

Study on Autonomous Obstacle Avoidance Based On Guidance

Zhizhang Hu^{1,*}, Guangdi Hu¹, Xiaoyuan Zhu², Zihao Lei¹, and Haipeng Wan²

¹School of Mechanical Engineering, Southwest Jiaotong University, No.999 Xi'an Road, Chengdu, China

²School of Information Science and Technology, Southwest Jiaotong University, No.999 Xi'an Road, Chengdu, China

Abstract. DJI Guidance is an entirely new visual sensor navigation system, which has a function of real-time transmission of the depth of field images and other data. It is a robust optical sensing platform which significantly promotes the development of unmanned aerial vehicle (hereinafter referred to as UAV). On this basis, we have further studied and developed Guidance in application to the field of UAVs to also cut a striking figure, by applying the threshold and binaryzation to the single frame of a static image generated from Guidance. Utilizing OpenCV's contour detection algorithm, we find all the closed contour information and re-plan the feasible region. Also, we specially re-project the depth information to achieve three-dimensional reconstruction of the image. Then combining the re-planned feasible domain with three-dimensional reconstruction, we can perform real-time obstacle avoidance and mapping. The results demonstrate that our improvement can efficiently complete the obstacle avoidance and real-time mapping in the specific circumstances.

INTRODUCTION

In recent years, more and more attention has been paid to UAVs as advanced technology, and the research of UAV autonomous obstacle avoidance has become one of the focuses of the study. For an obstacle avoidance system, the detection and location of obstacles is the first step and the most critical step to obtain the surrounding environment. Currently, the widely used methods of detection and location can be classified into two categories: one is based on UAV's simultaneously localization and mapping (hereinafter referred to as SLAM) [1] to surrounding environment. Thus the accurate location information of UAV and stereo relationship between UAV and obstacles is obtained. The other is based on precise location information provided by external facilities, such as global positioning system (hereinafter referred to as GPS) and motion detection system. Apparently, only the dynamic obstacle avoidance system can solve the UAV collision problem in the real world. In the realm of autonomous obstacle avoidance system, considering both of the overall cost and effectiveness, the visual sensor is a better choice. Guidance, introduced by DJI, is an innovative visual-sensing platform but there is few related research at present. In this paper, the obstacle avoidance of UAV is explored by use of DJI Guidance.

1. Feasible Region Decision

Utilizing binocular vision system of Guidance and image processing, location and distance information of

obstacles is obtained. In this way, UVA can decide the feasible region.

1.1 Binocular vision stereo correction

In view of the binocular vision technology, it is generally using two cameras to obtain two images of the same scene in different spatial locations, then matching feature points and obtaining internal parameters of cameras by calibration. In this way, the coordinated information from disparity map could be computed. When the two image plane is completely line aligned, the calculation of disparity is the most simple. However, due to the multiple unavoidable errors, it is impossible for the two cameras to complete line alignment accurately. Therefore, the stereo correction has become indispensable. By stereo correction, the image planes of two cameras can be projected back to the same plane. Since the front parallel planes are finite, there are bound to be constrained by selection, including maximum view overlapping and distortion minimization. There are many algorithms to help implement selection. In this paper, we choose OpenCV machine vision language as the development platform. There are two main algorithms: one is the Hartley algorithm, and the other is the Bouguet algorithm. In light of the calibration mode that Guidance can use, we choose the Bouguet algorithm that generates fewer distorted images generation[2] and maximizes image overlap. [3]

First, the rotation matrix and the translation vector between the two cameras of the binocular camera can be represented as follows:

* Corresponding author: kevinhoo@my.swjtu.edu.cn

$$\begin{aligned} R &= R_r(R_l)T \\ T &= T_r - R(T_l) \end{aligned} \quad (1)$$

Where: R denotes the rotation matrix to be calculated, and T denotes the translation vector to be computed. The subscript l and r refers to the parameters of the left camera and the parameters of right camera, respectively.

The spatial position relation is shown in Figure 1:

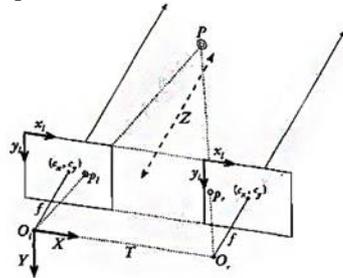


Fig. 1. Spatial Position Relationship.

