Proceedings of CONF-MPCS 2024 Workshop: Quantum Machine Learning: Bridging Quantum Physics and Computational Simulations

DOI: 10.54254/2753-8818/52/2024CH0160

Research on application method of intelligent driving technology based on monocular vision sensor

Zeyu Zhang

School of Mechanical and Electrical Engineering, Chongqing Jiaotong University, Chongqing, China

200165@yzpc.edu.cn

Abstract. With the development of driverless cars, intelligent driving technology is increasingly used in the automotive industry, monocular vision sensor plays an indispensable role in intelligent driving technology because of its simple structure, low cost and abundant information. This paper discusses and optimizes the application of the monocular vision sensor in intelligent driving. The basic principle and key technologies of the monocular vision sensor are described in detail. In the specific application of the monocular vision sensor, this paper focuses on the monocular vision sensor's depth learning network, multi-information fusion technology, improved target detection and tracking algorithm. Through in-depth research and analysis, a series of optimization strategies based on the monocular vision sensor, such as the FAST Region Convolutional Neural Network (FAST–RCNN) vehicle target detection method and improved Scale-Invariant Feature Transform (SIFT) algorithm, are proposed. Finally, this paper summarizes the intelligent driving technology based on the monocular vision sensor and suggests that the monocular vision sensor will play a more important role in intelligent driving technology. Future research shall focus on improving the accuracy of the algorithm, such as the development of end-to-end convolutional neural network fusion methods, the proposed depth multi-modal sensor fusion network, and so on.

Keywords: Monocular vision sensor, intelligent driving, SLAM, deep learning network, multi-information fusion technology.

1. Introduction

Intelligent driving technology has become a research hotspot in the field of modern transportation. As an important part of intelligent driving systems, vision sensor plays a vital role. This paper focuses on the application of the single-binocular vision sensor in intelligent driving technology, aiming at discussing its working principle, application method, and optimization strategy, and analyzing its future development trend.

By capturing the image information of the surrounding environment, the vision sensor can provide abundant spatial and temporal data support for the intelligent driving system. Monocular vision sensor is widely used in intelligent driving because of its simple structure and low cost.

The applications of monocular vision sensors in intelligent driving include environment perception and target detection, ranging and positioning, multi-task fusion and optimization, etc.

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/52/2024CH0160

Monocular vision is widely used in the environment perception of autonomous vehicles because of its simple structure and low computational cost. Monocular vision target recognition technology based on depth learning, such as Faster R-CNN and its improved version, can effectively improve the speed and accuracy of target recognition [1]. The simulation results show that the relative error is 2.71% in indoor environment and 3.81% in outdoor environment. Compared with the traditional algorithm, the proposed method can effectively improve the detection and ranging accuracy of the obstacles ahead, and its practicability and effectiveness are verified. This has important implications for the navigation control of autonomous vehicles under complex road conditions [2].

For the process of feature extraction and matching, a clustering obstacle detection method combining Scale-Invariant Feature Transform (SIFT) feature points is proposed. Firstly, the region of interest (ROI) of the image region is extracted, and the horizontal and vertical edge detection methods are used to identify the possible obstacles in the image region. Then, the following step is to calculate whether each ROI region contains obstacles, and remove the region that does not detect obstacles from the image. The remaining regions were used for the extraction of SIFT feature points. Then, K-means clustering is applied to these feature points to locate the obstacles accurately.

To reduce the influence of different size forward-looking targets on the depth of field estimation, a pyramid structure is adopted to preprocess the input image. During the follow-up training, the depth estimation problem was transformed into an image reconstruction problem, and a new loss function was designed, which could replace the true depth label [3].

Monocular vision simultaneous localization and map building (SLAM) provides a method to realize simultaneous localization and map building in an unknown environment. By relying only on monocular camera sensor data, this approach can achieve positioning and map construction. However, due to the relatively low computational complexity, monocular vision SLAM technology is widely used in small robots. Nevertheless, there are still some problems with this technique in terms of robustness and scale uncertainty [4].

Through the in-depth study of the application method of the monocular vision sensor in intelligent driving technology, this paper expects to further promote the development of intelligent driving technology, contribute to the construction of a safer, more intelligent and more efficient transportation system, and propose a series of optimization schemes. This paper not only sorts out the relevant theories in the field of intelligent driving but also provides a valuable reference for practical application.

