Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

DISPARITY REFINEMENT PROCESS BASED ON RANSAC PLANE FITTING FOR MACHINE VISION APPLICATIONS

R. A. Hamzah*, A. F. Kadmin, S. F. A. Ghani, M. S. Hamid and S. Salam

Faculty of Engineering Technology, Technology Campus, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia

Published online: 05 October 2017

ABSTRACT

This paper presents a new disparity map refinement process for stereo matching algorithm and the refinement stage that will be implemented by partitioning the place or mask image and re-projected to the preliminary disparity images. This process is to refine the noise and sparse of initial disparity map from weakly textured. The plane fitting algorithm is using Random Sample Consensus. Two well-known stereo matching algorithms have been tested on this framework with different filtering techniques applied at disparity refinement stage. The framework is evaluated on three Middlebury datasets. The experimental results show that the proposed framework produces better-quality and more accurate than normal flow state-of-the-art stereo matching algorithms. The performance evaluations are based on standard image quality metrics i.e. structural similarity index measure, peak signal-to-noise ratio and mean square error.

Keywords: computer vision; disparity refinement; image segmentation; RANSAC; stereo matching algorithm.

Author Correspondence, e-mail: rostamaffendi@utem.edu.my

doi: http://dx.doi.org/10.4314/jfas.v9i4s.13

1. INTRODUCTION

Stereo vision is active study for many years. This focus comprises methods for matching processing, acquiring, analyzing, understanding images, scene reconstruction, motion estimation, object recognition and image restoration [1]. The system of stereo vision consists of stereo camera system to estimate the depth. Two cameras are placed horizontally i.e., left and right positions. These two views produce stereo images that allowing the visual depth information to be recovered [2]. The difference between the views are used to plot or map the depth of the environment so call disparity map. The study of stereo matching algorithm is comprehensively studied problems in computer vision. The research trend is focusing on achieving high accuracy results. According to this, disparity map estimation algorithms can be considered as global and local methods. For local methods, the difference or disparity is calculated at a given image pixel which depends on numbers within a selected support window size. This method is also recognized as area based method. Unlike local method, global algorithm guesses the differences at one pixel using the disparity approximations at entirely the other pixels through energy minimization process. This process is to minimize a global cost function that combines data and smoothness terms. The smoothness cost compares computed disparities of neighboring pixels in disparity map, which greater difference in disparity will produce greater smoothness cost. According to [3], stereo matching usually achieves (subsets of) four steps as shown by Fig. 1 which: Step 1. Matching cost computation; Step 2. Cost (support) aggregation; Step 3. Disparity selection and optimization; Step 4. Disparity refinement. The first step measures the matches between two pixels, one is from left image and another one is from the right image. Some common used methods are sum of square difference, sum of absolute difference and normalize cross-correlation, etc. All local algorithms require step 2 and normally make plain smoothness assumptions by aggregating support. In order to include more texture data and decrease the affection of the noise and color inconsistence between the stereo images, the similarity measurements are aggregated in a local region such fix window, square window, shiftable window or weighted window. Therefore, this stage is also related to image filtering techniques. At this step, cost aggregation stage has great impact on the speed and accuracy of a stereo system. For global methods, this step is excluded in the developed algorithms since it is not using window-based techniques to aggregate the disparity values. In step 3, the disparity of each pixel can be projected by local methods like winner take all (WTA) approach or some other global methods such as belief propagation and graph cuts which include the flatness constraint between neighboring pixels.

After the three steps, preliminary disparity map can be produced with some outliers in the occlusion regions and textureless region. Step 4 is known as post-processing stage or refinement stage. Normally, at this stage, the methods are suggested to improve and to refine the disparity maps by various post-processing techniques such as cross checking, median filter, box filter, etc.

