Stereo matching algorithm using census transform and segment tree for depth estimation

Muhammad Nazmi Zainal Azali[1], Rostam Affendi Hamzah[2], Zarina Mohd Noh[1], Izwan Zainal Abidin[3], Tg Mohd Faisal Tengku Wook[2]

[1]Department of Computer Engineering, Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka, Durian Tunggal, Melaka, Malaysia
[2]Department of Electronics Engineering Technology, Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik, Universiti Teknikal Malaysia Melaka (UTeM), Durian Tunggal, Melaka, Malaysia
[3]Chief Executive Officer, Terra Drone Technology Malaysia Sdn Bhd, Bukit Jalil, Kuala Lumpur, Malaysia

Article Info

ABSTRACT

This article proposes an algorithm for stereo matching corresponding process that will be used in many applications such as augmented reality, autonomous vehicle navigation and surface reconstruction. Basically, the proposed framework in this article is developed through a series of functions. The final result from this framework is disparity map which this map has the information of depth estimation. Fundamentally, the framework input is the stereo image which represents left and right images respectively. The proposed algorithm in this article has four steps in total, which starts with the matching cost computation using census transform, cost aggregation utilizes segment-tree, optimization using winner-takes-all (WTA) strategy, and post-processing stage uses weighted median filter. Based on the experimental results from the standard benchmarking evaluation system from the Middlebury, the disparity map results produce an average low noise error at 9.68% for nonocc error and 18.9% for all error attributes. On average, it performs far better and very competitive with other available methods from the benchmark system.

Corresponding Author:

Rostam Affendi Hamzah
Department of Electronics Engineering Technology
Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik, Universiti Teknikal Malaysia Melaka (UTeM)
Durian Tunggal, Melaka, Malaysia
Email: rostamaffendi@utem.edu.my

1. INTRODUCTION

The stereo matching algorithm is a solution to the problem that occurs between the generated stereo image pairs. It is a widely produced technology in most applications especially in depth estimation and cyber-physical systems (CPS). This article explains the stereo matching algorithm used in the matching process to generate a disparity map for the usage in those applications. The disparity map result comprises depth information which will mostly be used as a leading application for depth estimation and CPS such as for object detection [1], 3D surface reconstruction process [2], [3], motion system [4], face detection [5], and industrial automation [6]. The stereo matching algorithm is important in determining each output produced, and it has a direct impact on the accuracy of 3D surface reconstruction. Stereo matching also determines how each stereo image's estimated location, measurement, and pixels are calculated. Furthermore, the production or applications utilizes stereo images is not easy to be manipulated, it is crucial for machine vision system that provides output similar to the human vision. When an item is closer to the viewer, the gap between the eyes becomes lax,

according to a study that investigated the usage of zooming images [7]. Furthermore, the author precise on the use of zooming images stated about the resemblance of human vision with machine vision when an item is closer to the viewer and the disparity between the eyes increases. In other words, in order to acquire accurate results, the matching procedure necessitates the use of a robust framework and algorithm to generate an accurate disparity map.

In current consideration of delivering accurate results, the algorithm framework must be precise to estimate the depth from disparity map. Therefore, this article proposes a new framework that must be able to reduce the noise and accurate for depth estimation. Furthermore, the map obtained is very important in the creation of 3D surface reconstruction [8]. Most studies conducted on developing the matching algorithms are depends on four main processes that was proposed by the Rhemann *et al.* [9] where step 1 formulates matching cost computation, step 2 formulates cost aggregation, step 3 expresses the disparity optimization (choose the normalization value with the level of disparity), and lastly step 4 formulates the disparity improvement (post-processing is used to fine-tune the final disparity map). By taking the basic methods of the stereo matching algorithm, that does not mean all the results produced are accurate. Some methods used have their pros and cons. There is also a method that displays quick operating time but minimal precision on the edges owing to incorrect window size assortment. The researchers face a difficulty in obtaining a correct results for the local method [10]. Based on the standard dataset from the Middlebury stereo evaluation online system [11], various algorithms have been proposed and evaluated in order to reduce the mismatching rate. The constructed dataset comprises of stereo images shown with varied ambient illuminations and many exposures, with and without a mirror sphere of the lighting conditions. Among the proposed algorithms that used the Middlebury datasets as their benchmarks are optimization for plane-based stereo [12], memory-efficient and robust [13], adaptive cross guided filter with weights [14], and edge-based disparity map estimation [15].