We set the R matrix to rotate the right camera image plane to the left camera image plane. In order to minimize the distortion, the R matrix needs to be divided into two parts, called synthetic rotation matrix R_l and R_r . Rotating each camera in half so that the camera can be coplanar, but line alignment is still not available, therefore a new matrix R_{rect} is needed, which transforms the left camera pole to infinity and aligns the pole line horizontally. In order to calculate R_{rect} , it is necessary to establish a rotation matrix starting with the pole e_1 . The main point (c_x, c_y) is the origin of the left image, so the direction of the pole can be:

$$e_1 = \frac{T}{\|T\|} \quad (2)$$

Without any other restrictions, the next vector e_2 must be orthogonal to e_1 . So one can choose the orthogonal direction of the main light, that is:

$$e_2 = \frac{\begin{bmatrix} -T_y T_x & 0 \end{bmatrix}^T}{\sqrt{T_x^2 + T_y^2}} \quad (3)$$

The third vector is the cross product of the first two vectors. In this way, the third vector is:

$$R_{rect} \begin{bmatrix} (e_1)^T \\ (e_2)^T \\ (e_3)^T \end{bmatrix} \quad (4)$$

From this, line alignment can be achieved by setting: $R_l = R_{rect} r_l, R_r = R_{rect} r_r$.

Eventually, one can calculate the projection matrix, which transforms the three-dimensional points under the coordinates into the two-dimensional points under two-dimensional homogeneous coordinate system:

$$P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ w \end{bmatrix} \quad (5)$$

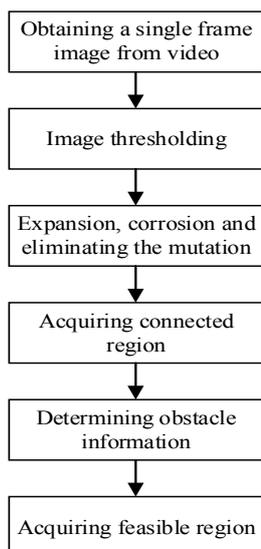
1.2 Feature Point Matching

In general, one feature in an image may have many candidate matching pairs in another image, and there is only one structure with the same name, so ambiguous matching occurs. [4] In this case, it is necessary to complete the matching according to a specified matching algorithm and experience. Simultaneously, the selection of feature points is critical, since it plays a critical role in the accuracy and rapidity of binocular distance measurement. If the feature points are too few, the judgment of contours of the objects would be inaccurate, and too many feature points will cause the operation speed to decrease and increase the probability of matching errors. Based on the OpenCV language, we tried two algorithms: 1. Block stereo matching algorithm that is, using the sum of absolute differences (hereinafter referred to as SAD) to find the matching points between the two images that have been corrected. [5] 2 SIFT feature extraction algorithm. SIFT features have scale invariance, that is, the matching effect is less affected by the angle of the shot and the change of light. [6] In order to compare these two algorithms, we use VisualStudio in a desktop computer with Intel Core i7-4790CPU (3.60GHz) to build a simulation test environment based on OpenCV. By comparing the samples of 10 pixels to 320*240, we found that the speed and effect of the SAD algorithm in the texture-rich region were better than that of SIFT algorithm. However, in the low texture region, the matching error rate of the SAD algorithm is high, the need for late interpolation correction kept enhancing, thus reducing the speed of operation. In contrast, the matching output of the SIFT algorithm are all within the acceptable range, and the average processing speed is 17% higher than the SAD algorithm. Therefore, we select SIFT algorithm to match feature points.

1.3 Image Processing

With image processing, the feasible region can be obtained, and the related process is shown in table one:

Table 1. Processing Process



For a scene, binocular cameras of Guidance obtains two images with a tiny difference of shooting angle simultaneously, with which the depth image could be computed. Depth image contains the distance information of each pixel, in which the more extensive the gray scale is, the closer the distance between UVA and obstacle is. For the video information returned by Guidance, it is processed through each single frame static image in this paper. The video captured by Guidance can obtain a single frame static image in real time and can be stored directly in the Mat structure of OpenCV so that it is convenient for further processing and calculation. In the laboratory, we used the Guidance binocular camera to capture two static images at the same time as shown in Figure 2 and Figure 3:



Fig.2. Left Camera Image



Fig.3. Left Camera Image

After the selection and matching of feature points in two pictures, the disparity image is obtained. Then the depth image can be computed from the disparity image and internal parameters of the binocular camera obtained

from calibration. The derivation of the relation between depth image and disparity image is shown in Figure 4 and related formula:

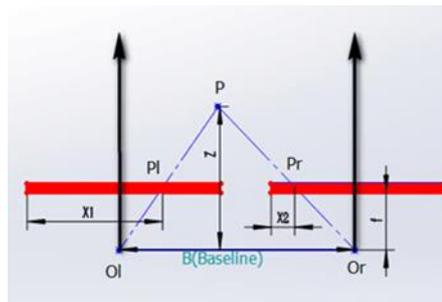


Fig.4. Relationship Derivation

In Figure 4, O_l and O_r respectively represent the position of two camera optical center, the obstacle P 's position in the two camera image plane are apiece 1 and 2, and disparity:

$$\Delta x = x_l - x_r \tag{6}$$

The mathematical expression of depth z can be obtained from similar triangle relations:

$$\begin{cases} \frac{x_l}{f} = \frac{x}{z} \\ \frac{x_r}{f} = \frac{x-b}{z} \end{cases} \tag{7}$$

Moreover, then depth z can be calculated:

$$z = \frac{bf}{x_l - x_r} \tag{8}$$

From (8) we can see that depth z is inversely proportional to the disparity Δx . Therefore, the distance information of the obstacle can be calculated by matching the feature points measured in the binocular vision. In the experiment, disparity image (Figure 5) and depth image (Figure 6) is obtained.

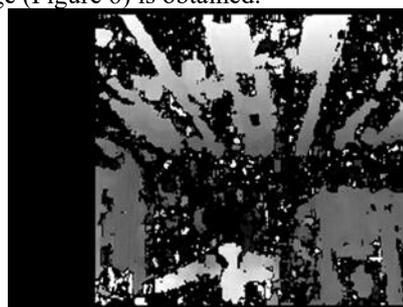


Fig.5. Disparity Image



Fig.6. Depth Image

In this paper, the obtained depth images are processed by thresholding. Since the grayscale information in the image represents the fixed distance quantity, we selected 125 (corresponding distance is 4.5m) and 60 (corresponding distance is 2.0 meters) as the threshold respectively for first and second thresholding processing to the same image after experiments. [7] Taking the instance of selecting 125 as the threshold to illustrate: During the processing, a pixel with a value higher than 125 is considered an obstacle-free object, i.e., at this point, the region is a feasible region and marked as 255. On the contrary, a pixel with the value that below 125 is considered an obstacle at that moment, labeled as 0 and as an infeasible region.

$$\begin{cases} f_s(i, j) = 255 & f_s(i, j) \geq 125 \\ f_s(i, j) = 0 & f_s(i, j) < 125 \end{cases} \quad (9)$$

By processing the image in real time, Guidance is able to obtain the feasible region without obstacle in the distance between 2m and 4.5m simultaneously, to realize the real-time monitoring and avoidance of the obstacle object. The purpose of threshold processing is to reduce the amount of data and improve the speed of real-time processing. In the experiment, processing speed has been enhanced by 7% compared to full gray scale processing.

Through the above threshold processing, we can get a binary image, as shown in Figure 7:



Fig.7. Binary Image

In Figure 7, one can see that there is some noise with erroneous distance information in the image, which can affect the judgment of the position of the obstacle information. Therefore, it is necessary to process the obtained binary image. In this paper, the mean filtering process is utilized. The mean filter can efficiently eliminate the salt and pepper noise and reduce the error distance information. Meanwhile, the filtering speed is faster and has better real-time performance. [8] By mean filtering, some of the noise in the image is effectively suppressed, as shown in Figure 8:



Fig.8. Filtered Image

In the image after the mean filtering, there is still a small area of unreasonable jump value in some regions, forming a slight mutation connected domain. The reason for this is the distance measurement error caused by light, shadow, texture and misalignment, which could be avoided by eliminating the smaller connection domain. [9] The size of the domain to be eliminated is related to the actual area corresponding to it. For example, in this paper, 125 is the threshold value, and the obstacle distance is 4.5m. In this way, 300 is selected as the size of the eliminating domain. After eliminating, the effect is shown in Figure 9.