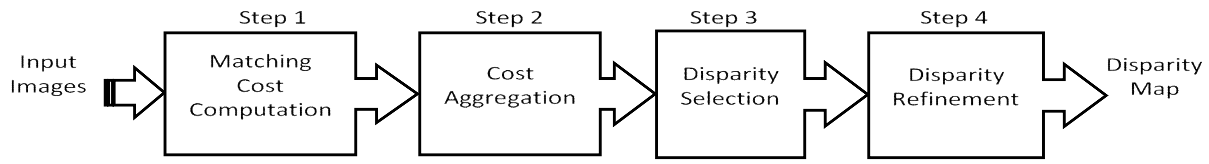

Fig.1. Stereo matching process

The purpose of disparity refinement stage is to reduce the noise and to improve the disparity maps. The refinement step consists of regularization and occlusion filling or interpolation process. The regularization process will reduce overall noise through filtering inconsistent and small variations of pixels on disparity maps. Interpolation process and occlusion filling are responsible for disparity approximations for these areas in which the disparity are unclear. Commonly, the occluded regions are filled with disparities from the textureless background area. The occlusion regions are detected by using left-right consistency check such implemented by [4-5]. The disparity refinement step combines local information from the local neighborhood near each measurement and a confidence metric. If the matching process rejects the disparities with a low confidence, the interpolation role estimates an approximation to the correct disparities from the local neighborhoods.

Typically, disparity map refinement methods such implemented by [6-7] reach an effective refinement of the original disparity map by manipulating the information provided by neighboring pixels. Two classical local disparity refinement techniques are Gaussian convolution and median filter. The Gaussian convolution combines a disparity estimation with those of its neighbors according to weights defined by a Gaussian distribution. This method mainly reduces the noise in the disparity map, but the Gaussian filter also reduces the amount of fine detail available in the final depth mapping. An effort done by [8] which imposed weight on Gaussian filter to improve the disparity maps by approximations the lost depth values using the nearby good depth pixels as a guide to prevent filtering across object boundaries. Median filter is another common filter which applied on disparity refinement. This filter is able to remove small isolated mismatches disparities via its edge preserving property and suitable with real time implementation due to low computational complexity. This filter selects the median value within the window pixels to acquire the final result of central pixel. In [9], the disparity maps refinement stage using median filtering is developed for real time stereo vision algorithm. Furthermore, the modification median filter by [10] which using constant time weighted technique. Ma's alteration archives high accuracy of removing noise and error while respecting the edge of disparity maps.

The diffusion technique performs a function similar to a Gaussian convolution, but an adaptation called anisotropic diffusion solves problems of the Gaussian convolution destroying edges and fine detail. Anisotropic diffusion apply smooth images without crossing any edges which implemented by [11] in their refinement disparity maps. Their approached have been improved by [12] through multi-resolution anisotropic diffusion. The disparity maps is downsampled by three different factors of resolution. Each resolution consists of 35 iterations of anisotropic diffusion. Another approach by [13] which employed two steps process to advance refine the estimated disparity maps. They have presented through color image guided depth matting process under the framework of Bayesian matting and 2D polynomial regression smoothing techniques. Nevertheless, the methods above are produced foreground fattening effect when computing the original disparity map [14].

In this paper, a new stereo matching algorithm has been proposed. The framework is added with BF and RANSAC plane fitting process to refine the disparity maps. Then plane fitting process is employed to obtain the disparity planes. This process will significantly enhances stereo matching performance, mainly in areas of low contrast and low texture. This new framework can be applied for any developing stereo matching algorithms at step 4 and able to deliver more accurate result. At the same time, this framework is able to enhance the structure of the output disparity map with comparison to the reference image (in this article is ground truth). The rest of this paper is organized as follows: Section 2 will discuss previous works on common techniques in disparity maps refinement. Section 3 explains briefly the proposed framework and plane projection. Section 4 describes about the quality assessment methods in this article through several standard quality metrics approaches. Experimental arrangements and results will be shown in section 5, followed by the conclusions from the proposed framework at Section 6.

2. METHODOLOGY

The proposed framework is shown by Fig. 2. Step 1 until step 3 have been explained in section 1. In common implementation of disparity map refinement as explained at section 2, step 4 only consists of filtering technique as shown by the dotted arrow (step 3 directed to filtering box at step 4). From this new proposed framework, the process of image segmentation to be added at step 4 with plane fitting process. The segmented image works as a reference image or mask to refine the final disparity map through the plane fitting process.

