International Journal of Engineering & Technology, 7 (4.11) (2018) 9-12

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Accurate Objects Detection Using Stereo Vision Sensor for Machine Vision Applications

RA. Hamzah[1]*, MGY. Wei[2], NS. Nik Anwar[2], AF. Kadmin[1], SF. Abd Gani[1], MS. Hamid[1]

[1]*Fakulti Teknologi Kejuruteraan Elektrik & Elektronik, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia*
[2]*Fakulti Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia*
**Corresponding author E-mail: rostamaffendi@utem.edu.my*

Abstract

This paper presents a new algorithm for object detection using a stereo camera system, which is applicable for machine vision applications. The propose algorithm has four stages which the first stage is matching cost computation. This step acquires the preliminary result using a pixel based differences method. Then, the second stage known as aggregation step uses a guided filter with fixed window support size. This filter is efficiently reduce the noise and increase the edge properties. After that, the optimization stage uses winner-takes-all (WTA) approach which selects the smallest matching differences value and normalized it to the disparity level. The last stage in the framework uses a bilateral filter, which is effectively further reduce the remaining noise on the disparity map. This map is two-dimensional mapping of the final result which contains information of object detection and locations. Based on the standard benchmarking stereo dataset, the proposed work produces good results and performs much better compared with some recently published methods.

Keywords: *matching algorithm; machine vision; object detection*.

1. Introduction

In recent years, stereo vision system has revealed a growing interest in machine vision applications such as 3D surface reconstruction [1], range estimation [2] and embedded system [3]. Fundamentally, machine vision technology requires information from an image. The information extracted can be a complex set of data for example position, orientation and identity of each object in an image. Hence, this paper proposes an algorithm to extract and collect this data based on stereo vision system. The developed algorithm will be processed the data and provide the extracted information such as the depth and coordinates of object detection. The algorithm process is also known as stereo corresponding algorithm. The improvement of this work creates the correspondence among a brace of images that will produced a disparity or depth map. The map contains the depth information and locations of the objects detection. The triangulation principle [4] is utilized to calculate the depth, which will be used in numerous applications in machine vision applications such as industrial automations augmented reality (AR) [5] and robotic arm [6]. Some of the applications are capable to be implemented in real time or near real time. It depends on the capability of the central processing unit which needs extra computational requirement.
In real situation, the robotic movement requires an accurate depth estimation which the disparity map needs to be at the highest level of precision. Furthermore, this depth information is also capable to be used in 3D superficial reconstruction for the AR which visualizes the real environmental conditions. Therefore, the disparity map approximation is the most challenging and important works in stereo vision study areas especially in the computer vision field. Recently, numerous research articles have been issued in this research area and excessive improvement has been flourished. Fundamentally, there are multiple stages were developed by [7] to build a stereo correspondence or matching algorithm. First, the matching cost computation which is to calculate the matching points between two images. Second, the cost aggregation stage reduces the noise after matching cost process. Third, the optimization and disparity selection stage which this step is to normalize the disparity values on each pixel on image. Last step is known as refinement stage or post-processing step to decrease the outstanding invalid pixels or noise on the final disparity map.
Generally, there are two foremost methods in stereo matching algorithm which are recognized as local and global methods [2]. The classification is held by the method on by what method the disparity is calculated. To define the disparity map based on the energy minimization method or function is used in global approach. The calculation is built on the flatness position from closer pixels, which uses the minimization of global energy function. One of the well-known method in global method is the Markov Random Field (MRF) that measures the energy function. Several established methods using the MRF were Belief Propagation (BP) [3] and Graph Cut (GC) [8], which applied the energy minimization technique. The BP technique employed the MRF by incessantly absolution pointers from present point to the nearby points or neighbors. Else, the GC method uses the MRF approach by maximizing flow rule and cut the smallest energy flow structure. Global methodologies obtain low precision on images with radiometric problem and the framework require complex calculation to process high resolution images [9].
Local approach processes using local structures or contents. This method works based on a support region in predefined sizes. There are many published methods using local-based approach or using window-based methods such as convolution neural network [10], multiple window, fixed window and adaptive window. The benefit of a local approach is low calculation on the computational requirement and acquires fast time execution. Local approaches use the Winner-Takes-All (WTA) strategy in the disparity selec-

