Optimization of Stereo Vision Depth Estimation using Edge-Based Disparity Map

Juan Du[1], James Okae[1]

[1]School of Automation Science and Control Engineering, South China University of Technology, China
dujuan@scut.edu.cn, jamesokae@yahoo.com

Abstract

Stereo Vision based depth estimation has become a major research topic due to its enormous importance in applications such as industrial automation, Advanced Driver Assistance Systems. Stereo Vision approach of estimating depth is usually a complex process due to issues such as lighting, reflections, camera distortion and alignment, image complexity, pixel noise etc. A simple but an efficient Edge-Based Disparity Map algorithm to optimize the performance of a stereo vision system and accurately relate computed disparity map to true depth information has been presented in this paper. Additionally, a comparative study between the proposed Edge-Based Disparity Map algorithm, conventional Block Matching algorithm and Semi Global Stereo Matching algorithm performances with varying block window size was investigated and discussed. MATLAB demonstration results obtained from the investigated algorithms for corresponding matching results and disparity calculations which are important building blocks for depth (distance) estimation tasks were evaluated based on accuracy and efficiency (computation time).

1. Introduction

The visual ability to perceive the world in 3D and calculate the distance (depth) of an object has become a critical concern due to its numerous impacts in many applications such as Industrial Automation[1], Advanced Driver Assistance Systems and Robot Navigation [2-3]. Stereo vision is one of the approaches used to obtain the 3D and depth information of objects [1-4]. Stereo Vision approach of estimating depth is usually a very complex process due to issues such as lighting, reflections, camera distortion and alignment, pixel noise etc. Finding an efficient solution is a great concern for researchers. Stereo vision system has two cameras displaced horizontally from one another to obtain two different views of a scene, in a manner similar to human binocular vision. This approach aims to implement the principle behind how the human brain deduces depth from a scene. By using two cameras with different view angles to obtain images from the same scene, it is possible to compute how far the scene is from the camera with a high degree of accuracy.

A typical stereo vision solution of estimating depth and extracting 3D object information usually implements the technique shown in Fig. 1 below. The approach requires stereo camera calibration [5-8] to establish a relation between image pixels and the real world dimensions.

It is not practical to obtain or keep a perfect stereo camera alignment even with high precision cameras in the setup of a stereo vision system. This challenge makes it difficult to locate corresponding points in stereo images and is usually known as correspondence problem. Correspondence problem is simplified by aligning the two cameras to be coplanar in the process called rectification [9-12]. Issues such as lighting, camera misalignment, pixel noise, image complexity, camera focus etc. also complicate correspondence problem making it difficult to obtain an accurate matching results. Contributions made to detect and match corresponding points include; Speeded UP Robust Features (SURF) [13] and Scale-Invariant Feature Transform (SIFT) [14]. Sum of Absolute Difference (SAD) [15], Sum of Square Difference (SSD) and Normalized Cross Correlation (NCC) [16].

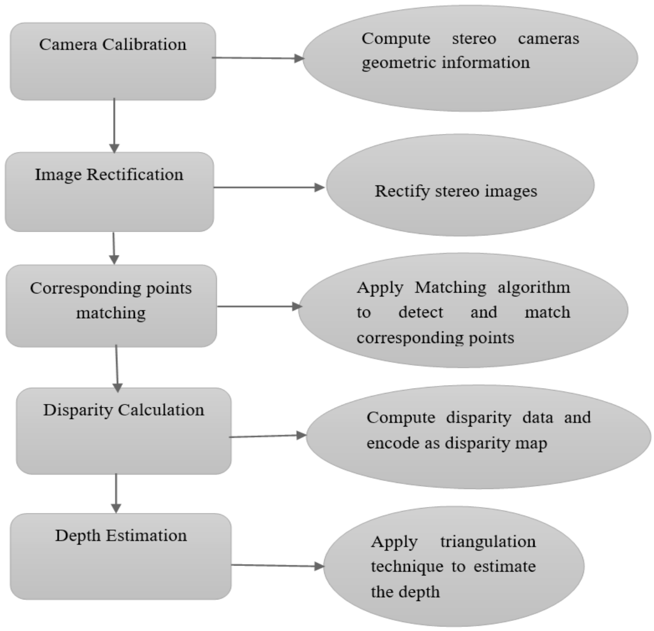

Fig.1. Step by step approach of the a typical stereo vision solution

The different view angles of the stereo cameras results in horizontal separation parallaxes of the stereo-pairs captured. The relative displacements or the horizontal shift between detected matched points are calculated and stored as disparity map. The disparity values together with information about camera parameters obtained from camera calibration are used to compute depth through a process called triangulation.

