environments.

4) Chilean Underground Mine Chilean underground mine dataset is the world's largest robotic dataset of underground copper mines, which is collected by a robot platform equipped with a stereo camera, 3D LiDAR, and 2D frequency-modulated continuous wave radar that traverses part of the 2-kilometer mine. It mainly contains stereo image data and high-resolution 3D LiDAR point cloud data. This dataset can be used for the initial development and evaluation of mine inspection robots, as well as for intelligent mining and material extraction. This dataset is beneficial for validating and benchmarking sensor fusion and SLAM systems in real-world environments with varying lighting and surface properties from a more general robotic navigation and mapping perspective.

5) The JRDB dataset, an egocentric dataset collected from the social mobile manipulation robot JackRabbot, provides benchmarking and training data for research on autonomous robot navigation and all social robot perception tasks in human environments. The dataset includes stereo cylindrical 360° RGB video at 15 frames per second, 64 minutes of annotated multimodal sensor data, 3D LiDAR point cloud, RGB-D video at 30 frames per second, and spherical images from a 360° fisheye camera. The dataset is annotated with over 2.4 million body bounding boxes and 600,000 body pose annotations and includes both indoor and outdoor environments.

The above datasets all contain standards for their respective applications, and due to the constantly changing work environment, intelligent robots need to adapt their existing knowledge such as feature from source domain to the target one. This kind of adaptation can be called domain adaptation. Domain adaptation methods for robotic perception enable robots to understand knowledge from public domain datasets and transfer them to the current work environment to complete specified tasks [26].

B. Sensor fusion methods for robotics perception

Sensor fusion in the robot environment perception module has become standard, and the most common applications are vision cameras, LiDAR, IMU and their fusion applications [27]. For the robot environment perception system, it is always faced with diverse and complex environmental information. Robustness and parallel processing capability are the most basic requirements of environment perception for information fusion algorithms. Different sensors have different working principles and collected data, with different adaptability to the environment. The perception method based on multi-sensor fusion can combine the advantages of various sensors, break the inherent limitations of a single sensor, and provide more accurate and reliable information for subsequent robot control. Multi-sensor fusion algorithms can be roughly divided into the following types:

1) The weighted average method is the simplest and most intuitive method. The principle of the method is to perform a weighted average of the data collected by each sensor according to the artificially specified weighting rules, and the final result is used as a fusion value. The weighted average method directly processes the data output by the sensor. But the weighting rules are greatly affected by the subjectivity, and the final fusion effect also has a corresponding change.

2) Kalman filter method. It is mostly used to combine low-level, real-time, dynamic, redundant data from multiple sensors. By measuring the historical state and current state of target, this method can estimate the current state of target [28]. In multi-sensor fusion, Kalman filter is often used as the basic algorithm. Although the Kalman filter has been around for more than half a century, the corresponding Kalman filter variant has proven to be one of the most effective multi-sensor fusion algorithms. The extended Kalman filter algorithm is a Kalman filter method that deals with nonlinear relationships.

For example, Wang et al. [29] used the extended Kalman filter method to fuse the encoder odometer speed, monocular vision poses, and inertial sensing unit information for robot positioning, expanding the application scenarios of monocular vision.

3) The Bayesian estimation method, based on the prior probability, combines new data information with prior information to generate a new probability, and performs multi-sensor fusion sensing tasks in a loop. The essence of Bayesian estimation is to obtain the optimal estimate of the parameter θ through Bayesian decision making to minimize the total expected risk. Dong et al. [30] proposed a variational Bayesian Cubature Information filter (VBCIF-QR) for nonlinear multi-sensor system with uncertain noise statistics, which facilitates multi-sensor information fusion. Experimental results show that VBCIF-QR algorithms perform better than conventional cubature Kalman filters.

4) DS (Dempster-Shafer) evidence theory, which is an extension of Bayesian inference method, does not need to know the prior probability of events, and introduces the concept of reliability function. In human–robot collaborations (HRC) environment, Li et al. [31] employed Dempster–Shafer evidence theory to fuse the results of the best three standing-posture recognition algorithms. The experimental results show that the D-S fusion algorithm can improve the recognition accuracy.

