Development of stereo matching algorithm based on sum of absolute RGB color differences and gradient matching

Rostam Affendi Hamzah¹, M. G. Yeou Wei², N. Syahrim Nik Anwar³

¹Fakulti Teknologi Kejuruteraan Elektrik and Elektronik , Universiti Teknikal Malysia Melaka, Malaysia ^{2,3}Fakulti Kejuruteraan Elektrik, Universiti Teknikal Malysia Melaka, Malaysia

ABSTRACT
This article presents local-based stereo matching algorithm which comprises a devel-
opment of an algorithm using block matching and two edge preserving filters in the framework. Fundamentally, the matching process consists of several stages which will produce the disparity or depth map. The problem and most challenging work for matching process is to get an accurate corresponding point between two images.
Hence, this article proposes an algorithm for stereo matching using improved Sum of Absolute RGB Differences (SAD), gradient matching and edge preserving filters.
It is Bilateral Filter (BF) to surge up the accuracy. The SAD and gradient matching
will be implemented at the first stage to get the preliminary corresponding result, then
the BF works as an edge-preserving filter to remove the noise from the first stage.
The second BF is used at the last stage to improve final disparity map and increase
the object boundaries. The experimental analysis and validation are using the Mid-
dlebury standard benchmarking evaluation system. Based on the results, the proposed
work is capable to increase the accuracy and to preserve the object edges. To make the
ment has been made to compare with other existing methods and it shows the proposed
work in this article perform much better.
Copyright © 2020 Insitute of Advanced Engineeering and Science.
All rights reserved.

Rostam Affendi Hamzah, Fakulti Teknologi Kejuruteraan Elektrik and Elektronik, Universiti Teknikal Malaysia Melaka, Malaysia. Email: rostamaffendi@utem.edu.my

1. INTRODUCTION

Computer vision is interdisciplinary field that comprises methods for acquiring, processing and analyzing and image understanding from digital images or videos. It is artificial intelligence to mimic the human visual system. Stereo vision is a part of them and the process to get the information such as object detection, recognition and depth estimation is called as stereo matching. This process starts with corresponding from one point on reference image to another point on the target image. These images can be two or more. In this article, the images are using from the stereo camera input which is also known as stereo images. The matching algorithm from the matching process produces disparity map. This map consists of depth information which is valuable for many applications such as virtual reality [1], 3D surface reconstruction [2], face recognition [3] and robotics automation [4-5]. The stereo baseline can be setup in a wide or short baseline [6] distance which depends on the applications. To determine the range or distance estimation, the triangulation function is applied to each of the pixel on the disparity map. Therefore, to get an accurate result, the matching process requires complex and challenging solution for depth or distance estimation. It requires precise function on the propose framework. Fundamentally, matching algorithm consists of multiple

stages which was proposed by Szeliski and Scharstein [7]. First stage, matching cost computes the preliminary matching point of stereo image. Second stage, the filtering is utilized to reduce the preliminary noise of the first stage. Then, disparity selection and optimization stage normalizes the disparity value each pixel on the image. Last stage is to refine the final result and also known as disparity map post-processing step.

In stereo matching development, there are two major approaches available in developing the algorithm framework. It is local methods as published in [8-10] and global method [11]. Mostly local methods use local properties or local contents using windows-based technique such as fixed windows implemented in [12-13], adaptive window [14], convolution neural network [15] and multiple windows [16]. In common, Winner-Takes-All (WTA) strategy is applied for local based optimization. It is low computational complexity and fast execution time [17-19]. Local method such implemented in [20] that used plane fitting technique to increase the accuracy at the final stage. This method also known as RANSAC that efficiently works on the low textured areas. However, the error still occurred on the object edges. Their method requires several iterations for plane fitting process. If wrong iterations, then it will affect the results. Commonly, local methods show fast running time, but low accuracy on the edges due to improper selection of windows sizes. Hence, to get an accurate result for the local approach is a challenge to the researchers.

