

Three-dimensional Reconstruction based on Quadrotors with Binocular Vision

Yong Shen

College of Electronics and Information Engineering
Tongji University
Shanghai, China
avrilsin@outlook.com

Youling Yu

College of Electronics and Information Engineering
Tongji University
Shanghai, China

Danyu Bi

College of Electronics and Information Engineering
Tongji University
Shanghai, China

Claudio Melchiorri

Dept. of Electrical, Electronic and Information Engineering
University of Bologna
Italy

Abstract—Three-dimensional reconstruction is used to get more detailed information on the current scenes such as the overlook of a city or a car. The process of reconstruction is based on binocular vision system. In this paper we apply binocular vision to quadrotors. As a result, more information is obtained in reconstruction when two quadrotors with cameras are searching in different circumstances. Video stabilization is used to guarantee the quality of images. Disparity map is computed by the semi-global stereo matching algorithm. The final reconstruction proves the proposed method to be effective.

Keywords—video stabilization; stereo matching; three-dimensional reconstruction; binocular vision.

I. INTRODUCTION

Last few years witnessed the rapid development of unmanned aerial vehicles among which quadrotor has been the most popular one. Quadrotor can take off and land vertically. Owing to its flexibility, quadrotor is widely applied to military filed and civilian domain. Most quadrotors choose monocular camera as onboard sensor. However monocular camera cannot acquire depth information. So to obtain more useful information, quadrotor formation that is already realized can be introduced.

In this paper, two quadrotors in formation are taken. Because the quadrotors are formed in parallel on same altitude, the two cameras make up a binocular vision system. We can make use of two images from binocular vision system to compute the corresponding depth information. This paper proposes a method which can reconstruct three-dimensional environment for quadrotors based on binocular vision. Video stabilization is used to preprocess the original video so that the effect of vibration of quadrotors is reduced.

II. VIDEO STABILIZATION

Cameras on quadrotors inevitably suffer vibration along quadrotors when they are flying. To make sure the high quality of images from cameras, it is necessary to use video stabilization. In the process of video stabilization, it is crucial to extract feature trajectories from a video. Optical flow is an approximation of the local image motion based on local derivatives in a given sequence of images. Considering the computing velocity and real time performance, I chose Lucas-Kanade method. It is commonly used to compute to compute optical flow. The Lucas-Kanade detector is a gradient base displacement detector and is the basis of many image pairing algorithms. It does not give an exact solution but a relatively good approximation [1]. The transformation from previous frame to current frame is calculated through using optical flow for all frames. There are three parameters in the transformation: dx , dy and da . dx and dy describe the position change of a pixel in the image plane. Meanwhile da refers to angle. The feature trajectories can be obtained by accumulating the transformation mentioned above. In a relatively stabilized video, features should move in a constant velocity in image space. Thus, it is required to smooth trajectories to get final results. Then Gaussian smoothing is adopted. Gaussian pyramid is computed in an iterative way to minimize the error of optical flow. Create a new transformation such that new transformation is the transformation rectified by smoothed trajectory. After applying the new transformation to the original video, the stabilized video consequently comes out.

III. EPIPOLAR CONSTRAINT

The epipolar constraint [2] states that the correspondence for a point belonging to the certain line across one of the image plane lies on a line on another plane. It is shown in Fig.1.

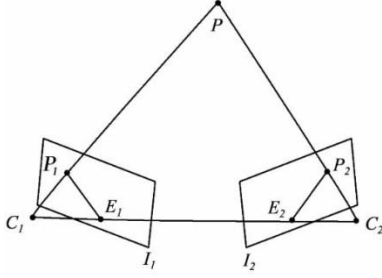


Fig.1. Epipolar constraint

Thus match precision is improved. Moreover, time consumed to search for a match point is shortened to a great extent.

For any point P in the world coordinate, there are two projection points respectively in two image planes: P_1 and P_2 . C_1 and C_2 are optical centers of two cameras. Line C_1P intersects image plane I_1 at point P_1 . On line C_2P , there are a lot of points whose projections on image plane I_2 lie at P_2 while their projections on image plane I_1 lie on line P_1E_1 .

IV. CALIBRATION

Calibration refers to obtaining some parameters through pictures taken by cameras. Intrinsic parameters of the two cameras can be found such as focal length, image center, parameters of lenses distortion and so on. Extrinsic parameters indicate relative position between two cameras. Rotation matrix R and translation vector T describe the relation between two cameras. Calibration is available both in OpenCV and Matlab. Chessboard works as calibration tool when it is placed in different angles. In this paper, OpenCV is used to implement calibration. Several stereo pairs of images from two cameras are taken as input. The result returns intrinsic and extrinsic parameters to be used in the following steps.

