Improvement of stereo matching algorithm based on sum of gradient magnitude differences and semi-global method with refinement step

R.A. Hamzah✉ and H. Ibrahim

A new stereo matching algorithm which uses improved matching cost computation and optimisation using the semi-global method (SGM) is proposed. The absolute difference is sensitive to low textured regions and high noise on the stereo images with radiometric distortions. To get over these problems, sum of gradient magnitude differences has been introduced at the first stage. This method is strong against the radiometric differences on the stereo images. Hence, this approach will reduce the error of preliminary data for stereo corresponding process. The SGM is used at the aggregation, and optimisation stage uses 16 different directions of 2D path. Additionally, the iterative guided filter is utilised at the refinement stage which minimises the errors and increases the accuracy. The proposed work produces accurate results and performs much better compared with some established algorithms based on the standard stereo benchmarking evaluation from the Middlebury and KITTI.

Introduction: Stereo 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. 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 applications. Thus, the disparity map estimation is the most important and challenging jobs in computer vision research area [1]. 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 optimisation)
- *Step 4:* Post-processing and refinement (i.e. to refine final disparity map)

There are three major optimisation methods which are known as global method, local method and semi-global method (SGM). The categorisation is supported by the method on how the disparity is computed. The global method uses energy minimisation 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 minimisation technique is one of the famous approaches in global methods. These methods were graph cut (GC) and belief propagation (BP) 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 releasing indicators from current point to the nearest points or neighbours. Global approaches show good accuracy, but they require high computational demand to process an image.

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 example multiple window, adaptive window, fixed window and convolution neural network. Local approaches employ only local contents or features. The advantage of local method is low computational demand and fast running time. The results are produced by using a minimum raw data from matching cost and the lowest disparity values will be selected. Local methods are also known as winner takes all strategy in their optimisation stage. However, the local methods quality is low, particularly in the area of low texture regions.

In current years, several SGM [3] approaches have been formulated to get a good corresponding map results. The SGM uses the advantages of global energy function and local method. The energy function with smoothness term is used in the framework. Fundamentally, the first term of the SGM uses sum of the matching cost over all the pixels. This term is the most important stage of the SGM-based stereo matching algorithm. It consists of preliminary corresponding differences between the stereo images. Some established methods are absolute differences (AD), sum of absolute differences (SAD), sum of squared differences (SSD), normalised cross-correlation (NCC) and census transform (CT). Hence, at this term the error must be at the lowest value to increase the accuracy. This Letter proposes a new stereo matching algorithm based on the SGM. This Letter uses newly proposed sum of gradient magnitude differences (GMD) at matching cost. Additionally, a process is added at the disparity refinement stage which uses the iterative guided filter (iGF) to reduce the error of final disparity map results.

Matching cost computation: The first stage of the algorithm is to determine the corresponding values between the stereo images. A modification of GM has been proposed at this stage. GMD has been introduced which uses a fix window size. Therefore, the GMD matching values of left and right images are given by

$$SG(x, y, d) = \frac{1}{w^2} \sum_{(x,y,d)\in w_g} \left| m_l(x, y) - m_r(x, y, d) \right| \tag{1}$$

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

Cost aggregation and disparity computation: The cost aggregation process is one of the important stage for SGM which optimises the stereo matching algorithm framework. This stage aggregates the matching cost based on the path directions of SGM and selects the disparity values. This stage demonstrates overall performance of the proposed work. Fundamentally, the SGM is defined based on the energy function which contains the matching cost volume and smoothness constraints as given by

$$\begin{aligned} E(d) = \sum_p SG(p, d) + \sum_{q\in w_q} P_1 T[|d_p - d_q| = 1] \\ + \sum_{q\in w_q} P_2 T[|d_p - d_q| > 1] \end{aligned} \tag{2}$$

where p is the pixel of interest coordinates (x, y), P_1 and P_2 represent the constant penalty for all neighbourhood pixels in w_q which the condition is always $P_2 \geq P_1$. The T denotes Kronecker delta function which provides 1 when the condition in the bracket is satisfied. To minimise the energy function in (2), the cost along the path with different directions is formulated as

