Disparity Map Algorithm Based on Edge Preserving Filter for Stereo Video Processing

R. A. Hamzah, M. S. Hamid, A. F. Kadmin, S. F. Abd Ghani, S. Salam
Fakulti Teknologi Kejuruteraan, Kampus Teknologi, Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
rostamaffendi@utem.edu.my

***Abstract*— This paper proposes a new local-based stereo matching algorithm for stereo video processing. Fundamentally, the Sum of Absolute Differences (SAD) algorithm produces an accurate results on the stereo video processing for the textured regions. However, this algorithm sensitives to low texture and radiometric distortions (i.e., contrast or brightness). To overcome these problems, the proposed algorithm utilizes edge-preserving filter which is known as Bilateral Filter (BF). The BF algorithm reduces noise and sharpen the images. Additionally, BF works fine on the low or plain texture areas. The proposed algorithm produces an accurate results and performs much better compared to some established algorithms on the standard benchmarking results of the Middlebury and KITTI dataset.**

***Index Terms*— Bilateral Filter; Computer Vision; Edge Preserving Filter; Video Processing.**

I. Introduction

Stereo matching algorithm process establishes the correspondence between a pair of images and produces a disparity map [1]. This map can be utilised for the depth estimation based on the triangulation concept which will be used for many applications such as robotics automation [2] and virtual reality [3]. An accurate disparity map makes robotic movements to operate in actual situations more precisely. Furthermore, the depth data applied for 3D surface reconstruction for augmented reality application. Thus, the disparity map estimation is one of the most important and challenging problems in computer vision research area. In recent years, many research papers have been published in this research area and great improvement have been succeeded. Four main steps were proposed by Scharstein and Szeliski [4] in their taxonomy to build up a stereo matching algorithm:

Step 1: Matching cost computation
Step 2: Cost aggregation
Step 3: Disparity optimization
Step 4: Disparity refinement

There are two major optimization methods which are known as global and local methods. The categorization is supported by the method on how the disparity is computed. Global method uses energy minimization function to determine the disparity map. The function is based on the smoothness confinement from nearer pixels which uses global energy function. The Markov Random Field (MRF) energy minimization technique is one of the famous approach in global methods.. These methods include Belief Propagation (BP) method [5] and Graph Cut (GC) method [6]. The GC technique employs the MRF approach which uses maximum flow rule and cut the minimum energy flow arrangement. Otherwise, the BP technique implemented MRF approach by continuously release indicators from current point to the nearest points or neighbours. Global approaches acquire good accuracy but they are high computational demand to process an image. Local method employs a support window or region-based on a predefined sizes. There are many published methods that related to the window-based such as multiple window [9], adaptive window [10], fixed window [7] and convolution neural network [8]. The advantage of local method is low computational demand and fast running time. The results produce by using a minimum raw data from matching cost and will be selected the lowest disparity values. Local methods are also known as Winner Takes All (WTA) strategy in their optimization stage. However, the local methods [11][12] quality are low, particularly in the area of low texture regions.

This paper suggests an original local-based stereo disparity map algorithm based on SAD and edge-preserving filter which is capable to produce accurate results [13]. The first step uses the SAD algorithm [14], the cost aggregation and final stages employ the BF with WTA strategy. This work is prepared as follows. Section II clarifies the suggested structure of stereo disparity map algorithm. Section III shows the investigational preparations and the results. The final part is the conclusion in Section IV.

II. Methodology

The first stage uses the Sum of Absolute Differences (SAD) as an area or block-based correspondence measurements. The second stage known as cost aggregation step will be implemented using a Bilateral Filter (BF) which also known as edge-preserving filter. After that, the WTA strategy is used at the disparity selection and optimization stage. The WTA uses minimum value from the cost aggregation step to be replaced with the minimum disparity value. The consistency checking step of valid and invalid pixels locations are implemented by using left-right checking technique. This checking process also produced the occlusion regions and the invalid pixels in the area of untextured regions. Then, the fill-in process is used to replace the invalid pixels with valid disparity values on the disparity map. The last stage consists of implementing the BF once again to reduce and remove the remaining noise which normally appear during the fill-in process.