5) Fuzzy logic reasoning. Fuzzy logic is multi-valued logic. It expresses the degree of authenticity by specifying a real number between 0 and 1, which is equivalent to the premise of the implicit operator, allowing the uncertainty in the process of information fusion of multiple sensors to be directly expressed in the inference process. Consistent fuzzy inference can be produced if some systematic approach is adopted to model the inference of the uncertainty in the fusion process. Cui [32] designed a mobile obstacle avoidance wheeled robot, which uses fuzzy logic model to navigate in static environment, and the simulation result proves the effectiveness of the algorithm.

6) Artificial neural network, created by simulating the structure and neural signals of the human brain, can simulate complex nonlinear mappings. Compared with other algorithms, deep neural networks (DNNs) have been widely used recently in robot environment perception and improved perception capabilities [26], which can learn features that are difficult to extract by general mathematical methods. However, deep neural network models based on supervised learning require a large number of labeled data to train, while the labeling work is manpower-costly and time consuming.

For complex environments such as tunnels and mines that are being excavated and constructed, there are usually many operators and related excavation equipment, which brings great challenges to the perception of autonomous mobile robots. The most important of which is to solve the following problems: how to perceive the surrounding environment, how to move to the target position and how to actively avoid obstacles. Precise positioning of mobile robots in the current environment is the key to solving these issues and is the most important part of the robot control system and the basic premise for completing the task.

C. Multi-modal perception applied on intelligent robot

Different types of sensors have its unique methods and principles to obtain information. Comprehensive processing of these information can make up for the insufficiency of a single sensor and meet the requirements of initial environmental perception. At present, the typical multi-modal perception methods for intelligent robot are listed below.

1) Fusion perception based on vision and LiDAR

Over the past few decades, SLAM techniques for multi-sensor fusion have made further progress. SLAM systems based on various sensors have been developed and applied in various fields. The feature-based fusion SLAM framework was established as early as 1990 [33]. The primary goal of SLAM is to resolve the issue of positioning and map construction of robots when working in complex environments, which is one of the mainstream research methods of current mobile robot positioning technology. Efficient and accurate positioning is the premise of control systems and path planning, especially in tunnels, mines, underground and other GPS-rejected environments, where real-time robot positioning is required to build incremental maps. According to the different environmental perception sensors of mobile robots, SLAM technology is mainly divided into two categories: visual SLAM and LiDAR SLAM [34]. LiDAR is the most commonly used SLAM sensor in structured scenes, and the SLAM algorithm framework and corresponding theories have been studied in depth [35]. However, due to the sparseness of the information provided by LiDAR, it brings a large error to the positioning effect of the robot in the unstructured environment [9], so pure LiDAR SLAM is not suitable for complex environments such as outdoor.

In complex scenes, a complete real-time positioning requirement cannot be achieved using a single vision or LiDAR sensor. The robustness and positioning accuracy of mobile robot can be significantly enhanced by using multi-sensor fusion SLAM technology. Song et al. [36] proposed a target relative localization method based on fusion of RGB-D depth camera and 2D LiDAR, introduced vision and depth tracking methods, and proposed an improved track fusion scheme to achieve robust two-dimensional localization. Básaca-Preciado et al. [37] combined 3D laser and camera vision to generate high-precision localization schemes through dynamic triangulation. Yang et al. [38] proposed a vision-and-LiDAR-based target detection and localization strategy to solve the problem of lack of object information in the process of changing the perspective of mobile robots. Sun et al. [39] proposed a structured environment corner feature extraction method based on sensor fusion of monocular vision and laser ranging data. This method employs a new expression of data-associative Extended Kalman Filter (EKF), which improves the accuracy of SLAM. High-accuracy SLAM is based on an optimal fusion of LiDAR and camera, so it is necessary to ensure precise calibration between these two types of sensors. In recent years, with the deepening of deep convolutional neural networks (CNN) in robotics applications, Park et al. [40]

proposed a CNN-based calibration method, which takes the LiDAR and camera differences as input and returns the calibration parameters. For real-time applications, this method provides a quick in-line calibration solution.