Another approach in stereo matching algorithm to produce the disparity map is global optimization method. Fundamentally, this method uses energy-based function which is known as Markov Random Field (MRF). The method in global optimization such as Belief Propagation (BP) [21] and Graph Cut (GC) [22] produce accurate results. Each pixel of interest calculation requires all pixel's energy in disparity map. It calculates neighboring or nearby pixels using maximum flow and the selection is made based on the minimum cut-off energy on the disparity map. The algorithms implemented using global optimization approach normally involve high computational requirement due to all pixel's energy calculation and absorption. Global methods involve iterations which increase the execution time each disparity map reconstruction. This article aims to produce accurate results and competitive with some established methods. The first function or stage will be implemented using improved Sum of Absolute Differences (SAD) [23] with gradient matching. Then, the second stage utilizes the edge preserving filter which is known as Bilateral Filter (BF) [24]. This filter is capable to remove noise and preserved object edges. The third stage is optimization based on WTA strategy. Last stage, the BF is applied once again to remove unwanted or remaining invalid pixels.

2. RESEARCH METHOD

The diagram of the proposed work is dispalyed by Figure 1. The stereo matching algorithm starts with STEP 1 to get the preliminary disparity map. The improved SAD has been proposed which the weighted technique is used on the block matching process. The combination of improved SAD with gradient matching in this article should be able to increase the effectiveness of corresponding process and accuracy. Then at STEP 2, the BP is utilized to reduce the noise and preserved the object edges. The BP is capable to efficiently remove noise on the low texture regions and sharping the object boundaries. The optimization uses WTA strategy which this method normalizes the floating points numbers and selects minimum disparity values on the disparity map. Final stage at STEP 4 is also using the BP but with the disparity values. This filter is a type of nonlinear filter and capable to improve final disparity map.



Figure 1. A flowchart of the proposed algorithm.

Int J Elec & Comp Eng	ISSN: 2088-8708	2377

2.1. Matching cost computation

The first stage of the proposed framework is using the weighted SAD. The preliminary disparity map is produced at this stage. Hence, robust function must be used to increase the effectiveness on the disparity map. The problem on matching process at this stage on the low texture regions must be at minimum. The weight is proposed at SAD to improve the values on the low texture regions. Thus, the consistency of the weight at the low texture region is capable to make the matching process accurate and reduces the mismatch or invalid pixels. The RGB values are used with the weight of sum of intensity differences on right image I_r and left image I_l which is given by (1):

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

where (x, y) are the coordinates pixel of interest with d represents the disparity value, W is the proposed weight, RGB channels numbers are i and w represents kernel of SAD algorithm. The second part is gradient matching components. It contains the magnitude differences from each image. There will be two directions that need to be calculated on this gradient differences. Vertical direction G_y and horizontal direction of G_x are the directions with the equations are given by (3) and (2):

$$G_y = \begin{bmatrix} 1\\0\\-1 \end{bmatrix} * Im \tag{2}$$

$$G_x = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} * Im \tag{3}$$

where Im is input image and * represents convolution operation on the gradient matching. The G_x and G_y are the gradient magnitude for m which is given by (4):

$$m = \sqrt{G_x^2 + G_y^2} \tag{4}$$

(5) is the gradient matching kernel G(x, y, d).

$$G(x, y, d) = |m_l(x, y) - m_r(x - d, y)|$$
(5)

The matching cost function at this stage is given by (6) where the input volume of SAD(x, y, d) and G(x, y, d) are combined together.

$$MC(x, y, d) = SAD(x, y, d) + G(x, y, d)$$
(6)

2.2. Cost aggregation

This second stage more likely to filter the preliminary disparity map from stage one. Normally the preliminary disparity map contains high noise and it must be removed. Some of invalid and uncertainties pixels are constructed during the matching process. Hence, at this stage the filter must be robust and is capable to remove high noise of invalid pixels and preserved the object boundaries. The BP is used due to strong preserving object edges and at the same time efficient to remove high noise especially on the plain color and low texture regions. (7) is the BF function used in this article.