tion and optimization step. The WTA utilizes a lowest raw data from aggregation step. Then, it will be normalized the raw data to the minimum disparity value. Recently, numerous local approaches have been developed. In [11] proposed a complex method using simulated support window. This method requires multiple estimations which increases the time execution. In [12] imposed a modification of normalized cross correlation (NCC) function at earlier step that decreases the accuracy of preliminary differences. Hence, their approach yields enormous noise and invalid pixels on the objects edges. Frequently, local approaches produce high error on the object edges because of inappropriate selection of the windows filter size. Therefore, the challenge is to get an accurate window size for local method to increase the accuracy.

2. Methodology

Figure 1 displays the framework of the proposed work. The first step uses pixel base correspondence method known as absolute differences. Then, a guided filter (GF) will be used at cost aggregation stage. After that, the WTA will be implemented at disparity optimization step to select the minimum disparity value. The final stage is using a bilateral filter (BF) to remove the remaining invalid disparity and increases the accuracy of final disparity map.

2.1. Matching Cost Computation

This stage is the first step of the algorithm structure, which estimates the matching differences of the preliminary data between stereo images. Therefore, this phase is the most important part in the algorithm structure. The method used at this step must be robust and is capable to avoid from too many noise. Thus, the proposed work uses the absolute differences (AD) function which treats each pixel individually and the properties of the pixel can fully utilized. The AD computes the horizontal direction along the grayscale values on each pixel of stereo image. The benefit of the AD is that this function is fast and capable to preserve the properties of each pixel. The AD function is represented in (1):

$$AD(x,y,d) = w(I_l(x,y) - I_r(x,y,d)) \tag{1}$$

where I represents the pixel intensity at the coordinates (x,y), d represents the disparity value of left to right image differences and w is the weight.

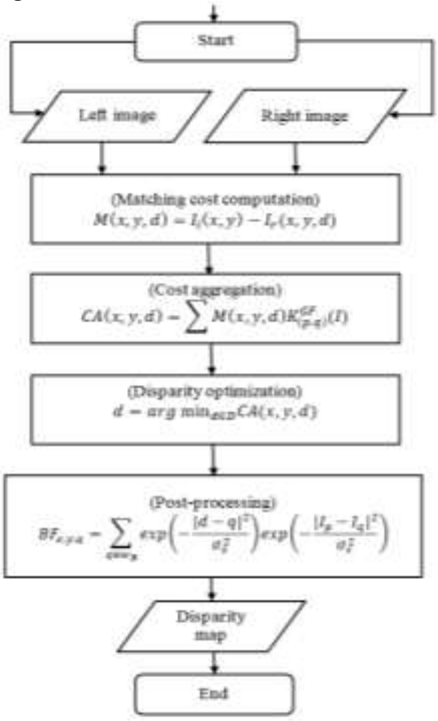

Fig. 1: A flowchart of the proposed algorithm.

2.2. Cost Aggregation

The cost aggregation is the second step of the local-based algorithm structure. This step decreases noise from the first stage to filter the preliminary result with more consistent matching differences before the next stage of optimization and disparity selection. This article recommends the usage of edge-preserving filter, where this filter is capable to remove the noise and preserved the object's edges detection. Hence, this work uses the guided filter (GF) which was developed by [13]. The filter kernel is shown in (2):

$$GF_{(p,q)}(I) = \frac{1}{w^2} \sum_{(p,q) \in w_c} \left(1 + \frac{(I_p - \mu_c)(I_q - \mu_c)}{\sigma_c^2 + \varepsilon}\right) \tag{2}$$

where {p,q,w,c,ε,σ,μ,I} denoted by { coordinates of (x,y), neighboring coordinates, window support size, center pixel of w, constant parameter, variance value, mean value, reference image (left input image)}. The GF is utilized in this work in line for fast execution of time processing, which the calculation is only depend on the image pixels. The GF is able to increase the accuracy at the object edges. The final equation of this stage is given in (3):

$$CA(x,y,d) = AD(x,y,d) GF_{(p,q)}(I) \tag{3}$$

where $AD(x,y,d)$ is the input of the first stage and $GF_{(p,q)}(I)$ represents the kernel of the GF with left image as a reference image in this article.