2. Multitasking integration and optimization

2.1. Multi-tasking framework MGNet

MGNet, or Multiview Geometry Network, is a multi-task framework designed specifically for monocular Geometry scene understanding. It consists of two core tasks: panoramic segmentation and self-monitoring monocular depth estimation. In the task of panoramic segmentation, the model can segment the image and obtain the full semantic information and instance-level scene details, while in the aspect of self-supervised monocular depth estimation, the model automatically measures the monocular video depth using the constraints of the depth camera and the geometric environment [5].

2.2. Multi-sensor information fusion

In recent years, many researchers have begun to explore the combination of visual information and other sensor information to improve the accuracy and robustness of location. These fusion schemes usually use multiple sensors on the same hardware platform to collect data, to achieve higher positioning accuracy and better robustness. A fusion SLAM algorithm of monocular vision and inertia based on extended Kalman filter (EKF) and graph optimization (VISLAM) is proposed [4]. The algorithm improves the localization precision and the cost ratio of computation in three-dimensional space and realizes the localization with low cost and high precision. The EKF-based monocular vision and inertial fusion odometer are implemented to obtain motion estimation with less delay. The global

DOI: 10.54254/2753-8818/52/2024CH0160

map-aided EKF feedback mechanism is introduced, and the positioning accuracy is further improved by solving the system of linear equations.

3. Optimization strategy of monocular vision sensor in intelligent driving

In this chapter, the optimization strategy of depth learning framework, multi-information fusion technology, target detection and tracking algorithm and SLAM technology are presented, to improve the accuracy and robustness of the intelligent driving vehicle.

3.1. Optimization of the deep learning framework

Advanced depth learning framework such as depth residuals network is used to improve the accuracy of image recognition. These frameworks can effectively improve the performance of the model by increasing the depth of the network and introducing residual learning mechanism. At the same time, the use of very deep convolution networks can be considered, these networks improve the accuracy of recognition through the use of the small-scale convolution kernel.

3.2. Optimization of multi-information fusion technology

There is a certain risk of failure when using visual sensors to detect vehicles within the visual range. This risk is even greater in the non-line-of-sight condition. To solve this problem, it is suggested to construct a perceptual fusion framework based on deep learning [6]. The model combines complementary potential embedding with many advanced fusion strategies, and can effectively fuse image, radar, acoustic, seismic and other sensing modes. This model significantly surpasses the traditional single-peak detection method and greatly improves the accuracy of vehicle tracking and detection under non-line-of-sight conditions.

The validity of the model is verified on the multi-mode ESCAPE datasets. The experimental results show that the proposed fusion technique improves the vehicle detection performance by 33.16% on average compared with the visual detection method under the condition of 30-42% Natural Language Operating System (NLOS). On the more challenging multimodal NuScene datasets, the systems developed based on this model are on average 22% better than the competing approach. These results fully demonstrate the superiority and practicability of the model.

A functional model combining multi-vision information fusion is proposed. The feature fusion module is the core of the functional model, which combines the information of multiple visual sensors effectively, thus greatly improving the accuracy and robustness of detection and tracking [6]. The experimental results show that the proposed method is superior to the traditional single-vision sensor method in vehicle detection and tracking.

3.3. Optimization of target detection and tracking algorithm

An improved FAST Region Convolutional Neural Network (FAST-RCNN) vehicle target detection method based on binocular vision ranging is proposed. With the rapid development of intelligent driving technology, the demand for high-quality vehicle target detection algorithms is increasing. In this paper, a method of distance measurement based on binocular stereo vision is proposed to solve the limitation of traditional algorithms in speed and accuracy. Experimental results show that compared with FAST-RCNN, the proposed algorithm can significantly speed up the vehicle detection process and shorten the time by 42 ms. At the same time, the accuracy of the algorithm is only 2.4%, which shows high accuracy and good real-time performance. This innovative algorithm not only effectively improves the accuracy of vehicle detection, but also provides an advanced data processing platform for autonomous driving systems.