$$W^{BF}(p,q) = \sum_{q \in w_B} exp\left(-\frac{|p-q|^2}{\sigma_s^2}\right) exp\left(-\frac{|I_p - I_q|^2}{\sigma_c^2}\right)$$
(7)

where p is the location pixel of interest at (x,y), w_B and q are window size of BF and neighboring pixels respectively. The σ_s denotes a factor of spatial adjustment and σ_c equals to similarity factor for the color detection. The p-q is spatial Euclidean interval and $|I_p - I_q|$ denotes the Euclidean distance in color space. Hence, (8) is the cost aggregation function of BF with the matching cost computation input.

$$C(p,d) = W^{BF}(p,q)MC(p,d)$$
(8)

2.3. Disparity optimization

This stage optimizes the disparity values on disparity map. The normalization is based on the minimum disparity values with the floating-point number which the WTA is selected in this article. The WTA is normally being used in the local based methods due to fast implementation. The WTA function is given by (9).

$$d_{x,y} = \arg\min_{d \in D} C(p,d) \tag{9}$$

where D represents a set of valid disparity values for an image and C(p, d) denotes the second stage of aggregation step. Fundamentally, after this stage the disparity map still contains noise or invalid pixels. Thus, this map needs to be improved and the last stage is will remove remaining noise.

2.4. Disparity refinement

The last stage of the algorithm framework is known as refinement or post processing stage. It has several continuous processes which starts with handling the occlusion regions, filling the invalid pixels and filtering final disparity map. The left-right consistency checking process is conducted to identify occlusion areas and some invalid pixels. Then, these invalid pixels are restored with valid pixel values through the filling process. Some of artifacts and unwanted pixels will be removed using the BF and at the same time preserved the object boundaries. The BF smoothes the final disparity map as indicates by (7).

3. RESULT AND ANALYSIS

This section explains about the disparity map results that will be represented by color-scale intensity. The different color tones show that the respected objects are mapped based on the disparity values and the distance sensor (i.e., stereo camera). Most probably the lighter intensity volume indicates that the object is closer to the sensor. The experimental analysis has been executed on a personal computer with Windows 10, 3.2GHz and 8G RAM. The input images are from the Middlebury stereo evaluation dataset [24] which contains 15 standard images and must be submitted online. These images are very complex, and each image consists of different characteristics and properties such as light settings objects depth, incoherence regions, different resolutions and low texture areas. The values of $\{w, \sigma_s, \sigma_c, w_B\}$ are $\{9x9, 17, 0.4, 11x11\}$.

Figure 2 shows a sample Jadeplant image (i.e., left and right) from the Middlebury training dataset with different brightness and high contrast. Generally, due to the brightness difference, these input images are very challenging to be matched. It contains different pixel values at the same corresponding point. However, the proposed algorithm is correctly discovered the disparity locations. The level of disparity contour are precisely assigned and object distance are well-recognized. Figure 3 shows the final disparity map results of 15 training images from the Middlebury dataset. The accuracy attributes for error evaluation are nonocc (non-occluded) and all error. The nonocc error is the error evaluation based on the non-occluded regions on disparity map while all error represents the all pixels' evaluation on an image of disparity map. Within these 15 images, Pipes and Jadeplant images are the most difficult images to be matched. These images comprise several piping lines and leaves with different sizes respectively. Yet, the propose algorithm can reconstruct almost accurate disparity map with clear discontinuities regions. Fundamentally, real images from the Middlebury are difficult and very challenging to get an accurate corresponding point. It was developed to test the robustness of an algorithm where same corresponding point maybe contains different pixel values. Additionally, each image contains difference characteristics such as plain color objects, shadow, discontinuity regions and occluded areas.

