TELKOMNIKA, Vol.17, No.1, February 2019, pp.261~267
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v17i1.9185

Stereo matching based on absolute differences for multiple objects detection

Rostam Affendi Hamzah*[1], Melvin Gan Yeou Wei[2], Nik Syahrim Nik Anwar[3]
[1]Universiti Teknikal Malaysia Melaka, Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik,
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
[2,3]Universiti Teknikal Malaysia Melaka, Fakulti Kejuruteraan Elektrikal,
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
*Corresponding author, e-mail: rostamaffendi@utem.edu.my

Abstract

This article presents a new algorithm for object detection using stereo camera system. The problem to get an accurate object detion using stereo camera is the imprecise of matching process between two scenes with the same viewpoint. Hence, this article aims to reduce the incorrect matching pixel with four stages. This new algorithm is the combination of continuous process of matching cost computation, aggregation, optimization and filtering. The first stage is matching cost computation to acquire preliminary result using an absolute differences method. Then the second stage known as aggregation step uses a guided filter with fixed window support size. After that, the optimization stage uses winner-takes-all (WTA) approach which selects the smallest matching differences value and normalized it to the disparity level. The last stage in the framework uses a bilateral filter. It is effectively further decrease the error on the disparity map which contains information of object detection and locations. The proposed work produces low errors (i.e., 12.11% and 14.01% nonocc and all errors) based on the KITTI dataset and capable to perform much better compared with before the proposed framework and competitive with some newly available methods.

Keywords*: absolute differences, bilateral filter, computer vision, guided filter, stereo matching*

1. Introduction

In recent years, stereo vision system has revealed a growing interest in machine vision applications such as 3D surface reconstruction [1], range estimation [2] and embedded system [3]. Fundamentally, machine vision technology requires information from an image. The information extracted can be a complex set of data for example position, orientation and identity of each object in an image. Hence, this paper proposes an algorithm to extract and collect this data based on stereo vision system. The developed algorithm will be processed the data and provide the extracted information such as the depth and coordinates of object detection. The algorithm process is also known as stereo corresponding algorithm. The improvement of this work creates the correspondence among a brace of images that will produced a disparity or depth map. The map contains the depth information and locations of the objects detection. The triangulation principle [4] is utilized to calculate the depth which will be used in numerous applications in machine vision applications such as industrial automations augmented reality (AR) [5] and robotic arm [6].

In real situation, the robotic movement requires an accurate depth estimation which the disparity map needs to be at the highest level of precision. Furthermore, this depth information is also capable to be used in 3D superficial reconstruction for the AR which visualizes the real environmental conditions. Therefore, the disparity map approximation is the most challenging and important works in stereo vision study areas especially in the computer vision field. Recently, numerous research articles have been issued in this research area and excessive improvement has been flourished. Fundamentally, there are multiple stages were developed by Scharstein and Szeliski [7] to build a stereo correspondence or matching algorithm. First, the matching cost computation which is to calculate the matching points between two images. Second, the cost aggregation stage reduces the noise after matching cost process. Third, the optimization and disparity selection stage which this step is to normalize the disparity values on

Received May 31, 2018; Revised October 30, 2018; Accepted November 26, 2018

each pixel on image. Last step is known as refinement stage or post-processing step to decrease the outstanding invalid pixels or noise on the final disparity map.

Generally, there are two foremost methods in stereo matching algorithm which are recognized as local and global methods [2]. The classification is held by the method on by what method the disparity is calculated. To define the disparity map based on the energy minimization method or function is used in global approach. The calculation is built on the flatness position from closer pixels which uses the minimization of global energy function. One of the well-known method in global method is the Markov Random Field (MRF) that measures the energy function. Several established methods using the MRF were Belief Propagation (BP) [3] and Graph Cut (GC) [8] which applied the energy minimization technique. The BP technique employed the MRF by incessantly absolution pointers from present point to the nearby points or neighbors. Else, the GC method uses the MRF approach by maximizing flow rule and cut the smallest energy flow structure. Global methodologies obtain low precision on images with radiometric problem and the framework require complex calculation to process high resolution images [9].