Fig.9. Image after Eliminating

For the optimized binarized image, all the obstacles can be marked out by the contour detection algorithm, and the size and position information of the obstacle can be calculated. The results are shown in Figure 10, with a rectangular mark out the obstacles:

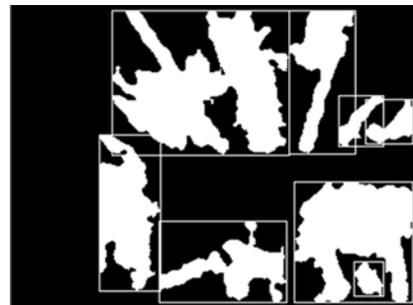


Fig.10. Contour Detection

In this image, all the obstacles of the location information can be obtained, and feasible region according to the size of the UAV in the distance that is within 4.5m can be obtained. Similarly, selecting 60 as the threshold of the image for the same treatment, feasible region within 2m is accessible. In this way, the real-time obstacle avoidance can be realized by simultaneously processing of continuous images in the real-time feedback of the UAV.

2 Three-dimensional Reconstruction

Due to the lack of environmental holism in the approach proposed earlier, implementing three-dimensional reconstruction (hereinafter referred to as 3D reconstruction) with feasible region detection to improve environmental wholeness is introduced.

2.1. 3D Reconstruction Technology

If the camera's internal parameter matrix is known, the reprojection matrix Q is available.

$$Q = \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 0 & f \\ 0 & 0 & \frac{-1}{T_x} & (C_x - C'_x)T_x \end{bmatrix} \quad (10)$$

Where: C'_x denotes the x coordinates of the main point in the right image, and the other parameters all come from the left image. If the main light intersects at infinity, then $C_x = C'_x$, and the lower right corner of the item is 0. [2] Present, given a two-dimensional homogeneous point and the disparity d associated with it, one can project this point into three dimensions:

$$Q \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} \quad (11)$$

The 3D coordinates are:

$$\left(\frac{X}{W}, \frac{Y}{W}, \frac{Z}{W} \right) \quad (12)$$

Based on this, 3D reconstruction can be achieved by spatial projection of the acquired depth information. In accordance with the processing order, the experimental results are as follows.



Fig.11. Left Camera Image

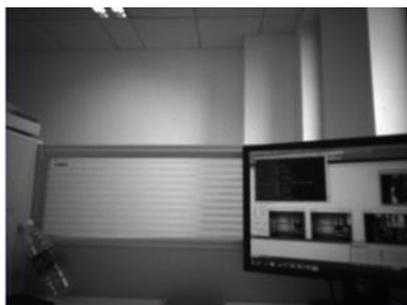


Fig.12. Right Camera Image



Fig.13. Disparity Image



Fig.14. Depth Image

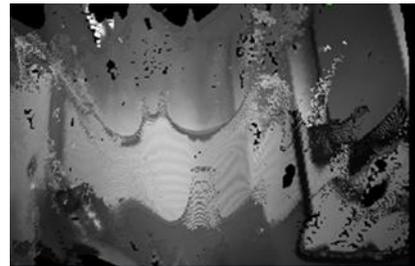


Fig.15. Result of 3-D Reconstruction

2.2. Accuracy Correction

Since the limited computing power of the Guidance itself can be very limited in response to the need for high precision reconstruction, according to the actual demand of flight and Guidance hardware conditions, after the experiment, we found the reconstruction accuracy can be reduced to a continuous reconstruction needs. The reason is that in the UAV operations, Guidance only needs to let the UAV know what in the flight environment is obstacles, as well as the obstacle and the UAV itself position relationship. There is no need to distinguish what obstacle is. At the same time, when each pixel of the disparity map is projected to the world coordinate system, knowing its precise location is unnecessary. What only needs to know is the region of location, into the treatment of some areas. In this study, the area length and width dimensions were 30 cm, and the height of the area was determined from the height histogram of the pixels falling in the area, and the risk of the region was determined according to the region and its coordinate distance. The implement of this method may result in the expansion of obstacle in the depth and the width maximum to 60cm so that the maximum diameter of the feasible region can be reduced by 60cm. Although the

diameter of the feasible region can be reduced, in determining the safety zone, the feasible region diameter must have been more than the fuselage diameter of a certain margin. This method not only simplifies the data processing but also improves the real-time performance without affecting the performance of obstacle avoidance of the UAV. After measurement, each 3D reconstruction period is 0.6 seconds, which can meet the security requirements of reconstruction.

