# Gradient Magnitude Differences and Guided Filter for Stereo Video Matching Algorithm

R. A. Hamzah, M. Saad Hamid, A. F. Kadmin, S. Fakhar Abd Ghani, K. A. A. Aziz, S. Salam, T. M. F. T. Wook

*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 stereo video matching algorithm which uses Gradient Magnitude (GM) differences and Guided Filter (GF). The radiometric and edges distortions are the problems that contribute to the quality of the results for stereo video matching algorithm. Hence, this article proposes an algorithm to reduce these problems. The first stage, the GM is utilized. The GM is strong against the radiometric distortion on an image due to different brightness on an image or between the stereo cameras. The second stage, the GF is used to improve the edges of object matching and is efficiently to remove the noise. Based on the standard benchmarking stereo dataset, the proposed work in this article produces good results and performs much better compared to before the proposed framework. This new algorithm is also competitive with some established methods in the literature.**

***Index Terms*— Computer Vision; Gradient Matching; Guided Filter; Stereo Matching Algorithm.**

## I. Introduction

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

- Step 1: Matched cost computation (i.e., to calculate corresponding points of stereo images)
- Step 2: Cost aggregation (i.e., to reduce the noise after Step 1)
- Step 3: Disparity selection (i.e., to select the disparity value and optimization)
- Step 4: Post-processing and refinement (i.e., to refine final final disparity map)

There are two major optimization methods which are known as global and local methods [3]. The categorization is supported by the method on how the disparity is computed. The 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 approaches in global methods. These methods were Graph Cut (GC) [4] and Belief Propagation (BP) [5] which implemented based on MRF approach. 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 the current point to the nearest points or neighbours. Global approaches acquire low accuracy on radiometric distortion and they are a high computational demand for processing an image [6].

The local method employs a support window or region-based on predefined sizes. There are many published methods that associated to support regions or using window-based techniques for examples multiple windows, adaptive window, fixed window and convolution neural network [7]. Local approaches employ only local contents or features. The advantage of the local method is low computational demand and fast running time. The results produced 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 quality are low, particularly in the area of high radiometric distortions [3].

In current years, several local approaches have been formulated to get good results. The virtual support window was introduced by Hu et al. [8] which proposed a complicated method to change the support windows. Inecke and Eggert [9] enforced an adjustment of normalized cross correlation (NCC) feature at the first stage that reduces the preliminary disparity map efficiently. However, this method produces high noise on the edges. Commonly, local methods demonstrate low accuracy on the edge detection regions due to improper chosen of the window sizes. Hence, this problem is a challenge to the researches to come through and acquire good accuracy.

This paper proposes a new stereo video corresponding algorithm which uses GM and edge-preserving filter GF. This proposed work follows the framework of local method. Hence, the matching cost computation uses the GM algorithm, the cost aggregation and final stage employ the GF with WTA strategy. This paper is arranged as follows. Section 2 describes the suggested algorithm framework of stereo video corresponding algorithm. Then, Section 3 explains the setup and arrangement of the experimental settings in this article and the final part is the conclusion and acknowledgement.

## II. The Propose Algorithm Framework

Figure 1 shows a diagram of the propose corresponding work in this article. The first stage uses the GM as a pixel base correspondence measurement. The second stage known as cost aggregation step will be implemented using the GF which also known as an edge-preserving filter. After that, the WTA strategy is used at the disparity selection and optimization stage. The WTA uses the 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 different brightness regions on images. 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 second GF to reduce and remove the remaining noise which normally appears during the fill-in process.