$$\begin{aligned} L_r(p, d) = SG(p, d) + \min\{&L_r(p - r, d), \\ &L_r(p - r, d - 1) + P_1, \\ &L_r(p - r, d + 1) + P_1, \\ &\min L_r(p - r, i) + P_2)\} \\ &- \min_D L_r(p - r, D) \end{aligned} \tag{3}$$

where r is set to 16 directions and D is a set of disparities. The cost for each disparity d at pixel p is calculated by sum of $SG(p, d)$ and the minimum path cost which also consider the penalty values. The final disparity value is selected at the minimum cost which given by

$$d(p) = \min_D \sum_r L_r(p, d) \tag{4}$$

Disparity refinement: The last step of the proposed algorithm is the disparity map refinement stage. It comprises several continuous processes that started with occlusion handling, managing the invalid pixels and removing remaining noise. The occlusion regions are determined by left-right consistency checking process. Then, these invalid pixels will be replaced by nearby valid corresponding pixels by the fill-in process. This replacing process creates unwanted artefacts on the disparity map. Hence, in this Letter the iGF is utilised to remove the unwanted artefacts. The filter kernel of the iGF $G_{p,q}(I_n)$ is given by

$$G_{p,q}(I_n) = \frac{1}{|w|^2} \sum_{q\in w_k} \left(1 + \frac{(I_{p,n-1} - \mu_{k,n-1})(I_{q,n-1} - \mu_{k,n-1})}{\sigma^2_{k,n-1} + \epsilon} \right) \tag{5}$$

where n is the iteration number, I_n denotes the guidance greyscale image, w_k represents the support window with the size of $r \times r$ and w is the number of disparity values in w_k. σ, μ and ϵ represent the variance, mean of the intensity values and control element for the smoothness term, respectively. The work uses the same approach in [4] for the guidance greyscale image during the iterations.

Experimental results: All of the analyses are implemented on the computer system of Window 10, 3.2 GHz processor and 8 GB memory. The standard database from the Middlebury at http://vision.middlebury.edu/stereo/eval3/ benchmarking stereo evaluation has been used to measure the accuracy. This benchmarking system provides 15 training images and could be evaluated based on the bad pixel percentage of all pixels (*all*) and non-occluded pixels (*nonocc*). Additionally, another standard benchmarking system is added in this Letter from the KITTI at http://www.cvlibs.net/datasets/kitti/eval to measure the accuracy based on the real stereo images. The parameter values used in this work were $\{w_g, w_k, n, \epsilon\}$ with the values of $\{11 \times 11, 9 \times 9, 3, 0.0001\}$.

Table 1 shows that the proposed GMD method performs much better compared with other state-of-the-art methods at matching cost computation. Fig. 1 shows the disparity map of Adirondack image for the methods in Table 1. The GMD produces good contour of disparity levels with lowest noise displayed on the image. Tables 2 and 3 display the results of the Middlebury and KITTI quantitative measurements. Based on these tables, the proposed work is more accurate at $r = 16$ which means more directions of SGM path will produce more accurate disparity estimation. Additionally, it shows that the proposed stereo matching algorithm in this Letter is competitive with some recently published methods in the Middlebury and KITTI databases.

Table 1: Comparison of the matching cost computation methods using the Middlebury training database

	Average of bad pixel (%)	
	All	Nonocc
Proposed work (GMD)	9.65	8.82
CT	10.53	9.78
NCC	13.23	12.18
SAD	17.41	16.25
SSD	18.73	17.15
AD	24.61	22.54

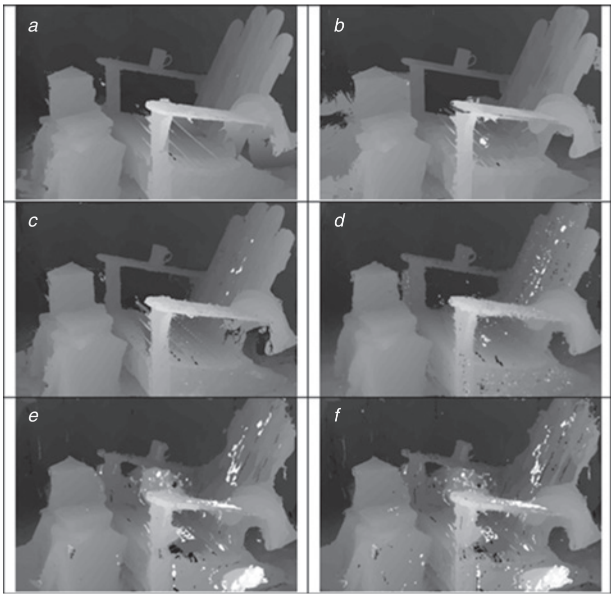

Fig. 1 *Results of Adirondack image from the Middlebury which display as follows*

a GMD
b CT
c NCC
d SAD
e SSD
f AD

Table 2: Comparison of the results of the proposed method with other published methods using the Middlebury training database

