IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 4 April 2023, accepted 14 May 2023, date of publication 18 May 2023, date of current version 30 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3277547

APPLIED RESEARCH

Embedded System Vehicle Based on Multi-Sensor Fusion

RUI TONG, QUAN JIANG, ZUQI ZOU, TAO HU, AND TIANHAO LI
Department of Electrical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Corresponding author: Rui Tong (202421583@st.usst.edu.cn)

This work was supported by the Open Project of the National and Local Joint Engineering Laboratory of High Energy Saving Motor and Control Technology of Anhui University under Grant KFKT202103.

ABSTRACT As intelligent driving vehicles came out of concept into people's life, the combination of safe driving and artificial intelligence becomes the new direction of future transportation development. Autonomous driving technology is developing based on control algorithms and model recognitions. In this paper, a cloud-based interconnected multi-sensor fusion autonomous vehicle system is proposed that uses deep learning (YOLOv4) and improved ORB algorithms to identify pedestrians, vehicles, and various traffic signs. A cloud-based interactive system is built to enable vehicle owners to master the situation of their vehicles at any time. In order to meet multiple application of automatic driving vehicles, the environment perception technology of multi-sensor fusion processing has broadened the uses of automatic driving vehicles by being equipped with automatic speech recognition (ASR), vehicle following mode and road patrol mode. These functions enable automatic driving to be used in applications such as agricultural irrigation, road firefighting and contactless delivery under new coronavirus outbreaks. Finally, using the embedded system equipment, an intelligent car was built for experimental verification, and the overall recognition accuracy of the system was over 96%.

INDEX TERMS Automatic driving, multi-sensor fusion, Internet of Things (IoT), edge intelligence, multi-object detection, YOLOv4.

I. INTRODUCTION

Nowadays automobiles have become an indispensable tool for people to travel. Moreover, the level of manufacturing and popularity of automobiles plays an important role in measuring a country's level of modernization and technology. With the continuous progress of science and technology, there are various potential safety hazards in the traditional manual driving vehicles. The continuous growth of car ownership has also caused road traffic congestion. In addition to increasing the safety awareness of traffic participants, all sectors of society hope to reduce the occurrence of traffic accidents and traffic congestion through technological progress, among which intelligent driving vehicles are the most concerned [1]. Compared with traditional manual driving, automatic driving has great advantages in safety, convenience and efficiency. Currently, autonomous vehicle have been put into trial operation in China and have achieved some success on specific roads. In recent years, the construction of smart cities and smart transportations have made intelligent driving a new direction of urban transportation development [2], [3] [4]. Automatic driving technology has become the main research direction of road traffic [5]. The paper [6] improves road traffic congestion by optimizing traffic signal lights. The paper [7] controls and predicts vehicle routes by optimizing and predicting traffic signals to alleviate traffic congestion. Intelligent transportation systems have been used in different countries to help solve traffic problems, the paper [8] has studied many countries' approaches to solving traffic problems involving their tools and technologies, and examines their different tools, technologies and applications. The paper [9] optimized the route through deep learning. The paper [10] reviews the latest technology of autonomous artificial intelligence. it pays special attention to Special attention data preparation,

The associate editor coordinating the review of this manuscript and approving it for publication was Alessandro Pozzebon.

feature engineering and automatic super parameter optimization. With more and more researches around intelligent transportation, the deep integration of intelligent driving technology, machine vision, environment awareness technology, and edge intelligent processing have become research hotspots. Unmanned driving technology has developed rapidly.

With the upgradation of the global industrial chain, the division and cooperation of global factories have promoted the rapid development of the automobile industry [11]. As of December 2019, China's car ownership had exceeded 250 million. The increase of car ownership not only causes traffic congestion, air pollution, and other problems but also brings new challenges to urban planning. Traffic accidents caused by automobiles also cause loss of life, health and property. The use of new energy vehicles reduces the dependence on fossil fuels [12] and reduces air pollution. Road traffic safety has become a serious issue. In 2018, there were about 60000 people died in traffic accidents in China, meanwhile there are about 1.24 million people died in traffic accidents each year all over the world. According to the China Highway, 86% of motor vehicle accidents were caused by illegal driving, where as the top five were caused by failing to give ways, driving without a license, speeding, drunk driving, and driving in the opposite direction. In the current environment, in addition to improving driver safety awareness and regulating civilized driving, all sectors of society hope to reduce the occurrence of traffic accidents and traffic congestion through technological progress. The paper [13] surveys and summarizes that most of the current traffic accidents were caused by human error, and the most prominent reason was fatigue driving. Automatic and auxiliary driving technologies can help reduce the occurrence of traffic accidents. In the 1950s, some developed countries began studying automatic driving.

Since the beginning of the 21st century, Internet of Things (IoT) technology, 5G applications, physical computing ability, dynamic vision technology, artificial intelligence technology, and automatic driving technology have been developed. The paper [14] completely abandons the concept of ''driver'' and introduces a autonomous vehicle based on deep learning and autonomous navigation. It uses a DRL-based urban environment automatic driving strategy, performs sensor fusion on two types of input sensor data, and trains the network to test the vehicle's automatic driving ability. This provides new possibilities for solving complex control and navigation related tasks. Compared with human-driving vehicles, intelligent driving vehicles offer higher reliability and faster reaction times, which can effectively reduce the number of traffic collisions. Based on automatic driving, this paper [15] proposed two DNN based deep learning models for the first time. It is used to measure the reliability of the driverless vehicle and its main on-board unit (OBU) components; however, this model increases the complexity of the control system and increases the system costs. In [16], The new challenges facing environmentally awareness technology are discussed from three perspectives: technology, the external environment, and application. The paper also points out the future development trends of environmental awareness technology. Environment sensing mainly uses fusion sensing to sense the driving environment around the vehicle (infrared sensing, radar system). The different sensors take full advantage to form a sensor system with complete information coverage. Deep learning is a new subject in the field of machine learning. At present, deep learning networks [17] and radio frequency technology [18] are used for target detection and image processing. As one of the cores of intelligent transportation, pedestrian detection is highly relevant to the industry and has important application value. Road sign recognitions are diverse, and specific instruction sign recognitions enable driverless vehicles to be used in night road patrols. Multi sensor fusion technology expands the application of driverless vehicles to agricultural irrigation, intelligent transportation, road fire protection and other places.

The Internet of Things (IoT) technology and the fifth-generation mobile network (5G) provide a large number of interfaces for edge intelligent devices through low-cost radio modules [19]. A large amount of local data that are difficult to store can be stored in the ECS through the Internet of Things technology. IoT technology allows intelligent edge devices to access and provide feedback to users. At the same time, client data are collected from the client and the remote-control command [20] is executed. the paper [21] uses a combination of the Internet and machine vision to remotely control the crawler.The user function of an intelligent driving system is designed based on the development of Internet of Things technology. The system significantly enhances the security of a vehicle through the combination of cloud-based unlocking and local fingerprint unlocking through small programs. This paper [22] proposed and implemented an integrated automatic login platform based on mobile fingerprint identification using block-chain theory, and constructed a convenient and integrated automatic login platform through smart phones, which has strong security.

Our contributions are mainly as follows:

1) We also considered the safety of driver-less vehicles from another perspective. As there is no driver, the instrumentalization attributes of the vehicles become increasingly obvious. How can owners absolute control their vehicles? Through IoT and wireless communication technologies, we provide a control scheme based on cloud remote data interactions. As the owner's permission is obtained through the cloud, the user of the vehicle is confirmed through fingerprint verification, which provides a reference for the accountability of driver-less vehicles after traffic accidents.
2) We designed a multi device management platform through the Internet of Things technology to realize the information connection between embedded devices, mobile WeChat applets and database platform. The intelligent recognition system and motion control system use different master chips, which allows the

intelligent processing system focus on the processing of machine vision. Compared to a single chip processing system, the processing time of machine vision is shorter.

3) In the image training phase, we added training for semi-occluded pedestrians and vehicles to improve the accuracy of image recognitions. We trained 5607 pictures in two categories.
4) Through data fusion among IoT technology, multiple chips and multiple sensors, the entire application system can realize complex functions and complete diversified control. Automatic driving vehicles can be used for agricultural irrigation, unmanned road patrols, road fire protections, and other applications.

The rest of the organizational structures of this paper are as follows: The second part introduces the related work; The third part introduces the system from the aspects of software, hardware and mechanical structure, and introduces the workflow of the entire IoT system; The fourth part introduces the experimental results. Finally, we conclude the work of this paper.

II. RELATED WORK

A. IMAGE PROCESSING

As an important branch of artificial intelligence, machine vision has significant applications in science and technology, medical treatment, the military industry, and other fields. Machine vision is receiving more and more attention as a fundamental component of emerging technologies such as automatic driving, unmanned driving, intelligent machine work, virtual reality and augmented reality [23], [24] [25].

1) IMPROVED ORB IMAGE PROCESSING

Feature extraction and image matching algorithms are the basis of many image algorithms that are widely used in image stitching, moving object detection, visual navigation, and other fields [26]. Oriented fast and rotated bridge (ORB) is a fast algorithm for extracting and describing feature points. Fast and brief are feature detection algorithms and feature vector creation algorithms, respectively, for extracting and describing features. The biggest feature of the ORB algorithm is its fast calculation speed. SIFT, a scale-invariant feature transformation, is a description used in the field of image processing. This description has scale invariance, can detect key points in an image, and is a local feature descriptor. The advantage of the SIFT algorithm is that its features are stable, it is invariant to rotation, scale transformation and brightness, and it is also stable to a certain extent to perspective transformation and noise; the disadvantage is that the real-time performance is not high, and the feature point extraction ability for smooth edge targets is weak. Speeded Up Robust Features (SURF) improves the method of feature extraction and description, and completes feature extraction and description in a more efficient way. The calculation time is only 1% of SIFT and 10% of SURF. SURF improves the way of feature extraction and description, and completes feature extraction and description in a more efficient way.