This article presents a new stereo matching algorithm based on census transform. It is based on the objective of analyzing the algorithm performances on the low texture regions, repetitive pattern, and discontinuity regions. The first stage of the proposed stereo matching algorithm is based on matching cost computation using census transform. Then the next stage reduces the noise of the images after cost computation is using the segment-tree method [16] where it is one of the methods that has a significant impact on a stereo system's speed and accuracy. The proposed algorithm is structured similarly to a local method. As a result, the winner-takes-all (WTA) strategy is being used in the optimization. The last stage in the proposed algorithm is to use weighted median filtering. The non-linear computational complexity of weighted median filtering to find the median value. In line with that, it also gives a great impact in implementing the output where it achieves in filtered images with sharper edges. This article has also been properly arranged where the next section will explain the methodology in the development of the proposed algorithm. After that, the third section will explain the experimental analysis of the disparity map using qualitative and quantitative measurements that have been evaluated on Middlebury benchmark. The last section represents the conclusion and acknowledgment that completes the description of the whole article.

2. RESEARCH METHOD

The framework presented in Figure 1 is a framework that displays all the methodology used in the proposed algorithm. The first step in the development of stereo matching algorithm is cost computation in order to obtain a preliminary disparity map. The input stereo images of left and right images will be matched or corresponded. This process will be implemented using census transform where this technique transforms the input images to the binary type images. After that, these binary images will be compared at the same pixel locations between left and right images respectively. Then it is followed by the second step where it works to minimize noise while preserving the object's edges. This article uses segment-tree method where it is capable of efficiently removing noise from the low texture regions as well as sharpening object borders. This technique is one of the most effective segmentation types [16] to increase the accuracy. The optimization stage uses the WTA strategy, which normalises the floating point numbers and replace it with the lowest disparity values on the disparity map. The WTA is fast and the normalization function for this strategy is almost accurate based on the study.

Budiharto *et al.* [2]. The last stage of effort additionally employs one of the edge-preserving filters known as weighted median filtering. This is a nonlinear filter type that can be used to improve and smooth the final disparity map. The pixel intensities on the image at this stage are the final value of disparity map that will be used in the depth estimation. The applications normally applied the triangulation principal from the disparity value to get the depth estimation.
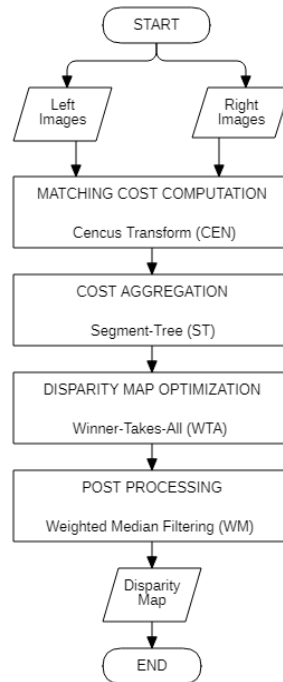
Figure 1. The proposed algorithm

2.1. Matching cost computation

The first stage done in completing the proposed algorithm is to use the census transform which is one of the important parts to ensure the preliminary data from stereo images to generate the disparity map. It is an important part from all other stages where it will calculate the corresponding points of the stereo images thus capable of avoiding unnecessary noise. Most of the problems occurred in the stage of matching process is in the low texture and repetitive regions that must be kept to a barely minimum error. Next census transform is a method that has been proposed based on the binary relative intensity from input images, it is resilient to variations in intensity. It is a method of non-parametric transform [17]. Where the $C(P)$ denotes the census transform of a pixel P. $C(P)$ converts the local neighborhood of a pixel P to a bit string that represents the set of neighboring pixels whose intensity is less than P's. The census transform function is defined by (1).

$$C(P) = \otimes_{[i,j] \in D} \ \xi(P, P + [i,j]) \tag{1}$$

Where the symbol $\otimes$ symbolises the concatenation, D the non-parametric window around P, and ξ the transform defined by (2).

$$\xi\{P, P + [i,j]) = \begin{cases} 1, if \ P > P + [i,j] \\ 0, \quad otherwise \end{cases} \tag{2}$$