Local approach processes using local structures or contents. This method works based on a support region in predefined sizes. There are many published methods using local-based approach or using window-based methods such as convolution neural network [10], multiple window, fixed window and adaptive window. The benefit of a local approach is low calculation on the computational requirement and acquires fast time execution. Local approaches use the Winner-Takes-All (WTA) strategy in the disparity selection and optimization step. The WTA utilizes a lowest raw data from aggregation step. Then, it will be normalized the raw data to the minimum disparity value. Recently, numerous local approaches have been developed. Hu et al. [11] proposed a complex method using simulated support window. This method requires multiple estimations which increases the time execution. Inecke and Eggert [12] imposed a modification of normalized cross correlation (NCC) function at earlier step that decreases the accuracy of preliminary differences. Hence, their approach yields enormous noise and invalid pixels on the objects edges. Frequently, local approaches produce high error on the object edges because of inappropriate selection of the windows filter size. Therefore, the challenge is to get an accurate window size for local method to increase the accuracy.

This article offers a novel stereo matching algorithm using pixel-based matching cost computation. Then, the aggregation stage utilizes a guided filter (GF) [13] with fixed windows size. The GF is capable to eliminate the noise and preserved the object edge. The propose algorithm follows the structure of a local method. Therefore, the optimization is using the WTA strategy. The final stage of the algorithm, a bilateral filter (BF) is employed. The BF is strong against the illumination changes and efficient in removing the noise. This article is arranged as follows. Section 2 describes the methodology or the planned algorithm framework. Then, section 3 explains the experimental setup and the discussion of the disparity map results with qualitative and quantitative measurement. The final part is the conclusion and the acknowledgement.

2. Research Method

Figure 1 displays the framework of the proposed work. The first step uses pixel base correspondence method known as absolute differences. Then, a guided filter (GF) will be used at cost aggregation stage. After that, the WTA will be implemented at disparity optimization step to select the minimum disparity value. The final stage is using a bilateral filter (BF) to remove the remaining invalid disparity and increases the accuracy of final disparity map.

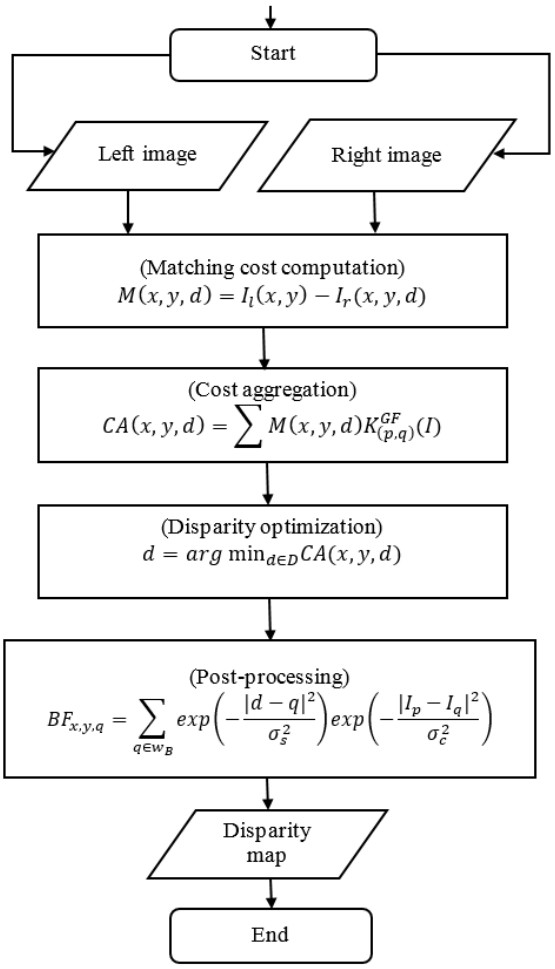

Figure 1. A flowchart of the proposed algorithm

2.1. Matching Cost Computation

This stage is the first step of the algorithm structure which estimates the matching differences of the preliminary data between stereo images. Therefore, this phase is the most important part in the algorithm structure. The method used at this step must be robust and is capable to avoid from too many noise. Thus, the proposed work uses the absolute differences (AD) function which treats each pixel individually and the properties of the pixel can fully have utilized. The AD computes the horizontal direction along the grayscale values on each pixel of stereo image. The benefit of the AD is that this function is fast and capable to preserve the properties of each pixel. The AD function is represented by (1) follows:

$$\mathrm{AD}(x,y,d) = I_l(x,y) - I_r(x,y,d) \quad (1)$$

where *I* represents the pixel intensity at the coordinates (x,y) and *d* represents the disparity value of left to right image differences.