2. Methodology

The propose depth estimation algorithm implemented the step by step technique shown in figure 1 above. The details of the various steps are described below.

Step I:

The calibration process is performed by taking multiple images of special calibration pattern (chess board) from which corners were detected and extracted for estimating camera parameters. Camera calibration computes projection matrices which show the relationship between object pixels positions in

3D world and 2D image plane as well as the spatial relationship between the two cameras as indicated below:

$$\lambda I = WC \quad (1)$$

Where λ = scalar C = camera projective transformation
W = 3D world homogeneous coordinates
I = Camera 2D homogeneous coordinates

$$I = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

$$W = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3)$$

$$C = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

Where r = rotation t = translation f = focal length

c = optical axis

Step II:

Calibrated rectification approach is applied in this work due to its accurate alignment of non-parallel stereo cameras to appear as if they are parallel. The rectification operation simplifies 2D correspondence problem to 1D correspondence problem, making it easier to locate corresponding points and calculate the disparity of the corresponding points.

Step III:

Disparity calculates the distance between detected corresponding points or the horizontal shift between detected matched points. Disparity values together with information about camera parameters obtained from camera calibration are used to compute depth. The accuracy of the disparity data depends on corresponding points matching results.

The figure below illustrates how the disparity of corresponding point is related to the depth of a 3D scene point. The figure is assumed to be accurately rectified or having both cameras mounted parallel to each other. The focal length **f** of the two cameras are same.

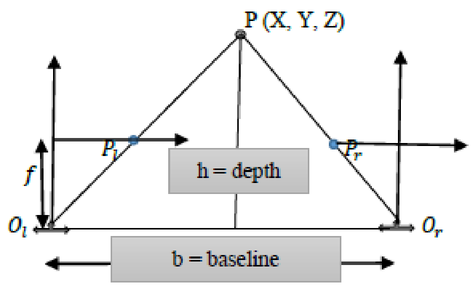

Fig. 2. A rectified stereo camera system

Let
X = x axis of the 3D point P
O_l= optical axis of the left camera
O_r= optical axis of the right camera
P = 3D point with coordinates X, Y, Z
P_l= projection of 3D point P in the left camera
P_r= projection of 3D point P in the right camera

By using similar triangle method to find the **X** coordinates for $\mathbf{P_l}$ and $\mathbf{P_r}$ in the left and right camera respectively.

$$P_l = \frac{f}{h} X \quad (5)$$

$$P_r = \frac{f}{h} \quad (6)$$

The disparity d which is the difference in the X coordinates for P_l and P_r is calculate as,

$$d = \frac{b}{h} f \quad (7)$$

Therefore the depth h of the 3D point P can be found as

$$\text{depth (h)} = \frac{b}{d} f \quad (8)$$

$$\Rightarrow \lim_{d \to 0} \frac{b}{d} f = \infty \quad (9)$$

It can be observed from equation (9) that, depth $h \to \infty$ as the disparity $d \to 0$. This means zero disparity indicates an object or a point at infinity distance from the camera.

This work further investigates and compares the performances of conventional Block Matching with SAD as evaluation cost function [15], Semi Global Matching [17] and the propose Edge-Based Disparity Map in computing the disparity map. The performance evaluation of these algorithms is based on accuracy and computation time. The brief description of these algorithms is shown below.

Block Matching (BM)

The block matching technique divides a reference image I_L into blocks and compare each block with a corresponding block in a target image I_R. The evaluation metric proposed in this work for block matching is SAD due to its simplicity, computation time and easy implementation with hardware [15]. This traditional BM approach is usually produces unreliable disparity results due to variations in lighting, different view angles, color etc.