A. Matching Cost Computation

The first stage of the algorithm is to determine the corresponding values between the stereo images. The SAD algorithm is based on the sum of intensity differences of double matching pixels in RGB frequencies on the right image I_r and left image I_l which is given by Eq. (1):

$$SAD(x,y,d) = \sum_{(x,y)\in w} |I_l^i(x,y) - I_r^i(x-d,y)| \quad (1)$$

where d signifies the difference value, (x, y) are pixel of interest coordinates, i and w denote as RGB channels number and a support region or windows of SAD algorithm. This stage produces raw disparity values of corresponding points and contains very large noise.

B. Cost Aggregation

This step is an important stage of algorithm development for local-based method which is minimizing the matching uncertainties. This stage reduces noise of preliminary disparity map and will demonstrated overall performance of local-based methods. The BF is designated since this filter was established to preserve the edges and decrease the noise. The BF kernel is given by Eq. (2):

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

where p represents the (x,y) coordinates pixel of interest in a support window, w_B and q are the BF support window and neighbouring pixels. The σ_s represents a spatial adjustment parameter and σ_c corresponds to the color similarity parameter. The $p\times q$ refer to spatial Euclidean distance and I_p-I_q is the Euclidean distance in color space. The aggregation cost at this step is given by Eq. (3):

$$C(p,d) = WM_{p,q}^{BF} SAD(p,d) \quad (3)$$

C. Disparity Selection and Optimization

To acquire an accurate disparity map, this work computes final disparity map by utilizing the WTA strategy. The WTA technique selects the minimum aggregated corresponding value for every pixel's location. The WTA equation is given by Eq. (4):

$$d_p = arg\ min_{d\in D} C(p,d) \quad (4)$$

where C(p,d) means the cost aggregation data and D represents a set of allows discrete disparity values. The disparity maps attained at this stage still encounter invalid pixels in the regions of occluded area.

D. Disparity Map Refinement

The last step of the corresponding algorithm comprises several continuous processes that started by occlusion handling, managing the invalid pixels and removing remaining noise. The occlusion regions are determined by left-right consistence checking process which contain as invalid pixels. Then, these invalid pixels will be replaced by valid corresponding pixels by fill-in process. This process uses nearby valid pixels to replace the invalid pixels in the occluded regions. This replacing process creates unwanted artifacts on the disparity map. To remove that artifacts, the BF is used as given the kernel by Eq. (5). The parameters of BF used the same settings as implemented at cost aggregation stage to increase the efficiency of final disparity map results.

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

All of the analyses are implemented on the computer system of Window 10, 3.2GHz processor and 8GB memory. The standard database from the Middlebury [15] benchmarking stereo evaluation has been used to measure the accuracy. This benchmarking system provide 15 training images and could be evaluated based on the bad pixel percentage of *all* and *nonocc* pixels. The parameter values used in this work were {w, σ_s, σ_c, w_B} with the values of {7×7, 17, 0.3, 11×11}. Table 1 and 2 show the quantitative results of the Middlebury dataset. It shows that the proposed work is more accurate than the work in [13], [16] and [17] for *nonocc* error. Additionally, for the all error attribute in Table 2, the proposed work produced the lowest error. The achievement of the proposed work has been analysed with other local algorithms in [15]. Based on the results, the proposed algorithm is among the lowest of average errors which indicates the competitive achievement of the proposed work.

III. Experimental Results

The experiments are carried out on the platform of Window 10 on desktop PC with 3.2GHz processor and 8GB memory. To evaluate the accuracy, the experimental images are using a standard quantitative online stereo benchmarking dataset from the Middlebury [15] and qualitative measurement based on the stereo video of autonomous navigation benchmarking dataset from KITTI [18]. The accuracy is measured from the bad pixel percentage of non-occluded pixel (*nonocc*) and all pixels (*all*). The parameter values used in this work were {w, σ_s, σ_c, w_B} with the values of {7×7, 17, 0.3, 11×11}. Table 1 and 2 show the quantitative results of the Middlebury dataset. It shows that the proposed work is more accurate than the work in [13], [16] and [17] for *nonocc* error.

