The Real-Time Depth Map Obtainment Based on Stereo Matching

Fei Wang, Kebin Jia, and Jinchao Feng[(✉)]

Multimedia Information Processing Group, College of Electronic Information
and Control Engineering, Beijing University of Technology, Beijing, China
fengjc@bjut.edu.cn

Abstract. The depth measurement according stereo vision is a very popular method of measuring the depth information. But the calculation in stereo matching is large, time-consuming and in large errors in matching, when used in real-time systems to obtain depth information are often ineffective. This paper improves one of the local matching algorithms, builds a complete real-time system for calculating depth image with different SAD (Sum of Absolute Differences) matching window in different texture regions. In this paper, we capture images from the Bumblebee2 stereoscopic camera which mounted on a small unmanned car, then use Matlab calibrating the camera and do the calculation to correct the raw images and do stereo matching in VS2010 to get the depth map. This method is simple, real-time performance and adaptability, and the quality of the depth map calculated from this method is somewhat improved.

Keywords: Depth information · Stereo matching · Texture · SAD · Real-time

1 Introduction

With the development of the robot industry, unmanned car industry and virtual reality industry, three-dimensional reconstruction of machine vision has become more and more popular. Depth information acquired is the most basic and most important part of three-dimensional graphics. The main approaches for depth acquisition include laser scanning, structured light and stereo [1]. Laser scanning, also called ToF (Time of Flight), rangefinders measure the time it takes for the light to travel to the objects and back. This method can get the precise, distant data, but the device is heavy, not flexible, large and expensive. The method of structured light uses a projector to illuminate the object with the structured light and get the back information. The optical encoder technology that used in Kinect is a kind of structured light. This approach can get accurate depth data but just for limited range and is vulnerable to the outside light. The stereo imaging method used in this paper calculates the depth of the image by two images at different angles. This approach is simple, flexible, affordable, of course, also can get the accurate data. The camera calibration is easier and more precise, after Zhang Zhengyou proposed camera calibration method, so the main issues on stereo vision lies on stereo matching. In the efforts of many scholars, the performance of stereo matching is better and better.

In the 1980s, a visual computing theory proposed by Marr applied to the binocular stereo matching started the exploration of stereo vision theory. Today, stereo matching can be divided into global and local stereo matching. In local stereo matching, Kim et al. [2] described the applications of variable window algorithm, as set forth correlation function improves the matching precision on depth discontinuity area. Yoon et al. [3] proposed an adaptive weighted cost of window aggregation method, which firstly to calculate a weight for each pixel in the window. The weight for pixel is depended on the chromatic aberration and spatial position difference between the current pixel and center pixel. This method can get high-quality disparity map. Because of the large shape of the selection window, and the high complex weight calculation, the performance of the algorithm is not so good. The algorithm that Zhang et al. [4] proposed distributes for each pixel horizontal and vertical two arms that orthogonal itself. This algorithm also can get high quality disparity map, but need compare the color of center pixel with any other pixel which cost a lot of time and can't satisfy the real-time requirement. The global matching algorithms usually used are dynamic programming stereo matching method, graph cuts stereo matching method and the belief propagation matching method. In this paper, we improve the BM (Block Matching) algorithm in OpenCV which belongs to local stereo matching. BM algorithm uses fixed SAD window for stereo matching, has a good real-time performance. The proposed algorithm in this paper firstly extracts the edge of images by Canny operator and then decides the size of SAD window according to the area (edge area or not) the pixel belongs to. The algorithm has low time complexity and good robustness, it improves the matching accuracy.

2 Camera Model and Calibration

The real world is a three-dimensional world. Despite some controversy in binocular stereo vision and monocular stereo vision, it is complicated and difficult to get the corresponding depth information relying on only one image. However, we can calculate the depth map easily through two images obtained from a calibrated stereo camera. In order to analyze images with geometry theory, we need to model the system of imaging, and then process them with geometry methods.

Four coordinate systems [5] are used in the stereo calibration include the world coordinate and camera coordinate, the image plane coordinate and pixel coordinate. Information in the world coordinate system switch to the camera coordinate system through an external calibration parameter matrix W (including rotation matrix R and translation matrix T), and then to image plane coordinate system through the inner calibration parameter matrix K. Assuming that P (X, Y, Z, 1) is a point in the world, and p (x, y, 1) is the corresponding point in the pixel coordinate, so we can get the equation $p = sKWP$ (s is the scale factor).

The point of the camera coordinate Pc (x_c, y_c, z_c), can be expressed as Pc = WP,

$$\begin{bmatrix} xc \\ yc \\ zc \end{bmatrix} = \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{1}$$

where R represents a rotation matrix, T represents a translation matrix. This transformation just between two three-dimensional coordinates.

Inner calibration parameter matrix consist of camera focal length f and the center coordinate of imaging plane c. p(u, v) is a point in image plane, we can get the equation:

$$
\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} fx & 0 & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} xc \\ yc \\ zc \end{bmatrix}
\tag{2}
$$

In this way, we can find the correspondence between image plane coordinate and world coordinate according to the two matrices. One of the most important purposes of calibration is to calculate the matrices and the other is to obtain the distortion coefficients of cameras.