2.3. Disparity Optimization

Generally, every image contains a set of disparity values. This stage utilizes the WTA strategy which uses minimum disparity value for every location on the disparity map. Hence, the WTA is most suitable approach to be used in this article. The formulation of the WTA is given in (4):

$$d(x,y) = \mathrm{argmin}_{d \in D} CA(x,y,d) \tag{4}$$

where D denotes the range of disparity on an image, d(x,y) is the selected disparity value at the location of (x,y) and $CA(x,y,d)$ is the value of the second stage.

2.4. Post-Processing

The post-processing step is the last stage of the algorithm framework. This stage is also known as filtering stage or disparity refinement stage which decreases the invalid disparity values and the outlier noise on the final results [14]. Fundamentally, the used of second filtering process at this step is to surge the effectiveness and accurateness of the final disparity map. The suggested work in this article is using the bilateral filter (BF), which works for the second filtering process. The final disparity result is given in (5):

$$BF_{(p,q)} = \sum_{q \in w_B} \exp\left(-\frac{|p-q|^2}{\sigma_s^2}\right) \exp\left(-\frac{|I_p - I_q|^2}{\sigma_c^2}\right) \tag{5}$$

where the (p,q) represent the pixel of interest and neighbor pixel coordinates, w_b denotes the BF window size, $|I_p - I_q|^2$ and $|p-q|$ are the intensity differences and spatial distance respectively, σ_s^2 and σ_c^2 are the spatial distance and color similarity parameters.

3. Results and Discussion

This section provides the experimental setup and results of the proposed work in this article. All of the experiment were executed using a personal computer with the features of CPU i7-5500 and 8G RAM. The standard benchmarking dataset have been used in this article from the Middlebury Stereo evaluation benchmark [15]. This article also uses the KITTI benchmarking dataset. The KITTI dataset was developed by [16], which consists of 200 training images. This dataset were recorded from real environment of autonomous vehicle navigation using a stereo vision system. Hence, it comprises very challenging and complex images for the evaluation of stereo matching algorithm. The performance is measured based on the all and nonocc error attributes of bad pixel percentages. The all error is the error of invalid disparity values on all pixels of the disparity map image. The nonocc error is the error of invalid disparity values on the non-occluded regions. The quantitative results in this article are evaluated based on the development kit provided by the KITTI. The parameters used in this experimental were $\{w_B, w_C\}$ with {11×11, 9×9}.