With referring to Figure 3, the disparity maps of low texture surfaces such as Motorcycle, MotorcycleP, Playtable and PlaytableP are well recreated with different depth and disparity contour. Other regions difficult to be matched are plain colour objects and shadow such as images of ArtL, Recycle, Piano and PianoL. These regions consist of similar pixel values and possibility to get wrong matching are very high. The disparity maps from the proposed work display almost accurate matching for these images. It shows that the proposed work is able to get correct matching pixels on these regions and robust against the plain colour areas. The quantitative measurement from the Middlebury online results are given in Tables 1 and 2. These results are produced by the Middlebury online benchmarking evaluation system with two error attributes as explained above. Some established methods are also included in these Tables to show the competitiveness of the proposed work. Overall, an average error measurement is assessed to rank the best results. For Table 1, the proposed method is ranked at top of the table with 6.11%, and Table 2 with 9.15%. It shows that the proposed work is competitive with other recently published methods and can be used as a complete algorithm. The proposed

work is rank at top compared to [15-17,19,25,26] for *nonocc* error. The weight average error is 6.11% where Jadepl, Playrm and Vintge images are the lowest error produced. For the *all* error attribute in Table 2, the proposed work is produced at 9.15% which is the lowest average error. It shows that the proposed work in this article is competitive with some established methods.



Figure 2. Different brightness and contrast of the input Jadeplant stereo images. Accurate disparity map result is produced by the proposed work compared to the image without the proposed work. The plant structures are clearly displayed and smooth disparity map can be notified



Figure 3. These final disparity images are from the Middlebury dataset which the results tabulated in Table 1 and Table 2

Table 1. The *nonocc* error results using the Middlebury dataset. The comparison results with other published methods

Algorithms	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlayP	Recyc	Shelvs	Teddy	Vintge	Weight Ave
Proposed Algorithm	3.66	4.23	19.54	3.11	3.87	5.01	11.02	6.32	5.11	24.85	5.10	3.33	7.52	2.21	7.90	6.11
SNCC [19]	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 [26]	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
MPSV [15]	3.83	6.00	19.70	5.85	5.53	5.68	34.30	9.59	5.86	15.30	4.20	4.59	13.00	3.70	14.30	8.81
ADSM [16]	13.30	6.10	15.00	3.67	5.67	7.08	20.60	6.57	13.20	23.10	3.55	5.76	17.20	3.05	10.10	8.95
DoGGuided [17]	15.20	9.57	27.10	5.64	8.31	8.09	32.40	9.67	14.00	24.50	5.32	5.56	16.20	4.15	15.00	12.00
BSM [25]	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

Development of stereo matching algorithm based on... (Rostam Affendi Hamzah)

Table 2. The <i>all</i> erro	or results using the N	Middlebury dataset.	These comparisons sho	w the competitiveness of
		the proposed m	ethod	

the proposed method																
Algorithms	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlayP	Recyc	Shelvs	Teddy	Vintge	Weight Ave
Proposed Algorithm	4.74	7.45	27.21	5.57	5.54	6.01	11.64	11.77	7.31	27.05	8.54	3.86	9.21	3.33	8.87	9.15
SNCC [19]	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 [26]	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
ADSM [16]	14.30	10.60	34.10	6.00	8.00	7.37	20.40	12.10	16.90	25.50	5.84	5.83	17.20	4.11	11.10	12.30
MPSV [15]	5.87	9.43	40.20	9.11	8.80	7.03	34.20	15.80	8.58	16.90	5.89	6.78	13.70	4.82	16.80	12.70
DoGGuided [17]	20.10	28.00	56.50	13.80	16.80	13.40	37.30	23.80	30.30	30.80	13.00	9.13	19.00	13.40	23.60	22.30
BSM [25]	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

To verify the potentiality of the proposed algorithm, the images from the KITTI [27] are also tested. These images are more difficult and challenging to be matched. It contains complex edges and structures such as shadow, plain color surfaces, high different contrast and brightness areas with large untextured regions. The experimental results are shown in Figure 4. The disparity map results show accurate disparity values estimation in grayscale. As for reference, the signage, a cyclist, trees and cars, are well-reconstructed with correct disparity level. It shows the proposed work in this article capable to work with difficult stereo images from real environment.



