A Review of Obstacle Avoidance Research based on Binocular Vision UAV

Cheng Liu
Jimei University, Xiamen, Fujian, China

Abstract: The widespread application of UAVs puts forward higher requirements for autonomous flight and safety, among which obstacle avoidance ability is particularly critical. Binocular vision technology has become a research hotspot in the field of UAV obstacle avoidance due to its advantages of passive perception, cost-effectiveness, and ability to provide dense depth information. This review systematically reviews the research status and development trend of UAV obstacle avoidance technology based on binocular vision.

Keywords: Binocular Imaging; Binocular Stereo Vision Imaging; Multi-Camera 3D Imaging; Eye Camera Ranging

1. Introduction

With its flexible mobility and low cost, UAV has been widely used in logistics and distribution, agricultural plant protection, power inspection, disaster rescue, geographical surveying and mapping, film and television aerial photography and other fields. As application scenarios extend to complex urban environments (such as high-rise gaps, indoor warehouses) and dynamic scenarios (such as crowds and traffic flows), the environmental uncertainty faced by drones has increased sharply. Autonomous obstacle avoidance capabilities have become the core requirements for ensuring flight safety and improving mission reliability.

Early drones relied on a single sensor for obstacle avoidance, but there were significant drawbacks: high accuracy but expensive, high power consumption, bulk weight not suitable for small drones, and difficult to perceive transparent obstacles such as glass.

Ultrasonic/infrared sensors: short working distance (usually < 5m), susceptible to environmental interference (temperature, humidity), only suitable for low-speed near-ground scenarios.

Monocular vision: Depends on motion to estimate depth (such as SLAM, optical flow), high computational complexity, uncertain absolute scale of depth estimation, and poor robustness of dynamic scenes.

Therefore, an active perception solution with high cost performance, environmental adaptability and real-time is needed.

Binocular vision imitates the principle of human binocular parallax, and calculates depth information by simultaneously collecting two images, which has unique advantages:

Passive Perception: Doesn't rely on active light sources, has strong concealment, and is suitable for military or sensitive areas.

Rich information: Provides RGB texture and dense depth map at the same time, and supports geometric semantic fusion perception.

Cost and power advantages: Consumer-grade binocular camera modules (such as IntelRealSense, ZED) are lightweight, low power consumption, and easy to integrate.

Moderate measurement accuracy: When the baseline design is reasonable (0.1-0.3m), the accuracy can reach the centimeter level within 10m, which meets the needs of most obstacle avoidance scenarios.

Despite significant progress in binocular visual obstacle avoidance, its practical application in complex scenarios still faces severe challenges:

Real-time bottleneck: The delay of dense depth estimation and planning algorithms on the embedded platform (< 100 ms) is difficult to meet the requirements of high-speed flight.

Inadequate environmental robustness:

Lighting upheaval (bright/dark light), weather disturbances (rain and fog), and untextured areas (white walls, sky) cause matching to fail.

Transparent/reflective objects (glass curtain walls, water surfaces) cause depth estimation errors.

Dynamic obstacle interaction: trajectory prediction and collaborative avoidance strategies for moving targets such as pedestrians and vehicles are not yet mature.

Therefore, in-depth research on high-robust, low-latency, and intelligent binocular visual obstacle avoidance technology has great theoretical and engineering value for promoting the autonomous upgrading of UAVs and expanding the application boundaries.

Binocular vision UAV obstacle avoidance research is the frontier of the interdisciplinary integration of computer vision (stereo vision), robotics (perception, planning, control), artificial intelligence (deep learning, reinforcement learning), multi-sensor fusion, embedded systems and other disciplines. It continues to put forward new theoretical challenges (such as perception under extreme conditions, real-time reliable decision-making, resource-constrained intelligence), and promotes the development and breakthrough of basic theories in these fields.

This technology is the key to unlocking the large-scale, safe, efficient, and autonomous application of drones in complex real-world scenarios. It directly improves flight safety, expands application boundaries, reduces costs and improves efficiency, and as a core enabling technology, it effectively promotes the upgrading of the UAV industry and the vigorous development of the low-altitude economy.