V. STEREO MATCHING ALGORITHM

Stereo matching algorithms can be divided into two categories: local matching algorithm and global matching algorithm. Local matching algorithm tries to find the points within the disparity range of the points to be matched. When the pair of points is chosen, calculate the similarity between them. The most similar pair is the final solution. Compared to global matching algorithm, local matching algorithm is much more straightforward. Global matching algorithm searches for disparity assignments that minimize an energy function over the whole stereo pairs by using a pixel-based matching cost.

Semi-global stereo matching such as semi-global block matching (SGBM) takes advantage of several one-dimensional constraints to replace a two-dimensional constraint. Its matching cost is also pixel-based.

A. Pixel-based matching cost computation

When two points respectively from two image planes are matched, there is a cost required to match them. The cost is

called pixel-based matching cost. To a certain point, match point with the smallest cost is the desired one. Birchfield and Tomasi proposed a matching cost calculation method which proved better robustness. The term $d(x_i, y_i)$ measures how unlikely it is that $I_L(x_i)$ and $I_R(y_i)$ are images of the same scene point [3].

The matching cost is calculated as

$$\bar{d}(x_i, y_i, I_L, I_R) = \max \{0, I_L(x_i) - I_{\max}, I_{\min} - I_L(x_i)\} \quad (1)$$

$$d(x_i, y_i) = \min \{\bar{d}(x_i, y_i, I_L, I_R), \bar{d}(y_i, x_i, I_R, I_L)\} \quad (2)$$

B. Cost aggregation

Pixel-based matching cost computation can be affected by many factors such as noise and similar areas. So sometimes smaller cost is calculated from the wrong match point. If the wrong match point is used, the corresponding disparity will be wrong. Consequently it will affect the points nearby, which will lead to more unexpected stereo point pairs. Therefore, semi-global stereo matching introduces an additional smoothness constraint into the calculation of the energy function. $E(D)$ [4] is regarded as the energy function dependent on the disparity image D :

$$E(D) = \sum_p (C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1]) \quad (3)$$

The first term is the sum of all pixel matching costs for the disparities of D . The second term adds a constant penalty P_1 for all pixels q in the neighborhood N_p of p , for which the disparity changes a little bit (that is, 1 pixel). The third term adds a larger constant penalty P_2 , for all larger disparity changes.

Dynamic programming is used to calculate the minimum of the energy function. Usually it can efficiently solve energy minimization in one-dimensional space. For a two-dimensional image, several different directions of dynamic programming are adopted to find the minimum of the energy function:

$$\begin{aligned} L_r(p, d) = & C(p, d) + \\ & \min \{L_r(p-r, d), L_r(p-r, d-1) + P_1, \\ & L_r(p-r, d+1) + P_1, \min \{L_r(p-r, i) + P_2\} \end{aligned} \quad (4)$$

The above formula shows the matching cost of point p with the disparity d in the direction of r . The final matching cost is the sum in all directions.

$$S(p, d) = \sum_r L_r(p, d) \quad (5)$$

C. Disparity computation

Two images from two cameras can be assumed as a base image and a match image. In the base image, the disparity d of the pixel p in the corresponding disparity image can be expressed as

$$D_{bp} = \min_d S(p, d) \quad (6)$$

In the match image, the disparity d of the match pixel q in the corresponding disparity image can be expressed as D_{mq} which means that the disparity image D_m can be determined from the same costs by traversing the epipolar line that corresponds to the pixel q of the match image. The final disparity map results from the comparison between D_{bp} and D_{mq} . When the difference between D_{bp} and D_{mq} is too large, the disparity of pixel p is considered invalid.

$$D_p = \begin{cases} D_{bp} & \text{if } |D_{bp} - D_{mq}| \leq 1 \\ D_{mq} & \text{otherwise} \end{cases} \quad (7)$$

Semi-global stereo matching applies different penalty gains according to different depth changes, which ensures smoothness constraint. The consistency check satisfies the uniqueness constraint in the process of determining the disparity map.

VI. THREE-DIMENSIONAL RECONSTRUCTION

To reconstruct the certain scene, it is critical to get the depth information. The distance between the camera lens and the object in front of it can be determined from the already obtained disparity map. Fig.2 shows the relationship.

O_R and O_T are the optical centers of the cameras. Point P projects on two image planes at P_R and P_T . X_R is the X axis coordinate of P_R on its image plane. X_T is the X axis coordinate of P_T . The distance between O_R and O_T is b . f expresses the focal length. $X_R - X_T$ is the disparity between the left image and right image. So given the disparity map, focal length and the baseline, the Z axis coordinate of point P can be easily computed [5].

$$Z = \frac{b \cdot f}{X_R - X_T} \quad (8)$$

Since the depth information is acquired, the spatial coordinates in the camera coordinate system can be calculated. A great number of three-dimensional coordinates form a database called point cloud. There are some invalid coordinates in point cloud. Before building the reconstruction, it is required to eliminate these invalid data which don't reflect the exact three-dimensional environment. Median filtering is adopted to remove discrete noise to ensure that the subsequent grid image does not appear excessive volatility. Make use of the rest coordinates to reconstruct the captured scene.