2.2. Cost Aggregation

The cost aggregation is the second step of the local-based algorithm structure. This step decreases noise from the first stage to filter the preliminary result with more consistent matching differences before the next stage of optimization and disparity selection. This article recommends the usage of edge-preserving filter where this filter is capable to remove the noise and preserved the object's edges detection. Hence, this work uses the guided filter (GF) which was developed by He et al. [13]. The filter kernel is shown by the (2):

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

where {*p,q,w,c,ε,σ,μ,I*} denoted by { coordinates of (*x,y*), neighboring coordinates, window support size, center pixel of w, constant parameter, variance value, mean value, reference image (left input image)}. The GF is utilized in this work in line for fast execution of time processing which the calculation is only depend on the image pixels. The GF is able to increase the accuracy at the object edges. The final equation of this stage is given by (3):

$$CA(x,y,d) = AD(x,y,d)GF_{(p,q)}(I) \quad (3)$$

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

2.3. Disparity Optimization

The third step of the structure is the disparity optimization and selection which reduces the data assortment on the disparity positions and represented it with disparity value. Commonly, the local based stereo matching algorithm is consuming the Winner-Takes-All (WTA) strategy [9]. The WTA usages the minimum value of $CA(x,y,d)$ and embodied the same location with the disparity value. The function of the WTA stage is given by (4):

$$d(x,y) = \arg\min_{d\in D} CA(x,y,d) \quad (4)$$

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

2.4. Post-processing

The post-processing step is the last stage of the algorithm framework. This stage is also known as filtering stage or disparity refinement stage which decreases the invalid disparity values and the outlier noise on the final results [14]. Fundamentally, the used of second filtering process at this step is to surge the effectiveness and accurateness of the final disparity map. The suggested work in this article is using the bilateral filter (BF) which works for the second filtering process. The final disparity result is given by the (5):

$$BF_{(p,q)} = \sum_{q\in w_B}\exp\left(-\frac{|p-q|^2}{\sigma_s^2}\right) exp\left(-\frac{|I_p-I_q|^2}{\sigma_c^2}\right) \quad (5)$$

where the (*p,q*) represent the pixel of interest and neighbor pixel coordinates, w_b denotes the BF window size, $|I_p - I_q|^2$ and $|p-q|$ are the intensity differences and spatial distance respectively, σ_s^2 and σ_c^2 are the spatial distance and color similarity parameters.

3. Results and Analysis

This section provides the experimental setup and results of the proposed work in this article. All of the experiment were executed using a personal computer with the features of CPU i7-5500 and 8G RAM. The standard benchmarking dataset have been used in this article from the Middlebury Stereo evaluation benchmark [15]. This article also uses the KITTI benchmarking dataset. The KITTI dataset was developed by Menze and Geiger [16] which consists of 200 training images. This dataset was recorded from real environment of autonomous vehicle navigation using a stereo vision system. Hence, it comprises very challenging and complex images for the evaluation of stereo matching algorithm. The performance is measured based on the *all* and *nonocc* error attributes of bad pixel percentages. The *all* error is the error of invalid disparity values on all pixels of the disparity map image. The *nonocc* error is the error of invalid disparity values on the non-occluded regions [17-18]. The difference between these two errors is *all* error calculates all pixel, but the nonocc only counted the pixels in the certain regions (i.e., nonoccluded regions). The quantitative results in this article are evaluated based on the development kit provided by the KITTI. Some other methods [19-22] used different methodology of evaluation to get the performance results.

