A New Structure of Stereo Algorithm Using Pixel Based Differences and Weighted Median Filter

Y. Gan[1], R.A. Hamzah[2], and N.S. Nik Anwar[1]

[1]Fakulti Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka
[2] Fakulti Teknologi Kejuruteraan Elektrik & Elektronik, Universiti Teknikal Malaysia Melaka

melvin.gyw92@gmail.com

Abstract—**This paper presents a new structure of developing a disparity map from stereo matching process. The stereo algorithm is applied here and the proposed framework has three stages which are matching stage with cost computation function, followed by second stage with optimization on the disparity map and finally, disparity map post processing or refinement step. First the procedure begins with matching cost computation stage where a combination of pixel based differences technique is applied. It is Gradient Matching (GM) and Absolute Difference (AD). Next, the algorithm proceeds to disparity optimization stage, where Winner-Takes-All (WTA) technique optimizes the disparity map by setting the lowest disparity value for each pixel. Then, weighted median (WM) filter is executed at the disparity refinement stage. Here, noise reduction and smoothening is implemented to the disparity map to form the final result. Finally, based on a standard benchmarking evaluation data set from the Middlebury, the proposed stereo algorithm has 38.0% accuracy for nonocc error and 41.9% accuracy for all error. In addition, the proposed work yields a better accuracy when compared to some of the works in the Middlebury Stereo.**

Keywords—disparity map; computer vision; stereo vision; stereo matching; median filter

I. INTRODUCTION

THE purpose of this paper is to propose a new framework for stereo algorithm, that will produce disparity map. Disparity map or also refer as depth map are then be used for applications such as 3D surface reconstruction in archaeological artifacts observation by Dellepiane et al. [1], 3D terrain reconstruction by Correal et al. [2]; which can be easily use for surveillance and exploration, object detection and depth estimation for industrial robotic, implemented by Dinham and Fang [3], applications in augmented reality system [4] implemented by Markavic et al. and many more. For this stereo algorithm, the three stages of taxonomy are applied for obtaining disparity map. These stages are the matching cost computation, disparity optimization and disparity refinement.

In this work, image pairs from Middlebury Stereo, a standard benchmarking is used. During the first stage; matching cost computation, the system begins with an image pairs; left image (reference image) and right image. During the matching stage, the pixel from the image pairs will correspond to form an initial disparity map. Next, disparity optimization process will optimize the function of each pixel in disparity map by setting a disparity level. Finally, at disparity refinement stage, the disparity map is further refine using post processing; Left Right (LR) consistency checking process and fill-in process to replace invalid pixel.

Base on the literature survey by Rostam and Haidi [5], there are many techniques that had been implemented for matching cost computation and each has it pros and cons. In this work, the matching process apply the combination of two techniques, Absolute difference (AD) and Gradient Matching (GM). AD algorithms required simple computation and have fast processing speed, but is easily

Article history: Manuscript received 30 April 2019; received in revised form 22 August 2019; Accepted 26 August 2019.

interrupt low texture region, resulting low accuracy for stereo vision. Therefore, with the pairing of AD and GM, it produces radiometric distortion reduction characteristic.

Disparity map optimization method mainly categorized into local, global and semi global method. The method classification is decided by the way the disparity is calculated. Each approach has their own pros and cons, for example, local approach is computational simple as the system only uses information from pixel that is close to each other, through the process of comparing, therefore requires shorter processing time making it more suitable for real time implementation. However, it has low accuracy due to high noise sensitivity and mismatches. Global methods aim to identified all disparities by minimizing the energy function at once [6], making this method complex, longer processing and expensive. Semi global method is the integration of both approach, making the result varies depending on the integration and techniques applied.

The propose stereo matching algorithm uses a common local method; Winner-Takes-All (WTA) technique. In the refinement stage, post processing is applied and an image edge preserving technique, weighted median filter with bilateral filter is used to clean up noise that presented in the final disparity map.

This purpose of the work is to propose a new framework for stereo vision to obtain disparity map. The algorithm uses the combination of Absolute Differences (AD) and Gradient Matching (GM) at the matching stage. Here, the disparities of both methods integrate with each other after each respective method corresponded the pixel of their left image to the pixel of their right image. Next, a local method that require less computational time, less complexity and cheaper - Winner-Take-All (WTA) technique is used for disparity optimization. Then, with post processing technique and Weighted Median (WM) filter, the final disparity map is refined by removing, reducing and smoothening the noises.