2.2. Cost aggregation

Cost aggregation is basically the second stage of the local-based algorithm in the proposed framework. Theoretically, this stage reduces noise after what has been done in the first stage. Based on the literature works at early stage, the selection of appropriate and robust technique of filtering at this step is capable to remove high noise. Almost 70% of noise can be detached and for sure the accuracy can be increased. As a result, the segment-tree [18] is selected to be used at this stage. The segment-tree is a type of image segmentations and this method is efficient to increase the accuracy. In fact, it is utilized because brighter intensity levels equate to greater contributions. This segment-tree is technically based on three process steps that are carried out with the reference colour or intensity image, where the pixels are arranged into a series of segments. Second, a tree graph is generated for each segment, and finally, these independent segment graphs are connected to form the segment-tree structure. With the segment-tree function, the following is defined where the reference image is modeled as a graph $G = (V, E)$, with a subset of edges E and E' chosen for the segment-tree $T = (V, E')$ as given by (3), (4) and (5), step 1 (initialization).

$$we = w(s,r) = |I(s) - I(r)| \tag{3}$$

Where the color or intensity reference image I is characterised as a linked, undirected graph $G = (V, E)$, where each node in V resembles to a pixel in I and each edge in E connects two adjacent pixels. The weight of an edge e linking pixels s and r is defined. The edges in E are sorted in a non-decreasing order based on the weights specified and a subtree is formed for each node in V. E' has no edges, step 2 (grouping).

$$wej \leq \min(Int(Tp) + \frac{k}{|Tp|}, Int(Tq) + \frac{k}{|Tq|}) \tag{4}$$

Where a full scan of the edge set E, the subtrees are merged into larger groupings. Let v_p and v_q represent the nodes connected by edge $ej\ E$. If v_p and v_q are from distinct subtrees and the edge weight w_{ej} meets a criterion. T_p and T_q are combined into a new subtree $T_{p,q}$. Simultaneously, e_j is incorporated in E'. The criterion that considers the relative dissimilarity of the two subtrees is expressed as (4). The highest edge weight in T_p is denoted by $Int\ (T_p\ )$, and k is a constant parameter. Each subtree corresponds to a visually coherent segment after visiting each edge in E. The edges of these subtrees (which are already gathered in E') are then eliminated from E, step 3 (linking).

$$Tp, q = (V_{p,q}, E_{p,q}) \tag{5}$$

Where more edges are chosen from E to connect the subtrees. Its role is to look for these edges in a subsequent scan of E. If an edge connects two different subtrees, the subtrees should be merged and the edge should be included in E'. The search ends when all of the trees have been combined into a single component.

2.3. Disparity map optimization

The next stage which is the third stage to complete the structure disparity map is disparity optimization. Technically, the disparity map obtained from the previous stage will be processed in a WTA strategy [19]. The minimum disparity value $CA\ (x, y, d)$ of each pixel of the map of disparity was applied and the same position was incorporated with the disparity value. The (6) is the WTA strategy.

$$d(x, y) = \arg\min_{d \in D} CA(x, y, d) \tag{6}$$

Where D represents the disparity range on the image, $d\ (x, y)$ indicates the chosen disparity value at the co-ordinates of (x, y) and $CA\ (p, d)$, the second-stage data which is the cost aggregation step. Essentially, the disparity map yet includes noise or erroneous pixels after this stage. This map requires to be enhanced and the leftover noise removed in the last step.

2.4. Post processing

The last stage after completing the algorithm structure is identified as the refinement disparity map or post-processing stage. It is a process where multiple ongoing procedures begin with the management of the occlusion regions, complete the pixels that are faulty, and filter the final disparity map. The method started with left-right inspection, progressed to fill in the invalid pixel and ended with the filtering using the weighted median filter (WM) with bilateral filter. This method began with an image of the left reference disparity map, which corresponded to the image of the right reference disparity map. Inconsistent values between the two are referred to be invalid disparity. The WM filtering equation is the final disparity map as indicates by (7).

$$|d_{LR}(p) - d_{RL}(p - d_{LR}(p))| \leq \tau_{LR} \tag{7}$$

The next step is to fill in invalid pixels when the left image is predetermined as an image reference. Valid pixel substitution here is done by the filling procedure which starts from left and then from right again. The invalid disparity is changed by the closest valid disparity value. This value should also be inserted on the same scan line. An example of this can be seen in (8).

$$d(p) = \begin{cases} d(p - i) \leq d(p + j), \\ d(p + j), otherwise. \end{cases} \tag{8}$$

$(p + j)$ denotes location of the first valid disparity on the right side, while $d(p)$ a disparity value at the location of p, while $(p + i)$ denotes the location of first valid disparity on the left side. This method produces undesirable streak artefacts. The weighted median filter with bilateral filter is used to remove the remaining noise from the disparity map. The (9) shows the for the bilateral filter $B\ (p, q)$.

$$B(p, q) = \exp\left(-\frac{|p-q|^2}{\sigma_s^2}\right) \exp\left(-\frac{|d(p-d(p)|^2}{\sigma_s^2}\right) \tag{9}$$