Additionally, for the *all* error attribute in Table 2, the proposed work produced the lowest average error 9.09%. The achievement of the proposed work has been analyzed with other local algorithms in [15]. Based on the analyzing results, the proposed algorithm is among the lowest of average errors which indicates the competitive achievement. Figure 1 shows the final disparity map images in color of the Middlebury dataset. These images are based on the quantitative results in Table 1 and 2 which uploaded to the Middlebury stereo evaluation database.

Table 1
The comparison results of *nonocc* error using the Middlebury dataset

Algorithms	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlayP	Recyc	Shelvs	Teddy	Vintge	Weight Ave
MC-CNN [11]	0.76	2.55	16.30	1.27	1.27	1.83	5.07	2.29	2.27	3.11	3.03	2.48	4.41	1.12	14.80	3.82
Proposed Algorithm	3.58	4.55	12.80	3.48	3.37	5.60	12.70	5.92	5.04	18.10	4.61	3.36	8.64	2.49	8.45	6.10
SNCC [13]	2.89	4.05	18.10	2.68	2.52	3.52	7.08	6.14	5.64	45.40	3.13	2.90	7.59	1.58	13.50	6.97
ELAS [16]	3.09	4.72	29.70	3.28	3.29	4.30	8.31	5.61	6.00	21.80	2.84	3.09	9.00	2.36	10.90	7.22
BSM [17]	7.27	11.40	30.50	6.67	6.52	10.80	32.10	10.50	12.50	24.40	12.80	7.42	16.40	4.88	32.80	13.40

Table 2
The comparison results of *all* error using the Middlebury dataset.

Algorithms	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlayP	Recyc	Shelvs	Teddy	Vintge	Weight Ave
Proposed Algorithm	4.79	7.68	27.30	6.26	6.12	6.67	13.20	11.30	7.73	20.10	6.10	3.90	8.99	3.60	9.64	9.09
SNCC [13]	3.63	6.78	39.80	5.12	5.11	4.65	8.23	11.80	8.05	45.60	4.36	3.29	8.10	2.55	14.80	10.40
ELAS [16]	4.08	7.18	52.80	5.39	5.45	4.96	9.00	10.70	7.94	23.20	3.83	3.78	9.46	3.34	11.60	10.60
MC-CNN [11]	4.24	18.70	34.10	7.21	7.22	6.00	9.35	13.50	18.30	9.71	9.37	4.64	6.62	9.35	21.60	11.80
BSM [17]	12.70	28.70	58.70	14.80	14.70	16.00	35.80	24.50	29.40	31.00	20.20	12.10	19.20	14.30	39.30	23.50