The accuracy of the proposed stereo algorithm is determined by sending the proposed results to a standard benchmarking dataset which is the Middlebury Benchmarking Stereo. In addition, fifteen sets of stereo images were provided, which are then operate on the proposed stereo algorithm for final disparity maps. The results are then uploaded back to online benchmark dataset, for quantitative result. The result shows the overall error and also error for each disparity map respectively.

The remaining of this article is prepared as follows. The methodology on the proposed algorithm is shown in section 2. Section 3, 4 and 5, each will present the taxonomy approach of the work and section 6 is the result of the experiment. Finally, the conclusion is summarized in section 7.

II. METHODOLOGY

The proposed stereo algorithm is separate into three different stages and the block diagram of the framework is presented in Fig. 1. During the first stage, the framework begins by matching the image pairs; left image and right image. During this process, the pixel from left image will correlate with pixel from the right image to form an initial disparity map. This process is executed respectively for both technique Absolute difference (AD) and Gradient matching (GM). The disparity maps of both techniques is then integrated, creating a new integrated initial disparity map.

At the second stage; disparity optimization stage, the disparity map is optimized with a common local technique, Winner-Take-All (WTA). Then, for third stage, the process continues with post processing that is Left Right (LR) consistency checking and the filling in process of invalid pixel. Weighted Median (WM) filter is that applied to secure the final disparity map.

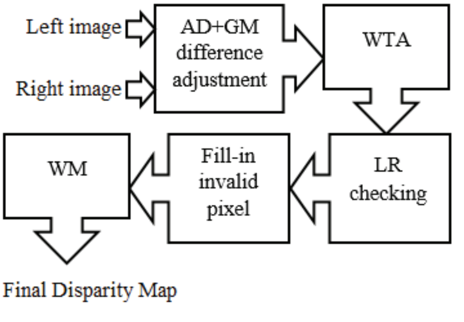

Fig. 1. Block diagram for proposed algorithm.

III. MATCHING COST COMPUTATION

Matching cost computation stage is where left image will correlate with right image resulting in a disparity map. For the proposed algorithm, pixel based differences techniques such as AD and GM are selected. AD technique is computationally simple and has fast processing speed, but it still has it challenges, such as, high error on surface with low texture. To overcome these, GM is proposed to integrate with AD. This is due to radiometric distortion reduction characteristic of GM technique. Implemented by Hamzah et al. [7], the matching cost started with absolute difference (AD), integrated with gradient matching (GM). Here, respectively, the pixels from left image correspond to the pixel from the right image to form the initial disparity map for both AD and GM. The equation for AD is shown in (1):

$$\mathrm{AD}(\mathrm{p},\mathrm{d}) = |(\mathrm{I_l}(\mathrm{p}) - \mathrm{I_r}(\mathrm{p},\mathrm{d})| \tag{1}$$

where p = (x,y), is the coordinate of the targeted pixel, while d is disparity or depth value. Then, representing left image I_l and representing right image. GM technique is integrated together with AD technique for enhancing the accuracy during matching process. AD technique implemented by Tan and Monasse [8] produces high distortion in disparity map, especially low texture region and error at boundaries. By integrating this both technique together, it is able to transcend the high distortion.

Next, through GM method implemented by Zhu and L_i. [9], the gradient is calculated and acquired for each stereo image. In (2) and (3), the equations represent the gradients values,

$$\mathrm{G_x} = [1 \quad 0 \quad -1] * \mathrm{I} \tag{2}$$

$$\mathrm{G_y} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} * \mathrm{I} \tag{3}$$

where G_x represent the direction horizontally and G_y represent the direction vertically. Next, I define the selected image and * denote the convolution operation. The value from G_x and G_y. is then applied to gradient magnitude, m, which is shown in (4):