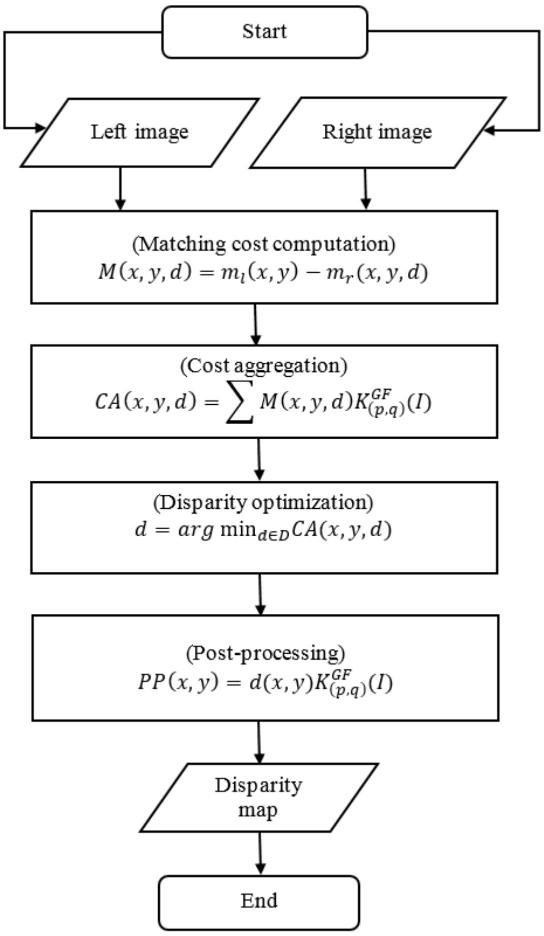



Figure 1: A flowchart of the proposed algorithm.

### A. Matching Cost Computation (MCC)

The MCC is the first stage of the algorithm framework which calculates the preliminary data of stereo matching differences for every pixel. Hence, this stage is the most important part of the algorithm framework. The approach used at this stage must be good and robust to avoid too many noises. Therefore, the proposed work uses the Gradient Magnitude (GM) differences. The GM calculates the direction along the grayscale values on the stereo images based on the gradient magnitude. The gradient values can be utilized to replicate image edges. The advantage of the GM is robust against the image noise that contains the local brightness variations between the image views. The GM represents by the gradient masks on an image *I(x,y)* which is shown by Equation 1 and 2 as follows:

$$G_x = [1 \quad 0 \quad -1] \circledast I \tag{1}$$

$$G_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \circledast I \tag{2}$$

where $I$ represents the image $I(x,y)$, $G_x$ and $G_y$ are the gradient masks of horizontal and vertical directions respectively. The $\circledast$ is the convolution sum operation. By using both of the components in Equation 1 and 2, the magnitude value $m$ is defined by Equation 3 as follows:

$$m = \sqrt{G_x^2 + G_y^2} \tag{3}$$

Therefore, the gradient magnitude differences or GM matching value of left and right images are given by Equation 4:

$$M(x,y,d) = m_l(x,y) - m_r(x,y,d) \tag{4}$$

where $(x,y)$ is the pixel of interest coordinates, $d$ represents the disparity value, $m_l$ and $m_r$ are the magnitude values of left and right images respectively.

### B. Cost Aggregation (CA)

The CA is the second stage of the local-based algorithm framework. This stage reduces noise from the MCC data to make the results more consistent before at the next stage of optimization and disparity selection. The proposed work in this article uses the GF. The GF was developed by He et al. [9] which is capable of reducing the noise and to preserve the edges of matching objects. The GF kernel is given by Equation 5:

$$K^{GF}_{(p,q)}(I) = \frac{1}{w^2}\sum_{(p,q)\in w_c}\left(1 + \frac{(I_p - \mu_c)(I_q - \mu_c)}{\sigma_c^2 + \varepsilon}\right) \tag{5}$$

where $\{w,p,q,c,I,\mu,\sigma,\varepsilon\}$ represented by {window support size, coordinates of $(x,y)$, neighboring coordinates, center pixel of $w$, reference image (left input image), mean value, variance value, constant parameter}. The GF is used in this article due to fast processing which relies on the image pixels and better filtering effect near the object edges. The final equation of this stage is given by Equation 6:

$$CA(x,y,d) = M(x,y,d)K^{GF}_{(p,q)}(I) \tag{6}$$

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

### C. Disparity Optimization and Disparity Selection (DODS)

The third stage of the framework is the DODS which minimizes the data selection on a location and represented it with disparity value. Generally, the local based stereo video matching algorithm is using Winner-Takes-All (WTA) strategy [11]. The WTA used the minimum value of $CA(x, y, d)$ and represented the same location with the disparity value. The DODS stage is given by Equation 7:

$$d(x, y) = \arg\min_{d \in D} CA(x, y, d) \tag{7}$$

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

### D. Post-processing (PP)

The PP is the last stage of the algorithm framework. This stage is also known as disparity refinement stage which reduces the noise and invalid disparity values on the results from the DODS [12]. Fundamentally, the used of second filtering process at this stage is to increase the efficiency and accuracy of the disparity map. The proposed work in this article is using the GF which works for the second filtering process. The final disparity result is determined from the Equation 8:

$$PP(x, y) = d(x, y) K^{GF}_{(p,q)}(I) \tag{8}$$

where the $d(x, y)$ is the disparity value at the coordinate of (x,y) and $K^{GF}_{(p,q)}(I)$ is the GF kernel with the left reference image $I$.

## III. Experimental Results

This section presents the experimental results and discussion on the performance of the proposed work. All of the experiment was implemented using a personal computer with the features of CPU i7-5500, 8G RAM and GPU GTX550. The standard benchmarking dataset have been used from the KITTI Vision Benchmark. This dataset was developed by Menze and Geiger [11] which consists of 200 training images. The stereo images and videos were recorded from the real environment of autonomous vehicle navigation using a stereo vision system. Hence, it contains very complex and challenging images for stereo matching algorithm processes. The performance is measured based on the *nonocc* and *all* error attributes of bad pixel percentages. The *nonocc* error is the error of invalid disparity values on the non-occluded regions. The *all* error is the error of invalid disparity values on all pixels of the disparity map image. The quantitative results in this article are measured from the development kit provided by the KITTI. The finest parameters of $\{w,\varepsilon\}$ are equal to $\{13\times13, 0.0001\}$ that have been used in the proposed algorithm.

Table 1 shows the quantitative results of the proposed framework with some published methods. As for comparison before and after the proposed work, the TestB is the algorithm without the proposed framework which uses a common algorithm with absolute differences matching technique. The proposed algorithm is ranked at the top of the table with 11.63% and 13.41% of *nonocc* and *all* errors respectively. If compared with the TestB algorithm, the proposed work reduces the *nonocc* and *all* errors with 17.75% and 20.15% respectively. The second algorithm produced by 3D, and followed by Hu, NCC, GC and BP.

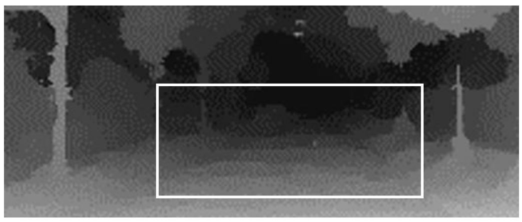
Table 1
The results of the KITTI training dataset.

| Algorithm | *nonocc* error (%) | *all* error (%) |
|---|---|---|
| Proposed work | 11.63 | 13.41 |
| 3D [1] | 12.75 | 14.22 |
| Hu [8] | 13.55 | 16.65 |
| NCC [9] | 14.11 | 15.79 |
| GC [4] | 14.34 | 15.98 |
| BP [5] | 15.67 | 17.84 |
| TestB | 29.38 | 33.56 |

Figure 2 shows the effectiveness of the proposed work with the radiometric difference on an image. The brightness changes of the tree's shadow in a white rectangle are well-detected on the disparity map. The different contour of disparity levels is precisely displayed on the result. Figure 3 shows the sample results of five continuous stereo video images from the KITTI dataset. It can be seen that the trees (in white circles) are missing on the disparity maps for the

TestB algorithm. However, the trees of the disparity maps for the proposed framework (Prop) are well-detected at the same positions. It shows that the proposed work in this article is robust against the radiometric difference and capable of preserving the object edges.