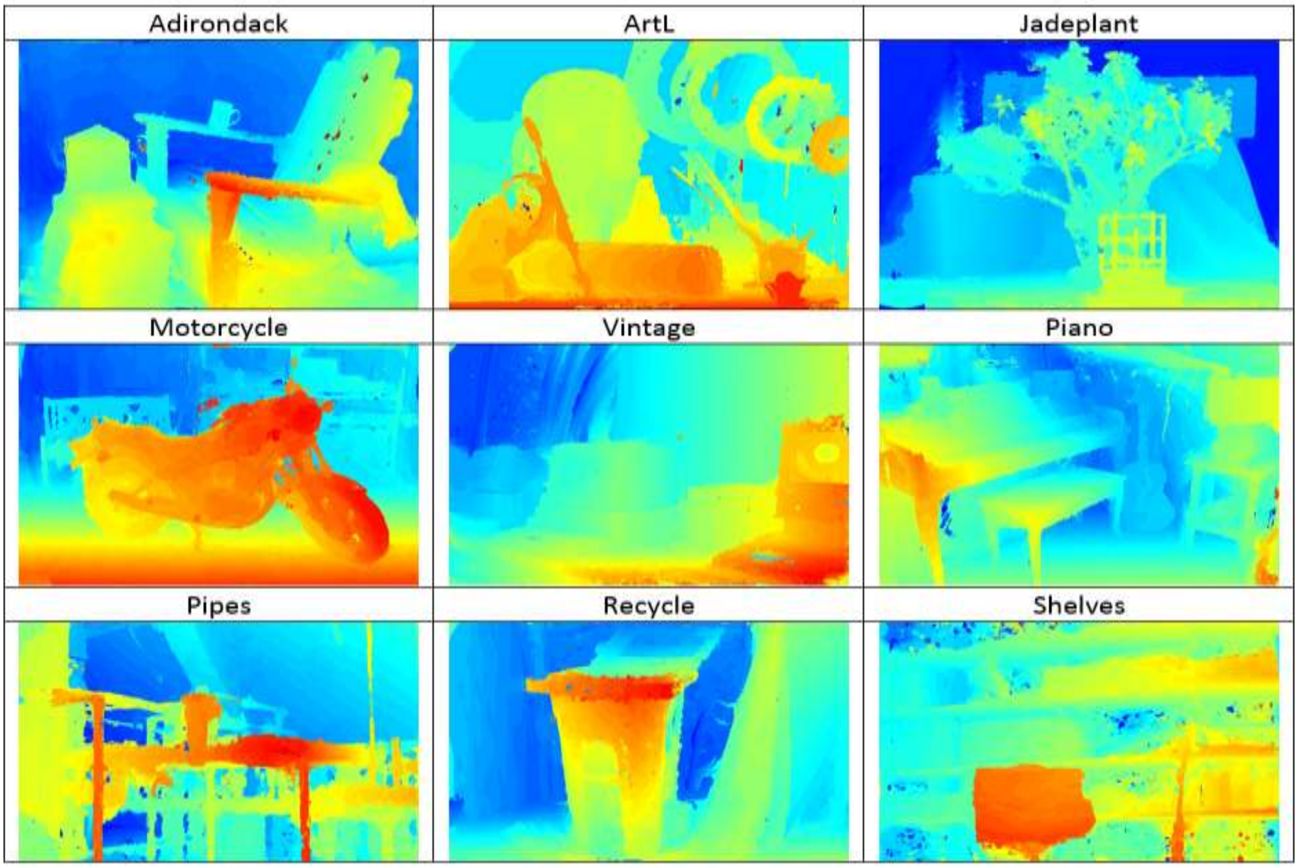

Fig. 2: The results from the Middlebury Stereo Evaluation dataset.

Figure 2 shows the results of objects detection using the Middlebury images. The objects are well-detected and reconstructed on the map such as a mug in the Adirondack, the brushes in ArtL, a tree in Jadeplant, a motorcycle in Motorcycle, the personal computer in Vintage, piano and a chair in Piano, a recycle box in Recycle and a chair in the Shelves image. The complex images such as the pipes structures in the Pipe image are also well-reconstructed. It shows that the proposed framework in this article is capable to detect and reconstruct the detected objects. Table 1 shows the average results of the Middlebury training dataset. It shows that the proposed work rank at top of the table and followed by ELAS, MPSV, ADSM, DoGGuided and TestG. All methods in Table 1 are displayed in the Middlebury database.

Table 2 displays the quantitative results of the proposed framework with some published methods. As for comparison before and after the proposed work, the TestG is the algorithm without the proposed framework which uses common algorithm with absolute differences matching technique. The proposed algorithm is ranked at top of the table with 12.11% and 14.01% of nonocc and all errors respectively. If compared with the TestG algorithm, the proposed work reduces the nonocc and all errors with 18.20% and 20.47% respectively. The second algorithm produced by 3D and followed by Hu [11], NCC [12], GC [8] and BP [3].

Table 1: The average results of the Middlebury training dataset.

Algorithm	Nonocc Error (%)	All Error (%)
Proposed work	7.05	10.34
ELAS	7.22	10.60
MPSV	8.81	12.70
ADSM	8.95	12.30
DoGGuided	12.00	22.30
TestG	28.42	30.10

Table 2: The average results of the KITTI training dataset.