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

The modulus of gradient magnitude, m in (4) is done separately on the image pair; left image, m_l and right image, m_r. With the gradient displacement from the direction of x and static position from the direction of y, gradient matching cost, GM(p,d) is compute, which is shown in (5):

$$\mathrm{GM}(\mathrm{p},\mathrm{d}) = |\mathrm{m_l}(\mathrm{p}) - \mathrm{m_r}(\mathrm{p},\mathrm{d})| \tag{5}$$

where m_l denotes the gradient value from left image and m_r, denotes the gradient value from the right image. By integrating of AD and GM technique, matching cost function M(p,d) is form, presenting in (6):

$$\mathrm{M}(\mathrm{p},\mathrm{d}) = \mathrm{AD}\ (\mathrm{p},\mathrm{d}) + \mathrm{GM}\ (\mathrm{p},\mathrm{d}) \tag{6}$$

IV. DISPARITY OPTIMIZATION

In this stage, a common local technique; Winner-Take-All (WTA) is utilize as it reduce complexity, which implemented by S. Lee et al. [10]. WTA normalized the disparities of each pixel in depth map. Here, with disparity value obtain from matching cost computation, the lowest value is assigned to each pixel at the disparity map. The formulation for the WTA technique is shown on (7):

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

where D is the range of disparity for an image, d (x,y) is the selected disparity value at the position of (x,y) and M(p,d) is the value from the matching cost function; stage one.

V. DISPARITY REFINEMENT

During this stage, the disparity will receive regularization and interpolation [5]. Regularization is normally done using filter where invalid pixels are remove or replace while interpolation fill in the invalid pixels. The process begins with post processing, LR checking and the fill-in process. The Left Right (LR) checking method, implemented by Hamzah et al. [7] and Kordelas et al. [11] corresponds the disparity map from left reference with disparity map at right reference. Mismatched disparities between those two are determine as flawed or invalid. The equation for LR checking is presented in (8):

$$|d_{LR}(p) - d_{RL}(p - d_{LR}(p))| \leq \tau_{LR} \quad (8)$$

Next, the filling in process of invalid pixel was executed by conditioning the left image as reference. Here, the filling in or replacement process for corrupted pixel begin from left side of the disparity map and later progressed to the right. Then, the corrupted disparity is rewrite by valid disparity value that are closest to the corrupted disparity. Furthermore, the valid disparity value must be presence on same scanning line.