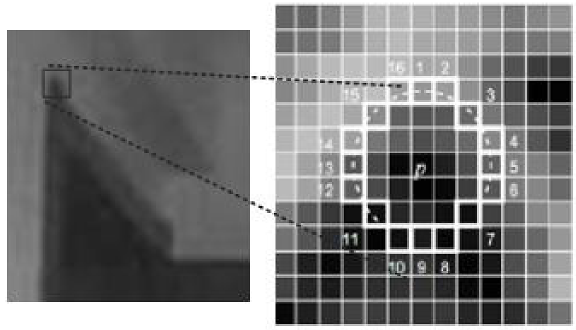

FIGURE 1. Schematic diagram of Fast detection algorithm.

This is mainly because fast is used to accelerate the extraction of feature points. This provides a more rapid response for automatically driving vehicles to recognize road signs. In [27], the SIFT, SURF, and ORB feature descriptors were compared and analyzed experimentally. Compared with other advanced works, the best results can be obtained by crossing SIFT, SURF, and ORB methods with random forest classification models. The brief algorithm was used to calculate descriptors. The unique representation of the binary string of the descriptor not only saves storage space but also greatly shortens the matching time.

After confirming the maximum number of key points to be found, you need to complete the extraction of key points. The following is an introduction to key point extraction:

1) The first step is to locate all key points in the training image, confirm the maximum number of key points to be searched, and extract fast feature points using a fast algorithm. If there is a large difference between a pixel and sufficient pixels in its surrounding neighborhood, the pixel may be a feature point. The confirmation process in the algorithm is as follows:

The value of the threshold parameter H of the selected feature point is determined. For any pixel point P on the image, fast compares 16 pixels in the circle with point P as the center in Figure 1. If there are more than 8 pixels on the circle whose gray value is less than LP - H (LP is the gray value of point P) or whose gray value is greater than LP + H, then pixel P is selected as the key point.

2) The second step is to help orb achieve scale invariance through pyramid. Orb algorithm itself does not have scale invariance. This part of the function is set in OpenCV by setting pyramid extraction rate scale and pyramid level. The original image is reduced to n images according to the extraction rate, and the scaled image is shown in Formula 1

$$I' = \frac{1}{scale^k}(k = 1, 2, \ldots, n) \tag{1}$$

3) The third step is to define the direction of feature points, realize the rotation invariance of orb, and use the gray centroid method to calculate the direction angle of feature

points. An image area, the element expression of the corresponding 2 × 2 matrix is:

$$m_{pq} = \sum_{x,y} X^p y^q I(x, y) \tag{2}$$

X, Y is the coordinate value, and I (x, y) is the pixel value. The centroid of the image area is:

$$C = (\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}) \tag{3}$$

The direction of the feature point can be obtained, thus achieving rotation invariance.

$$\theta = arctan(m_{\Delta\Gamma}, m_{\Gamma\Delta}) \tag{4}$$

1) Step 1: the image is smoothed by Gauss check to enhance the noise resistance of the image.

2) Step 2: in the field of each feature point, compare the pixel sizes of the two for binary assignment.

$$\tau(p; x, y) = \begin{cases} 1 & \text{if p(x)< p(y)} \\ 0 & \text{if p(x)>= p(y)} \end{cases} \tag{5}$$

In the formula, P (x) and P (y) are the pixel values of random points x and y respectively. In this way, the binary assignment of each feature point will form a binary code, which forms a binary feature point descriptor.

3) Step 3: to satisfy the invariance of scaling, the brief algorithm constructs the image pyramid.

2) MODEL TRAINING

Early target detection was mostly simple shape objects, or template matching technology was used to realize the recognition of simple objects. The traditional target detection technology had low accuracy and cumbersome feature extraction steps. Different algorithms were needed for different objects. With the development of image processing technology, deep learning has been applied in image classification [28], [29].

Convolutional neural network (CNN) is one of the most widely used neural networks in image recognition at present. Some researchers try to apply CNN to target testing and obtain good results. The target testing method based on deep learning develops rapidly. There are two kinds: two-stage target object detection mode and single-stage target object detection mode. The two-stage object detection method generates a large number of candidate regions in the image, and then classifies and regresses each region. Girshick et al. [30] proposed R-CNN, whose performance is significantly superior to other traditional target detection models, and initiated a new era of target detection based on deep learning. The paper [31] proposed a convolution network method based on fast region (Fast R-CNN) for object detection, The paper [32] uses a pedestrian detection method based on genetic algorithm to optimize XGBoost training parameters for pedestrian recognition. The single-stage target detection method is a method that transforms the target detection problem into a regression problem. YOLO (you only look once) network divides the input image into grids of equal size, and each grid predicts candidate boxes and category information of two scales to complete the detection, which can quickly complete the target detection. In Figure 2,describe the process of AI development and training.

The detection rate of CNN algorithm is low and can not meet the real-time detection of road conditions. The characteristic of YOLO algorithm is that only one CNN operation is required for the image, so that the target position can be selected in the output layer. YOLOv4's backbone feature extraction network is CSPDarkNet53, which makes a little improvement on Darknet53 and uses CSPNet (cross stage partial networks) for reference. CSPNet solves the problem of repeating gradient information in network optimization in other large convolutional neural network framework backbone, and integrates the gradient changes into the feature map from beginning to end, thus reducing the parameter quantity and flops value of the model, ensuring the inference speed and accuracy, and reducing the model size.

Data of prediction box position:

$$b_x = \sigma(t_x) + c_x \tag{6}$$
$$b_y = \sigma(t_y) + c_y \tag{7}$$
$$b_w = p_w + e^{c_w} \tag{8}$$
$$b_h = p_h + e^{c_h} \tag{9}$$

Here, bx and by are the coordinates of the center point of the prediction frame; A_2 and A_2 is the width and height of the prediction frame; C_x, C_y is the coordinate of the center point of the first verification frame; p_w, p_h are the width and height of a priori frame;(t_x t_y p_h t_h) is the location information of the prediction box; σ is the activation function, representing the probability between [0,1]. The calculation formula of confidence is:

$$confidence = Pr(object) \times IOU^{truth}_{pred} \tag{10}$$

Pr(object) confirms whether the target exists, $IOU^{truth}_{pred}$is the interaction ratio between the prediction frame and the real frame.

B. FINGERPRINT VERIFICATION

Fingerprint identification is controlled by EAIDK-310's retrieval of the allowed data from the cloud. When core processor obtains the allowed signal from the cloud, it sends a signal to the fingerprint module. After receiving the signal, the fingerprint module enters the detection mode and compares it with the stored fingerprint after receiving the input. At this stage, the fingerprint is verified by using AS608, simulating the door part in the actual use of the vehicle. The simulation of our car door is through the fingerprint module. The paper [33] considers the uniqueness of fingerprint information and applies fingerprint verification to transactions in the financial market. Article [34] applies AS608 to the data acquisition end of the access control system.

This design uses template matching method to extract fingerprint image on the basis of thinning image. The template

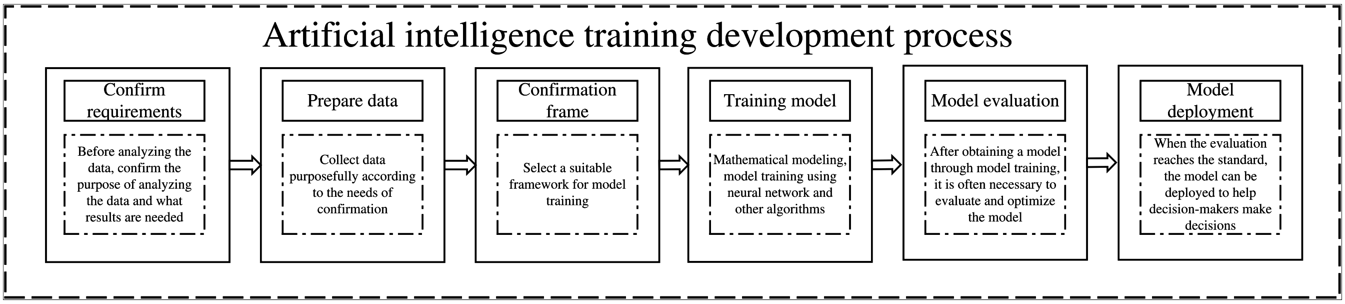

FIGURE 2. Artificial intelligence training.

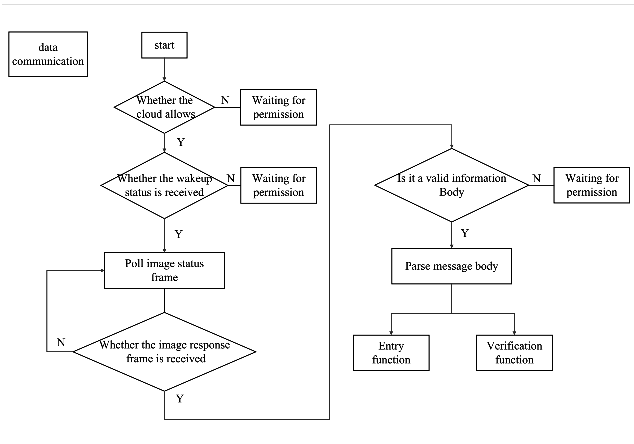

FIGURE 3. Fingerprint verification data communication.

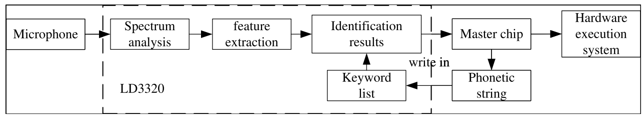

FIGURE 4. Voice module workflow.

matching method has high efficiency. It mainly extracts the feature points of fingerprint, including endpoint and bifurcation point. Based on the statistical analysis of 8 neighborhood points, the end points and bifurcation points are obtained. Among all the forms of the 8 neighborhoods, 8 can basically meet the conditions of endpoint characteristics, and 9 can also basically meet the conditions of bifurcation points. The fingerprint module is divided into two modes: verification and input.