Semi-Global Matching (SGM)

Semi global matching method performs pixel-wise matching utilizing mutual information and the approximation of a global smoothness constraint [17]. This algorithm provides accurate stereo matching especially at object boundaries and is robust against illumination changes. This is a unique combination of the global and local matching algorithms for accurate pixel-wise matching. The SGM calculates disparity with an accurate sub-pixel. However SGM is very slow for some real time application hence requires optimization technique. Details of the SGM algorithm can be seen in [17].

Edge-Based Disparity Map

In order to facilitate matching stereo image points and improve the reliability of the disparity results at shorter computation time, a simple and efficient version of the conventional Block Matching algorithm is propose. Edge-Based Disparity Map algorithm combines the traditional SAD algorithm with edge detection algorithm to enhance the reliability of the results obtained from similarity finding of corresponding points in a block. Edge is an important feature for matching. Edge information of object are extracted from the rectified images. These indicate the boundaries of the object and its surfaces. The extracted edge information reduces the amount of input data to be processed by removing less important data while keeping the properties of the object structure. Morphology filtering operation is applied to the edge detected image to enhance the shape of the object by expanding its boundaries. The Edge-Based strategy reduces the disparity map computation time and makes corresponding points detection simple and accurate. The work proposes canny edge detection method [19-20] since it was found to extract high-quality edges from images and also detects true weak edges. A block diagram of the Edge-Based Disparity Map Algorithm is shown in figure 3 below.

Step IV:

Triangulation principle is applied in estimating the depth of the objects. When disparity data is combined with the results from the camera calibration, it is possible to recover the depth of the object by utilizing the relation shown in equation (8) above.

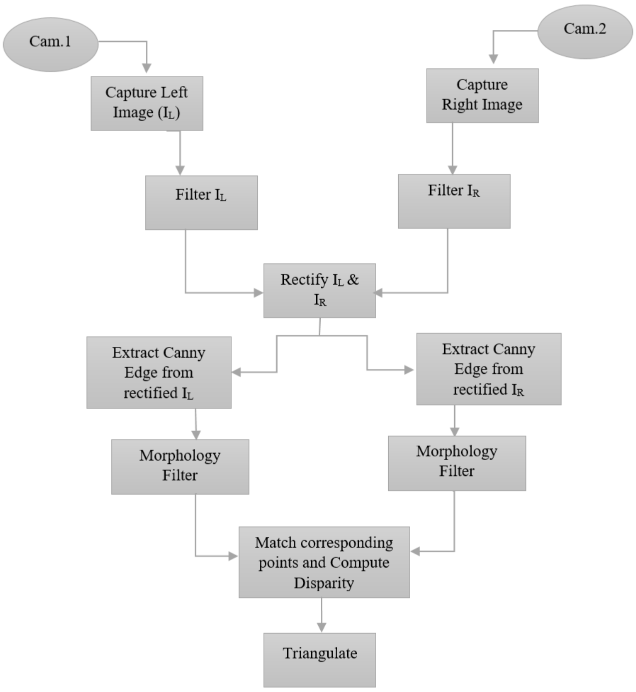

Fig. 3. Block Diagram of Proposed Edge-Based Algorithm

3. Simulation Results and Discussions

Camera Calibration

Camera parameters were acquired by taking multiple stereo images of special calibration pattern (chessboard). The overall re-projection error obtained from the calibration process was 0.63 pixels as shown in fig 4 below. This re-projection error shows the distance between the corner points detected in the captured images and the corresponding 3D world points projected into the 2D image.

Fig.4. Results from camera calibration

Rectification

The geometric transformation results from the camera calibration process is applied to the captured stereo images with non-parallel optical axis to obtain a rectified stereo images with epipolar lines parallel to the horizontal axis. The rectification results simplify the corresponding points matching problem. The corresponding points of the rectified stereo images have identical vertical coordinates as shown in fig 5 below.

Disparity Map

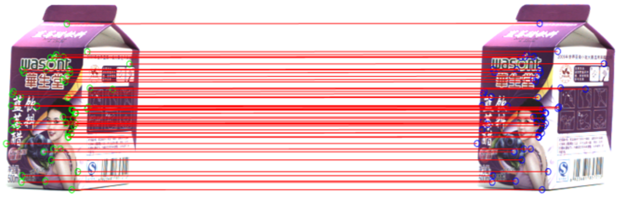

Fig 5. A rectified stereo images

One of the several results of disparity map computation from the simulation tests have been shown below.