Where pixels of interest (p, q) are denoted, and $|p - q|$ refers to spatial euclidean and $|d(p) - d(q)|2$ to euclidean. The spatial distance and colour parameters are $\sigma_s{}^2$ and $\sigma_s{}^2$ are similiar. Higher weighted was applied to the filter since it is a form of edge preserving filter that improves the disparity map accuracy. The weighted of $B\ (p, q)$ is transformed into the sum of histogram $h\ (p, d_r)$, resulting in (10).

$$h(p, d_r) = \sum_{q \epsilon w_p | d(q) == d_r} B(p, q) \tag{10}$$

Where d_r is the disparity range and w_p is the window size with the radius $(r \times r)$ at the centred pixel of p. The median value of $h\ (p, d_r)$ given by (11) determines the final disparity value WM.

$$WM = med\{d | h(p, d_r)\} \tag{11}$$

3.   RESULTS AND ANALYSIS

The proposed algorithm's result disparity maps are tested using the standard benchmarking dataset known as the Middlebury Benchmark [20]. Furthermore, the analytical experiment was carried out on a personal computer with Windows 10 specifications, a CPU i7 8700 @ 3.2 GHz, an RTX 2070 super, and 16 GB RAM. The Middlebury dataset provides fifteen standard input images that must be uploaded online as a result file that is ready to be compiled in.zip format. The images presented are quite complicated in shape, and each image comprises variety of attributes and properties such as light surrounding of objects depth, disorganisation regions, varying outcomes, and low texture areas. The values of $\{M, \sigma_s, \sigma_c, W_B, W_{MF}\}$ are $\{13 \times 9, 17, 0.3, 9 \times 9, 13 \times 13\}$. In this study, performance is tested and assessed using all and $nonocc$ error properties of $avgerr$ pixel percentages. An invalid disparity value on a non-occluded region is called a $nonocc$ error. While all error is referred to the error caused by erroneous disparity values across all pixels in the image's disparity map.

In the selection of determining the accuracy of the image, the following shows one example that can be shown one example of the image of Adirondack in Figure 2(a) and Figure 2(b), this can be compared between the proposed algorithm with the algorithm proposed by Fakhar $et\ al.$ [21]. Clearly shows with the proposed algorithm is to generate a more accurate result with less noise. Figure 3 depicts all of the displays that have been developed and evaluated using the Middlebury dataset. There are fifteen images produced as a result of the evaluation, and all correctness is defined by all error and non-occluded error ($nonocc$). All errors are measured on the basis of all image pixels and non-occluded pixels without the occluded region of disparity map within 15 images provided, in all the images there are some images that are quite difficult to match such as Jadeplant and Playtable. From the image, it contains some difficulties from the perspective of tables and leaves of varying dimensions. However, despite these difficulties, the proposed algorithm is capable of reconstructing a substantially accurate disparity map with identifiable discontinuities.

Technically, the image obtained from Middlebury is a difficult image to process accurately depending on the corresponding point. The original image that is developed may contain differences in pixel values and it is even developed to test the robustness of an algorithm where the same relevant point is used. In addition, each image can be viewed from various perspectives in terms of objects of a particular color, shadows, discontinuity regions, and obscured areas. Figure 3 shows the results from the KITTI 2015 dataset which these three inputs left reference images are labelled as #000002_10, #000003_10 and #000004_10 in the database. These images are captured from the real environment. The disparity map results are displayed in grayscale color. The results show clear depth estimation with almost accurate objects detected. High grayscale intensity means the object is closed to the stereo camera and for the lowest grayscale value indicate the objects are faraway.

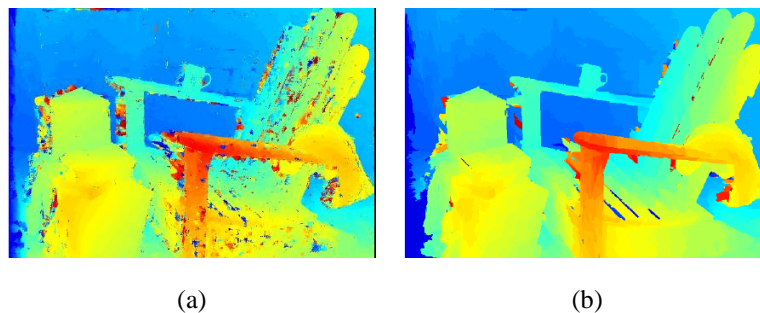(a)   (b)

