Review of Dynamic Obstacle Avoidance for Autonomous Driving Based on Binocular Vision

Danni Xu *

College of Computer Science, Beijing University of Technology, Beijing, China

* Corresponding author: xdn159@emails.bjut.edu.cn

Abstract. With the rapid development of automatic driving technology, the vehicle's ability to accurately perceive and quickly respond to the surrounding environment has become the key to realize safe and efficient driving. Within this landscape, binocular vision technology has distinguished itself as a game-changer. Its capacity to capture real-time, three-dimensional environmental data offers preeminent advantages in tackling one of autonomous driving's most critical challenges: dynamic obstacle avoidance. This paper begins by exploring the application of binocular vision in Mars rovers, illustrating its pivotal role in autonomous obstacle avoidance. Building on this foundation, this paper then delves into the fundamental principles of binocular vision, including depth calculation methods and the overall system architecture. Then, this paper discusses in depth the dynamic obstacle avoidance strategies in autonomous driving, which are closely dependent on the real-time and accurate environmental information provided by the binocular vision system. In the concluding sections, this paper shines a spotlight on the primary hurdles confronting current dynamic obstacle avoidance technologies. These challenges range from adapting to monotonous environments like arid desert, to pushing the boundaries of hardware capabilities, and refining algorithms to achieve split-second decision-making. This paper provides valuable references and lessons for the practical application of automatic driving perception technology.

Keywords: Autonomous driving, binocular vision, dynamic obstacle avoidance.

1. Introduction

Autonomous vehicles have led a ground-breaking revolution in the field of transportation, aiming to enhance the road operational efficiency and driving safety [1]. The obstacle perception system, a core component of intelligent driving, is responsible for detecting obstacles and understanding road structures [2]. In the face of the intricate real-world environment and the demand for high-level intelligence, the obstacle perception system must exhibit superior accuracy, robustness, and density. Particularly when confronted with dynamic obstacles such as pedestrians, vehicles abruptly changing lanes, and cyclists, the performance requirements become exceptionally stringent. Given that obstacle avoidance is one of the most formidable challenges for self-driving vehicles, it is imperative for these autonomous systems to autonomously detect the presence of any obstructions in their trajectory and devise efficient strategies to effectively navigate around them [3].

Currently, most autonomous driving solutions use monocular vision systems. These systems, however, are highly dependent on deep learning as they cannot directly acquire depth information. This not only leads to high training time costs but also makes them prone to missed detections and results in relatively weak generalization capability for environment perception. Another acquisition solution uses LiDAR, but its high integration cost sets an obstacle to the mass production and application of self-driving cars [4]. In addition, the software level also relies on High-Definition map (HD map), which are expensive and only cover a limited number of cities.

In contrast, binocular vision solutions directly perceive geometric information such as depth and relative position of objects in three-dimensional (3D) space through disparity, showing significant advantages in terms of efficiency and reliability. At the cost level, the integration cost of binocular vision camera is lower than that of LiDAR; at the same time, the solution can still achieve good scene generalization with low data accumulation requirements, and can directly identify the scene and upload the point cloud to the in-vehicle control system, which effectively saves the ordering cost of