Algorithm	Nonocc Error (%)	All Error (%)
Proposed work	12.11	14.01
SSW	12.75	14.22
Hu	13.55	16.65
NCC	14.11	15.79
GC	14.34	15.98
BP	15.67	17.84
TestG	30.31	34.48

Figure 3 displays the sample results of objects detection using the KITTI dataset. The first image shows the cars and a tree, which the proposed work is capable to detect these objects on the disparity map. The signage indicator, a tree and a car are also well-detected in the second image. The moving objects such as a cyclist in the third image is well-reconstructed. The fourth and fifth images show the cars are well-reconstructed on the disparity map. The different contour of disparity levels based on the different color scheme are precisely displayed on each result.

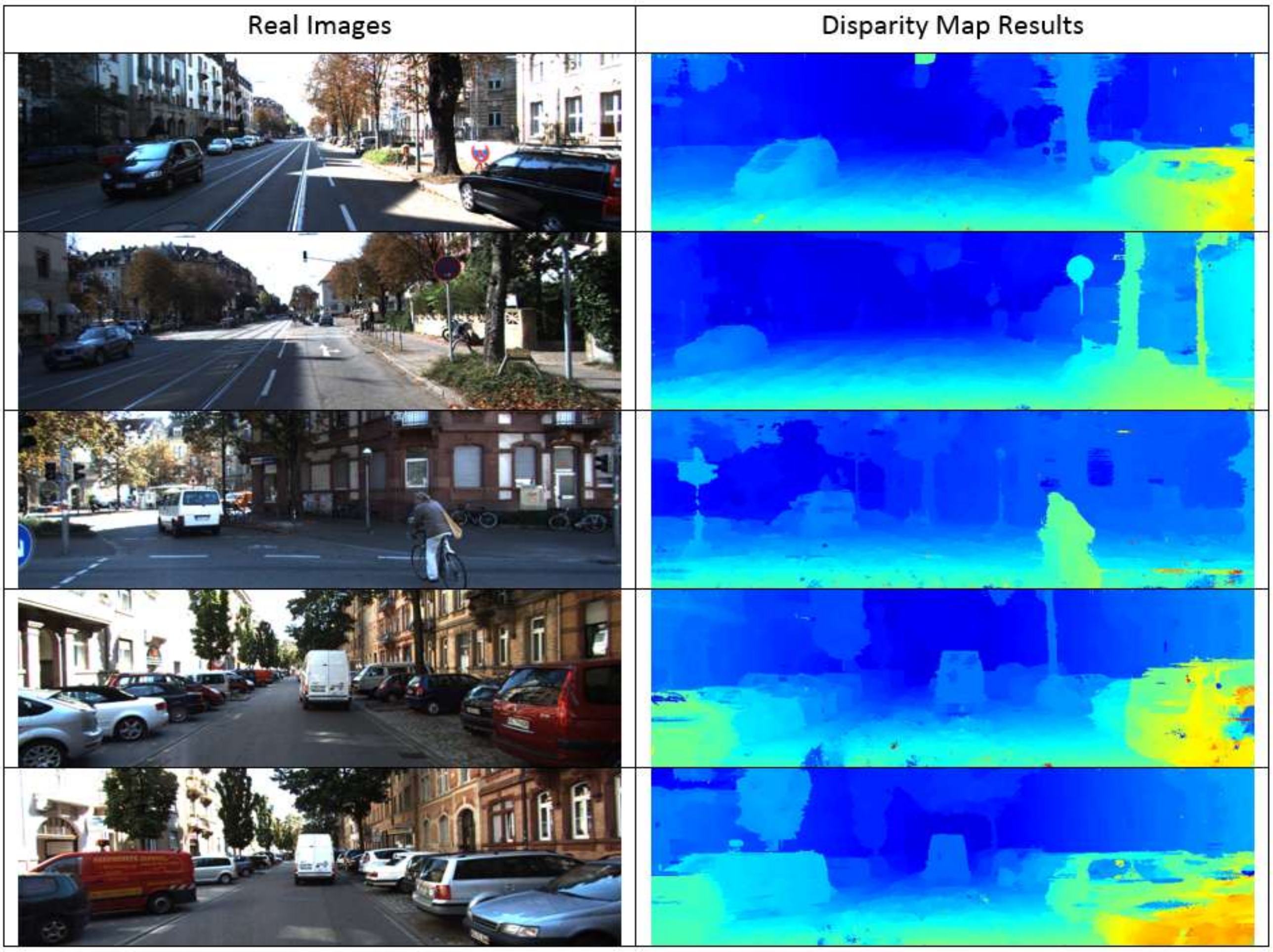

Fig. 3: Sample results of objects detection from the KITTI dataset.