Figure 2 shows the results of objects detection using the Middlebury images. The objects are well-detected and reconstructed on the maps such as a mug in the Adirondack, the brushes in ArtL, a tree in Jadeplant, a motorcycle in Motorcycle, the personal computer in Vintage, piano and a chair in Piano, a recycle box in Recycle and a chair in the Shelves image. The complex images such as the pipes structures in the Pipe image are also well-reconstructed. The chairs are well discovered on the Playroom and Playtable images. It shows that the proposed framework in this article is capable to detect and reconstruct the detected objects.

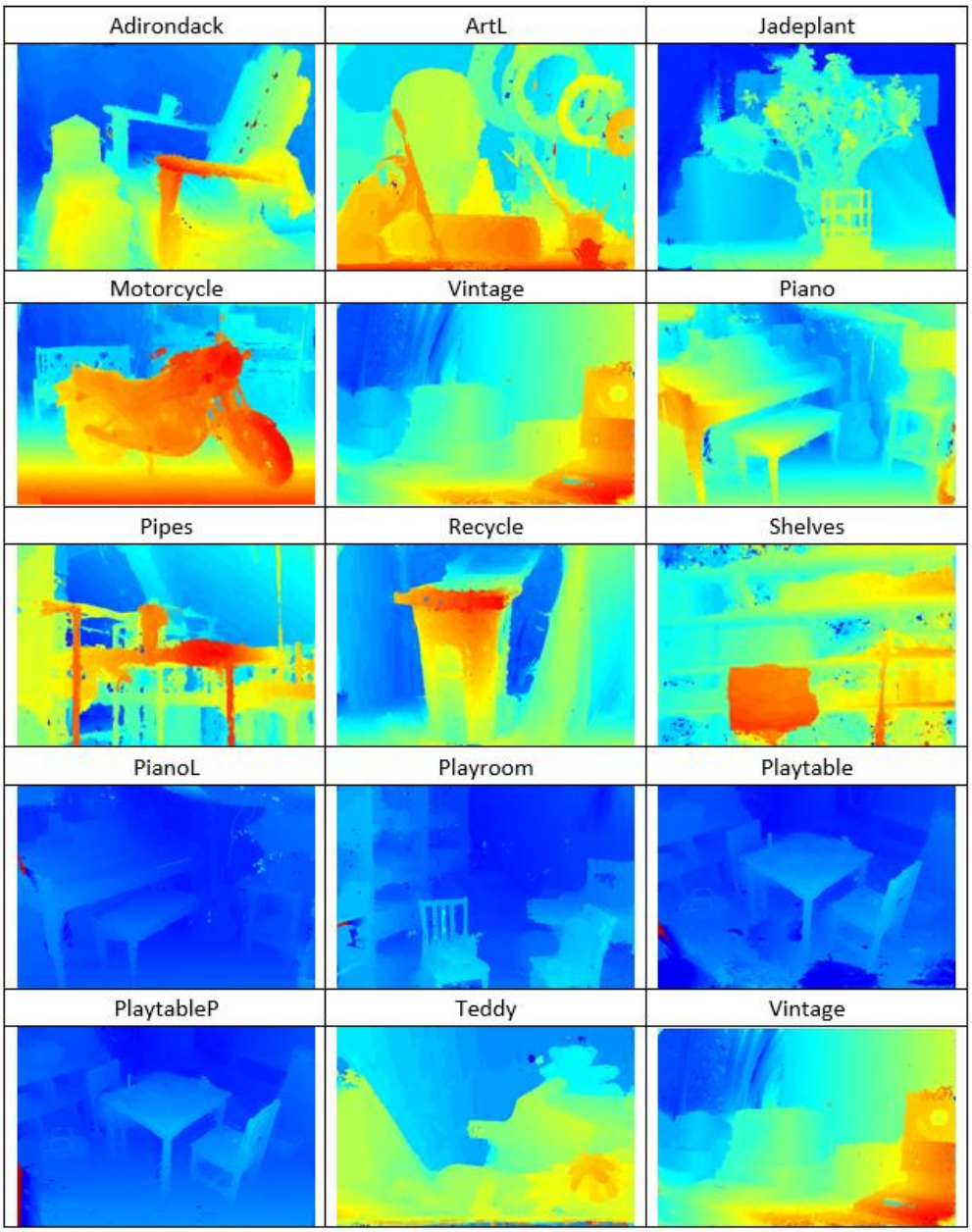

Figure 2. The results from the Middlebury Stereo Evaluation dataset

Table 1 displays the quantitative results of the proposed framework with some published methods. As for comparison before and after the proposed work, the TestG is the algorithm without the proposed framework which uses common algorithm with the first stage using interpolation matching technique. Second stage is using box filter with WTA strategy at optimization step. The last stage uses median filter with fixed windows size. The proposed