The left and right images in Fig. 6 indicate results obtained from the described conventional Block Matching and Semi-Global Block Matching algorithms respectively. The Fig. 7 is the result obtained from applying the proposed Edge-Based Disparity Map algorithm. The SGBM provided dense disparity map compared with the conventional BM and Edge-Based Disparity Map Algorithm.

The average computation speed and depth estimation results of these algorithms when run on Intel(R) core ™ i5-6400 CPU@ 2.70GHz 2.71GHz processor are shown in table 1 and table 2 respectively. The size of the input stereo images is 2592 × 1944. These algorithms were tested with different window block sizes and their influence on the computation time has been indicated in table 1. The best depth estimation results from these algorithms with varying window block sizes are shown in Table 2. The Edge-Based Disparity Map algorithm indicated less average computation time compared with the conventional BM and SGBM algorithms as shown in Table 1.

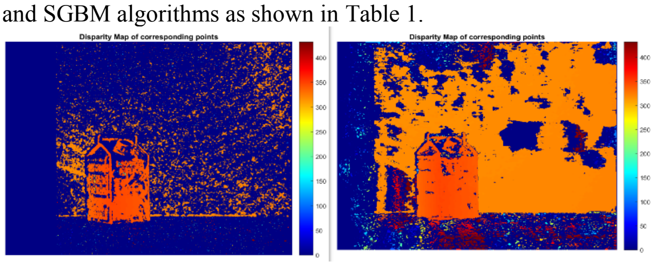

Fig 6. Disparity map results for BM and SGBM algorithms

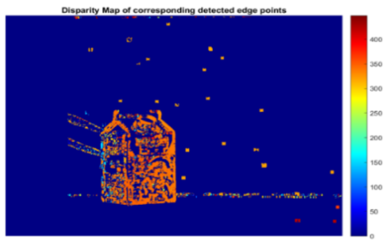

Fig 7. Disparity map results for Edge-Based Disparity Map Algorithm.

Table1. Average computation time from investigated algorithms

Block Size	Edge-Based Disparity Map Computation Time (secs)	BM Computation Time (secs)	SGM Computation Time (secs)
11×11	0.2708428	0.2819046	13.390921
13×13	0.2781360	0.2785018	13.327272
15×15	0.3105502	0.3198698	13.232021

17×17	0.3541398	0.3699456	13.2019232
19×19	0.4094594	0.4383026	13.1585786

Table 2. Depth Estimation Results from investigated algorithms

Test No	Actual Depth (cm)	Edge-Based Disparity Map Estimated Depth (cm)	SGM Estimated Depth (cm)	BM Estimated Depth (cm)
1	95	95.10	95.13	95.19
2	124	124.22	124.24	124.25
3	130	130.18	130.28	130.33
4	135	135.23	135.25	135.29
5	150	150.38	150.40	150.45

Depth Estimation

The performance evaluation of the investigated algorithms was based on the computation time duration for disparity map calculation and accuracy of the depth estimation as indicated in the Table 1 and Table 2. The SGM algorithm provided dense disparity map calculation but was found to be very slow and will not be convenient for some real time applications. The Edge-Based Disparity Map algorithm revealed the shortest computation time and also produced the best depth estimation results.

The simulation demonstration also revealed that, the selection of the block window size is critical for obtaining an accurate disparity map. The simulation revealed noisy disparity map for block window sizes smaller than the optimal size. Conversely, the disparity map was found to be blurred for block window sizes larger than the optimal window size. Additionally, smaller block window sizes compared with larger block window sizes, proved to work well in localizing object boundaries with high degree of accuracy. The stereo cameras positions were also found to be an influence for obtaining a good disparity map.

4. Conclusion

A simple but efficient algorithm to perform stereo based depth estimation has been presented in this work. A comparative study between Edge-Based Disparity map, conventional Block Matching and Semi Global Matching algorithms for stereo corresponding point matching and disparity map calculation which are key operations for depth estimation and many other computer vision tasks have also been shown. The results indicate that, the proposed Edge-Based Disparity Map algorithm is faster than the conventional Block Matching algorithm and the Semi Global algorithm, therefore good for real time applications and also provides high accurate++ depth estimation results.