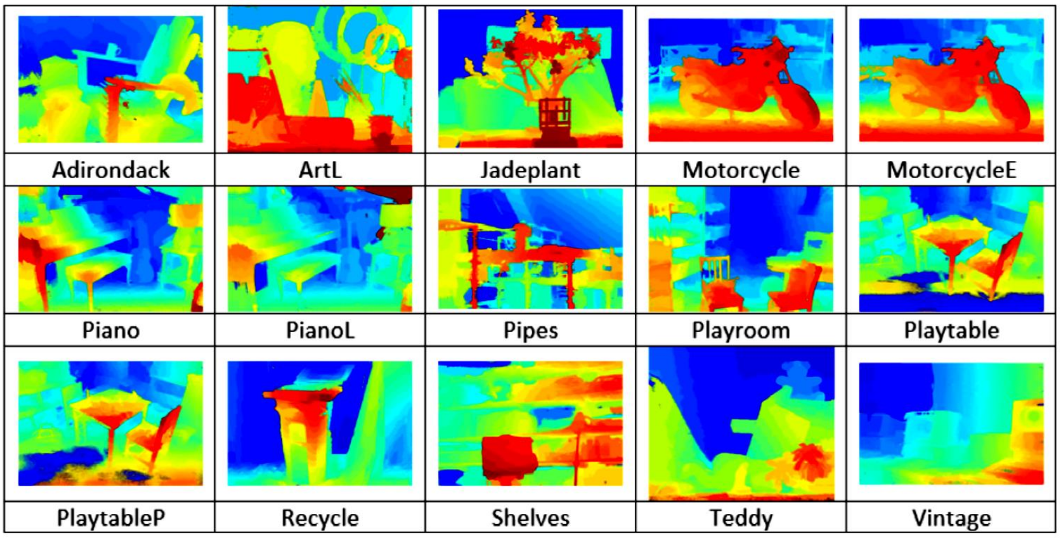

Figure 1: The disparity map results of the Middlebury dataset

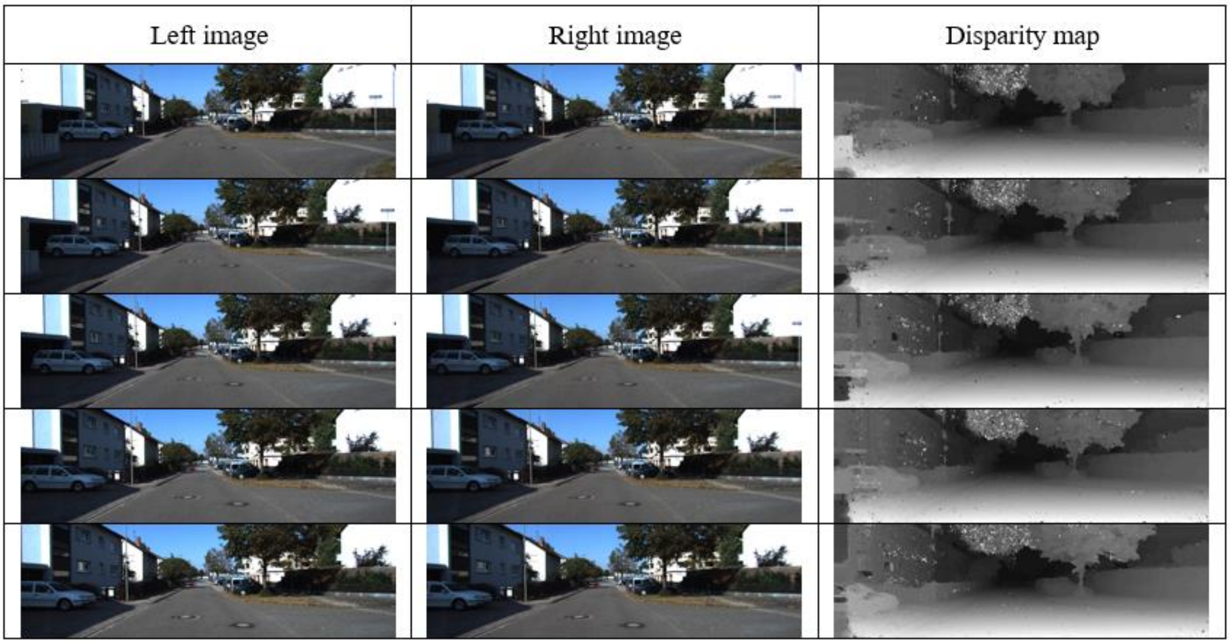

Figure 2: Five continuous frames of stereo video and the grayscale disparity map results of the KITTI dataset.

Figure 2 shows sample results of the KITTI stereo video dataset. These real images are processed based on frame by frame image which was recorded and captured from the stereo video of autonomous vehicle navigations. Based on the results of five continuous frames, the disparity maps produced by the proposed work are smooth and clear. An accurate object detections are perceived such as road surface, trees, cars and building. It shows that the proposed work is robust against the low texture and plain color surfaces. The contour of the disparity map results show accurate level of depth recognition

IV. CONCLUSION

In this paper, the local-based stereo matching algorithm for stereo video was presented. The algorithm used the combination of SAD algorithm and edge-preserving filter which were able to produce accurate results based on the standard benchmarking dataset. Furthermore, the proposed framework is competitive with some established algorithms as shown in Table 1 and 2. The proposed work is also well capable to work with real environment of vehicle navigation which was displayed in Figure 2.

ACKNOWLEDGEMENT

This work was supported by Ministry of Higher Education of Malaysia with account no. RAGS/1/2015/ICT01/UTEM/03/B00115 and the Universiti Teknikal Malaysia Melaka.