2. Research Direction

2.1 Binocular Imaging Research

First, the camera imaging uses a pinhole imaging model**,** which collects an inverted image of the object, and its size is proportional to the size of the object, which is related to the focal length of the camera. After the image taken by the monocular camera is determined by the detection algorithm, the image pixel coordinates need to be further converted into physical coordinates to accurately obtain the physical position of the target object, and the actual physical distance represented by a single pixel needs to be calculated, that is, the proportional relationship between the image size and the object size is obtained, and four coordinate systems are established between them, and the connection between them is studied in this chapter. Due to the fact that the manufacturing technology does not reach the degree of the ideal pinhole imaging model, the real world does not fully conform to the theoretical lens, and the physical surface of the general camera lens is not completely theoretical, resulting in the distance between the two pixels in the captured image, and the proportional relationship between the two points corresponding to the actual physical space. In binocular cameras, the third-dimensional distance can be calculated according to the parallax of the matching point, which requires that the imaging plane of the two cameras is the same plane, the plane is strictly perpendicular to the optical axis of the two cameras, and the acquired image meets the conditions of row alignment [1].

2.2 Two Structures of Binocular Stereo Vision Imaging System

In binocular stereo vision imaging system, the left and right cameras can have a variety of different positions. The following two situations are mainly discussed. One is that the position and orientation of the left and right cameras are not required to be placed arbitrarily, so that the optical axes of the two cameras are not parallel but intersect at a certain point in space. The other situation is to require the left and right cameras to be parallel to each other, so that the two images taken are coplanar and perpendicular to the optical axis, and make the image plane horizontal coordinate axis collinear, generally called this system is a binocular stereo vision system with parallel alignment configuration, its plane diagram is as shown in the figure, the advantage of the system is that the camera imaging model is simple, if the model parameters of the two cameras in the system are the same, the imaging model of the system can be regarded as a combination of two monocular camera imaging models of cockroaches; If the plane of the two images is coplanar and perpendicular to the optical axis, and the horizontal axis of the image plane is collinear, then the outer pole is at infinity, and the row of the image array is the pole line, The three-dimensional matching corresponding points can be carried out along the corresponding rows, which can reduce the dimensionality of the matching search range, thereby greatly reducing the amount of matching operations [2]

2.3 Multi-Camera 3D Imaging System

The multi-camera 3D imaging system adds more camera angles on the basis of the binocular system, thereby alleviating the problem of data loss caused by the blind spot of the camera's field of view. In recent years, scholars in related fields have designed and built a series of multi-camera 3D imaging systems with different effects suitable for different scenarios using different system structures and imaging methods. These imaging systems can be roughly divided into multi-camera 3D imaging systems based on passive vision and multi-camera 3D imaging systems based on active vision according to different imaging principles. The multi-camera 3D imaging system based on passive vision mainly adopts a passive 3D imaging scheme based on multi-view stereo matching [3-4]. Dang et al. [5] built a ten-camera 3D imaging system with 5 cameras built on two curved frames to collect multi-view images of the bovine body, and finally realized the complete 3D imaging of the bovine through the global matching of multi-view images. Yang et al. [6] proposed a multi-camera depth estimation system based on semantic segmentation, which first obtains the depth map between adjacent cameras through multi-camera calibration and stereo matching, then splices the multi-camera images through feature extraction and matching, and then uses the semantic segmentation network to identify and segment the object, and establishes an object height library. The accuracy of the system is susceptible to mismatching issues. For three-dimensional imaging of coded fringe structured light, Deng et al. [7] built a four-nocular structured light 3D imaging system using a 5G power module as the object, using a DLP projector as the active light source to project sine bar structured light, using the multi-wavelength phase expansion method to achieve high-precision phase calculation, and finally realizing multi-eye imaging through phase matching. Yu Jinmiao [8] designed a four-camera three-dimensional imaging system based on sinusoidal stripe structured light using an aero engine blade as the object, using a digital projector as the active light source to achieve monocular three-dimensional imaging through phase solution, and then splicing the point clouds from four perspectives to achieve multi-view imaging. In order to solve the problems of existing multi-camera 3D imaging systems relying on large overlapping angles between cameras and cumbersome calibration, Feng Chuang et al. [9] combined multiple single-camera and single-projector structured light systems to build a multi-structured light 3D imaging system, and optimized the system parameters by reconstructing standard parts, so that the system can achieve high-precision 3D imaging without overlapping viewing angles. Coded striped structured light 3D imaging systems using DLP projectors as active light sources can quickly achieve 3D imaging of objects, but due to the high cost of the instrument and limited brightness and resolution, it is susceptible to ambient light interference in large-scene applications, and the accuracy and stability are poor. In contrast, line lasers have low price and energy concentration, so they have higher accuracy and robustness in large industrial scenarios with more complex environments. For multi-eye laser 3D imaging, Zhang Boxiao [10] designed a four-eye structured light color 3D imaging system, which uses a line laser as the light source and scans through a conveyor belt, and the four cameras are divided into two sets of binocular 3D imaging and four sets of monocular imaging through binocular stereoscopic matching and laser triangulation respectively. However, the system's reliance on conveyor belt scanning limits the system's field of view and application scenarios.