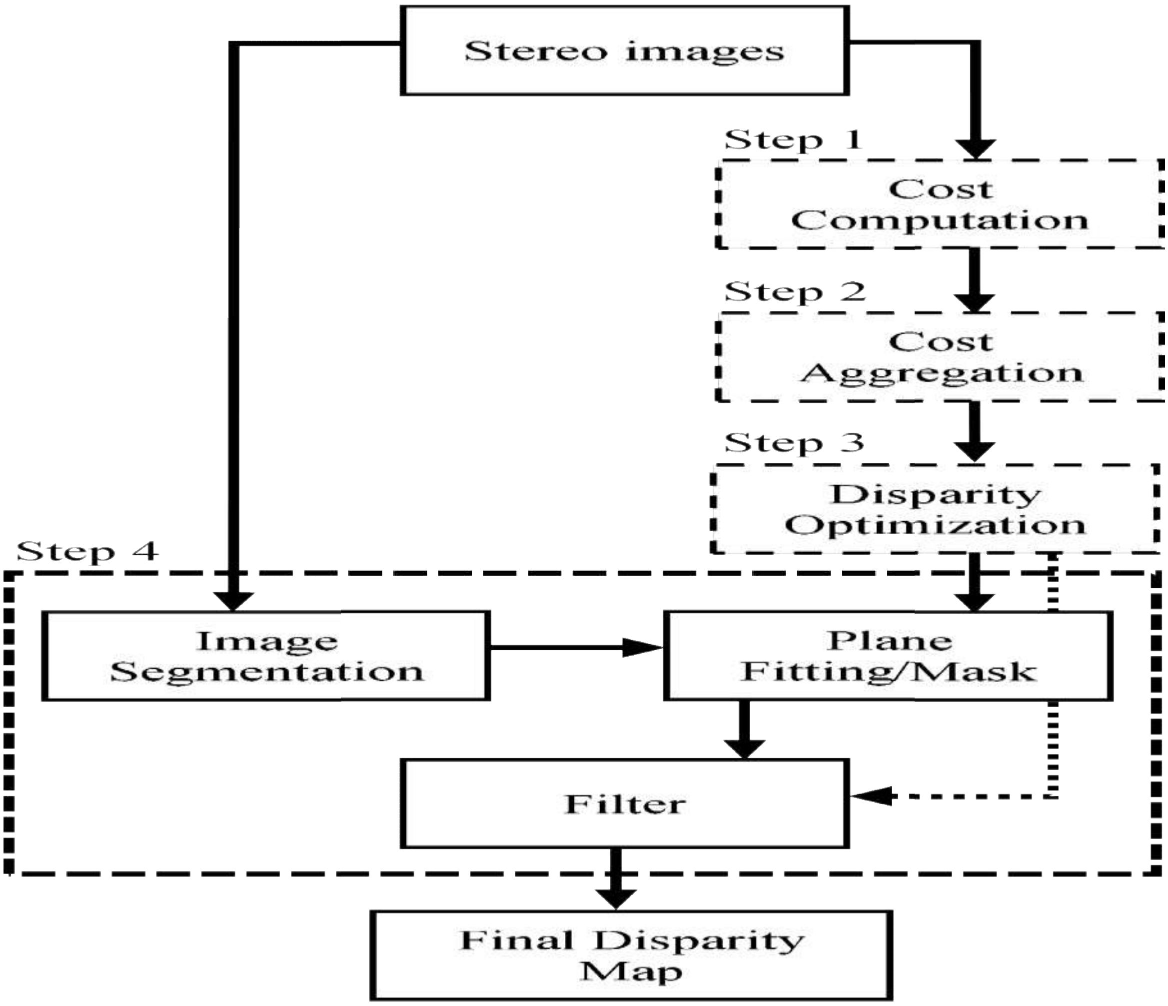

Fig.2. A framework of the proposed work

The process of segmentation uses two steps. Firstly, the edges are selected through the weights on each edge measure the dissimilarity between pixels intensity. Let edge E, i.e. (v_i, v_j) ϵ E corresponding to pairs of neighboring pixels of v_i and v_j. Each (v_i, v_j) ϵ E has a weight of w(v_i, v_j) which is a non-negative measure of the dissimilarity between neighboring elements of v_i, and v_j.. Any edge components in E are segmented from the conditions given as follows. The edges between two pixels in the same component should have relatively low weights and edges between pixels in different components should have higher weights. These two groups will formed the segmentation region of CϵS, which C is a component in segmentation region of S.

Secondly, evaluating the boundary between two components in segmentation process. The components are compared with Dif(C_1, C_2) at given Equation (1) to evaluate if there is indication for a boundary between a pair of components, the minimum value of $w((v_i, v_j))$ is selected. The purpose is to determine the segmentation group within these two regions.