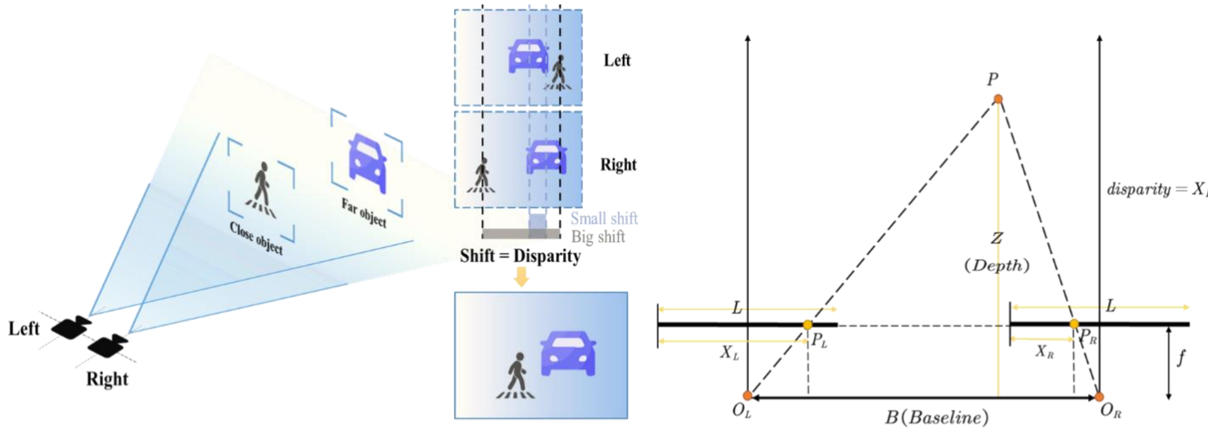
HD map. Presently, companies such as Mercedes-Benz, BMW, and DJI have incorporated binocular vision solutions.

This essay focuses on analyzing the research progress and trends in dynamic obstacle avoidance for autonomous driving based on binocular vision. It begins by briefly describing the basic principles of binocular vision, laying the foundational understanding for subsequent discussions. Secondly, it delves into the dynamic obstacle avoidance strategies employed in autonomous driving systems utilizing binocular vision, exploring the methodologies and algorithms developed to travel complex environments. Thirdly, the essay examines the major challenges faced in implementing effective dynamic obstacle avoidance based on binocular vision, highlighting the technical and practical hurdles that researchers and developers must overcome. Finally, the essay concludes by summarizing the key strategies, providing perspectives on the future directions and potential advancements in the field of dynamic obstacle avoidance for autonomous driving utilizing binocular vision technology.

2. Binocular Vision

The Mars Exploration Rover (MER), the earliest vehicle to employ binocular vision for driving assistance, was equipped with a stereoscopic system for exploring Mars's uncharted terrain. Subsequent rovers, notably the Zhu Rong, have also incorporated stereo binocular vision for navigation and obstacle avoidance. This adoption stems from its high reliability in unknown environments, enhancing driving safety [5].

Similarly, binocular vision holds significant application value for autonomous driving. Fig. 1 (a) shows the schematic of distance measurement based on binocular vision in autonomous driving. In Fig. 1 (b), two cameras are fixed on the same straight line, P is the target point, and its imaging points on the left and right cameras are PL and PR, respectively. OL and OR are the optical centres of the two cameras, respectively. The focal lengths of the two lenses are both f. The straight line connecting the two optical centres is called the system baseline, and its length is B. The same spatial point corresponds to two phase-plane coordinate points, and the positional difference between these two points is called the disparity. The distance Z from a spatial point P to the imaging plane is called the depth of the point.



(a) (b)

Figure 1. Binocular vision working principle [6].

Fig. 1(b) illustrates the projection of the model onto the xy-plane. Based on this projection, a relationship between disparity and depth is calculated, adhering to the geometric principles depicted in Fig. 1(b).

In ΔPO_LO_R, from the relationship of triangular similarity, it can be observed that:

$$\frac{B}{Z} = \frac{\overline{P_L P_R}}{Z-f}. \tag{1}$$

It is easy to see that:

$$\overline{P_L P_R} = B - \left[\left(X_L - \frac{L}{2}\right) + \left(\frac{L}{2} - X_R\right)\right]. \tag{2}$$

From this, one can obtain:

$$\frac{B}{Z} = \frac{B-\left[\left(X_L-\frac{L}{2}\right)+\left(\frac{L}{2}-X_R\right)\right]}{Z-f}. \tag{3}$$

Solving the equation, the depth can be derived as:

$$Z = \frac{f \cdot B}{X_L - X_R} = \frac{f \cdot B}{D}. \tag{4}$$

The formula for disparity D is as follows:

$$D = X_L - X_R. \tag{5}$$

The binocular depth extraction process involves several steps. It begins with acquiring binocular images from two cameras with a fixed baseline. These images are then pre-processed for quality enhancement. Next, features are extracted and matched across the two views. Following this, disparity maps are computed to quantify depth differences. These maps are then optimized for improved accuracy. Finally, they are converted into depth maps for various applications. The entire process is outlined in Fig. 2.