2) Fusion perception based on vision and millimeter wave radar

Millimeter-wave radar has been widely used in the automotive industry and is also common in advanced driver assistance system (ADAS) [41]. But there are few methods for robot state estimation using radar. Different from the environment perception methods based on vision and LiDAR, millimeter-wave radar has good robustness in complex environments. Cai et al. [42] exploited the properties of low-cost monolithic millimeter-wave radar to measure radial velocity through Doppler effect and fine-grained radar cross section (RCS), and proposed a position recognition framework called AutoPlace. Experiment on nuScenes [43] proved the robustness and achieved significant performance gain. Chen [12] designed a spherical robot equipped with millimeter-wave radar and vision sensors and proposed a fusion scheme of radar detection instead of Region Candidate Network (RPN). In addition, Chen performed feature-level fusion of vision and MmWave radar and used the extended Kalman filter algorithm to achieve object detection and tracking. Huang et al. [44] proposed a cross-modal contrastive learning representation method using simulated and real data, enabling robots to navigate autonomously in smoke-filled environments and comparable results to live LiDAR methods in smoke-free environments. Qian et al. [15] used a millimeter-wave radar active obstacle avoidance method based on binocular vision to improve the active obstacle avoidance and adaptive control capabilities of millimeter-wave radar. Wang et al. [14] proposed a robust target detection and classification algorithm based on the fusion of millimeter-wave radar and camera. The approximate location of the target was detected by fused sensors, and the corresponding region of interest (ROI) was accurately calculated. The joint classification network extracts the micro-Doppler features (from texture of ROI images) and the time-spectrum features. The proposed fusion network model (called RCF Faster RCNN) has been tested to prove the high accuracy and robustness of the model. In the visual degradation environment, Kramer et al. [45] proposed a method for estimating the speed of a mobile robot, which integrates the system-on-chip (SoC) millimeter-wave radar with an IMU. The experiments show that the effect is comparable to the visual inertial measurement method with good visual conditions.

To sum up, multi-modal sensors fusion is a robust solution in perceiving the extreme or unstructured environments.

IV. Conclusion

This paper summarizes the multi-modal perception technology for intelligent robots in complex environments, and concludes that multi-sensor fusion is not a single technology, but an interdisciplinary comprehensive theory and method, which is also a complex, comprehensive, and still evolving research field.

1) Thanks to the market demand for unmanned driving and the rapid progress of corresponding software and hardware technologies, radar technology is constantly being improved and upgraded, but most commercial radars are only suitable for unmanned vehicles. Perception radars for autonomous mobile robots performing specified tasks are not common. Therefore, in terms of hardware, the research and development of sensors should be the focus. How to make sensors more universal and how to better overcome the influence of different environmental factors are the key issues to be solved.

2) In terms of data collection, the research on intelligent robots applied to complex environments such as tunnels and mines is basically carried out in their own specific projects, and multi-sensor data collection is mainly concentrated on visual sensors and LiDAR. For example, the rock datasets of tunnel driving face and the datasets of mine ore are all non-public. The improvement and promotion of the related work of the transportation department and the construction department will provide a better environment for open source data collection. Since the public domain datasets are few, and the quality, quantity, types and collection scenarios of each dataset are different, especially in complex environments, which makes it difficult to formulate a standard for measuring the pros and cons of different fusion algorithms.

3) The performance evaluation standards of multi-sensor fusion algorithms are different. Different fusion algorithms are generally based on different datasets, and most fusion algorithms are not open source, which makes it hard to come up with a single theory of fusion and an effective generalized fusion model and algorithm. Establishing a unified multi-sensor fusion algorithm performance evaluation standard is one of the main development trends in the future.

4) In complex environments, the upper limit of computing power of the computing core unit carried by intelligent robots is low, and there are many designated computing tasks, such as robot positioning in environmental perception, path planning and movement, and designated target recognition and detection tasks. In practical applications, these tasks often need to be performed synchronously, but the computing power of the computing unit cannot satisfy the simultaneous operation of multiple sensing algorithms. Therefore, designing a multi-task learning algorithm or a model compression method would be a good choice to solve the problem of limited computing resources of the device.

References

[1] Gao C. Research on Present Situation and Development Trend of Industrial Robot in China. Modern Information Technology, 2019.

[2] Premebida C, Ambrus R, Marton Z C. Intelligent robotic perception systems. Applications of Mobile Robots. London, UK: IntechOpen, 2018.

[3] Xiang Zhou, Li Shenggen, Xiao Zhenghang, Cui Chang. Study on Unmanned Driving System in Tunnel Construction Scenario. Construction Mechanization, 2022,43(05):15-18.

[4] Sun Lining, Xu Hui, Wang Zhenhua, Review on Key Common Technologies for Intelligent Applications of Industrial Robots. Journal of Vibration, Measurement & Diagnosis, 2021, 41(02): 211-219+406.