To improve the efficiency and reliability of the algorithm, an improved SIFT algorithm is proposed. Firstly, the feature descriptor is reduced from 128-dimension to 24-dimension, which makes the feature space closer to the image space. Then, the trilinear interpolation method is introduced to enhance the relationship between the feature points, thus significantly reducing the possibility of mismatching. Finally, a random sampling consistency algorithm is used to remove the mismatches.

DOI: 10.54254/2753-8818/52/2024CH0160

The experimental results show that compared with the original SIFT algorithm, the new algorithm improves the matching speed and accuracy significantly while keeping a higher number of feature points, especially, it performs well in the condition of angle change and illumination change, which can meet the requirement of real-time SLAM.

3.4. Real-time data processing and optimization

In order to improve the perception ability of intelligent vehicles to the vehicle information in the urban complex environment, a vehicle detection method combining radar and vision is proposed. By introducing the process of target tracking, the accuracy of vehicle position and velocity estimation is enhanced. An adaptive vehicle detection method based on target depth of field is designed, which takes into account the vehicle's driving speed, distance and visual angle, etc., the adaptive recognition of different types of vehicle information is realized. Finally, a target tracking method based on the combined framework of kernel correlation filter and extended Kalman filter (KCF-EKF) is proposed to track the vehicle accurately. The experimental results show that the proposed method shows good reliability and robustness in many traffic environments and weather conditions [7].

3.5. The latest improved SLAM algorithm based on monocular vision

In order to improve the accuracy and speed of feature matching, the traditional Oriented FAST and Rotated BRIEF (ORB) algorithm is improved, the scale space theory is introduced, and the possible range of feature points is predicted according to the prior information of robot motion, this method can effectively reduce the operation time and improve the matching accuracy [8].

To solve the problem of scale uncertainty in monocular visual SLAM, researchers begin to explore how to integrate other sensor data into monocular visual SLAM. For example, the methods of combining monocular images with wheel odometer data and inertial measurement unit (IMU) data can improve positioning accuracy and robustness to some extent.

4. Future trends in intelligent driving based on monocular vision

In this chapter, the development trend of monocular data forehead sensors in intelligent driving technology and the latest application methods, such as camera auto-exposure algorithm, end-to-end convolutional neural network fusion method, depth multi-mode sensor fusion network and hybrid multi-sensor fusion framework, are introduced.

4.1. Automatic camera exposure algorithm based on feature point detection

An innovative camera automatic exposure algorithm based on feature point detection is proposed, which aims at obtaining images with rich texture and clear details by adjusting exposure time [9]. In the current self-driving vehicles, most of the camera-based automatic exposure algorithms are only suitable for static scenes. However, in dynamic environments, these algorithms often fail to achieve the desired results. By detecting the feature points and analyzing their positions on the image, the camera shutter speed can be used to adjust the exposure parameters, to obtain clear and rich details of the image.

4.2. End-to-end convolutional neural network

In order to improve the precision of environment perception and depth estimation, adopts a binocular stereo matching algorithm to extract dense parallax information [9]. Because binocular vision system has significant uncertainty, and lidar system is relatively sparse, so it is necessary to combine the parallax images of the two systems for fusion processing. Based on this, a new end-to-end convolutional neural network is designed for the fusion of binocular parallax and lidar parallax. This method not only enhances the accuracy of environment perception, but also improves the accuracy of depth estimation.

DOI: 10.54254/2753-8818/52/2024CH0160

4.3. Depth multimodal sensor fusion network

A novel deep fusion network is proposed, which aims at robust fusion of sensor data under severe weather conditions without relying on a large amount of ground truth data [10]. The self-adaptive single-time model is used to measure entropy to drive feature fusion. After training the clean data set, the network model can demonstrate its effectiveness on a wide range of validation data sets. With this approach, it is expected that existing sensor data can be utilized, thereby significantly reducing the amount of marking required. This innovative framework is expected to be an important milestone in the development of data fusion technologies in intelligent transportation systems in the future.