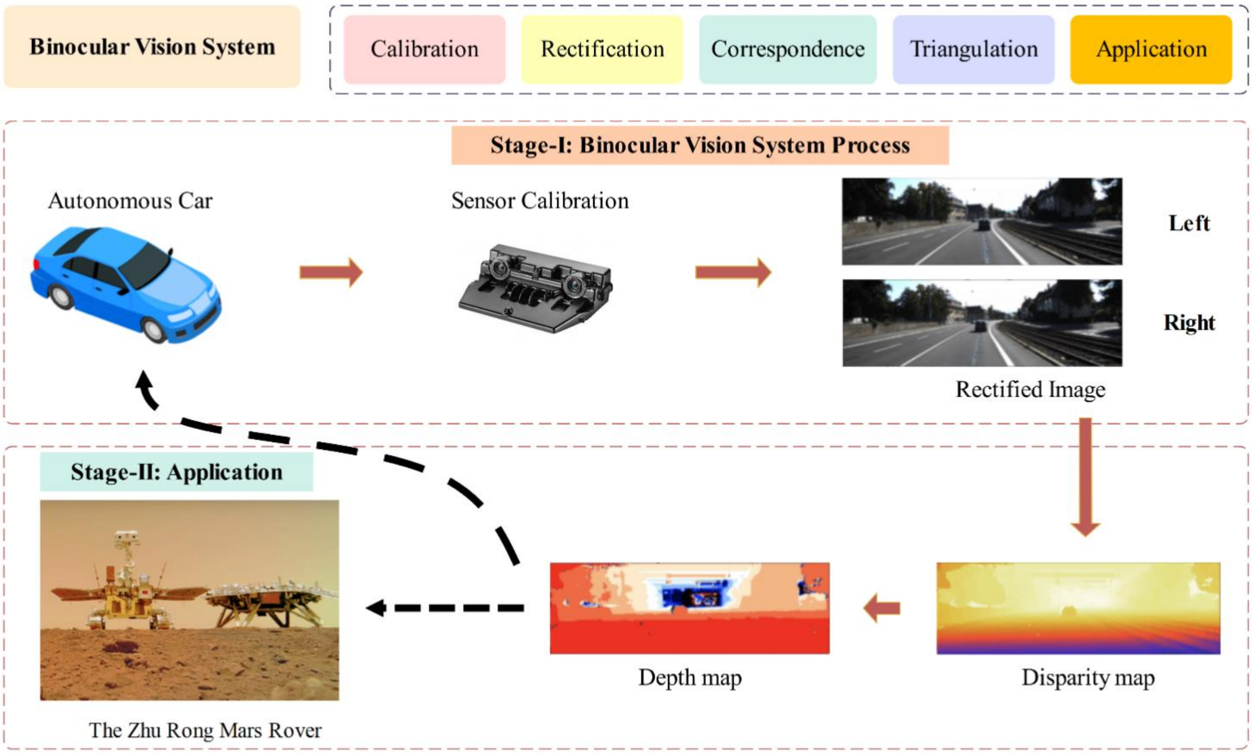


Figure 2. Overview of a stereo vision system [7] [8].

Binocular System Calibration: The core purpose of binocular system calibration revolves around accurately identifying the camera's intrinsic and extrinsic parameters, alongside its relative pose, leveraging an effective imaging model. This endeavor seeks to forge a precise mapping between spatial coordinates' object points and their corresponding projections onto the imaging surface.

Image Acquisition and Correction: The process of image acquisition necessitates the synchronized capturing of scene imagery through the binocular system. Nevertheless, discrepancies from the ideal model arise due to the specific lens characteristics of the camera and potential manufacturing imperfections. Consequently, it becomes vital to rectify the captured images, leveraging the calibration parameters derived from the preceding calibration stage, to ensure compatibility with the computational framework.