Since calibration with Matlab [6] is simpler than OpenCV, and gains wider recognition, we use Zhang calibration method in Matlab to do stereo calibration. Firstly, use stereo camera to capture images of calibration target from different angles. Then input the images into Matlab for calibration and copy the data into VS2010 for more operation.

3   Image Correction

Image distortion [7] from camera is divided into radial distortion and tangential distortion, the former refers to the deviation distance of ideal pixel position and the actual pixel position to the center of the image, mainly caused by the lens surface defects; the latter refers to the deviation angle from the ideal pixel position and the actual pixel position in the polar coordinate system, mainly due to the lens and the imaging plane is not parallel. Wherein the radial distortion can be divided into negative radial distortion (barrel distortion) and a positive radial distortion (pincushion distortion).

Model the radial, tangential distortion and establish an objective function to fit it. The distortion model:

$$
\begin{aligned}
u' &= u(1 + K1r^2 + K_2r^4 + K_3r^6) + 2P_1uv + P_2(r^2 + 2u^2) \\
v' &= v(1 + K1r^2 + K_2r^4 + K_3r^6) + 2P_2uv + P_1(r^2 + 2v^2)
\end{aligned}
\tag{3}
$$

Where K represents radial distortion coefficient, P for tangential distortion coefficient, r denotes the radius which is $\sqrt{u^2 + v^2}$, and (u', v') is the ideal coordinate in the image plane coordinate.

Objective function:

$$
\min F = \sum_{i=1}^{N} (u - u')^2 + \sum_{i=1}^{N} (v - v')^2
\tag{4}
$$

This function uses least square method which is used to solve the nonlinear distortion model directly, and simplifies the problem of solving this. We can do this in camera calibration and get parameter series {k1, k2, k3, p1, p2}, then put them into VS2010.

We use the cvStereooRectify function in OpenCV to correct the camera inner parament and make imaging plane of camera in geometry is parallel with Bouguet epipolar constraint method. Then send the obtained parameters to cvInitUndistortRectifyMap (), and get undistort rectify map which can save time in gaining corrected images next time. Finally give the map to cvRemap () and redraw the corrected image.

4   Obtaining the Depth Map

Binocular depth calculation is based on the principle of parallax [8]. For a point in space, its position in the image will be different, due to different shooting angle from stereo camera. The distance Z could be calculated with triangle similarity principle if the two image planes of the stereo camera are parallel. As shown in Fig. 1, we can get the equation

$$\frac{||T|| - (x_l - x_r)}{Z - f} = \frac{||T||}{Z} \Rightarrow Z = \frac{||T||f}{x_l - x_r} \tag{5}$$

where x_l, x_r are the abscissas of p_l, p_r, $||T||$ is the optical center distance which can be gained in the stereo calibration (Fig. 2).

Since the real-time requirements, local stereo matching method is used in the application. StereoBM algorithm calculates similarity with SAD, the biggest similarity point as the stereo matching point:

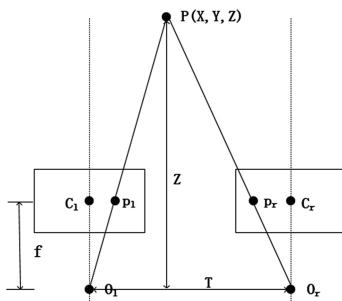$$SAD(x, y, d) = \sum_{i=-m}^{m} \sum_{j=-n}^{n} |I_l(x+i, y+j) - I_r(x+i+d, y+j)| \tag{6}$$

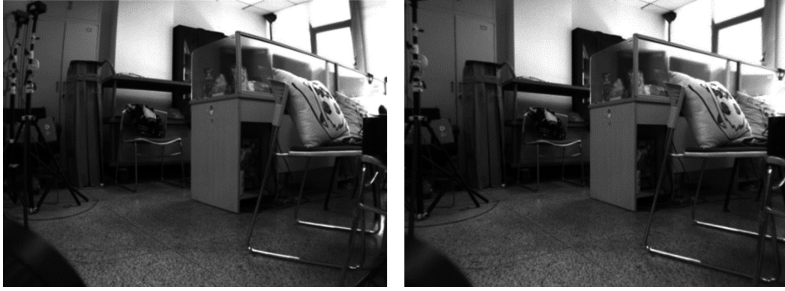**Fig. 1.** Binocular camera imaging schematic

Fig. 2. Images before (left) and after (right) correcting

Fig. 3. Depth map from small (left), medium (middle), large (right) SAD window

where d is the parallax value. And d which can make the SAD as the minimum is the real parallax value (Fig. 3).

We can set the state value include pre-filter setting, SAD window size etc., and use findStereoCorrespondenceBM () to get the parallax value for calculating the depth.