Figure 4. The disparity map results from the KITTI training dataset. This article utilizes images from the number of #000004_10-#000007_10

4. CONCLUSION

In this work, the combination of SAD algorithm based RGB color and gradient matching are producing accurate results. The second stage where edge preserving filter is utilized. The BF at the aggregation stage is capable to filter high noise and conserve the object boundaries of the preliminary disparity map. The WTA strategy was implemented at the optimization stage to normalize the floating points numbers to the disparity values. The second edge preserving filter was used at the last stage of the proposed work using the same BF. This nonlinear filter removed remaining noise and increase the efficiency of final disparity map. Overall, these edge preserving filters used in the proposed framework were able to remove noise especially on the low texture regions and able to preserve the object edges as shown by Figure 2. The quantitative measurement from the standard benchmarking Middlebury system also demonstrated low average errors were produced by the proposed framework at 6.11% and 9.15% of non-occluded and all pixel errors respectively. The training images are shown by Figure 3. From real images of the KITTI, the proposed work was also demonstrated accurate results.

ACKNOWLEDGEMENT

This project was sponsored by a grant from the Universiti Teknikal Malaysia Melaka with the Number: JURNAL/2018/FTK/Q00008.

REFERENCES

- S. F. Gani, et al., "Development of portable automatic number plate recognition (ANPR) system on Raspberry Pi," *International Journal of Electrical and Computer Engineering*, vol. 9, pp. 1805–1813, 2019.
- [2] R. A. Hamzah, et al., "Stereo Matching Algorithm for 3D Surface Reconstruction Based on Triangulation Principle," in International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2016, pp. 119–124, 2016.
- [3] S.F. Gani, et al., "Development of a portable community video surveillance system, *International Journal* of *Electrical & Computer Engineering*, vol. 9, pp. 1814–1821, 2019.
- [4] D Scharstein, et al., "A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms," in Stereo and Multi-Baseline Vision, 2001.(SMBV 2001). Proceedings. IEEE Workshop on, 2001, pp. 131–140.
- [5] R. Christian, et al., "Dense Wide-Baseline Scene Flow From Two Handheld Video Cameras," *IEEE Transaction on Circuit and Systems for Video Technology*, vol. 26, pp. 1–18, 2016.
- [6] Q. Yang, "A Non-local cost aggregation Method for Stereo Matching," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1402–1409.
- [7] H. Asmaa, et al., "Fast Cost-volume Filtering for Visual Correspondence and Beyond," *IEEE Transactions* on *Pattern Analysis and Machine Intelligence*, vol. 35, pp. 504–511, 2015.
- [8] H. Ibrahim, et al., "Stereo matching algorithm based on illumination control to improve the accuracy," *Image Analysis & Stereology*, vol. 35, pp. 39–52, 2016.
- [9] L. Qian, et al., "Stereo Matching Algorithm Based on Ground Control Points Using Graph Cut," in *Image and Signal Processing (CISP), 2014 7th International Congress on, 2014, pp. 503–508.*
- [10] W. Sih-Sian, et al., "Efficient Hardware Architecture for Large Disparity Range Stereo Matching Based on Belief Propagation," in *Signal Processing Systems (SiPS)*, 2016 IEEE International Workshop on, 2016, pp. 236–241.
- [11] H. Hirschmüller, et al., "Real-time Correlation-based Stereo Vision with Reduced Border Errors, International Journal of Computer Vision, vol. 47, pp. 1-3, 2002.
- [12] K. Jedrzej, et al., "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, pp. 94–104, 2013.
- [13] Q. Yang, et al., "Fast Stereo Matching Using Adaptive Guided Filtering," Image and Vision Computing, vol. 32, pp. 202–211, 2014.
- [14] J. Zbontar and Y. LeCun, "Computing The Stereo Matching Cost with a Convolution Neural Network," in IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1592–1599.
- [15] B. Jean-Charles et al., "Morphological Processing of Stereoscopic Image Superimpositions for Disparity Map Estimation," *Hal-01330139*, pp. 1–17, 2016.
- [16] M. Bing et al., "Accurate Dense Stereo Matching Based on Image Segmentation Using an Adaptive Multi-Cost Approach, Symmetry, vol. 8, p. 159,2016.
- [17] K. Masamichi, et al., "High Accuracy Local Stereo Matching Using DoG Scale Map," in *Fifteenth IAPR International Conference on Machine Vision Applications*, 2017, pp. 258–261.
- [18] W. Hu, et al., "Virtual Support Windows for Adaptive-weight Stereo Matching," in Visual Communications and Image Processing (VCIP), 2011 IEEE, 2011, pp. 1–4.
- [19] E. Nils and E. Julian, "Anisotropic Median Filtering for Stereo Disparity Map Refinement.," in VISAPP (2), 2013, pp. 189–198.
- [20] K.A.A. Aziz, et al., "A pixel to pixel correspondence and region of interest in stereo vision application", in 2012 IEEE Symposium on Computers & Informatics (ISCI), pp. 193–197, 2012.
- [21] M.G.Y. Wei, et al., "Stereo matching based on absolute differences for multiple objects detection," *Telkomnika*, vol. 17, pp. 261-267, 2019.
- [22] A. H. Hasan, et al., "Disparity Mapping for Navigation of Stereo Vision Autonomous Guided Vehicle," in Soft Computing and Pattern Recognition, 2009. SOCPAR'09. International Conference of, 2009, pp.