algorithm is ranked at top of the table with 12.11% and 14.01% of *nonocc* and *all* errors respectively. If compared with the TestG algorithm, the proposed work reduces the *nonocc* and *all* errors with 18.20% and 20.47% 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	12.11	14.01
SSW [17]	12.75	14.22
Hu [11]	13.55	16.65
NCC [12]	14.11	15.79
GC [8]	14.34	15.98
BP [3]	15.67	17.84
TM [18]	19.65	21.21
TestG	30.31	34.48

Figure 3 displays the sample results of objects detection using the KITTI dataset. The first image shows a car and the trees which the proposed work is capable to detect these objects on the disparity map. The contour of the color mapping on the detected objects are well-recognized on the disparity map result. The second until the fifth images show the cars are also well-reconstructed on the disparity map. The moving objects such as a lorry and a car in the last image are well-detected and reconstructed on the final result. In average, the different contour of disparity levels based on the different color scheme are precisely displayed on each result in this figure. It shows that the proposed algorithm is capable to work with real images as shown by the results using the KITTI dataset.

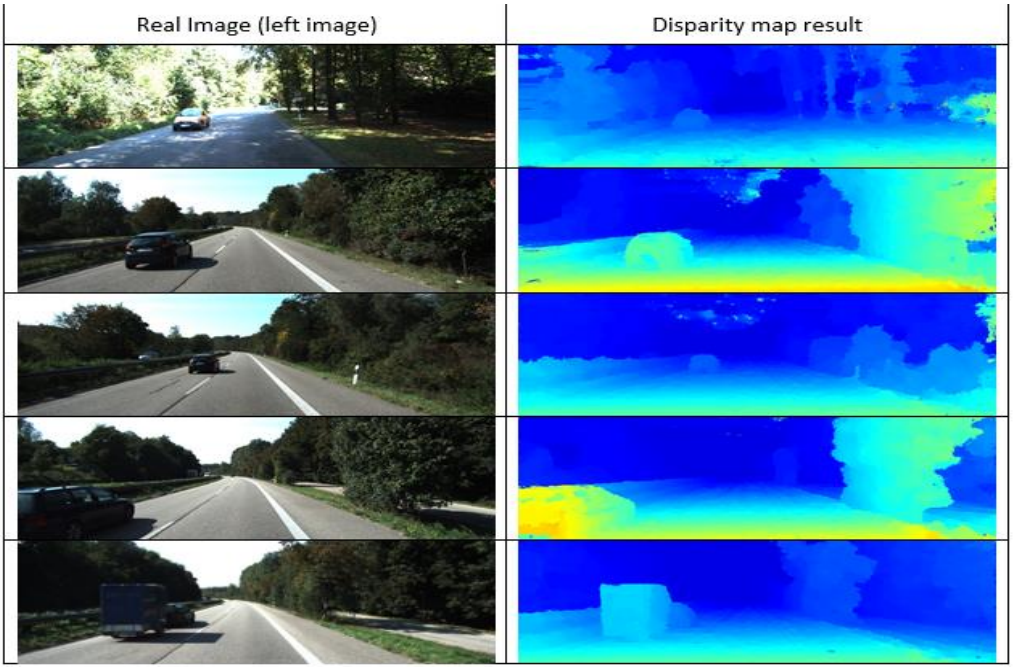

Figure 3. Sample results of objects detection from the KITTI dataset

4. Conclusion

A new framework of stereo matching algorithm was presented in this article. This algorithm was capable to detect any object based on the input from the stereo vision sensor. This algorithm is also capable to detect and robust against the moving objects as shown by the results in the KITTI dataset. Additionally, the proposed framework is able to increase the accuracy compared with TestG as tabulated in Table 1. The framework is also competitive with some established methods as tabulated in the same table.