After the fingerprint data is collected by the acquisition module, it is transmitted to the STM32 board through the RS232 serial port; The STM32 board performs comprehensive processing on the whole fingerprint data, including comprehensive evaluation of its image and data. At the same time, the user data is formed after extracting feature points and processing, and the fingerprint template is stored in flash. Number them to complete template storage. The user data is matched with the fingerprint template. If the matching is successful, the user is verified successfully. After the verification is successful, the vehicle is allowed to perform other control.

The control function of the fingerprint module is mainly completed by the communication task and the monitoring task. The communication task in Figure 3 is a separate thread, which queries and receives the frame data of the fingerprint module in real time. When the effective frame data is received, the data will be repackaged into a message body and sent to the monitoring task through the message mechanism between tasks; The monitoring task is another separate thread, which processes the message body sent by the communication task in real time. After receiving the effective message body, it will send the fingerprint to the input function module or the verification function module according to the message command content.

C. VOICE INTERACTION SYSTEM

Automatic speech recognition (ASR) is a key technology in human-computer interaction, which helps specific people realize the intelligent interaction between machines and people [35], [36]. For a long time, speech recognition has not become the main form of human-computer communication because of its unstable recognition accuracy and no good interaction experience. However, with the application of multi-core processor, Pascal framework graphics processor and high-performance solid-state disk, as well as the application of deep learning technology, we can train a larger data set to generate a more complex speech model [37], [38], [39]. The paper [40] uses cloud analysis to realize the voice interaction of early teaching machines. The paper [41] first introduces the basic ideas, advantages and disadvantages of HMM (hidden markov model)-based model and end-to-end models, and points out that end-to-end model is the development direction of speech recognition. The emergence of these methods has significantly improved the accuracy of speech recognition and made this technology enter the lives of the majority of users.

Speech recognition ASR technology is a technology based on key word list recognition. It is only necessary to set the list of key words to be recognized, and transmit these key words to the LD3320 in the form of characters to recognize the key words spoken by the user in Figure 4. The user is not required to perform any recording training. The most important practical significance of ASR technology is to provide a voice-based user interface (VUI) that is separated from keys, keyboards and mice, so that users can operate products more quickly and naturally. The speaker independent speech recognition technology (ASR) is to analyze a large number of speech data through linguists' speech models, scientists establish mathematical models, and extract the detailed features of primitive speech through repeated training, as well as the feature differences among various primitives. It is possible to obtain the speech characteristics of each primitive in the sense of statistical probability optimization. At last, senior engineers convert the algorithm and voice model into hardware chips for application in embedded systems.

The interactive voice control system is designed with VG009 and LD3320. The voice recognition chip LD3320 is used to process the voice signal received by the external microphone, extract the voice feature, match the key words in the stored key word list, and then match the words with the highest score from the key word list as the final output. The word list is input in the form of Pinyin string through the MCU. The LD3320 has two parts of speaker independent speech recognition, namely, the search engine module and the complete feature library. It does not need an external auxiliary chip and can realize a complete solution for speech recognition by a single chip. In consideration of the user experience of the whole intelligent system, the processing results are not only sent to the main control chip to realize the switching between multiple functions, but also the intelligent interactive system is realized with VG009.

D. MULTI SENSOR FUSION

Although the performance of the sensors is constantly improving, it is still difficult to meet the requirements of safe and stable driving of intelligent driving vehicles using a single sensor. Multi sensor fusion can solve the defect of insufficient performance of a single sensor, increase the information factors collected, provide more considerations for decision-making, and have stronger rationality. It is a major development trend of intelligent driving vehicles. The paper [42] realizes obstacle detection on more complex off-road surfaces through sensor fusion technology. Multi sensor fusion can be divided into decision-making level fusion, feature level fusion and data level fusion according to different fusion media. For example, Tesla's autopilot system integrates millimeter wave radar and camera to realize L2 level automatic driving. Baidu Apollo unmanned vehicle realizes unmanned driving through the integration of GPS, camera, infrared sensor and other data. The paper [43] discusses the application of multi-sensor fusion for unmanned vehicles.

In order to broaden the use of autonomous vehicles, we use multi-sensor fusion to realize the tracking mode, following mode and voice interactive control mode of autonomous vehicles. In Figure 5, The switching of these working modes is controlled through voice interaction. These working modes have greatly broadened the use scenarios of unmanned vehicles, making them suitable for the transportation of unmanned ports, contact-less distribution under the epidemic, agricultural irrigation, road fire prevention, etc. Intelligent identification combined with comprehensive sensing of sensors can help vehicles achieve more accurate control, and the combination of infrared and pedestrian identification can effectively ensure the safety of vehicles.

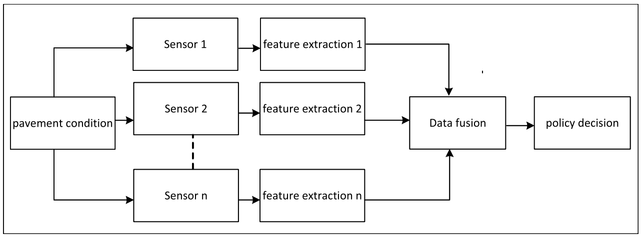

FIGURE 5. Sensor data fusion.

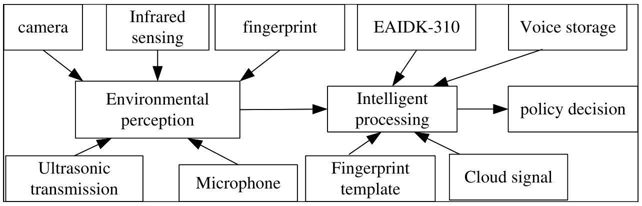

FIGURE 6. Sensor fusion of automatic vehicle system.

III. SYSTEM ARCHITECTURE

As far as we know, the multi-sensor fusion autonomous vehicle system based on cloud interconnection is an environment sensing system. The automatic driving function can be realized completely independently, so that the whole intelligent driving vehicle does not need a driver. In this section, we provide the overall design framework of the system.

A. SYSTEM OVERVIEW

The design comes from the automatic driving of intelligent control, cloud remote command and edge intelligent processing. The hardware structure of the overall platform is shown in Figure 7. The entire auto drive system includes an intelligent control system and a data management system. The intelligent control system is divided into intelligent detection part, motion control part, voice interaction system and fingerprint recognition system; The intelligent detection part detects the road surface information through the camera and recognizes the corresponding road surface information to issue corresponding instructions. The motion control part is equipped with infrared sensors and ultrasonic sensors. A multi-sensor fusion is formed to realize diversified control of intelligent vehicles. The intelligent detection part uses image recognition and processing technology to detect traffic signs, pedestrians, vehicles and traffic lights. The voice interaction system realizes interactive voice control for the

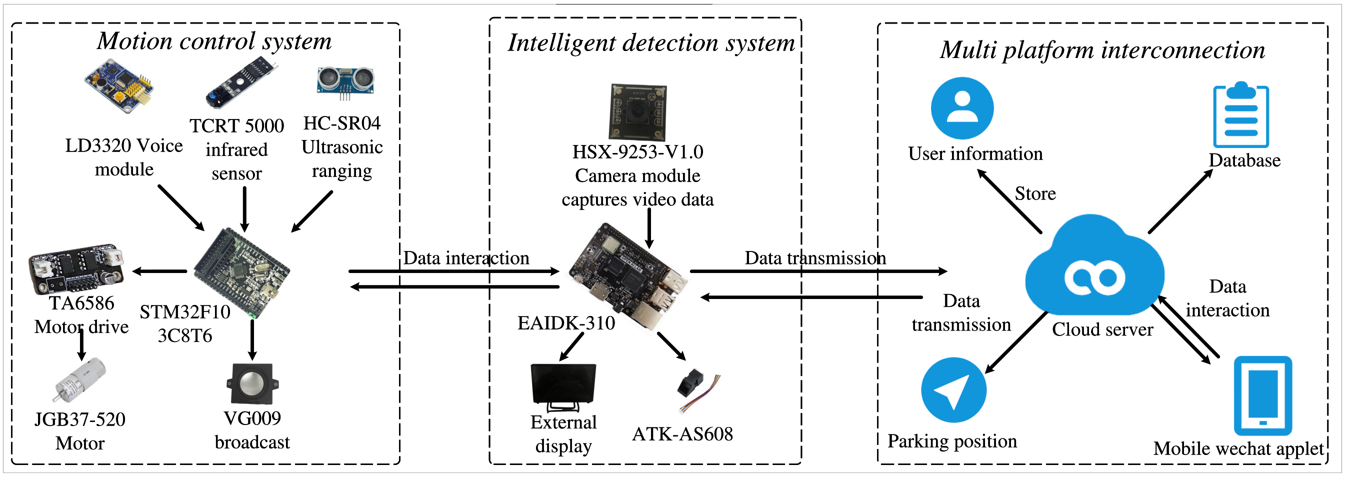

FIGURE 7. **The architecture of the proposed system.**

vehicle. Confirm the identity of the driver-less vehicle user through the communication between the fingerprint identification system and the cloud data. The sharing of driver-less vehicles is promoted through the interconnection of WeChat applet on mobile phone.

1) Motion control system: the motion control system is designed to complete the processing instructions of the intelligent platform. The motion control system includes STM32f103c8t6, TA6586 motor drive board, Tcrt5000 infrared sensor and DC motor. The motion control system integrates the data of various sensors, receives the signals of the intelligent platform, executes the processing results of EAIDK-310, and completes the mechanical functions of the vehicle. The whole motion control system is an execution system for multi-sensor fusion of road surface.

2) Intelligent detection system: the whole intelligent control system includes EAIDK-310, LCD display screen, camera and other modules. As the core of the whole system control, EAIDK-310 realizes the real-time detection of road information, directly interacts with the cloud information through WiFi, and detects the cloud control signal online in real time. EAIDK-310 is equipped with a high-performance ARM processor, with a main frequency of 1.3GHz. At the same time, its on-board running memory is as high as 1GB. Its rich peripheral interfaces include WiFi, USB, Ethernet and HDMI interfaces, providing good convenience for the interaction between the entire intelligent system and users. EAIDK-310 preloaded embedded deep learning framework Tengin supports the direct deployment of Caffe / tensorflow / Python / mxnet / onnx / Darknet and other training framework models, supports network performance optimization strategies such as layer fusion and quantification, provides a unified API (C / Python / JNI) interface, and provides self-defined operators for extended interfaces. EAIDK-310 can be used in a variety of intelligent detection and has a wide range of practicability.