575–579.

- [23] R. A. Hamzah et al., "Visualization of image distortion on camera calibration for stereo vision application," in *Control System, Computing and Engineering (ICCSCE)*, 2012 IEEE International Conference on, 2012, pp. 28–33.
- [24] S. Daniel and S. Richard, "Middlebury Stereo Evaluation Version 3 (Accessed date : November 2019, http://vision.middlebury.edu/stereo/eval/references.," 2019.
- [25] K. Zhang, et al., "Binary Stereo Matching," in Pattern Recognition (ICPR), 2012 21st International Conference on, 2012, pp. 356–359.
- [26] A. Geiger et al., "Efficient Large-scale Stereo Matching," in Asian Conference on Computer Vision, 2010, pp. 25–38.
- [27] M. Menze and G. Andreas, "Object Sflow for Autonomous Vehicles," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3061–3070.

BIOGRAPHIES OF AUTHORS



Rostam Affendi Hamzah was graduated from Universiti Teknologi Malaysia where he received his B.Eng majoring in Electronic Engineering. Then he received his M. Sc. majoring in Electronic System Design engineering from the Universiti Sains Malaysia in 2010. In 2017, he received PhD majoring in Electronic Imaging from Universiti Sains Malaysia. Currently he is a lecturer in the Universiti Teknikal Malaysia Melaka teaching digital electronics, digital image processing and embedded system.



Melvin Gan Yeou Wei was born in 1992, Melvin Gan Yeou Wei graduated from Universiti Teknikal Malaysia Melaka where he received his B.Eng majoring in Electrical in 2017. Currently, he is pursuing a Master Degree in the Universiti Teknikal Malaysia Melaka.



Nik Syahrim Nik Anwar was born in 1981, Nik Syahrim graduated from University of Applied Science Heilbronn, Germany where he received his Diplom in Mechatronik und Mikrosystemtechnik majoring in Mechatronics in 2006. In 2010 he received his M. Sc. majoring in Mechatronics from the University of Applied Science Aachen, Germany. In 2018, he received PhD majoring in Electrical from the Universiti Sains Malaysia. Currently he is a lecturer in Universiti Teknikal Malaysia Melaka teaching Electrical and Mechatronics subjects.