2.3. 3D Global Map Building

The perceived information obtained during the UAV flight is relative to the body freezing coordinates, so the three-dimensional information generated by the current reconstruction is the three-dimensional data under the local coordinates generated at the time of reconstruction. Therefore, the global information of each time can be integrated into a global map by providing the global coordinates of the global position when the UAV is reconstructed. [10] Considering that the reconstruction cycle is short, the map missing parts can be directly connected between each reconstruction cycle, and the constructed map primarily conforms to the real three-dimensional scene in the space. In this way, SLAM is available.

2.4. Conjunction of Obstacle Detection and 3D Reconstruction

Implementing obstacle detection in conjunction with 3D reconstruction could realize high-performance path planning. Path planning is to find a safe and efficient collision avoidance path for the autonomous vehicle between the given starting point and the target point in the concurrent scene. [11] Due to the limited computing power of Guidance equipment and the specific requirements of the actual UAV flight operation, the path planning strategy of Guidance requires security first, followed by low computation. We use the feasible domain detection combined with 3D reconstruction for path planning. In each local 3D reconstruction cycle (0.6 seconds), matching the connected feasible domain with the 3D modeling completed in the same cycle. In this way, Guidance could mark the feasible region obtained from the binary image in the 3D environment, then calibrate the feasible region with 3D information. If the feasible region is judged as safe in its corresponding 3D environment, it can be regarded as a safe flight airspace.

3 Conclusions

In this paper, we propose our improvement and research based on Guidance multi-information sensor. First, we focus on the processing of depth image and binary image, then re-plan the feasible region. Additionally, we imitate the principle of human eye imaging with the utilization of Guidance's binocular camera system to complete the three-dimensional reconstruction of the front space, so as to be more effective for obstacle avoidance. Finally, combining three-dimensional reconstruction with

feasible region planning, we propose a new idea for obstacle avoidance. Because of the low precision matching of the feasible region obtained from binary image and three-dimensional reconstruction, the actual avoidance effect will be limited, which is our deficiency at this stage. How to match accurately is the breakthrough point of our future work. Simultaneously, combining neural network algorithm and genetic algorithm to optimize path planning is also the focus of prospective research.

Acknowledgements

This work has been supported by 2016 National-Funded Student Research Training Program under the Grant Numbers 201610613031.

References

1. Artieda, J., Campoy, P., Correa, J. F., & Olivares, M. Visual 3-d slam from UAVs. *Journal of Intelligent & Robotic Systems*, **55**(4-5), 299-321. (2009)
2. Bradski, G. R., & Kaehler, A. *Learning OpenCV - computer vision with the OpenCV library: software that sees*. (DBLP. 2008)
3. Xin, X.. *Embedded System Design Based on FPGA*. (CMP. 2005)
4. Xiaohua, W.. *Research on 3D Reconstruction Technology Based on Computer Vision*. (Doctoral dissertation, SDUST, 2004)
5. Yang, Q., Rong, X., Yu, X., & Guan, C. Intensity Weighted Stereo Matching Algorithm. *Affective Computing and Intelligent Interaction*, **137**,93-100.(2012)
6. Bingxiong L. An improvement of matching criteria and method of SIFT feature and its application. (Doctoral dissertation, DLUT, 2010)
7. Tony Lindeberg. Scale-space theory: a basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, **21**(2), 224-270, (1994)
8. Kejun, W., Xinyan, X., Zhen, R.. Highly efficient mean filtering algorithm. *Application Research of Computers*, **27**(2), 434-438, (2010)
9. Zhizao, L., Yingwu, Z., & Zhongkai, Zhen. Optimizing Algorithm of Labeling Connected Components in Binary Images. *Journal of Anqing Normal University(Natural Science Edition)*, **16**(4), 34-39. (2010)
10. Huahua, C. *Research on the Key Technology of Vision-based Navigation: Stereo Vision and Path Planning*. (Doctoral dissertation, ZJU, 2005)
11. Toogood, R., Hong, H., & Chi, W.. Robot path planning using genetic algorithms. *IEEE International Conference on Systems*. **1.1**,489-494. (1995)