3) Voice interaction system: the whole voice control system is based on a dialogue between LD3320 and VG009, which is an interaction mode designed based on user experience.

4) Fingerprint identification system: ATK-as608 fingerprint module is used for fingerprint identification to collect fingerprints. The chip has built-in DSP arithmetic unit and integrated fingerprint identification algorithm, which can efficiently collect fingerprint images. The data collected by the fingerprint module is stored in the CPU.The ATK-as608 has a fingerprint wake-up function. When a finger is detected on the fingerprint detection module, the wake-up pin on the module sends a high level back to the IO port of the CPU. After receiving the high level, the CPU starts to receive the fingerprint data.

B. MECHANICAL STRUCTURE

In Figure 8,the purpose of mechanical design is to flexibly integrate all parts to facilitate the realization of various functions, the measurement angle of sensors, the sensor layout of the whole vehicle, and the convenience of human-computer interaction. The integrated platform retains the original intelligent vehicle model, and has been transformed and innovated. The design of the mechanical structure of the automatic driving vehicle includes the design of the body, the layout of the hardware structure, the design of the sensor and the human-computer interaction.

1) INTELLIGENT VEHICLE BODY DESIGN

In the design of the whole system, the control system of the chassis mobile platform is the most important, which is the core of the stable operation of the whole system. The level of the control system is directly related to the intelligent level of the intelligent platform. The design strategy of the control system also determines the functional characteristics, application scope and expansibility of the whole design system. There are many kinds of operating structures of mobile

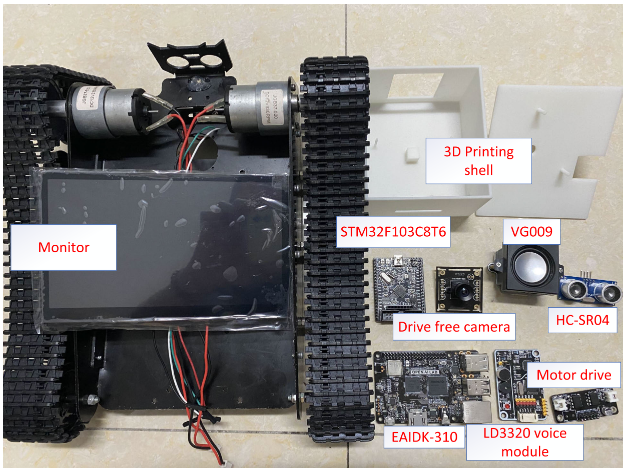

FIGURE 8. Schematic diagram of each part of the vehicle.

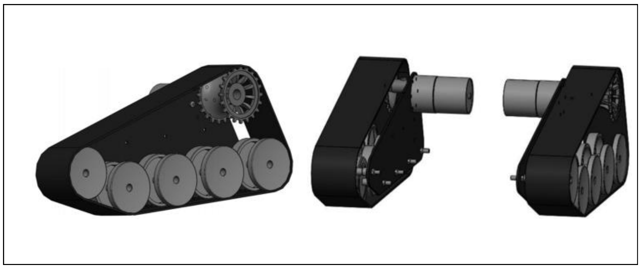

FIGURE 9. Mechanical structure and position of crawler.

chassis. At present, the commonly used ones are wheel type, crawler type, leg type and the combination of the above-mentioned structures.

Compared with other chassis, crawler chassis has the advantages of good carrying capacity, low overall cost and simple control. Crawler traveling mechanism is widely used in construction machinery, tractors and other field vehicles. The poor traveling conditions require that the traveling mechanism has sufficient strength and rigidity, and has good traveling and steering ability. The crawler is in contact with the ground, and the drive wheel is not in contact with the ground. When the motor drives the drive wheel to rotate, the drive wheel continuously rolls up the track from the rear under the action of the drive torque of the reducer through the engagement between the teeth on the drive wheel and the track chain. The ground part of the crawler gives the ground a backward force, and the ground correspondingly gives the crawler a forward reaction, which is the driving force to drive the machine forward. When the driving force is enough to overcome the traveling resistance, the roller rolls forward on the upper surface of the crawler, so that the machine travels forward. The front and rear crawlers of the crawler traveling mechanism of the whole machine can be turned independently, so that the turning radius is smaller.

In this paper, the mobile platform uses DC reduction motor to realize its movement ability, and it needs to move forward,

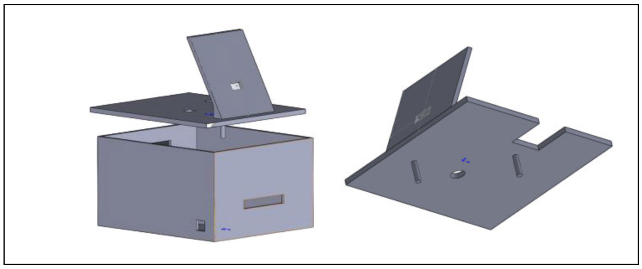

FIGURE 10. Overall structure of 3D printing shell.

backward, horizontally and diagonally. The voltage and current consumed by the DC motor in this operating state are relatively large and the change of its forward and reverse rotation needs to be realized by changing the polarity of the applied voltage.

2) HARDWARE STRUCTURE AUXILIARY LAYOUT DESIGN

Considering the fixing of STM32 and other devices such as motor drive board, the fixed structure of these components was designed using solid works, and the additive manufacturing was carried out using 3D printing. The overall structure is shown in the figure 10.

A set of fixed structure is established in the camera part. An external vertical plate with a width of 50mm, a length of 60mm and a thickness of 3mm is established on the upper surface of the box. A hole is opened on the plate to lead out the camera connecting line to ensure the stable and normal operation of the camera connection. At the same time, the plate is inclined at an angle of 80 Â° with the upper surface of the box. While the camera is fixed, the rear language command is introduced into the sound transmission hole of the voice module to facilitate voice control. The voice control module is fixed by the positioning hole on the PCB, and the corresponding fixed hole position is designed in the additive manufacturing, and the sound transmission hole is opened at the microphone position of the voice module on the upper cover of the box to facilitate the external control of voice transmission. The fingerprint module is fixed on the side of the box, and a hole is opened on the side plate to facilitate identification. The STM32 and the motor drive board are fixed by the positioning holes on the PCB, placed at the bottom of the box and a boss is added at the bottom to realize the stable installation of the board. The relevant control lines are led out through the openings on the four sides of the box.

C. SOFTWARE SYSTEM

The software platform is designed by structured programming method, and the software system is divided into three parts. It is divided into intelligent image processing, multi-sensor data fusion system and multi device management platform. The EAIDK-310 platform for intelligent processing is installed with the Linux fedora28 system and adopts the embedded deep learning framework Tengine We deploy the model based on YOLOV4 training on EAIDK-310 and cooperate with the improved orb algorithm to complete the image processing of the road surface. Multi sensor fusion is realized by STM32c8t6. Data docking among embedded devices, WeChat applet and database platform can be realized on the multi device platform.

In the user interface, we directly use LCD to display the road surface information collected by the camera, and display the recognition results to enhance the interaction between the intelligent device and the user. Due to the enhanced instrumentality of the driver-less device, we have developed a WeChat applet to facilitate the user's remote control instructions, display the driving path and parking location through interaction with the cloud data stream. The voice interaction system uses LD3320 to interact with the VG009 playback module and sends the control signal to the motion control system in Figure 12. The input of voice commands and the synthesis of voice play are based on the upper computer of the computer and written into the memory in advance.

The operation flow is as follows:

1) The user inputs the vehicle model and password through the mobile terminal login interface and requests the cloud terminal to confirm the login.
2) Enter the console and send the permission signal to the cloud. After the signal is successfully sent to the cloud, a pop-up window will be displayed to indicate whether the signal is successfully sent.
3) When EAIDK-310 receives the signal and is not allowed, the system is in a standstill state and as608 is in a sleep state.
4) After receiving the permission signal, the as608 is awakened. After receiving the information of the person entered in advance, the fingerprint opens the switch of the simulated door lock. Thus, the cloud local dual authentication is realized. It ensures the owner's control over his / her vehicle. The drive free camera starts to work and enters the detection environment.
5) After reaching the destination, the applet can remotely close the allow button, EAIDK-310 exits the detection state, and the camera exits the working state. At the same time, send the current parking position to the cloud, and you can view the location of the vehicle through the applet.

EAIDK-310 development board supports RJ45 interface and can provide 100m Fast Ethernet connection function. In Figure 11, the VNC connection supported by EAIDK-310 development board is adopted in the debugging process of this project. The specific settings are as follows:

1) IDENTIFICATION OF ROAD SIGNS, PEDESTRIANS AND VEHICLES

a: IMPROVE ORB

Because orb algorithm does not need deep learning and can complete self-renewal quickly, the applicability of this algorithm is more extensive and convenient. It can cooperate with the tracking function to complete the distribution task of the unmanned port and the automatic cruise of the specific road section.

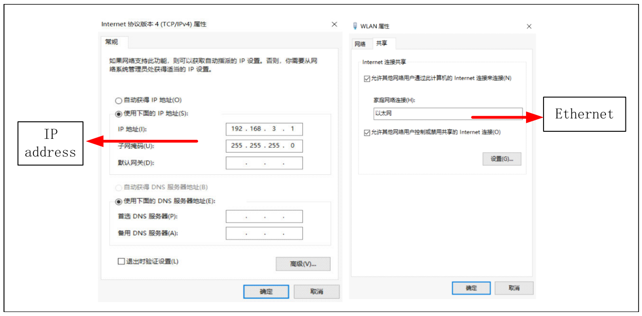

FIGURE 11. VNC connection settings.