$Dif C1,2 = min w((vi,vj)$ (1)

If the condition is true, then C_1 and C_2 will merge as a region otherwise the C_1 and C_2 remain as a different segment. The selection comparison is calculated as given by Equation (2):

$DC1C2 = true\ false\ if\ D(C1C2 < MintC1C2 otherwise$ (2)

The minimum difference $Mint(c_1,c_2)$ value of from Equation (2) as implemented by [15] is given by Equation (3) as follows:

$MintC1C2=m(Int\ (C1)+\tau C1, IntC2+ \tau(C2))$ (3)

The value of Int(C) represents a maximum value in a component. The threshold function τ given by Equation (4) is based on the size of the components, which k is a constant parameter and Z_c is the size of the component. The constant parameter sets a scale of observation, which a large scale causes for larger components in size. Smaller components are acceptable when there is a sufficiently small difference between neighboring components.

$\tau C = kZc$ (4)

2.1. Plane Fitting and Re-Projection

Based on the assumption that pixels in a segment are segmented to region or coplanar, plane estimation is one of the most effective ways for disparity refinement [16]. Before the plane fitting and re-projection take place, the size of the reference image must be equal to the size of the disparity map. RANSAC method estimates plane coefficients on segments defined on the reference image. RANSAC is selected in this work due to its minimization algorithm that can eliminate the outliers during the iterations [17]. The accuracy is projected by calculating the number of points within the segment that become the targeted plane projection. The plane fitting is succeeded with the points of the disparity map after the plane projection process.

2.2. Quality Measurement

In this paper, the evaluations are based on mean square error (MSE), peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). These methods of evaluations are tested on the final output of disparity map with grey-level (8 bits) images. Given Equation (5) and Equation (6) a reference image (ground truth) of r and test image of t, both having size of M×N, the MSE and PSNR between r and t which implemented and defined by [18] as follows:

$MSEr, = 1MNi = 1Mj = 1(r\text{-}t)2$ (5)

$PSNRr, = 10\ log10(2552MSEr,t)$ (6)

The PSNR value approaches infinity as the MSE approaches zero. This means that a greater PSNR value provides a greater image quality. The SSIM was developed by [18], and this method is used widely to measure the similarity between two images. SSIM is designed by modelling any images distortion as a combination of three factors that are loss of correlation, luminance distortion and contrast distortion. SSIM is defined as given Equation (7):

$\mathrm{SSIM}(\mathrm{r},\mathrm{t}) = \mathrm{l}(\mathrm{r},\mathrm{t})\mathrm{c}(\mathrm{r},\mathrm{t})\mathrm{s}(\mathrm{r},\mathrm{t})$ (7)

$lr,=2\mu r\ \mu t+C1\mu t2+\mu r2+C1$ (8)

$cr,=2\sigma r\ \sigma t+C2\sigma t2+\sigma r2+C2$ (9)

$sr,=\sigma tr+C3\sigma t\sigma r+C3$ (10)

The luminance evaluation function given Equation (8), which measures the closeness of the two images' mean luminance of μ_t and μ_r. This factor is maximal and equal to 1 only if $\mu_t = \mu_r$. The second term is the contrast evaluation function given by Equation (9), which measures the closeness of the contrast of the two images. The contrast is measured by the standard deviation σ_r and σ_t with maximal and equal to 1 only if $\sigma_r = \sigma_t$. The third term is the structure comparison function given by Equation (10), which measures the correlation coefficient. The σ_{rt} is the covariance between r and t. The value of SSIM output is positive and if the value is 0 means no correlation between images. If the output values is 1 then r = t. The positive constants C_1, C_2 and C_3 are used to avoid a null denominator.

3. RESULTS AND DISCUSSION

In this research, to show the improvement of the proposed framework, stereo images are using the standard datasets from Middlebury that can be downloaded at [18]. These datasets have been widely used for the benchmarking process in stereo matching algorithms development. The framework has been tested on two algorithms namely segment tree (ST) [19] and non-local aggregation (NL) [20]. The source codes of these two algorithms are provided by the algorithm's authors and can be downloaded from the internet. Additionally, these two algorithms have been modified to be used with different filtering techniques at refinement stage. This modification is to verify the new framework efficiency. Weighted median filter [21], guided filter [22], trilateral filter [23] and box filter [24] have been used with these two selected algorithms. The value of constant parameter k from Equation (3) is set about 300 in this experiment.