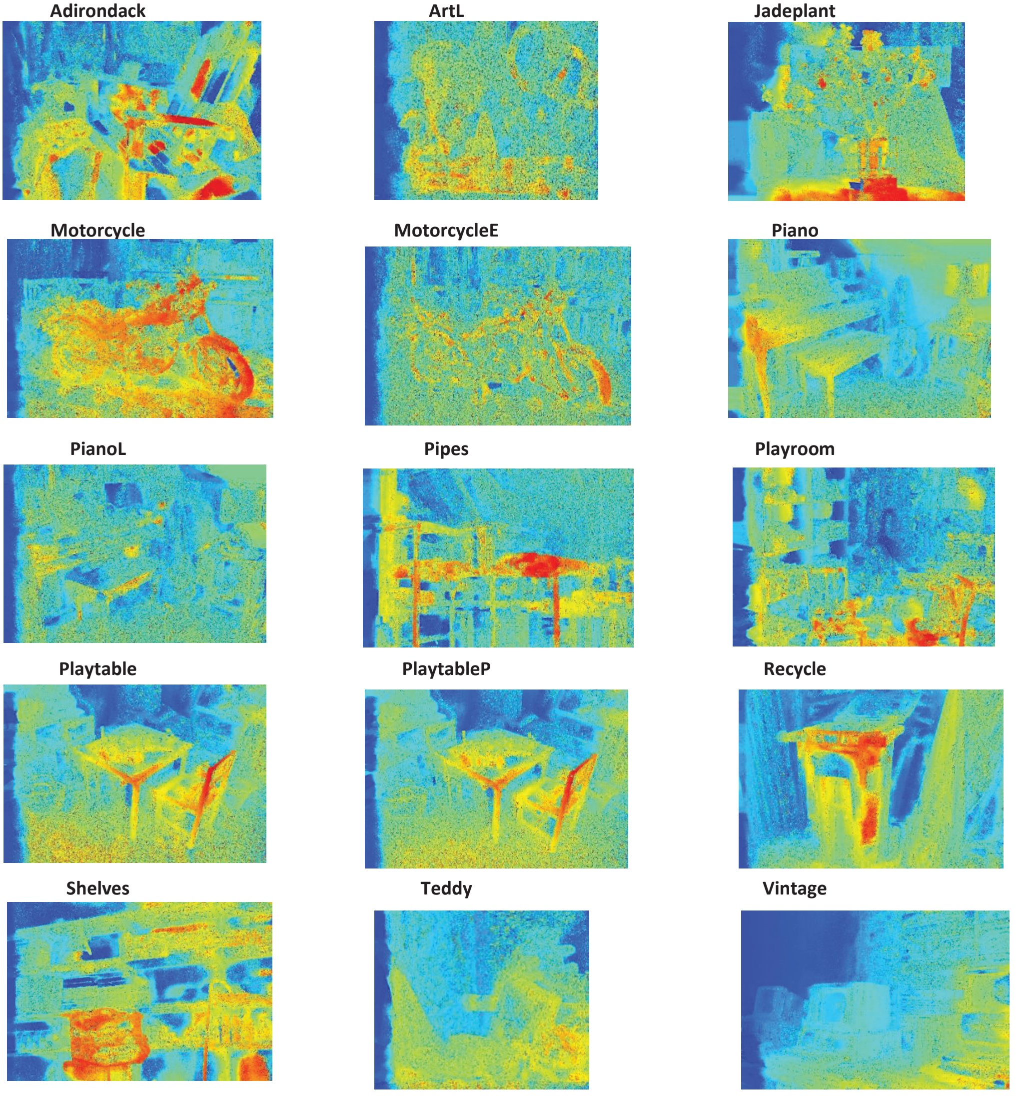

Fig. 2. The disparity map results of the Middlebury Training Dataset.

The equation of the fill-in process is shown in (9):

$$d(p) = \begin{cases} d(p-i), & d(p-i) \leq d(p+j), \\ d(p+j), & \text{otherwise.} \end{cases} \quad (9)$$

where d(p) represent the disparity at the position p, (p-i) represent the position of the first valid disparity at the left side and (p+j) represent the position of the first valid disparity at the right side. Post processing resulting undesirable line or dotted artefacts on the disparity map. To filter out some noises located within the disparity map, the process was continued by applying weighted median filter with bilateral filter. Bilateral filter equation B(p,q) is presented in (10):

$$B(p,q) = exp\left(-\frac{|p-q|^2}{{\sigma_s}^2}\right) exp\left(-\frac{|d(p)-d(q)|^2}{{\sigma_c}^2}\right) \quad (10)$$

where (p,q) is the target pixel , and $|d(p) - d(q)|^2$ is Euclidean and |p-q| represents spatial Euclidean. σ_s^2 and σ_c^2 both respectively represent the value of spatial distance and color similarity parameters. To improve the features of disparity map, an edge preserving filter is recommended which is the bilateral filter. The stereo algorithm uses a weighted median which has been implemented by Ma et al. [12]. Here, implemented by Tan and Monase [8], a higher weight is applied to the filter especially to disparities that have a same value and spatially close depending on the adjustment in sigma. The weighted of bilateral filter B(p,q) is later transform into sigma of histogram $h(p,d_r)$ which result in (11).

$$h(p, d_r) = \sum_{q \epsilon w_p | d(q) == d_r} B(p,q) \quad (11)$$

where, d_r refer the range for disparity while w_p is the mask (window size) with the radius (r × r) at centered p. The WM is the value of final disparity on the map and the $h(p,d_r)$ is given by (12):

$$WM = med\{d | h(p, d_r)\} \quad (12)$$

VI. RESULT AND DISCUSSION

In this work, the experimental test images are using standard benchmarking evaluation system for stereo algorithm which is the Middlebury Stereo [14]. The Middlebury Stereo dataset provided fifteen set of stereo image where each set image comes in pairs, right image and left image. The proposed stereo algorithm is executed based on a personal computer (CPU i7 and 8G RAM). The experiment focuses on nonocc error and all error, provided by the quantitative measurement from the Middlebury Stereo. The accuracy of nonocc error is obtained only from the regions with non-occluded areas in the disparity map. The nonocc calculation are based on the regions with clear pixel intensity. As for the accuracy of all error, it calculates all pixels in disparity map. It includes occluded and non-occluded regions. The final quantitative result is obtained through the analysis of Middlebury Stereo. Fig.2 shows the qualitative result, which consists of fifteen set of disparity map, that are operated by the proposed stereo algorithm.