4.4. Hybrid multi-sensor fusion framework

A novel hybrid multi-sensor fusion pipeline architecture is proposed to support autonomous vehicle environment perception in road segmentation, obstacle detection and tracking [11]. In this framework, a full-convolutional neural network (FCNX) and EKF are used to estimate the nonlinear state of the system. The goal is to create a more cost-effective, lightweight, modular, and robust fusion system solution that can maintain sensor performance even when it fails or underperforms. For road detection, the FCNX algorithm is used to improve accuracy. Compared with the traditional road detection based on a monocular vision sensor, the algorithm improves the detection accuracy and keeps the real-time efficiency on the embedded system. Test results on more than 3k road scenarios show that FCNX performs better in a variety of environment scenarios. In addition, the real-time performance of the algorithm is proved by experiments, and the real-time processing capability is further verified by the actual sensor data.

5. Conclusion

The application of the monocular vision sensor in intelligent driving technology is discussed in this paper. For an in-depth analysis, this research introduces the imaging principle, calibration method, ranging principle and application of vision sensor, and its combination with depth learning, and analyzes the key role and advantages of the vision sensor in intelligent driving.

The vision sensor provides a solid foundation for intelligent driving technology with its rich information acquisition ability and increasingly mature technology. Monocular vision sensor has a wide application potential in target detection and recognition, road scene understanding and so on. At the same time, the continuous progress of artificial intelligence technology provides strong support for the application of the vision sensor in intelligent driving. With the introduction of deep learning and machine learning, the vision sensor can identify the target accurately, estimate the depth and process the complex road scene in real time. The combination of these technologies not only improves the performance of intelligent driving system, but also opens up a new way for its future development. However, the vision sensor also faces some challenges and limitations in intelligent driving, such as low computing speed, non-dimensional invariance and data fusion with other sensors. Future research needs to explore these problems and seek effective solutions.

Looking into the future, with the development of technology and the increasing demand for intelligent driving, the application of the vision sensor in intelligent driving will be more extensive and in-depth. It is expected to see more innovative research and practice that will drive the development of intelligent driving technologies and bring more convenient and safe experiences for human travel.

In a word, the monocular vision sensor is an important part of intelligent driving technology, its development and application will provide strong support for the progress of the intelligent driving field. Through continuous exploration and innovation, it is believed that vision sensors will play an even more important role in the future development of intelligent driving technology.

References

[1] Chen Yupeng. Research on automatic driving monocular vision target recognition technology based on depth learning [D]. Jilin University, 2019.

DOI: 10.54254/2753-8818/52/2024CH0160

[2] Xu Yang, Zhao Yanan, Gao Li et al.. Monocular vision-based vehicle detection and tracking [J]. Laser journal, 2020,41(05): 18-22.
[3] Ding Meng, Jiang Xinyan. Scene depth estimation method based on monocular vision in advanced driving assistance Système d' aide à la conduite, à l' exploitation et à la maintenance [J]. Journal of Optics, 2020,40(17): 137-145.
[4] Quan Meixiang. Research on SLAM algorithm based on multi-sensor information fusion [D]. Harbin Institute of Technology, 2021.
[5] Markus Schön, M. Buchholz et al. “MGNet: Monocular Geometric Scene Understanding for Autonomous Driving.” IEEE International Conference on Computer Vision (2021). 15784-15795.
[6] Debashri Roy, Yuanyuan Li et al. “Multi-Modality Sensing and Data Fusion for Multi-Vehicle Detection.” IEEE transactions on multimedia (2023). 2280-2295.
[7] Zhao Wangyu, Li Bijun, Shan yunxiao, etc.. Hybrid millimeter wave radar and monocular vision for front vehicle detection and tracking [J]. Journal of Wuhan University (Information Science Edition) , 2019,44(12) : 1832-1840.
[8] Hu Wei, Liu Xingyu. An improved SIFT algorithm for image matching in unidirectional SLAM [J]. Electro-optic and control, 2019,26(05): 7-13.
[9] Chen Kun. Road environment sensing technology based on binocular vision and lidar fusion [D]. Zhejiang University, 2020.
[10] Mario Bijelic, Tobias Gruber et al. “Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather.” Computer Vision and Pattern Recognition (2019). 11679-11689.
[11] Babak Shahian Jahromi, Theja Tulabandhula et al. “Real-Time Hybrid Multi-Sensor Fusion Framework for Perception in Autonomous Vehicles.” Italian National Conference on Sensors (2019).