REFERENCES

[1] A. H. A. Hasan, R. A. Hamzah, and M. H. Johar, “Range estimation in disparity mapping for navigation of stereo vision autonomous vehicle using curve fitting tool,” *IJVIPNS*, vol. 9, no. 9, pp. 5–9, 2009.

[2] R. A. Hamzah, H. Ibrahim, and A. H. A. Hassan, “Stereo matching algorithm for 3D surface reconstruction based on triangulation principle,” in *Information Technology, Information Systems and Electrical Engineering (ICITISEE), International Conference on*. IEEE, 2016, pp. 119–124.

[3] I. Vedamurthy, D. C. Knill, S. J. Huang, A. Yung, J. Ding, O.-S. Kwon, D. Bavelier, and D. M. Levi, “Recovering stereo vision by squashing virtual bugs in a virtual reality environment,” *Phil. Trans. R. Soc. B*, vol. 371, no. 1697, pp. 1–13, 2016.

[4] D. Scharstein, R. Szeliski, and R. Zabih, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” in *Stereo and Multi-Baseline Vision, 2001.(SMBV 2001). Proceedings. IEEE Workshop on*. IEEE, 2001, pp. 131–140.

[5] R. A. Hamzah, H. Ibrahim, and A. H. A. Hassan, “Stereo matching algorithm based on per pixel difference adjustment, iterative guided filter and graph segmentation,” *Journal of Visual Communication and Image Representation*, vol. 42, pp. 145–160, 2017.

[6] S.-S. Wu, C.-H. Tsai, and L.-G. Chen, “Efficient hardware architecture for large disparity range stereo matching based on belief propagation,” in *IEEE International Workshop on Signal Processing Systems (SiPS)*,. IEEE, 2016, pp. 236–241.

[7] Q. Liang, Y. Yang, and B. Liu, “Stereo matching algorithm based on ground control points using graph cut,” in *7th International Congress on Image and Signal Processing (CISP)*, IEEE, 2014, pp. 503–508.

[8] H. Hirschmuller, P. R. Innocent, and J. Garibaldi, “Real-time correlation-based stereo vision with reduced border errors,” *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 229–246, 2002.

[9] J. Kowalczuk, E. T. Psota, and L. C. Perez, “Real-time stereo matching on cuda using an iterative refinement method for adaptive support-weight correspondences,” *IEEE transactions on circuits and systems for video technology*, vol. 23, no. 1, pp. 94–104, 2013.

[10] Q. Yang, P. Ji, D. Li, S. Yao, and M. Zhang, “Fast stereo matching using adaptive guided filtering,” *Image and Vision Computing*, vol. 32, no. 3, pp. 202–211, 2014.

[11] J. Zbontar and Y. LeCun, “Computing the stereo matching cost with a convolutional neural network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1592–1599.

[12] W. Hu, K. Zhang, L. Sun, J. Li, Y. Li, and S. Yang, “Virtual support window for adaptive-weight stereo matching,” in *Visual Communications and Image Processing (VCIP), 2011 IEEE*. IEEE, 2011, pp. 1–4.

[13] N. Einecke and J. Eggert, “Anisotropic median filtering for stereo disparity map refinement.” in *VISAPP (2)*, 2013, pp. 189–198.

[14] R. A. Hamzah and H. Ibrahim, “Literature survey on stereo vision disparity map algorithms,” *Journal of Sensors*, vol. 2016, 2015.

[15] D. Scharstein and R. Szeliski, “Middlebury stereo evaluation - version 3 (accessed date: March 2017, http://vision.middlebury.edu/stereo/eval/references.”

[16] A. Geiger, M. Roser, and R. Urtasun, “Efficient large-scale stereo matching,” in *Asian conference on computer vision*. Springer, 2010, pp. 25–38.

[17] K. Zhang, J. Li, Y. Li, W. Hu, L. Sun, and S. Yang, “Binary stereo matching,” in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 356–359.

[18] G. Andreas, P. Lenz, and R. Urtasun. "Are we ready for autonomous driving? The KITTI vision benchmark suite." In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3354-3361, 2012.