Table 1, Table 2 and Table 3 show the measurement result of before and after new framework implementations. The tables include the tested algorithms of modified ST and NL algorithms. The quality metrics measurement results indicate the new framework have improved all the methods measured. The results from the columns in new framework could refine the accuracy of disparity maps. The MSE values are decreased while implementing the new framework. Which means the errors are reduced. The PSNR and SSIM values are improved to verify the quality and structural of the disparity map. It consists of reference images and ground truth of three datasets (Venus, Teddy, Cones). This is to validate the quality metrics results and real output of disparity maps.

Table 1. Quality metrics result for Cones dataset based on segment tree and non-local aggregation algorithms

Image = Cones	**Normal Flow of Stereo Matching (Before)**			**New Framework With Segmentation Stage (After)**		
	MSE	**PSNR**	**SSIM**	**MSE**	**PSNR**	**SSIM**
ST + Weighted Median	392.09	22.19	0.920	390.78	22.21	0.930
ST + Guided Filter	408.38	22.02	0.918	407.40	22.03	0.926
ST + Trilateral	411.79	21.98	0.925	406.20	22.04	0.931
ST + Box Filter	393.23	22.18	0.928	390.78	22.21	0.929
NL + Weighted Median	396.90	22.14	0.920	396.82	22.15	0.929
NL + Guided Filter	407.43	22.03	0.922	405.77	22.04	0.927
NL + Trilateral	405.11	22.05	0.930	404.45	22.06	0.932
NL + Box Filter	394.37	22.17	0.927	391.38	22.20	0.929

Table 2. Quality metrics result for Teddy dataset based on segment tree and non-local aggregation algorithms

Image = Teddy	**Normal Flow of Stereo Matching (Before)**			**New Framework With Segmentation Stage (After)**		
	MSE	**PSNR**	**SSIM**	**MSE**	**PSNR**	**SSIM**
ST + Weighted Median	497.01	21.16	0.910	485.90	21.26	0.916

ST + Guided Filter	596.63	20.37	0.906	585.89	20.45	0.914
ST + Trilateral	594.10	20.39	0.915	583.65	20.46	0.921
ST + Box Filter	556.38	20.67	0.915	548.02	20.74	0.920
NL + Weighted Median	496.84	20.81	0.918	487.56	21.25	0.919
NL + Guided Filter	573.76	20.54	0.913	562.68	20.62	0.919
NL + Trilateral	571.53	20.56	0.922	560.67	20.64	0.926
NL + Box Filter	538.75	20.81	0.918	529.19	20.89	0.923

Table 3. Quality metrics result for Venus dataset based on segment tree and non-local aggregation algorithm

Image = Venus	**Normal Flow of Stereo Matching (Before)**			**New Framework With Segmentation Stage (After)**		
	MSE	**PSNR**	**SSIM**	**MSE**	**PSNR**	**SSIM**
ST + Weighted Median	15.65	36.18	0.976	12.94	36.89	0.983
ST + Guided Filter	15.86	36.12	0.977	13.82	36.72	0.982
ST + Trilateral	14.43	36.53	0.983	12.94	37.00	0.985
ST + Box Filter	13.43	36.84	0.983	12.26	37.24	0.984
NL + Weighted Median	14.33	36.56	0.978	13.41	36.85	0.984
NL + Guided Filter	13.87	36.71	0.980	13.01	36.90	0.984
NL + Trilateral	12.21	37.26	0.986	11.81	37.40	0.986
NL + Box Filter	11.82	37.40	0.985	11.69	37.45	0.985

4. CONCLUSION

A new framework for stereo matching algorithm has been introduced that performs significantly enhances stereo matching performance. By incorporating segmentation and plane fitting process, it possible to achieve improved accuracy as shown by comparison result in tables at section 5. There are three contributions while implementing this new framework. Firstly, the MSE values decrease. It indicates the errors can be minimized. Secondly, the PSNR values increases at most of the output stage. It shows that the higher quality output is produced. Finally, the SSIM values are increased demonstrate that the final structural disparity output are better than without the new framework implementation. This new framework well-matched to be implemented and applied to the tested stereo matching algorithms.