Binocular Matching: The establishment of correspondence between the pixels of the left and right binocular images is achieved based on the analysis of grayscale features. This process ultimately results in the acquisition of corresponding parallax images, which are essential for further depth analysis.

Depth Information Acquisition: The depth information of the scene is recovered from the parallax image, utilizing the underlying principles outlined in this section. This process involves the analysis of the parallax image to extract depth cues and generate a comprehensive depth map of the scene.

3. Dynamic Obstacle Avoidance Strategies

3.1. Vision-IMU Detection and Range Method Based Obstacle Avoidance Strategy

Wang proposes an obstacle avoidance method that integrates multiple technologies, including binocular vision imaging and an advanced detection method called Vision-IMU based detection and range method (VIDAR) [9]. Firstly, the imaging model and calibration of binocular camera were used to obtain accurate internal and external parameters. Then, VIDAR combined with Maximally Stable Extremal Regions (MSER) algorithm was used to extract the obstacle feature points, and the obstacle was accurately segmented by morphological closure operation, K-clustering, monocular and binocular vision. In terms of vehicle obstacle avoidance, an obstacle avoidance model based on the active binocular platform is designed. The obstacle detection of the target lane is realized through the rotation of the binocular camera, and the vehicle is controlled to change lanes or go straight according to the detection results. In the emergency braking scenario, an emergency braking system based on Time-To-Collision (TTC) model is constructed. The research results show that the method can realize the real-time detection and accurate segmentation of generalized obstacles in a large range under complex scenarios. Meanwhile, the effectiveness and feasibility of the designed obstacle avoidance model are verified through simulation and real vehicle verification, and the safe driving requirements of vehicles under different road conditions are met.

3.2. Obstacle Avoidance Strategy Based on Pedestrian Detection and Automatic Emergency Braking

Divaramakrishnan Rajendar et al. propose a stereo-vision based pedestrian detection and collision avoidance system for autonomous vehicles [10]. The system employs a dual-camera setup positioned at a fixed distance from each other to comprehensively scan the environment ahead of the vehicle. Leveraging stereo vision technology, it captures depth information of the surrounding scenery, enabling precise pedestrian detection and subsequent computation of a critical stopping distance. At the core of this architecture lies an autonomous emergency braking system (AEBS) controller algorithm, which meticulously calculates the safe braking distance between pedestrians and the vehicle. Should this distance fall below a predefined threshold (3.3 meters), the system seamlessly activates the AEBS, effectively mitigating the risk of collision. Extensive experiments have validated the superiority of this approach, demonstrating remarkable detection accuracy and collision avoidance capabilities, even under challenging scenarios such as complex environments, adverse lighting conditions, and low image quality, thereby contributing significantly to the reduction of traffic accidents.

3.3. YOLOv5s Deep Learning Based Obstacle Avoidance Strategy

Zhang et al. introduce an innovative target detection algorithm that harnesses the power of deep learning to tackle the complex challenges of dynamic road hazard avoidance in self-driving vehicles [11]. They use an advanced version of the YOLOv5s model, enhanced with several cutting-edge techniques to improve its performance. This optimized model is trained on a dataset consisting of self-made images and the KITTI dataset, resulting in improved accuracy for small object detection and overall performance. The research focuses on enhancing the vehicle's visual perception capabilities for real-time obstacle detection and tracking using a stereo camera setup. The optimized

YOLOv5s model is integrated with depth information from stereo vision to provide precise spatial positioning of detected hazards. This enables the vehicle to navigate safely by avoiding both stationary and moving obstacles. The proposed system is evaluated through simulations and real-world road tests, demonstrating its effectiveness in maintaining stable vehicle speed and lateral acceleration profiles during hazard avoidance maneuvers. The optimized model exhibits faster convergence during training and achieves better precision and means average precision (mAP) compared to baseline YOLO models. This results in more accurate and robust decision-making for dynamic obstacle avoidance, thereby enhancing the overall safety and reliability of the autonomous driving system.