(a) left reference image



(b) disparity map result

Figure 2: The sample result of brightness difference (a) with the tree's shadow (b) disparity map result

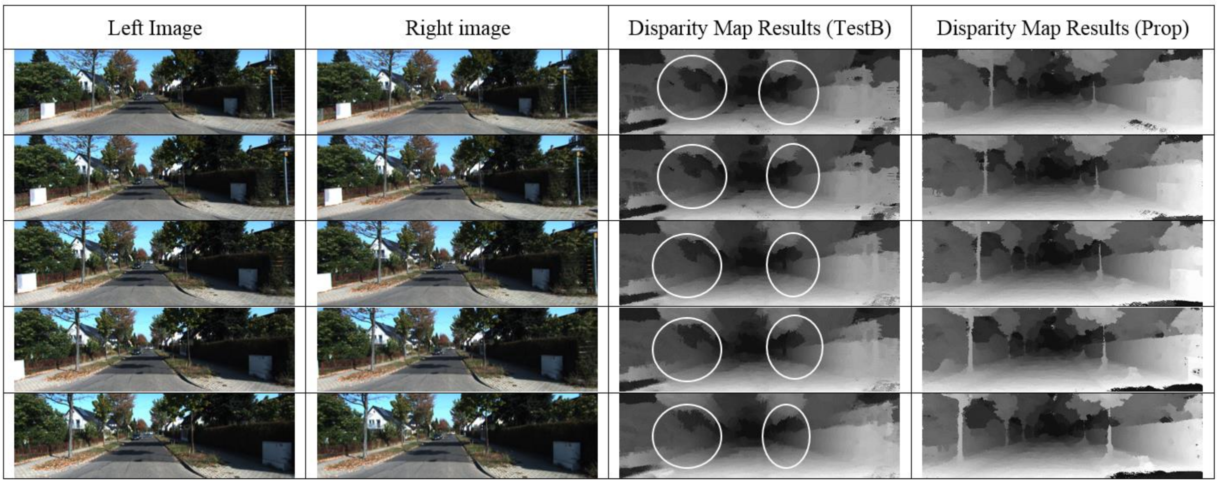



Figure 3: The sample of disparity map results using the KITTI dataset.

## IV. Conclusion

A new framework of stereo matching video processing is presented in this article. The GM is robust against the radiometric difference which is able to identify the regions with difference brightness. Additionally, the GF is able to increase the efficiency and preserve the edges of an object. Thus, this combination increases the accuracy compared with TestB or without the proposed work. The framework is also competitive with some established methods as tabulated in Table 1.

## Acknowledgment

This work is supported by a grant from the government of Malaysia with the reference number RAGS/1/2015/ICT01/FTK/03/B00115 and the Universiti Teknikal Malaysia Melaka.

## References

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

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

[3] R. A. Hamzah and H. Ibrahim, “Literature survey on stereo vision disparity map algorithms,” Journal of Sensors, vol. 2016, pp. 1-23, 2016.

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

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

[6] R. A. Hamzah, K.A.A. Aziz, and A.S.M. Shokri, “A pixel to pixel correspondence and region of interest in stereo vision application,” In IEEE Symposium on Computers & Informatics (ISCI), 2012, pp. 193-197.

[7] 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.

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

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

[10] K. He, J. Sun, and X. Tang, “Guided image filtering,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 6, pp. 1397–1409, 2013.

[11] 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.

[12] M. Menze and A. Geiger, “Object scene flow for autonomous vehicles,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3061–3070.

[13] R.A. Hamzah, A.F. Kadmin, S.F.A. Ghani, M.S. Hamid, and S. Salam, “Disparity refinement process based on RANSAC plane fitting for machine vision applications,” Journal of Fundamental and Applied Sciences, 9(4S), pp. 226-237, 2017.

[14] R. A. Hamzah, M. S. Hamid, H. N. Rosly, and N. M. Z. Hashim, “An Aligned epipolar line for stereo images with multiple sizes ROI in depth maps for computer vision application,” International Journal of Information and Education Technology, 1(1), pp. 15-19, 2011.