4. Conclusion

A new framework of stereo matching algorithm was presented in this article. This algorithm was capable to detect any object based on the input from the stereo vision sensor. This algorithm is also able to detect and robust against the moving objects as shown by the results in the KITTI dataset. Additionally, the proposed framework is able to increase the accuracy compared with TestG as tabulated in Table 1. The framework is also competitive with some established methods as tabulated in Table 1.

Acknowledgement

This work is supported by a grant from the Universiti Teknikal Malaysia Melaka with the number PJP/2018/FTK(13C)/S01632.

References

[1] Hamzah R A, Hamid M S, Kadmin A F, Ghani S A, Salam S. Disparity map algorithm based on edge preserving filter for stereo video processing. Journal of Telecommunication, Electronic and Computer Engineering, 2018, 1210(1-7), 59-62.
[2] Hamzah R A, Aziz K A A, and Shokri A S M. A pixel to pixel correspondence and region of interest in stereo vision application. Proceedings of the IEEE Symposium on Computers and Informatics, 2012, pp. 193-197.
[3] Wu S S, Tsai H, and Chen L G. Efficient hardware architecture for large disparity range stereo matching based on belief propagation. Proceedings of the IEEE International Workshop on Signal Processing Systems, 2016, pp. 236–241.
[4] Hamzah R A, Ibrahim H, and Hassan A H A. Stereo matching algorithm based on per pixel difference adjustment, iterative guided filter and graph segmentation. Journal of Visual Communication and Image Representation, 2017, 42, 145–160.
[5] Gudis E, van der Wal G, Kuthirummal S, Chai S, and Kumar R. Stereo vision embedded system for augmented reality. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2012, pp. 15-20.
[6] Balter M L, Chen A I, Maguire T J and Yarmush M L. Adaptive kinematic control of a robotic venipuncture device based on stereo vision, ultrasound, and force guidance. IEEE Transactions on Industrial Electronics, 2017, 64(2), 1626-1635.
[7] Scharstein D, Szeliski R, and Zabih R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Proceedings of the IEEE Workshop on Stereo and Multi-Baseline Vision, 2001, pp. 131–140.
[8] Liang Q, Yang Y, and Liu B. Stereo matching algorithm based on ground control points using graph cut. Proceedings of the International Congress on Image and Signal Processing, 2014, pp. 503–508.
[9] Hamzah RA, Kadmin AF, Hamid MS, Ghani SF, Ibrahim H. Improvement of stereo matching algorithm for 3D surface reconstruction. Signal Processing: Image Communication, 2018, 65, 165-172.
[10] Zbontar J and LeCun Y. Computing the stereo matching cost with a convolutional neural network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1592–1599.
[11] Hu W, Zhang K, Sun L, Li J, and Yang S. Virtual support window for adaptive-weight stereo matching. IEEE Visual Communications and Image Processing, 2011, pp. 1–4.
[12] Einecke N and Eggert J. Anisotropic median filtering for stereo disparity map refinement. Proceedings of the International Conference on Computer Vision Theory and Applications, 2013, pp. 189–198.
[13] He K, Sun J, and Tang X. Guided image filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(6), 1397–1409.
[14] Hamzah R A, Kadmin A F, Ghani S F A, Hamid M S, and Salam S. Disparity refinement process based on RANSAC plane fitting for machine vision applications. Journal of Fundamental and Applied Sciences, 2017, 9(4S), 226-237.
[15] Scharstein D and Szeliski R. Middlebury stereo evaluation - version 3. http://vision.middlebury.edu/stereo/eval/references.
[16] Menze M and Geiger A. Object scene flow for autonomous vehicles. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3061–3070.
[17] Hamzah R A, Ibrahim H. Improvement of stereo matching algorithm based on sum of gradient magnitude differences and semi global method with refinement step. Electronics Letters, 2018, 54(14), 876–878