5. ACKNOWLEDGEMENTS

This work was supported by Ministry of Education Malaysia with account no. RAGS/1/2015/ICT01/UTEM/03/2 and Universiti Teknikal Malaysia Melaka.

6. REFERENCES

[1] Szeliski R. Computer vision: Algorithms and applications. : Springer Science and Business Media, 2010

[2] Hamzah R A, Ibrahim H. Literature survey on stereo vision disparity map algorithms. Journal of Sensors, 2016, 2016:1-23

[3] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, 2002, 47(1-3):7-42

[4] Yang Q, Wang L, Yang R, Stewénius H, Nistér D. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(3):492-504

[5] Heo Y S, Lee K M, Lee S U. Joint depth map and color consistency estimation for stereo images with different illuminations and cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(5):1094-1106

[6] Hamzah R A, Ibrahim H, Hassan A H. Stereo matching algorithm based on per pixel difference adjustment, iterative guided filter and graph segmentation. Journal of Visual Communication and Image Representation, 2017, 42:145-160

[7] Hamzah R A, Rahim R A, Noh Z M. Sum of absolute differences algorithm in stereo correspondence problem for stereo matching in computer vision application. In 3rd IEEE International Conference on Computer Science and Information Technology, 2010, pp. 652-657

[8] Vijayanagar K R, Loghman M, Kim J. Real-time refinement of Kinect depth maps using multi-resolution anisotropic diffusion. Mobile Networks and Applications, 2014, 19(3):414-425

[9] Michael M, Salmen J, Stallkamp J, Schlipsing M. Real-time stereo vision: Optimizing semi-global matching. In IEEE Intelligent Vehicles Symposium, 2013, pp. 1197-1202

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

[11] Banno A, Ikeuchi K. Disparity map refinement and 3D surface smoothing via directed anisotropic diffusion. Computer Vision and Image Understanding, 2011, 115(5):611-619

[12] Vijayanagar K R, Loghman M, Kim J. Real-time refinement of Kinect depth maps using multi-resolution anisotropic diffusion. Mobile Networks and Applications, 2014, 19(3):414-425

[13] Zhang S, Wang C, Chan S C. A new high resolution depth map estimation system using stereo vision and depth sensing device. In IEEE 9th International Colloquium on Signal Processing and its Applications, 2013, pp. 49-53

[14] Wang Y C, Tung C P, Chung P C. Efficient disparity estimation using hierarchical bilateral disparity structure based graph cut algorithm with a foreground boundary refinement mechanism. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(5):784-801

[15] Felzenszwalb P F, Huttenlocher D P. Efficient graph-based image segmentation. International Journal of Computer Vision, 2004, 59(2):167-181

[16] Xu L, Au O C, Sun W, Li Y, Li J. Hybrid plane fitting for depth estimation. In IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2012, pp. 1-4

[17] Lima D A, Vitor G B, Victorino A C, Ferreira J V. A disparity map refinement to enhance weakly-textured urban environment data. In 16th IEEE International Conference on Advanced Robotics, 2013, pp. 1-6

[18] Scharstein D, Szeliski R. Middlebury stereo evaluation-version 3. Vermont: Middlebury College, 2015

[19] Mei X, Sun X, Dong W, Wang H, Zhang X. Segment-tree based cost aggregation for stereo matching. In IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 313-320

[20] Yang Q. A non-local cost aggregation method for stereo matching. In IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1402-1409

[21] Zhang Q, Xu L, Jia J. 100+ times faster weighted median filter (WMF). In IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2830-2837

[22] Hamzah R A, Ibrahim H, Hassan A H. Stereo matching algorithm for 3D surface reconstruction based on triangulation principle. In IEEE International Conference on Information Technology, Information Systems and Electrical Engineering, 2016, pp. 119-124

[23] Yu Y, Dong G, Wang J. Despeckling trilateral filter. In IEEE 10th IVMSP Workshop, 2011, pp. 42-47

[24] Pires B R, Singh K, Moura J M. Approximating image filters with box filters. In 18th IEEE International Conference on Image Processing, 2011, pp. 85-88

How to cite this article:
Hamzah RA, Kadmin AF, Ghani SFA, Hamid MS, Salam S. Disparity refinement process based on ransac plane fitting for machine vision applications. J. Fundam. Appl. Sci., 2017, 9(4S), *226-237*.