Future work will reveal factors influencing optimal block window size selection and how adaptive based algorithm can be implemented in selecting an optimal block window sizes for disparity map calculation with high degree of accuracy while maintaining shorter computation time for real-time applications.

5. References

[1] O. Jong-Kyu, L. Chan-Ho L. "Development of a stereo vision system for industrial robots", International conference on control, automation and systems, 2007, pp. 659-663.

[2] I. Kim, et al., "An embodiment of stereo vision for mobile robot for real time measuring distance and object tracking.", International Conference on Control, Automation and Systems, 2007, pp.1029-1033.

[3] G. Michalos et al., Robot Path Correction Using Stereo Vision System, "45th CIRP Conference on Manufacturing System." pp.352-357, 2012

[4] A. H. Rostam, M.A.H Afifah, I.M.S Sani, "The solution of stereo correspondence problem using block matching algorithm in stereo vision mobile robot", second international conference on computer research and development, 2010, pp.733-737.

[5] Z. Zhang, "A flexible new technique for Camera Calibration", IEEE Transaction on pattern analysis and machine intelligence. Vol. 22, No. 11 2000, pp. 1330-1334.

[6] J.Y. Bouguet, "Camera Calibration Toolbox for Matlab." Computational vision at the California institute of Technology.

[7] J. Heikkila and O. Silven. "A Four Step Camera Calibration Procedure with Implicit Image Correction." IEEE International Conference on Computer Vision and Pattern Recognition. 1997.

[8] Y.M Wang, Y. Li, J.B. Zheng, "A camera calibration technique based on OpenCV", 3rd International conference on information science and interaction sciences, 2010, pp.403-406

[9] L. Guo-Yu, Chen-Xu, "A Robust Epipolar Rectification Method of stereo pairs", International Conference on Measuring Technology and Mechatronics Automation, 2010, pp. 322-326.

[10] Xi Liu et al. "A method of stereo images rectification and its application in stereo vision measurement", second IITA international conference on geoscience and remote sensing, 2010, pp. 169-172

[11] P. Perek, D. Makowski, A. Napieralski, "Efficient uncalibrated rectification method for stereo vision systems", 23rd International Conference Mixed Design of Integrated Circuits and Systems, 2016, pp. 89-92.

[12] V. Cesar et al., "Uncalibrated image rectification for coplanar stereo cameras", International Conference on Computer Vision Theory and Applications (VISAPP), 2014, vol.3, pp.648-655.

[13] H. Bay, T. Tuytelaars, L.Van Gool, "Speeded-up robust features (SURF)", Computer vision and Image Understanding", 110:3, 2008, pp.346-359

[14] D.G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, 60, 2, 2004, pp.91-110.

[15] Konolige, K., Small Vision System: Hardware and Implementation, Proceedings of the 8th International Symposium in Robotic Research, pp. 203-212, 1997

[16] M. B. Hisham et al, "Template Matching using sum of squared difference and normalized cross correlation", IEEE Student Conference on Research and Development, 2015, pp.100-104

[17] Hirschmuller, H., Accurate and Efficient Stereo Proceedings by Semi Global Matching and Mutual information, International Conference on Computer Vision and Pattern Recognition, 2005.

[18] V. A. Deepambika, S.L. Arunlal, "Feature-based stereo correspondence algorithm for the robotic arm applications", International Journal of Current Engineering and Technology, 2013, pp. 2163-2166.

[19] R. H. Thaher, Z. K. Hussein, "Study and Comparison of Various Image Edge Detection Techniques" International Journal of Image Processing vol. 3, 2009, pp.1-11.

[20] J. Canny, “A Computational Approach to Edge Detection”, IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6), pp. 679-698, 1986.

*This work was supported by ①The Natural Science Foundation of Guangdong Province, China (Grant No. 2016A030313454)；②Fundamental Research Funds for the Central Universities (Grant No. 2015ZM135)；③Science and Technology Project of Nansha, Guangzhou, Guangdong(Grant No. 2015KF012)；④Science and Technology Project of Guangzhou (Grant No. 201707010061)；⑤Science and Technology Project（International Cooperation Research） of Nansha, Guangzhou, Guangdong(Grant No. 2016GJ012)；⑥The National 02 Major Projects(Grant No. 2014ZX02503)