Table I. The average results of the middlebury training dataset

Algorithm	all error (%)	nonocc error (%)
Proposed work	41.9	38.0
AVERAGE_ROB [15]	70.8	72.7
MEDIAN_ROB [15]	78.0	79.9

Table 1 shows the average results of the Middlebury evaluation errors. Based on this table, the proposed algorithm is ranked higher when compared with AVERAGE_ROB and MEDIAN_ROB. The accuracy is about 38.0% and 41.9% for nonocc and all. The ranking is then followed by AVERAGE_ROB and MEDIAN_ROB. By comparing with AVERAGE_ROB, ranked second, the propose algorithm has decreased in nonocc error and all error, which are by 34.7% and 28.9% respectively. Meanwhile, for MEDIAN_ROB, ranked last in accuracy, the proposed algorithm has more accurate in nonocc and all error by differences of 41.9% and 36.1%.

VII. CONCLUSION

A new framework; a local based stereo matching algorithm is proposed in this paper. It is using pixel base differences and weighted median filter. In addition, according to the results in Table 1, the proposed framework has higher accuracy for nonocc error and all error when compared with AVERAGE_ROB and MEDIAN_ROB. It shows that the proposed structure is capable to be used as a complete algorithm and competitive with some published methods.

ACKNOWLEDGMENT

The authors are in gratitude due to the financial support provide by Universiti Teknikal Malaysia Melaka for this article through PJP/2018/FTK(13C)/S01632.

REFERENCES

[1] M. Dellepiane , N. Dell'Unto, M. Callieri ,S. Lindgren, & R. Scopigno, "Archeological excavation monitoring using dense stereo matching techniques," Journal of Cultural Heritage, 14(3), pp. 201–210, 2013.

[2] R. Correal, G. Pajares, & J.J Ruz, "Automatic expert system for 3D terrain reconstruction based on stereo vision and histogram matching," Expert Systems with Applications, 41(4), pp. 2043–2051, 2014.

[3] M. Dinham & G. Fang, "Autonomous weld seam identification and localisation using eye-in-hand stereo vision for robotic arc welding. Robotics and Computer-Integrated Manufacturing," 29(5), pp. 288–301, 2013.

[4] M. Dosen, S. Cipriani, C. Popovic, D. & D. Farina, "Stereovision and augmented reality for closed-loop control of grasping in hand prostheses," Journal of neural engineering, 11(4), 046001–046009. Matsuo, T., Fujita, 2014.

[5] R. A. Hamzah and H. Ibrahim, "Literature survey on stereo vision disparity map algorithms," J. Sensors, vol. 2016. Art. no. 8742920, 2016.

[6] G.-S. Hong and B.-G. Kim, "A local stereo matching algorithm based on weighted guided image fltering for improving the generation of depth range images," Displays, vol. 49, pp. 80–87, 2017.

[7] RA. Hamzah, H Ibrahim, AH Abu Hasan, "Stereo matching algorithm based on per pixel difference adjustment, iterative guided filter and graph segmentation," J.Vis. Commun. Image R.42, pp. 145-160, 2017.

[8] P. Tan and P. Monasse, "Stereo disparity through cost aggregation with guided filter," Image Process. On Line (IPOL) 4, pp. 252–275, 2014.

[9] S. Zhu, Z. Li, "Local stereo matching using combined matching cost and adaptive cost aggregation," KSII T. Internet Inf. 9 (1), pp. 224–241, 2015.

[10] S. Lee, J.H. Lee, J. Lim, I.H. Suh, "Robust stereo matching using adaptive random walk with restart algorithm" Image Vision Comput. 37, pp. 1–11, 2015.

[11] G.A. Kordelas, D.S. Alexiadis, P. Daras, E. Izquierdo, "Content-based guided image filtering, weighted semi-global optimization, and efficient disparity refinement for fast and accurate disparity estimation," IEEE Trans. Multimedia 18 (2), pp. 155–170, 2016.

[12] Z. Ma, K. He, Y. Wei, J. Sun, E. Wu, "Constant time weighted median filtering for stereo matching and beyond," in: IEEE International Conference on Computer Vision, pp. 49–56, 2013.

[13] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešic´ , X. Wang, P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in: Pattern Recognition, pp. 31–42, 2014.

[14] Scharstein D and Szeliski R. Middlebury stereo evaluation - version 3 (accessed date: July 2018, http://vision.middlebury.edu/stereo/eval3/)