[5] Nao Jiti. Robotic Perception Completion Plan. Big Data Era, 2020(10):58-63.

[6] Alencar F A R, Rosero L A, Filho C M, Osório F S, Wolf D F. Fast Metric Tracking by Detection System: Radar Blob and Camera Fusion. 2015 12th Latin American Robotics Symposium and 2015 3rd Brazilian Symposium on Robotics. Uberlandia, Brazil: IEEE, 2015: 120–125.

[7] Wei Z, Zhang F, Chang S, et al. MmWave Radar and Vision Fusion for Object Detection in Autonomous Driving: A Review. Sensors, 2022, 22(7): 2542.

[8] Kim H, Choi Y. Location estimation of autonomous driving robot and 3D tunnel mapping in underground mines using pattern matched LiDAR sequential images. International Journal of Mining Science and Technology, 2021, 31(5): 779–788.

[9] Xu X, Zhang L, Yang J, et al. A Review of Multi-Sensor Fusion SLAM Systems Based on 3D LIDAR. Remote Sensing, 2022, 14(12): 2835.

[10] Kim H, Choi Y. Comparison of Three Location Estimation Methods of an Autonomous Driving Robot for Underground Mines. Applied Sciences, 2020, 10(14): 4831.

[11] He X, Gao W, Sheng C, et al. LiDAR-Visual-Inertial Odometry Based on Optimized Visual Point-Line Features. Remote Sensing, 2022, 14(3): 622.

[12] Chen Xiaohang. Target Detection and Tracking Based on Fusion of Millimeter-wave radar and vision for Spherical Robot. Zhejiang University, 2021.

[13] Muhammad S, Nardi D, Ohno K, Tadokoro S. Environmental sensing using millimeter wave sensor for extreme conditions. 2015 IEEE International Symposium on Safety, Security, and Rescue Robotics. IEEE, 2015: 1-7.

[14] Wang Z, Miao X, Huang Z, Luo H. Research of Target Detection and Classification Techniques Using Millimeter-Wave Radar and Vision Sensors. Remote Sensing, 2021, 13(6): 1064.

[15] Qian Bo, Chen Jie, Xu Qi, Wang Hongxing, Huang Zheng. Research on Active Obstacle Avoidance Method Based on Fusion of Binocular Vision and Millimeter Wave Radar. Microcomputer Applications, 2022, 38(07):48-50 +54.

[16] Kramer A. Radar-Based Perception For Visually Degraded Environments. University of Colorado at Boulder, 2021.

[17] Guan D, Cao Y, Yang J, Cao Y, Yang Y M. Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection. Information Fusion, 2019, 50: 148-157.

[18] Watanuki R, Horiuchi T, Aodai T. Vision-based behavior acquisition by deep reinforcement learning in multi-robot environment. ICIC Express Letters, Part B: Applications, 2020, 11(3): 237-244.

[19] Paral P, Chatterjee A, Rakshit A. Vision Sensor-Based Shoe Detection for Human Tracking in a Human–Robot Coexisting Environment: A Photometric Invariant Approach Using DBSCAN Algorithm. IEEE Sensors Journal, 2019, 19(12): 4549–4559.

[20] Nobis F, Geisslinger M, Weber M, Betz J, Lienkamp M. A Deep Learning-based Radar and Camera Sensor Fusion Architecture for Object Detection. 2019 Sensor Data Fusion: Trends, Solutions, Applications. Bonn, Germany: IEEE, 2019: 1–7.

[21] Peynot T, Scheding S, Terho S. The Marulan Data Sets: Multi-sensor Perception in a Natural Environment with Challenging Conditions. The International Journal of Robotics Research, 2010, 29(13): 1602–1607.

[22] Carlevaris-Bianco N, Ushani A K, Eustice R M. University of Michigan North Campus long-term vision and lidar dataset. The International Journal of Robotics Research, 2016, 35(9): 1023–1035.

[23] Kramer A, Harlow K, Williams C, Heckman C. ColoRadar: The direct 3D millimeter wave radar dataset. The International Journal of Robotics Research, 2022, 41(4): 351-360.

[24] Leung K, Lühr D, Houshiar H, et al. Chilean underground mine dataset. The International Journal of Robotics Research, 2017, 36(1): 16–23.