3.4. Binocular LGMD-Based Obstacle-Avoidance Strategy

Inspired by the neural structure and processes underlying human cognition, particularly the visual, auditory, and tactile systems, as well as knowledge gained from daily driving tasks, Gary J.W. Xu et al. developed a high-level cognitive system for integrating collision sensing and avoidance [12]. This bio-inspired cognitive approach leverages the robustness, self-adaptability, and low computation consumption observed in practical driving scenes. Specifically, Zhang was inspired by the lobula giant movement detector (LGMD) neuron in locusts, which exhibits a strong ability to detect moving objects. Recognizing the high sensitivity of LGMD neurons in perceiving objects in motion, Zhang adapted this mechanism to promote vehicles' ability to detect approaching objects. Based on the insights from LGMD and considering two human factors—central fixation bias and binocular vision—Zhang designed a collision avoidance strategy. This strategy models central fixation bias and integrates it with the LGMD-inspired mechanism, utilizing left and right LGMDs to process images from the vehicle's front-facing cameras. The result is a system that, upon detecting a potential collision, initiates direction selection and collision body distance prediction simultaneously, enhancing the vehicle's ability to navigate safely.

4. Key Challenges

4.1. Environmental Adaptation Challenges

The binocular vision system is highly sensitive to ambient lighting. Variations in illumination angle and intensity between the two camera lenses can cause significant brightness disparities in the captured images. These discrepancies pose a formidable challenge to the stereo matching algorithm, as they can introduce distortions and inconsistencies that hinder accurate correspondence establishment between image pairs. Furthermore, the binocular stereo vision method relies on the matching of visual features to reconstruct the 3D structure of a scene. Consequently, the system encounters substantial difficulties in monotonous environments devoid of discernible textures, such as vast sky, white walls, or arid deserts. In these scenarios, the lack of distinctive features can lead to substantial matching errors or even complete failures in establishing reliable correspondences, ultimately compromising the overall performance and reliability of the binocular vision system.

4.2. Hardware Challenges

The implementation of binocular camera systems in automotive applications poses several hardware challenges. Firstly, there are stringent requirements for high accuracy in measuring the distance and speed of surrounding objects, necessitating binocular cameras with high resolution and excellent distortion control. Secondly, amidst the intricate and dynamic driving environment, stability and reliability are paramount, enabling the system to withstand factors like varying light conditions and dust accumulation. At the same time, the vehicle-mounted binocular camera needs to maintain a high degree of stability throughout the entire life cycle of the vehicle, which has a high demand for the synergistic optimization of structural materials, optical components and production processes. Lastly, high-performance computing chips are indispensable for acquiring and swiftly processing these high-resolution images, enhancing computational speed and ensuring the real-time, efficient functioning of the binocular vision system.

4.3. Performance and Algorithmic Enhancements

Firstly, the core challenge in binocular vision technology lies in the intricate stereo matching process, which demands precise alignment of corresponding points between left and right images. Scenes characterized by uneven illumination, weak textures, or repetitive patterns pose significant obstacles to this accuracy, necessitating the development of high-precision algorithms to enhance disparity precision. Secondly, the substantial computational volume and complexity inherent in binocular vision pose challenges for productization and miniaturization due to stringent performance requirements. The need for pixel-level matching and robust error rejection strategies exacerbates these issues. To achieve efficient implementation, algorithms must continually be optimized to minimize computational complexity while maintaining or enhancing matching accuracy. Thirdly, high-performance computing solutions are paramount. In the limited vehicle space, it needs to realize the high-precision distance measurement and avoidance function of vehicles, pedestrians and non-standard obstacles at a long distance, and also needs to have the processing capability of dealing with complex scenarios such as close-range congestion, which requires the computing platform to have strong algorithmic processing capability to meet the high standard requirements of real-time and accuracy. Ultimately, Binocular vision algorithms must possess extremely high computational efficiency and real-time response capability, as self-driving cars need to perceive their environment and make obstacle avoidance decisions within milliseconds.