A subset of key point descriptions is established for the template map (reference map) and the real-time map (observation map). The target recognition is accomplished by comparing the key point descriptors in the two-point set. The similarity measure of key point descriptors with 128 dimensions adopts Euclidean distance.

Key point descriptors in template diagram:

$$R_i = (r_{i1}, r_{i2}, \ldots r_{i128}) \tag{11}$$

$$S_i = (s_{i1}, s_{i2}, \ldots s_{i128}) \tag{12}$$

Similarity measure of any two descriptors:

$$d(R_i, S_i) = \sqrt{(\sum_{j=1}^{128}(r_{ij} - s_{ij})^2)} \tag{13}$$

To obtain the paired key descriptor $d(R_i, S_i)$ needs to satisfy:

$$\frac{S_i}{S_p} < Threshold \tag{14}$$

S_i is the closest point to R_i in the real-time map, and SP is the next closest point to R_i in the real-time map.

After confirming that the number of feature points reaches the threshold value, it indicates that the road sign is recognized. After recognition and multi-sensor fusion processing, the automatic driving car will complete the corresponding action in Figure 13.

b: PEDESTRIAN AND VEHICLE DETECTION BASED ON YOLOv4

Due to the particularity of the use scene, the identification of pedestrians and vehicles is the master condition of road safety. For the detection of pedestrians and vehicles, the use of feature point detection can not meet the needs well due to factors such as obstacles and environment. We have made a variety of data sets of vehicles blocked and pedestrians with complex actions. It is a complex work to develop custom labels every time the image is preprocessed. It needs to introduce an efficient image annotation method. The boundary box is used to realize image matching and accelerate the

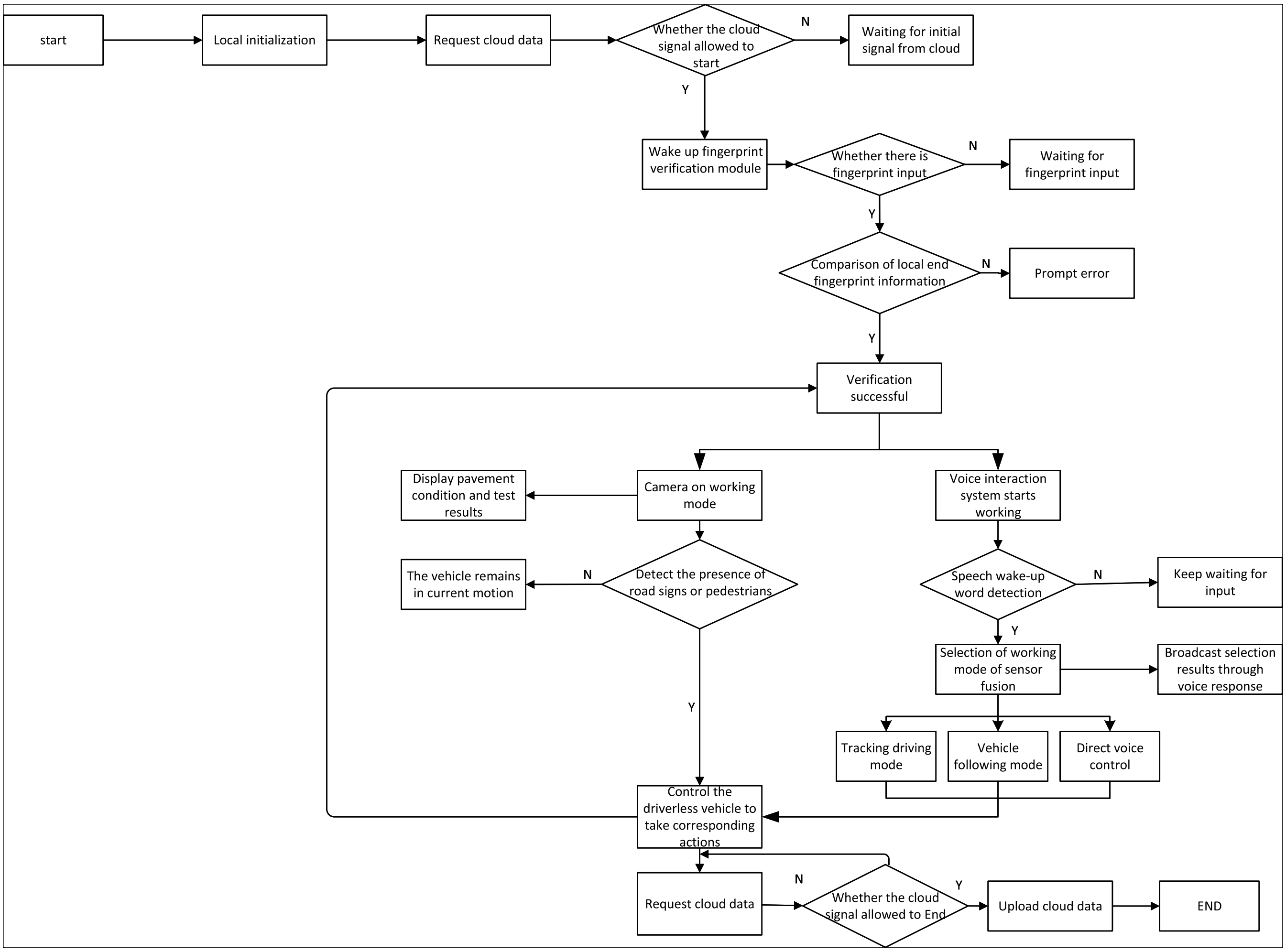

FIGURE 12. The flow chart for controlling the system.

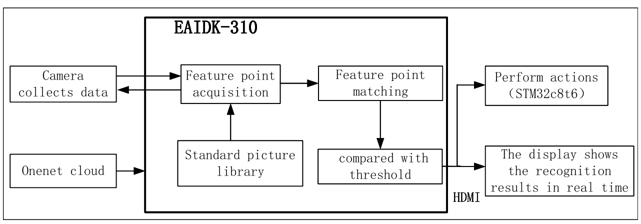

FIGURE 13. Identification process.

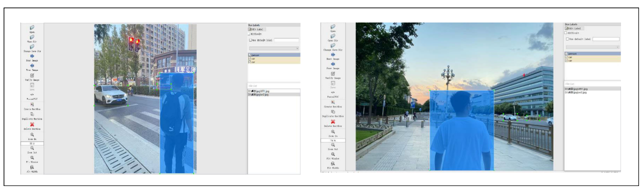

FIGURE 14. Pedestrians and cars are marked in the figure. As an image annotation tool, labelimg can save the generated annotations as an XML file in Pascal VOC format without secondary conversion.

transformation of images in computer vision. In this paper, rectangular box is selected for image annotation, and labelimg (a labeling tool for image preprocessing) is selected for image annotation. The annotation of the data set is shown in the figure14.

In order to obtain good recognition accuracy, a large number of data sets are needed. We constructed a data set of omni-directional pedestrian and vehicle detection. The collected images need to consider pedestrian and vehicle detection under various road conditions. We found a total of 5607 pavement conditions. According to the ratio of training set and test set 9:1. In this experiment, the size of picture input is 800×800 for training, and the initial learning rate is 0.001. The learning rate is a parameter determined by the programmer. A high learning rate means that the action of weight updating is larger, so it may take less time for the mode to converge to the optimal weight. However, if the learning rate is too high, the jump will be too large and not accurate enough to achieve the best. This value is usually 0.1-0.001. The data set is shown in the figure 15.

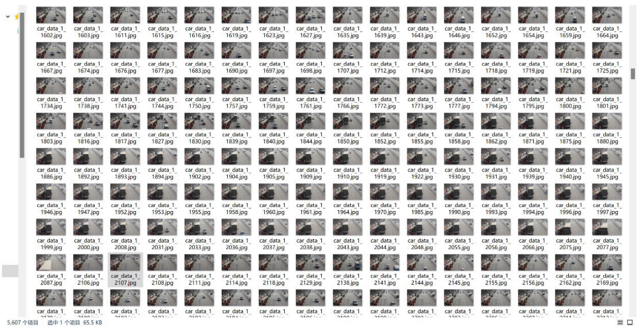

FIGURE 15. The data on the samples in the database.

TABLE 1. Type of call out box.

Number	Category	Ratio of training set to test set
1	Car	9:1
2	Person	9:1

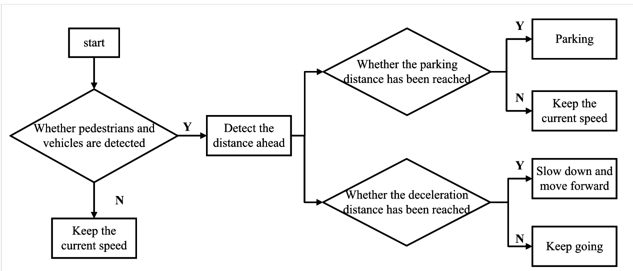

FIGURE 16. System architecture of pedestrian and vehicle identification.

We have illustrated the flow chart of pedestrian recognition in Figure 16.

2) MULTI-PLATFORM INTERCONNECTION

Before planning to build multi-platform interconnection, we need to consider the overall design of the traditional software and hardware integration platform. The design platform divides the data flow transmission into three corresponding parts, the control end facing users, the edge intelligent device end, and the data storage transfer station. The user control terminal is the channel for direct interaction between the intelligent system and the user, and needs to be convenient to use, complete in function and easy to control. The edge intelligent device is the execution carrier of the task. Due to the restriction of local storage, the data supplier needs to upload the data generated in the process of use. The data storage transfer station is used to analyze, classify and store the uploaded data.