	Average of bad pixel (%)	
	All	Nonocc
Proposed work ($r = 16$)	9.65	8.82
Proposed work ($r = 8$)	10.11	9.21
Proposed work ($r = 4$)	12.05	11.45
[5]	12.30	8.97
[6]	12.70	8.81
[7]	22.30	12.00

Table 3: Comparison of the results of the proposed method with other published methods using the KITTI database

	Bad pixel (%) based on D1-bg	
	All	Nonocc
Proposed work ($r = 16$)	12.20	11.87
Proposed work ($r = 8$)	15.28	14.78
Proposed work ($r = 4$)	17.02	16.85
[8]	19.61	18.76
[9]	25.01	24.67
[10]	45.83	45.46

Conclusion: In this work, the SGM-based stereo matching algorithm was proposed. The algorithm used the combination of window based GMD algorithm and refinement stage using the iGF which is capable of producing accurate results based on the standard benchmarking datasets. Furthermore, the proposed framework is competitive with some recently established algorithms as shown in Tables 2 and 3.

Acknowledgments: This work was supported by the Universiti Teknikal Malaysia Melaka and Universiti Sains Malaysia.

© The Institution of Engineering and Technology 2018
Submitted: *23 October 2017* E-first: *6 June 2018*
doi: 10.1049/el.2017.3956
One or more of the Figures in this Letter are available in colour online.

R.A. Hamzah (*Fakulti Teknologi Kejuruteraan, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia*)

✉ E-mail: rostamaffendi@utem.edu.my

H. Ibrahim (*School of Electrical & Electronic Engineering, Universiti Sains Malaysia, Seberang Perai Selatan, Malaysia*)

References

1 Hamzah, R.A., and Ibrahim, H.: 'Literature survey on stereo vision disparity map algorithms', *J. Sens.*, 2016, **2016**, pp. 1–23
2 Scharstein, D., and Szeliski, R.: 'A taxonomy and evaluation of dense two-frame stereo correspondence algorithms', *Int. J. Comput. Vis.*, 2002, **47**, pp. 7–42
3 Hirschmuller, H.: 'Stereo processing by semiglobal matching and mutual information', *Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, pp. 328–341
4 He, K., Sun, J., and Tang, X.: 'Guided image filtering', *Trans. Pattern Anal. Mach. Intell.*, 2013, **35**, (6), pp. 1397–1409
5 Ma, N., Men, Y., Men, C., *et al.*: 'Accurate dense stereo matching based on image segmentation using an adaptive multi-cost approach', *Symmetry*, 2016, **8**, (12), p. 159
6 Bricola, J.C., Bilodeau, M., and Beucher, S.: 'Morphological processing of stereoscopic image superimpositions for disparity map estimation'. 2016, hal-01330139
7 Kitagawa, M., Shimizu, I., and Sara, R.: 'High accuracy local stereo matching using DoG scale map'. Int. Conf. on Machine Vision Applications, Nagoya, Japan, May 2017, pp. 258–261
8 Richardt, C., Kim, H., Valgaerts, L., *et al.*: 'Dense wide-baseline scene flow from two handheld video cameras'. Int. Conf. on 3D Vision, Stanford, CA, USA, October 2016, pp. 276–285
9 Peña, D., and Sutherland, A.: 'Disparity estimation by simultaneous edge drawing'. Asian Conf. on Computer Vision, Taipei, Taiwan, November 2016, pp. 124–135
10 Yang, Q.: 'A non-local cost aggregation method for stereo matching'. Conf. on Computer Vision and Pattern Recognition, Providence, RI, USA, June 2012, pp. 1402–1409