2.4 Eye Camera Ranging

Binocular camera calibration: Binocular camera calibration is the basic link in the binocular vision system, which ensures that the system can accurately recover three-dimensional spatial information from the image. The calibration process involves determining the internal and external parameters of the camera and correcting lens distortion. The Zhang Zhengyou camera calibration method is a single-plane checkerboard camera calibration method proposed by Professor Zhang Zhengyou in 1998, which uses a binocular camera to collect multiple sets of checkerboard images by continuously moving the position of the checkerboard grid prepared in advance, and uses the stereoCameraCalibrator toolkit in MATLAB to calibrate the binocular camera, export the results, and extract the inner and outer and distortion parameters of the binocular camera [11].

3. Summary and Discussion

The camera imaging adopts a pinhole imaging model, and the actual physical distance of the pixels needs to be calculated to convert the image pixel coordinates into physical coordinates after the monocular camera is imaging, which involves four coordinate systems, and there is camera distortion due to manufacturing technical problems, so the relevant calibration model is introduced for the focal length and distortion coefficient. Binocular cameras can calculate the third-dimensional distance based on the parallax of the matching point, but there are strict conditions. The content focuses on binocular and multi-camera imaging and binocular ranging, and first expounds the binocular imaging research: the monocular camera is based on the pinhole imaging model, and needs to establish four coordinate systems to realize the transformation of pixel coordinates to physical coordinates, and it is also necessary to obtain the focal length and distortion coefficient through the calibration model to correct the proportion gap caused by lens manufacturing errors. Then, two structures of binocular stereo vision imaging system are introduced, in which the parallel alignment configuration has the advantages of simple model and dimensionality reduction to match the search range. Then, it is explained that the multi-camera 3D imaging system increases the viewing angle on the basis of binocular to reduce data loss, and is divided into two categories: passive and active vision. Finally, the binocular camera ranging is mentioned, pointing out that calibration is the basic link, it is necessary to determine the internal and external parameters and correct the distortion, and the Zhang Zhengyou camera calibration method is commonly used to collect multiple sets of images by moving the checkerboard, and complete the calibration and extract parameters with the help of the MATLAB toolkit.

References

[1] Zhang Zhan. Research on Key Technologies of 3D Reconstruction Based on Binocular Vision[D].University of Chinese Academy of Sciences (Shenyang Institute of Computing Technology, Chinese Academy of Sciences), 2019.

[2] Luo Gui'e. Research on Binocular Stereoscopic Vision Depth Perception and Three-Dimensional Reconstruction [D]. Central South University, 2012.

[3] Luo H, Zhang J, Liu X, et al. Large-scale 3d reconstruction from multi-view imagery:A comprehensive review[J]. Remote Sensing, 2024, 16(5): 773.

[4] Liu Z, Bai H, Sa G, et al. Design of multi-view vision system for real-time suturetracking and 3d reconstruction[C]// Proc. Int. Conf. Mech. Design. 2021.

[5] Dang C G, Lee S S, Alam M, et al. Korean Cattle 3D Reconstruction from Multi-View3D-Camera System in Real Environment[J]. Sensors, 2024, 24(2).

[6] Yang F, Liming Z, Yi Z, et al. Multi-camera System Depth Estimation[C]// 2022 IEEE6th Information Technology and Mechatronics Engineering Conference (ITOEC). IEEE,2022, 6: 1202-1207.

[7] Deng Fangwei, Huang Shijun. High-precision 3D structured light measurement technology for 5G power module[J]. ZTE Technology Journal, 2024,30(05): 75-80.

[8] Yu Jinmiao. Research on Three-dimensional Shape Measurement Method of Leaves Based on Multi-eye Stereoscopic Vision[D]. Wuxi: Jiangnan University, 2023.

[9] Feng Chuang, Luo Mingling, Deng Gaoxu, et al. Three-dimensional measurement method of multistructured light based on reference standard[J]. China Laser, 2024, 51(17): 174-183.

[10] Zhang Boxiao. Research on multi-eye structured light color 3D imaging technology[D]. Handan: Hebei University of Technology, 2023.

[11] Cui Dapeng, Li Min. Research on real-time ranging system based on binocular stereoscopic vision[J].Heilongjiang Science,2025,16(18):118-121.