[25] Martín-Martín R, Patel M, Rezatofighi H, et al. JRDB: A Dataset and Benchmark of Egocentric Robot Visual Perception of Humans in Built Environments. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[26] Chen H, Hu X, Xu Y. Domain Adaptation from Public Dataset to Robotic Perception Based on Deep Neural Network. 2020 Chinese Automation Congress (CAC). IEEE, 2020: 6218-6222.

[27] Wang Jinke, Zuo Xingxing, Zhao Xiangrui, Lv Jiajun, Liu Yong. Review of multi-source fusion SLAM: current status and challenges. Journal of Image and Graphics, 2022, 27(02): 368-389.

[28] Kalman R E. A new approach to linear filtering and prediction problems. 1960.

[29] Wang Liling, Li Sen, Ma Dong. Multi-sensor Fusion Mobile Robot Localization Based on Monocular Sparse Method. Machine Tool and Hydraulics, 2021, 49(24): 17–22.

[30] Dong X, Chisci L, Cai Y. An adaptive variational Bayesian filter for nonlinear multi-sensor systems with unknown noise statistics J. Signal Processing, 2021, 179: 107837.

[31] Li G, Liu Z, Cai L, Yan J. Standing-posture recognition in human–robot collaboration based on deep learning and the dempster–shafer evidence theory J. Sensors, 2020, 20(4): 1158.

[32] Cui W. Multi-sensor Information Fusion Obstacle Avoidance based on Fuzzy Control. 2019 IEEE 2nd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE). IEEE, 2019: 198-202.

[33] Leonard J J, Durrant-Whyte H F. Directed Sonar Sensing for Mobile Robot Navigation. Boston, MA: Springer US, 1992.

[34] Wang Haixia, Wu Qingfeng, Wu Xiangbin, Zhang Qiuyi. Application Research of Multi-sensor Fusion in Robot Position Perception. Electromechanical Engineering Technology, 2020, 49(12): 89-91.

[35] Ren Weijian, Gao Qiang, Kang Chaohai, Huo Fengcai, Zhang Zhiqiang. Overview of Synchronous Positioning and Mapping Technology for Mobile Robots. Computer Measurement and Control, 2022, 30(02): 1-10+37.

[36] Song H, Choi W, Kim H. Robust vision-based relative-localization approach using an RGB-depth camera and LiDAR sensor fusion. IEEE Transactions on Industrial Electronics, 2016, 63(6): 3725-3736.

[37] Básaca-Preciado L C, Sergiyenko O Y, Rodríguez-Quinonez J C, et al. Optical 3D laser measurement system for navigation of autonomous mobile robot. Optics and Lasers in Engineering, 2014, 54: 159–169.

[38] Yang F, Liu W, Li W, et al. A Novel Object Detection and Localization Approach via Combining Vision with Lidar Sensor. 2021 IEEE 4th International Conference on Electronics Technology. IEEE, 2021: 167-172.

[39] Sun F, Zhou Y, Li C, Huang Y. Research on active SLAM with fusion of monocular vision and laser range data. 2010 8th World congress on intelligent control and automation. IEEE, 2010: 6550-6554.

[40] Park K, Kim S, Sohn K. High-Precision Depth Estimation Using Uncalibrated LiDAR and Stereo Fusion. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(1): 321–335.

[41] Dickmann J, Klappstein J, Hahn M, et al. Automotive radar the key technology for autonomous driving: From detection and ranging to environmental understanding. 2016 IEEE Radar Conference. IEEE, 2016: 1-6.

[42] Cai K, Wang B, Lu C X. AutoPlace: Robust Place Recognition with Single-chip Automotive Radar. International Conference on Robotics and Automation. IEEE, 2022: 2222-2228.

[43] Caesar H, Bankiti V, Lang A H, et al. nuScenes: A multimodal dataset for autonomous driving. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11621-11631.

[44] Huang J-T, Lu C-L, Chang P-K, et al. Cross-Modal Contrastive Learning of Representations for Navigation Using Lightweight, Low-Cost Millimeter Wave Radar for Adverse Environmental Conditions. IEEE Robotics and Automation Letters, 2021, 6(2): 3333–3340.

[45] Kramer A, Stahoviak C, Santamaria-Navarro A, Agha-mohammadi A, Heckman C. Radar-inertial ego-velocity estimation for visually degraded environments. 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020: 5739-5746.