The self-driving vehicle designed in this paper uses the network module of EAIDK-310 as the channel for uploading the application information of edge intelligent processing equipment, and the cloud provides the possibility for remote data interaction and storage. In Figure 20,At the remote user end, we design an interactive page through the user's mobile

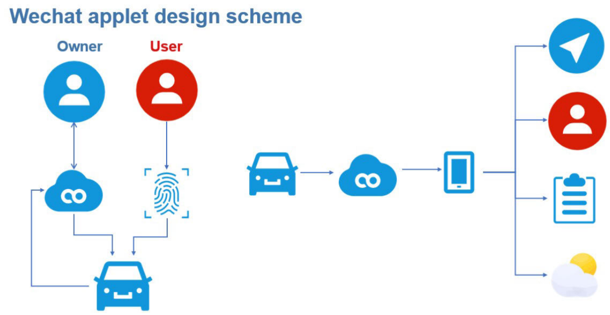

FIGURE 17. Overview of applet page design.

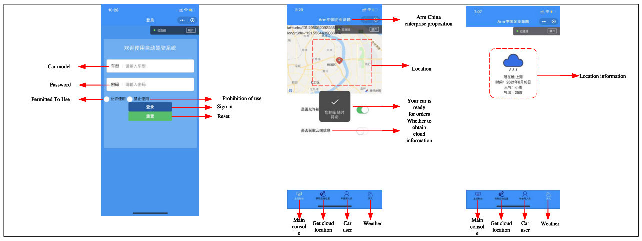

FIGURE 18. WeChat applet login interface, cloud unlock, weather view.

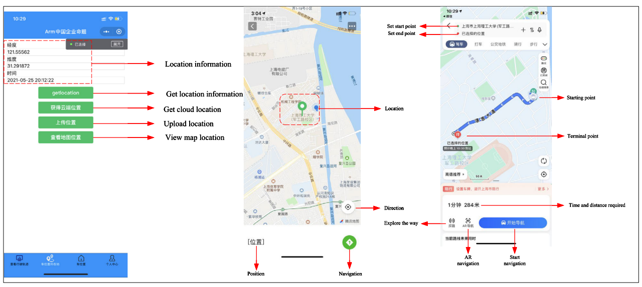

FIGURE 19. Parking position viewing and navigation through external API.

WeChat applet to provide convenience for remote control and facilitate the user to understand the use of the vehicle.

WeChat applet, developed by Tencent in China based on WeChat ecosystem, is an application level program that does not need to be downloaded. The combination of WeChat applet and unmanned user terminal realizes the visibility of data on the mobile terminal and greatly facilitates the user's use experience.

Before planning and designing each page, we learn from the early user login mode to identify users and distinguish the vehicles that need to be controlled. As a cloud platform connection station facing the actual user needs, WeChat applet collects and uploads the control data flow of the user during the interaction with the cloud, extracts the corresponding user data flow from the cloud for display, helps the user to remotely control the vehicle, view the parking position, And remind the user of the weather conditions through the external API in Figure 19.

Onenet is a PAAS (Platform as a service) Internet of things open platform created by China Mobile. It can greatly facilitate developers to connect devices, quickly complete the

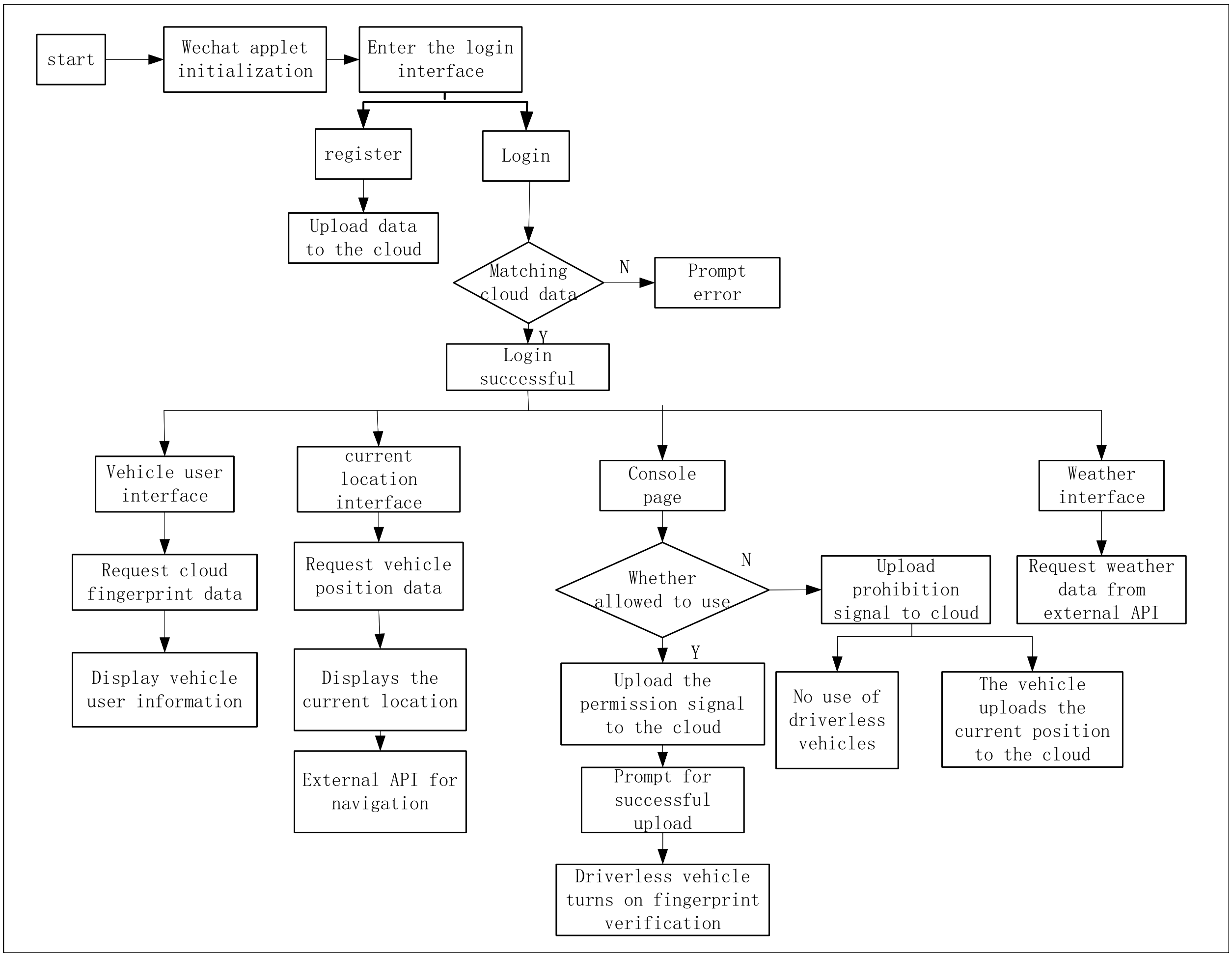

FIGURE 20. The flow chart for controlling the system.

deployment of intelligent hardware, and help the linkage of edge computing hardware. This design scheme stores and matches the control data flow, location data and user information of WeChat applet and EAIDK-310 through Onenet platform in Figure 21 to realize intelligent interaction and remote control of the whole intelligent system. Our cloud data access users need to provide corresponding data stream keys.

IV. EXPERIMENT AND RESULTS

The experiment evaluates the test results and verifies the accuracy of target detection (road signs, pedestrians), voice interaction and fingerprint verification in the actual working environment. At the same time, the control process of the whole system is verified. Since the design involves the participation of users, and considering the adaptability and individual differences of users, we conducted repeated experiments on different users for many times.

For the function of the system, we have carried out overall verification.Of course, there have been many traffic accidents since the launch of assisted driving. However, with the improvement of sensor accuracy and the coordinated

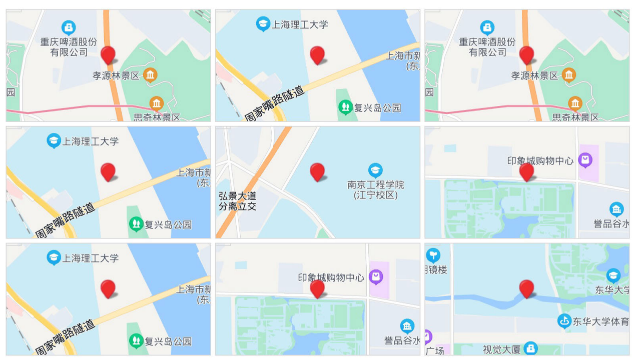

FIGURE 21. The cloud parses the stored parking position data.

processing of communication equipment, unmanned driving has been well applied in the field of medium and low speed. In order to avoid traffic accidents, we guarantee the priority of obstacle avoidance function during operation through the equipped ultrasonic sensors. Setting the priority of control signal is described in Table 2. The priority setting greatly ensures the safety attributes of the vehicle. Our entire article

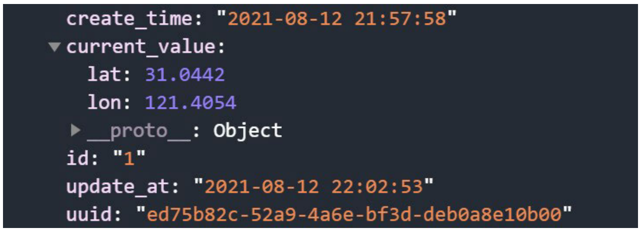

FIGURE 22. The user terminal grabs the parking position from the cloud.