By analyzing the depth maps above, the smaller SAD window can make the depth map have more edge information, but with more noise and mismatch in smooth area. With the increase of the SAD window size the effect on smooth area gets better, but the time program consumed gradually increased and the edge region gradually blurred. So this paper combines the idea of [9], using the Canny operator to get the edge of the image, and then processing the image with small SAD window in edge area and big SAD window in non-edge area (Fig. 4).

First of all, we use canny function to get the edge of single view, and then obtain the edge area map with mask. Finally, according to the map determine the size of sad window the current point to use, and get the final disparity and depth map. This test is done in different texture scenes, and compared with the depth image.

As shown in Fig. 5, the depth map in figure (a) is gained with the smaller SAD window and can be seen the depth information more accurate in edge portion where is rich of texture, but with more mismatching points in regions with lower degree of texture;
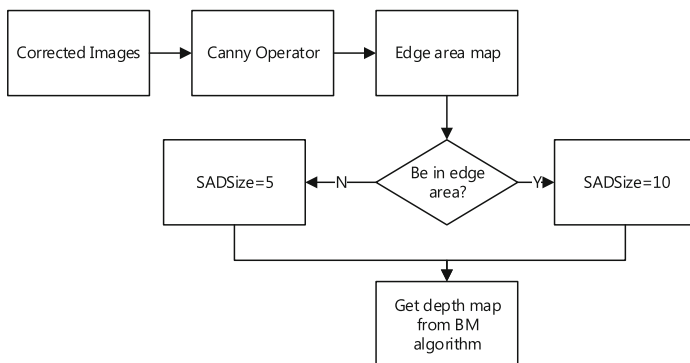
Fig. 4. Matching algorithm flowchart
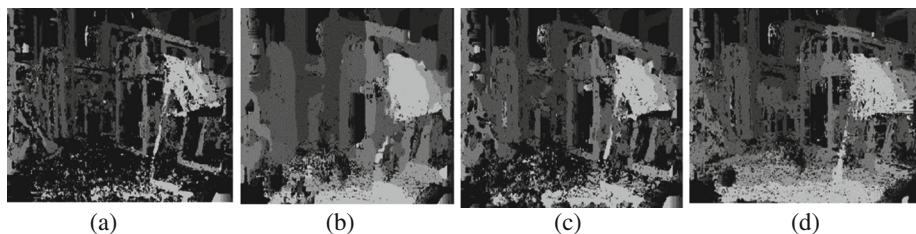

| (a) | (b) | (c) | (d) |

Fig. 5. General scene depth map

figure (b) is from larger SAD window, although the effect in low texture regions is better, many depth information for edge areas are lost, and cost a lot of time in matching; figure (c) is the optimal depth map from fixed SAD window; figure (d) is depth image obtained by the algorithm is proposed in this paper, and it has the best performance. We even can see chair behind the desk, and the floor is much better than others.

5 Conclusions

The system realizes the real-time acquisition of depth map with the binocular unmanned vehicle. To this end, the principle and implementation method of stereo camera calibration, image correction and stereo matching are researched. We calibrated the stereo camera in Matlab, and do image correction and acquisition of depth map in VS2010 with C++. This paper improves the original algorithm in OpenCV. The quality of depth map is improved on the basis of real-time and can adapt to different environment. But there is much space to improve the accuracy of the depth map, and need to continue exploring and researching.

References

1. Se, S., Jasiobedzki, P.: Stereo-vision based 3d modeling for unmanned ground vehicles. In: Proceedings of SPIE - The International Society for Optical Engineering, vol. 6561(1), 65610X-65610X-12 (2007)
2. Kim, G.B., Chung, S.C.: An accurate and robust stereo matching algorithm with variable windows for 3d measurements. J. Mechatron. **14**(6), 715–735 (2004)
3. Yoon, K.J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. J IEEE Trans. Pattern Analy. Mach. Intell. **28**(4), 650–656 (2006)
4. Zhang, K., Lu, J., Lafruit, G.: Cross-based local stereo matching using orthogonal integral images. J IEEE Trans. Circ. Syst. Video Technol. **19**(7), 1073–1079 (2009)
5. Liu, F.C., Xie, M.H., Wang, W.: Stereo calibration method of binocular vision. J. Comput. Eng. Des. **32**(4), 1508–1512 (2011)
6. Wang, Z.Z., Zhao, L.Y., Liu, Z.Z.: Binocular stereo vision distance measurement system based on a combination of matlab and openCV. J. Tianjin Univ. Technol. **1**, 13 (2013)
7. Lin, H.Y., Jian, S., Liu, Y.M., Rong, C.: Camera calibration technique based on rectification of image aberration. J. Jilin Univ. **37**(2), 433–437 (2007)
8. Wang, H., Zhiwen, X.U., Xie, K., Jie, L.I., Song, C.: Binocular measuring system based on openCV. J. Jilin Univ. **2**, 13 (2014)
9. Guo, L.Y., Sun, C.Y., Zhang, G.Y., Wu, J.H.: Variable window stereo matching based on phase congruency. Appl. Mech. Mater. **380–384**, 3998–4001 (2013)