5. Conclusion

With the rapid development of science and technology, autonomous driving technology has become an important trend in the future transport field. The promise of autonomous driving extends far beyond mere convenience. While it certainly liberates drivers from the tedium of navigating traffic, its true potential lies in revolutionizing road safety. By eliminating human error, which are the underlying cause of a vast majority of road accidents, autonomous vehicles have the power to save countless lives. Moreover, this technology is catalyzing a sweeping transformation of the automotive industry, driving innovation and efficiency across the board. At the heart of this autonomous revolution lies binocular vision technology. Much like how our two eyes work in tandem to perceive depth, this system equips self-driving cars with a pair of 'eyes'. This dual-camera setup allows the vehicle to construct a detailed 3D map of its surroundings, forming the cornerstone of precise obstacle detection and avoidance. The binocular vision system is able to capture and process the road scene in real time and generate high-precision depth maps, which provides strong data support for subsequent obstacle identification, tracking and prediction. When it comes to dodging obstacles on the ground, this paper delves into a variety of binocular vision-based obstacle avoidance strategies. They naturally blend the depth perception and obstacle characteristics from the binocular vision with the car's own movement and planned route. The result is a vehicle that can nimbly recognize and sidestep moving obstacles, much like an expert driver with swift reflexes. However, although binocular vision shows great potential in dynamic obstacle avoidance for autonomous driving, its development still faces many challenges. For example, complex environments such as light variations, extreme weather, and hardware conditions as well as limitations in real-time computation and processing capabilities are challenges for the further development of binocular vision technology. In response to these challenges, algorithm optimization, sensor fusion, hardware upgrades and many other aspects need to be continuously developed in order to promote the continuous progress of autonomous driving technology. As technology continually advances and research deepens, it is anticipated that more efficient and reliable dynamic obstacle avoidance strategies leveraging binocular vision will emerge in the future, offering robust support for the safe, intelligent, and efficient operation of autonomous vehicles.

References

[1] Tijani A. Obstacle avoidance path design for autonomous vehicles – a review. Technium: Romanian Journal of Applied Sciences and Technology, 2021, 3 (05): 64 - 81.

[2] Chen Long, et al. Milestones in autonomous driving and intelligent vehicles—part ii: Perception and planning. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53 (10): 6401 - 6415.

[3] Jain Sanjyoti, Malhotra Isha. A review on obstacle avoidance techniques for self-driving vehicle. International Journal of Advanced Science and Technology, 2020, 29 (06): 5159 - 5167.

[4] Zhang Xinyu, et al. A scene comprehensive safety evaluation method based on binocular camera. Robotics and Autonomous Systems, 2020, 128: 103503.

[5] Mao Xiaoyan, M Zhifu, C Jianxin, et al. Parallel Design and Implementation of Stereo Vision Algorithm of Zhurong Mars Rover. Journal of Deep Space Exploration, 2022, 9 (2): 202 - 210.

[6] Stereo vision. Available: https://en.smartereye.com/pages/stereo-vision.

[7] First Scientific Images of Mars Landing by Tianwen-1 Rover Unveiled. Available, 2021: https://www.cnsa.gov.cn/n6758823/n6758838/c6812123/content.html.

[8] Cohen Jeremy. Stereo Vision vs Sensor Fusion - Which approach is better. Available: https://www.thinkautonomous.ai/blog/pseudo-lidar-stereo-vision-for-self-driving-cars/.

[9] Wang Liming. Research on Vehicle obstacle avoidance method based on binocular vision. Ph.D. dissertation, Shandong University of Technology, 2023.

[10] Rajendar S, Rathinasamy D, Pavithra R, et al. Prediction of stopping distance for autonomous emergency braking using stereo camera pedestrian detection. Materials Today: Proceedings, 2022, 51: 1224 - 1228.

[11] Qiu C, Tang H, Yang Y, et al. Machine vision-based autonomous road hazard avoidance system for self-driving vehicles. Scientific Reports, 2024, 14 (1): 12178.

[12] Xu G J W, Guo K, Park S H, et al. Bio-inspired vision mimetics toward next-generation collision-avoidance automation. The Innovation, 2023, 4 (1).