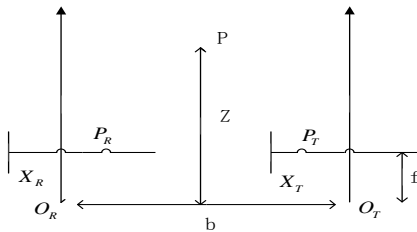


Fig.2. Depth and disparity

VII. EXPERIMENT AND RESULTS

In the experiment, the hardware platform consisted of two quadrotors with monocular cameras, and software platform included OpenCV function library, Visual Studio 2010 and Matlab 2008. AR.Drone was the quadrotor I used in the experiment. OpenCV was taken to process image. Matlab was used to reconstruct the scene.

Calibration was carried out beforehand. When the two cameras were stationary, several pictures were taken from two different points of view. In the scenes to capture, a chessboard was placed at different angles in front of two cameras. To make sure the precision of stereo calibration, a certain number of pictures were preferred. In this paper, seventeen pairs of images were used.

The flowchart of calibration process is shown in Fig.3. Only pictures with correct number of chessboard corners were used. When rough positions of corners were found, the next step was to find precise position of subpixel corners. Getting adequate information, we were able to carry out stereo calibration. After that, intrinsic and extrinsic parameters were found out. The calibration result was in TABLE I.

If calibration results don't meet the requirements, the error will propagate. We don't expect such occasion to happen so we need to check calibration quality. Because the output fundamental matrix implicitly includes all the output information, we can check the quality of calibration using the epipolar geometry constraint. The value of reprojection error is 1.10865 while the value of average reprojection error is 0.58781. It is apparently observed that both values are very small, which means the final calibration result is valid.

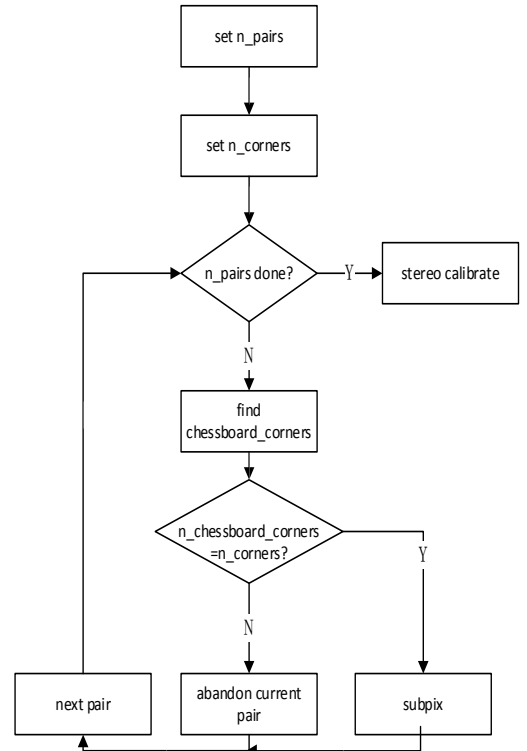


Fig.3. Calibration flowchart

TABLE I. CALIBRATION RESULT

Intrinsic parameters	Camera matrix1	$\begin{bmatrix} 677.89 & 0 & 327.27 \\ 0 & 677.89 & 219.52 \\ 0 & 0 & 1 \end{bmatrix}$
	Camera matrix2	$\begin{bmatrix} 677.89 & 0 & 319.63 \\ 0 & 677.89 & 215.30 \\ 0 & 0 & 1 \end{bmatrix}$
	Distortion vector1	$[0.33 \ -3.03 \ 0 \ 0 \ 0 \ 0 \ -4.95]$
	Distortion vector2	$[0.24 \ -1.53 \ 0 \ 0 \ 0 \ 0 \ -1.43]$
Extrinsic parameters	Rotation matrix	$\begin{bmatrix} 0.99 & -0.01 & 0.03 \\ 0.01 & 0.99 & 0.005 \\ -0.03 & -0.005 & 0.99 \end{bmatrix}$
	Translation vector	$\begin{bmatrix} -45 \\ 0.29 \\ 0.39 \end{bmatrix}$

Then slightly shake cameras and take pictures. Owing to video stabilization, the images obtained were not affected seriously. In the experiment, we took two pictures of the sofa with a dictionary on it. The information from stereo calibration was consequently used to remove lens distortions and turn the stereo pair of images in standard form, which was called rectification. The rectification results were shown in Fig.4 and Fig.5. The disparity map requires the rectified images from two cameras. The discussed algorithm SGBM was applied to compute disparity map which was Fig.6.