Acknowledgement

This work is supported by a grant from the Universiti Teknikal Malaysia Melaka with the reference number PJP/2018/FTK(13C)/S01632.

References

[1] Ibrahim H, Hassan AHA. *Stereo matching algorithm for 3D surface reconstruction based on triangulation principle*. International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). 2016: 119–124.
[2] Aziz KAA, Shokri ASM. *A pixel to pixel correspondence and region of interest in stereo vision application*. IEEE Symposium on Computers & Informatics (ISCI). 2012: 193-197.
[3] Wu SS, Tsai H, and Chen LG. *Efficient hardware architecture for large disparity range stereo matching based on belief propagation*. IEEE International Workshop on Signal Processing Systems (SiPS). 2016: 236–241.
[4] Ibrahim H, Hassan AHA. 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.
[5] Gudis E, van der Wal G, Kuthirummal S, Chai S, Kumar R. *Stereo vision embedded system for augmented reality*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2012: 15-20.
[6] Balter ML, Chen AI, Maguire TJ, Yarmush ML. Adaptive kinematic control of a robotic venipuncture device based on stereo vision, ultrasound, and force guidance. *IEEE Transactions on Industrial Electronics.* 2017; 64(2): 1626-1635.
[7] Scharstein D, Szeliski R, Zabih R. *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*. IEEE Workshop on Stereo and Multi-Baseline Vision. 2001: 131–140.
[8] Liang Q, Yang Y, Liu B. *Stereo matching algorithm based on ground control points using graph cut*. International Congress on Image and Signal Processing (CISP). 2014: 503–508.
[9] Hamzah RA, Kadmin AF, Hamid MS, Ghani SF, Ibrahim H. Improvement of stereo matching algorithm for 3D surface reconstruction. *Signal Processing: Image Communication*. 2018; 1(65): 165-172.
[10] Zbontar J, LeCun Y. *Computing the stereo matching cost with a convolutional neural network*. IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1592–1599.
[11] Hu W, Zhang K, Sun L, Li J, Yang S. Virtual support window for adaptive-weight stereo matching. *IEEE Visual Communications and Image Processing (VCIP).* 2011: 1–4.
[12] Einecke N, Eggert J. Anisotropic median filtering for stereo disparity map refinement. *VISAPP*. 2013: 189–198.
[13] He K, Sun J, Tang X. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2013; 35(6): 1397–1409.
[14] Kadmin AF, Ghani SFA, Hamid MS, Salam S. Disparity refinement process based on RANSAC plane fitting for machine vision applications. *Journal of Fundamental and Applied Sciences.* 2017; 9(4S): 226-237.
[15] Scharstein D, Szeliski R. Middlebury stereo evaluation - version 3 (accessed date: May 2018, http://vision.middlebury.edu/stereo/eval/references."
[16] Menze M, Geiger A. *Object scene flow for autonomous vehicles*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015: 3061–3070.
[17] Hamid MS, Rosly HN, Hashim NMZ. 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.* 2011; 1(1): 5-19.
[18] Hamzah RA, Rahim RA. *Depth evaluation in selected region of disparity mapping for navigation of stereo vision mobile robot*. 2010 IEEE Symposium on Industrial Electronics & Applications (ISIEA). 2010: 551-555.
[19] Winarno E, Harjoko A, Arymurthy AM, Winarko E. Face recognition based on symmetrical half-join method using stereo vision camera. *International Journal of Electrical and Computer Engineering*. 2016; 6(6): 2818.
[20] Budiharto W, Santoso A, Purwanto D, Jazidie A. Multiple moving obstacles avoidance of service robot using stereo vision. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2011; 9(3): 433-44.
[21] Xi HX, Cui W. Wide Baseline Matching Using Support Vector Regression. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2013; 11(3): 597-602.
[22] Paramkusam AV, Arun V. A Survey on block matching algorithms for video coding. *International Journal of Electrical and Computer Engineering*. 2017; 7(1): 216.