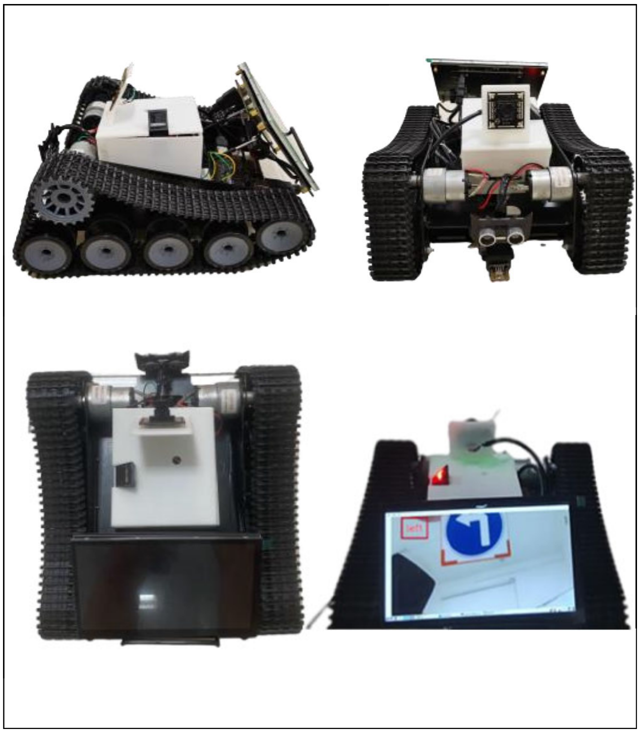

FIGURE 23. Self-driving vehicle.

is based on the cloud interconnection, and the entire system is simulated on embedded devices. We carried out a comprehensive experiment on smart cars in Yangpu District, Shanghai, focusing on the accuracy verification of interactive devices with many human factors. We tested cloud access using scripts. The communication between EAIDK-310 and the motion control system was tested through automatic data transmission. Under the condition of ensuring the accuracy of image recognition, voice interaction and fingerprint verification, the whole system rarely had problems when the signal was good.

We made assumptions about the experimental environment:

1)We use the embedded intelligent car to simulate the real vehicle, and use the motion control system to simulate the motion of the real vehicle.

2)The signal coverage of the whole experimental environment is good.

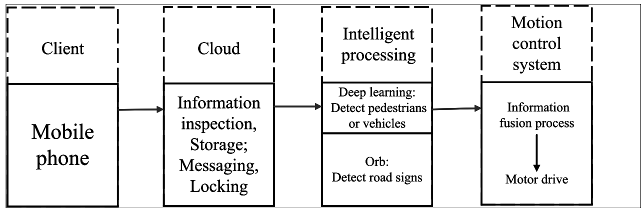

FIGURE 24. Data stream.

TABLE 2. Prioritization of control signals.

Vehicle movement	Signal priority
Parking state	lock signal >Obstacle avoidance signal > Pedestrian vehicle detection > Road sign recognition > Voice control
Vehicle running	Obstacle avoidance signal>Pedestrian vehicle detection > Road sign recognition > Cloud lock signal >Voice control

TABLE 3. User fingerprint test results.

User test number	Average test time	Test times	Error times	Signal priority
1	1.5s	300	21	93%
2	1.8s	300	17	94.33%
3	1.3s	300	12	96%
Average accuracy				94.44%

3)The special EAIDK-310 development board is used for visual processing, and the road section information is obtained in real time through the camera.

4)The WeChat applet we used for testing has not been released publicly and is only used for debugging.

Data flow transmission is shown in Figure 24:

A. FINGERPRINT IDENTIFICATION

The fingerprint module is debugged by the upper computer input detection. We input the fingerprint templates of the three main authors of this design and store them in the local end of the vehicle. The fingerprint template matching detection at the local end and the cloud cooperate to realize the double verification of the unmanned vehicle. As shown in Table 3.In order to verify the accuracy of fingerprint recognition, we entered three user templates, and each user conducted 300 tests.

B. VOICE INTERACTION SYSTEM

The wake-up waiting is set to 5s. After wake-up, the indicator stays on and waits for the input of voice commands. The setting of wake-up words greatly improves the efficiency of voice recognition. In the voice interaction system, we use the serial port to output commands for on-demand. Voice interaction files we use iFLYTEK's online voice synthesis tool to synthesize voice. There are two types of wake-up: voice wake-up and serial wake-up. When the correct fingerprint

TABLE 4. Voice interaction and serial port instructions.

Voice category	Voice feedback	Test times	Error times	Recognition accuracy	Serial communication characters	Action command
Hello,Xiaozhi	Master, you should be full of vigor today	300	9	97%	a	Awaken
Request a left turn	Turning left	300	23	92.33%	b	Turn left
Stop immediately	Stopped	300	5	98.33%	d	Parking
Request reverse	Backing up for you	300	8	73.33%	e	Reverse
Turn on tracking mode	Tracking mode turned on	300	9	97%	f	Enter tracking mode
Turn on following mode	Follow mode turned on	300	10	96.66%	g	Turn on follow mode
Average accuracy				96.19%		

TABLE 5. Image recognition.

Road sign type	Test times	Error times	Accurate Recognition	Recognize reaction time
Turn left	300	13	95.66%	2s
Turn right	300	14	95.33%	2s
Parking	300	5	98.33%	2s
Speed limit	300	9	97%	2s
Reverse	300	9	97%	2s
Average accuracy			96.67%	

FIGURE 25. The voice command is written into the LD3320 in the form of Pinyin string through the upper computer.

input is detected, it is wake up by the serial port (if the cloud allows).

C. IMAGE RECOGNITION EXPERIMENT

In the road sign recognition test, about 5 kinds of traffic signs' icons were collected, and feature points in the figure 26 were extracted by orb rapid recognition. The recognition of single road signs is shown in the figure below, and the recognition accuracy of road signs is shown in the figure below Table 5.

Important parameters in the detection of evaluating pedestrians and vehicles are as follows: TP refers to the detection of

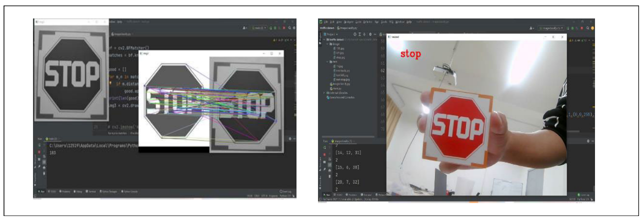

FIGURE 26. Feature point extraction and recognition effect.

vehicles or pedestrians, and there are also vehicles or pedestrians in the actual picture (positive samples are correctly identified as positive samples); TN is no model detected and there is no model in the sample (negative samples are correctly identified as negative samples); FP is the number of negative samples identified as positive samples; FN is a positive sample and is recognized as a negative sample; The recall rate (R) is the probability that the trained model can predict travelers and vehicles; The accuracy rate (P) is the ratio of the correctly detected target to the total number of targets; AP is the average precision of each category, and mAP is the average of all categories of APS.

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

TABLE 6. Evaluation index of vehicle and pedestrian detection.

Category	Recall	Precision	AP
Car	89.65%	94.15%	96.85%
Person	82.22%	93.24%	93.24%
mAP		95.04%	

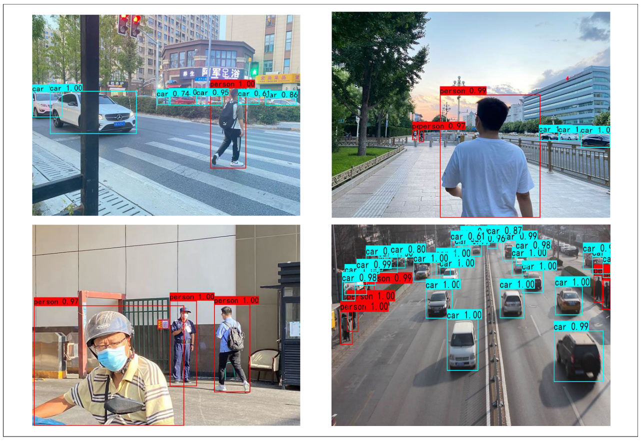

FIGURE 27. Pedestrian and vehicle detection.

$$AP = \int_0^1 P(R)dR \tag{17}$$

$$mAP = \frac{(\sum_{i=1}^{\epsilon} AP_i)}{c} \times 100\% \tag{18}$$

In the figure27,the recognition result shows that it has a good recognition effect on pedestrians and vehicles under general road conditions. However, the recognition accuracy of vehicles with severe occlusion will be reduced. For our 800 × 800 pixel image, the time for checking each frame in visual processing is 0.09s. In the actual test, we think the feedback time of the whole communication equipment is within 1s.

V. CONCLUSION

In this study, a set of auto-drive systems were designed based on cloud interconnections through EAIDK-310, STM32F103C8T6, and the interaction between the WeChat applet and cloud data, the double verification of the remote and local ends of the car is realized to ensure the control rights of the car owner. The double verification of the car remotely and locally to ensure the owner's absolutely control. The sharing concept of unmanned driving is promoted through the interaction with cloud location data flow. This paper introduces the overall design scheme, the work flow of the system, the design of the embedded system, the identification algorithm adopted, the model deployment, the construction of the cloud platform and page design of the WeChat applet. Finally, the specific functions are tested through physical construction to complete the implementation of the whole cloud networking control system.

The feasibility of the entire system was verified in this study. In subsequent experiments, we considered connecting with a physical vehicle. It is necessary to further consider the accuracy of various sensors to avoid traffic accidents and to create a more complete cloud and local storage system to cope with the subsequent large number of visits.