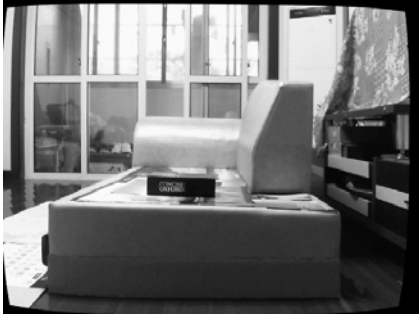


Fig.4. Rectified left image.



Fig.5. Rectified right image



Fig.6. 3D reconstruction

In the disparity map, the lighter color part was closer to the cameras while the deeper part was further. So we were able to get depth information from disparity map. The black points in the disparity map referred to points which were not matched successfully. Since there were just few black points in the disparity map, it could give us valid depth information.

Since the disparity map and the camera parameters were got, it was relatively easy to compute the point cloud. After processing the point cloud data, the three-dimensional reconstruction was built by Matlab. The Fig.7 showed the side view of the sofa.

It can be observed that there was a hump in the blue area. It was where the dictionary was placed. The backrest was in the area around $(-1000, -200)$ in plane Depth-Height. The reconstruction actually reflected the captured scene.

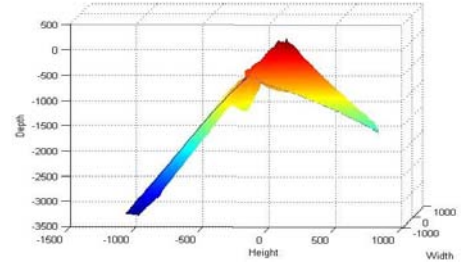


Fig.7. 3D reconstruction

VIII. CONCLUSION

This paper proposes a method which can reconstruct the three-dimensional environment for quadrotors based on binocular vision. Firstly video stabilization is implemented in order to guarantee the quality of images, which is the foundation of the following works. After calibration, algorithm SGBM is adopted to compute disparity map from rectified images. Disparity map is used to calculate the point cloud data. Finally the scene captured in the previous steps is reconstructed with the processed point cloud. The experiment has shown an acceptable performance of the binocular vision when we apply it to quadrotors in formation.

Future work includes improving and accelerating the stereo matching algorithm. Introducing more point cloud processing is another way to increase the precision of reconstruction.

REFERENCES

- [1] Zamudio Z, Lozano R, Torres J and Campos E. Stabilization of a helicopter using optical flow[C].2011 8th International Conference on Electrical Engineering, Computing Science and Automatic Control. 2007.
- [2] Gao B and Ma L Z. Semi-Global stereo matching combined with structure constraint[J]. Computer Application and Software.2009, 26(2): 244-247.
- [3] Birchfield S and Tomasi C. Depth discontinuities by pixel-to-pixel stereo[J]. International Journal of Computer Vision, 1999, 35(3): 269-293.
- [4] Hirschmüller H. Stereo processing by semiglobal matching and mutual information[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2008, 30(2): 328-341.
- [5] Wang J, Zeng Q, Xie F and Sun W Y. Research on 3D reconstruction of ATV's driving environment based on binocular vision[C].Electronics, Computer and Applications, 2014 IEEE Workshop on. IEEE, 2014: 556-559
- [6] Liu S, Yuan L, Tan P and Sun J. Steadyflow: Spatially smooth optical flow for video stabilization[C].Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014: 4209-4216.
- [7] Li X, Qin K, Yao P, et al. 3D surface reconstruction based on binocular vision[C].Mechatronics and Automation (ICMA), 2014 IEEE International Conference on. IEEE, 2014: 1861-1865..
- [8] Lee K Y, Chuang Y Y, Chen B Y and Ming O. Video stabilization using robust feature trajectories[C].Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009: 1397-1404..
- [9] Y.Lee and F.Fang. 3D Reconstruction of Polyhedral Objects from Single Parallel Projects using Cubic Comer [J]. Computer Aided Design, 2011, 43: 1025-1034.
- [10] L.Cao and J.Liu. What the Back of the Object Looks Like: 3D Reconstruction from Line Drawings without Hidden Lines [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(3): 507—517.
- [11] Z.Zhang. A flexible new technique for camera calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11), pp. 1330-1334.
- [12] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3-D machine vision metrology using off-the shelf TV camera and lenses, Journal of Robotics and Automation, 1987, 3(4), pp. 323-344.