REFERENCES

[1] L. Forlano, "Cars and contemporary communications| stabilizing/destabilizing the driverless city: Speculative futures and autonomous vehicles," *Int. J. Commun.*, vol. 13, p. 28, Jun. 2019.
[2] C. Legacy, D. Ashmore, J. Scheurer, J. Stone, and C. Curtis, "Planning the driverless city," *J. Transp. Rev.*, vol. 39, no. 1, pp. 84–102, Jan. 2019.
[3] E. F. Z. Santana, G. Covas, F. Duarte, P. Santi, C. Ratti, and F. Kon, "Transitioning to a driverless city: Evaluating a hybrid system for autonomous and non-autonomous vehicles," *Simul. Model. Pract. Theory*, vol. 107, Feb. 2021, Art. no. 102210.
[4] F. Shatu and M. Kamruzzaman, "Planning for active transport in driverless cities: A conceptual framework and research agenda," *J. Transp. Health*, vol. 25, Jun. 2022, Art. no. 101364.
[5] Y. Wiseman, "Driverless cars will make passenger rail obsolete [opinion]," *IEEE Technol. Soc. Mag.*, vol. 38, no. 2, pp. 22–27, Jun. 2019.
[6] I. Tomar, I. Sreedevi, and N. Pandey, "State-of-art review of traffic light synchronization for intelligent vehicles: Current status, challenges, and emerging trends," *Electronics*, vol. 11, pp. 2079–9292, Feb. 2022.
[7] L. Zhang, M. Khalgui, and Z. Li, "Predictive intelligent transportation: Alleviating traffic congestion in the Internet of Vehicles," *Sensors*, vol. 21, no. 21, p. 7330, Nov. 2021.
[8] U. Makhloga, "Improving India's traffic management using intelligent transportation systems," Tech. Rep., 2022.
[9] Á. Fehér, S. Aradi, and T. Bécsi, "Online trajectory planning with reinforcement learning for pedestrian avoidance," *Electronics*, vol. 11, pp. 2079–9292, Jul. 2022.
[10] P. Radanliev and D. De Roure, "Review of the state of the art in autonomous artificial intelligence," *AI Ethics*, vol. 18, p. 62, Jun. 2022.
[11] N. Shigeta and S. E. Hosseini, "Sustainable development of the automobile industry in the United States, Europe, and Japan with special focus on the vehicles' power sources," *Energies*, vol. 14, no. 1, p. 78, Dec. 2020.
[12] K. Rajagopalan, B. Ramasubramanian, S. Velusamy, S. Ramakrishna, A. M. Kannan, M. Kaliyannan, and S. Kulandaivel, "Examining the economic and energy aspects of manganese oxide in Li-ion batteries," *Mater. Circular Economy*, vol. 4, no. 1, pp. 1–22, Dec. 2022.
[13] D. Prasad, A. Anand, V. A. Sateesh, S. K. Surshetty, and V. Nath, "Accident avoidance and detection on highways," in *Microelectronics, Communication Systems, Machine Learning and Internet of Things*, 2022, pp. 513–528.
[14] A. R. Fayjie, S. Hossain, D. Oualid, and D. Lee, "Driverless car: Autonomous driving using deep reinforcement learning in urban environment," in *Proc. 15th Int. Conf. Ubiquitous Robots (UR)*, Jun. 2018, pp. 896–901.
[15] G. Karmakar, A. Chowdhury, R. Das, J. Kamruzzaman, and S. Islam, "Assessing trust level of a driverless car using deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4457–4466, Jul. 2021.
[16] Q. Chen, Y. Xie, S. Guo, J. Bai, and Q. Shu, "Sensing system of environmental perception technologies for driverless vehicle: A review of state of the art and challenges," *Sens. Actuators A, Phys.*, vol. 319, Mar. 2021, Art. no. 112566.
[17] P. Fergus and C. Chalmers, "Deep reinforcement learning applied deep learning," Tech. Rep., 2022, pp. 255–264.
[18] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Trans. Mobile Comput.*, vol. 13, no. 4, pp. 907–920, Apr. 2014.
[19] S. Kumar, P. Tiwari, and M. Zymbler, "Internet of Things is a revolutionary approach for future technology enhancement: A review," *J. Big Data*, vol. 6, no. 1, pp. 1–21, Dec. 2019.

[20] C. Suppatvech, J. Godsell, and S. Day, "The roles of Internet of Things technology in enabling servitized business models: A systematic literature review," *Ind. Marketing Manage.*, vol. 82, pp. 70–86, Oct. 2019.
[21] S. Wang, S. Zhang, R. Ma, E. Jin, X. Liu, H. Tian, and R. Yang, "Remote control system based on the Internet and machine vision for tracked vehicles," *J. Mech. Sci. Technol.*, vol. 32, no. 3, pp. 1317–1331, Mar. 2018.
[22] J.-H. Huh and K. Seo, "Blockchain-based mobile fingerprint verification and automatic log-in platform for future computing," *J. Supercomput.*, vol. 75, no. 6, pp. 3123–3139, Jun. 2019.
[23] T. U. Rehman, M. S. Mahmud, Y. K. Chang, J. Jin, and J. Shin, "Current and future applications of statistical machine learning algorithms for agricultural machine vision systems," *Comput. Electron. Agricult.*, vol. 156, pp. 585–605, Jan. 2019.
[24] M. L. Smith, L. N. Smith, and M. F. Hansen, "The quiet revolution in machine vision—A state-of-the-art survey paper, including historical review, perspectives, and future directions," *Comput. Ind.*, vol. 130, Sep. 2021, Art. no. 103472.
[25] J. Zhao, B. Liang, and Q. Chen, "The key technology toward the self-driving car," *Int. J. Intell. Unmanned Syst.*, vol. 6, no. 1, pp. 2–20, Jan. 2018.
[26] C. Luo, W. Yang, P. Huang, and J. Zhou, "Overview of image matching based on ORB algorithm," *J. Phys., Conf. Ser.*, vol. 1237, no. 3, Jun. 2019, Art. no. 032020.
[27] M. Bansal, M. Kumar, and M. Kumar, "2D object recognition: A comparative analysis of SIFT, SURF and ORB feature descriptors," *Multimedia Tools Appl.*, vol. 80, no. 12, pp. 18839–18857, May 2021.
[28] H. Fujiyoshi, T. Hirakawa, and T. Yamashita, "Deep learning-based image recognition for autonomous driving," *IATSS Res.*, vol. 43, no. 4, pp. 244–252, Dec. 2019.
[29] K. Xia, H. Fan, J. Huang, H. Wang, J. Ren, Q. Jian, and D. Wei, "An intelligent self-service vending system for smart retail," *Sensors*, vol. 21, no. 10, p. 3560, May 2021.
[30] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[31] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[32] Y. Jiang, G. Tong, H. Yin, and N. Xiong, "A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters," *IEEE Access*, vol. 7, pp. 118310–118321, 2019.
[33] T. Sangeetha, M. Kumaraguru, S. Akshay, and M. Kanishka, "Biometric based fingerprint verification system for ATM machines," *J. Phys., Conf. Ser.*, vol. 1916, no. 1, May 2021, Art. no. 012033.
[34] Z. Changhong and X. Reneng, "Intelligent laboratory access control system based on ZigBee technology," in *Proc. Int. Conf. Virtual Reality Intell. Syst. (ICVRIS)*, Jul. 2020, pp. 688–690.
[35] C. Xiong, "Design of intelligent garbage classification bin based on LD3320," in *Proc. Int. Conf. Signal Process. Mach. Learn. (CONF-SPML)*, Nov. 2021, pp. 7–10.
[36] J. Kowalski, A. Jaskulska, K. Skorupska, K. Abramczuk, C. Biele, W. Kopeć, and K. Marasek, "Older adults and voice interaction: A pilot study with Google home," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–6.
[37] A. Lee, K. Oura, and K. Tokuda, "MMDAgent—A fully open-source toolkit for voice interaction systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8382–8385.
[38] D. Wang, X. Wang, and S. Lv, "An overview of end-to-end automatic speech recognition," *Symmetry*, vol. 11, no. 8, p. 1018, Aug. 2019.
[39] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "SpecAugment: A simple data augmentation method for automatic speech recognition," 2019, *arXiv:1904.08779*.
[40] K. Xia, X. Xie, H. Fan, and H. Liu, "An intelligent hybrid–integrated system using speech recognition and a 3D display for early childhood education," *Electronics*, vol. 10, no. 15, p. 1862, Aug. 2021.
[41] P. Fu, D. Liu, and H. Yang, "LAS-transformer: An enhanced transformer based on the local attention mechanism for speech recognition," *Information*, vol. 13, no. 5, p. 250, May 2022.
[42] J.-W. Hu, B.-Y. Zheng, C. Wang, C.-H. Zhao, X.-L. Hou, Q. Pan, and Z. Xu, "A survey on multi-sensor fusion based obstacle detection for intelligent ground vehicles in off-road environments," *Frontiers Inf. Technol. Electron. Eng.*, vol. 21, no. 5, pp. 675–692, May 2020.
[43] Z. Wang, Y. Wu, and Q. Niu, "Multi-sensor fusion in automated driving: A survey," *IEEE Access*, vol. 8, pp. 2847–2868, 2020.

RUI TONG was born in China, in 1998. He received the bachelor's degree in electrical engineering and automation from Shanghai Dianji University, in 2020. He is currently pursuing the master's degree in electrical engineering with the University of Shanghai for Science and Technology, Yangpu, Shanghai. His research interests include predictive control and parameter identification.

QUAN JIANG was born in China, in 1963. He received the B.Eng. degree in electrical engineering from the Hefei University of Technology, China, in 1983, and the M.Eng. and Ph.D. degrees in electrical engineering from Southeast University, China, in 1986 and 1991, respectively. He is currently the Head and a Professor with the Department of Electrical Engineering, University of Shanghai for Science and Technology, Shanghai, China. He has rich research experience in PM motors and drives, switched reluctance motors and drives, and dc motors. He has published more than 110 academic articles. He is the coauthor of two books. He was granted 20 patents in the USA, China, Singapore, and Japan. His current research interests include the design, control, and testing of electric machines, electric drives, power electronics, finite element analysis of electromagnetic fields, and applications of micro-controller, DSP, and FPGA devices.

ZUQI ZOU was born in China, in 1997. He received the bachelor's degree in electrical engineering and automation from Shanghai Dianji University, in 2020. He is currently pursuing the master's degree in electrical engineering with the University of Shanghai for Science and Technology, Yangpu, Shanghai. His research interest includes motor control.

TAO HU was born in China, in 1997. He received the bachelor's degree in electrical engineering and automation from Changzhou University, in 2020. He is currently pursuing the master's degree in electrical engineering with the University of Shanghai for Science and Technology, Yangpu, Shanghai. His research interests include embedded technology and the control method of permanent magnet synchronous motors.

TIANHAO LI was born in China, in 1998. He received the bachelor's degree in electrical engineering and automation from the Henan University of Science and Technology, in 2021. He is currently pursuing the master's degree in electrical engineering with the University of Shanghai for Science and Technology, Yangpu, Shanghai. His research interests include embedded technology, the application of power electronics technology in power systems, the modulation and control strategies of multilevel converters, and the development of high-power power electronics devices.