Figure 2. The example shown in the disparity map shows the accuracy used in the Adirondack: (a) another proposed algorithm [19] and (b) proposed algorithm
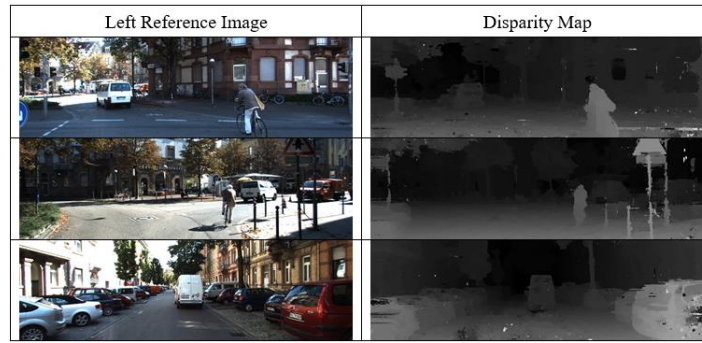
Figure 3. The example shown in the disparity map shows the accuracy used in the Adirondack

Figure 4 shows the disparity map images and the results have been improved containing low textures surfaces such as recycle, Adirondack, and piano. The contours with varying depth and dispersion are also clear and visible. Apart from that, the worst of them all, namely Jadeplant, Playtable, vintage, and PianoL show that plain color objects and shadows are difficult to match. Regions that match similar pixel values make it possible to get wrong matching is very significant. From the results uploaded to Middlebury benchmark, it is collected as a whole the results of the proposed method along with other published methods based on quantitative measurements in Table 1 and Table 2. From these results as well, the results produced by Middlebury benchmark are described previously and in the tables are compared each result produced by published methods. It is the competitiveness of the proposed work that shows the level of effectiveness of the proposed algorithm. The proposed algorithm is at the second top rank of the table which is 9.68% for *nonocc* errors and 18.9% for *all* errors. On average, which Table 1 that is *nonocc* error, it can be compared with other proposed algorithms where disparity stereo geometry cross aware (DSGCA) is in third ranked followed by pixel pair based guided filter (PPEP-GF), sample photoconsistency (SPS), image edge brightness image (IEBIMst), semi-global matching 1 (SGBM1), and multi-windows noise effected (MANE) are the lowest which is 11.9% error. While in Table 2 shows all the errors that are compared and shows semi-global matching 1 SBGM1 is in third ranked followed by MANE, double guided (DoGGuided), dynamic filter (DF), random-normalised cross control (R-NCC), and binary streo matching (BSM) are at the lowest which is 23.5% error. It is clear that the table shown can be competitive with other published work and this comparison is shown in detail. The method Kong *et al.* [14] is the lowest error based on the results in Table 1 and Table 2 respectively. However, the proposed method in this article still produced good results and image for the PianoL compared to area cross region guided filter (ACR-GIF-OW) [22].

Table 1. Performance comparison of *nonocc* error from the Middlebury images

Algorithms	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlayP	Recyc	Shelvs	Teddy	Vintge
ACR-GIF-OW [23]	5.78	3.01	3.91	11.2	2.81	2.91	4.95	27.1	4.59	5.49	12.3	2.58	2.50	12.6	1.86	6.58
Proposed algorithm	9.68	3.94	7.46	18.5	4.66	4.43	5.97	21.8	7.24	6.92	34.4	13.2	4.20	11.8	3.91	20.1
DSGCA [23]	9.75	3.25	5.95	18.9	3.60	3.41	7.17	21.1	7.23	9.36	29.4	7.94	3.80	14.7	3.51	39.7
SPS [24]	10.4	3.57	5.34	22.8	3.11	3.15	9.34	22.9	6.78	12.5	9.70	7.64	6.27	22.3	1.52	52.6
IEBIMst [25]	11.1	26.1	4.67	41.9	2.72	4.99	5.69	17.5	5.47	12.9	14.8	3.26	4.99	16.4	2.64	10.4
SGBM1 [26]	11.3	18.3	7.45	15.7	3.48	29.1	6.51	38.4	5.37	12.8	13.5	3.24	3.44	15.1	3.00	11.1
MANE [27]	11.9	6.58	5.81	20.7	4.52	4.31	10.6	20.9	8.62	15.0	34.7	10.5	5.50	20.2	3.12	46.5

Table 2. Performance comparison of *all* error from the Middlebury images

| Algorithms	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlayP	Recyc	Shelvs	Teddy	Vintge
ACR-GIF-OW [22]	9.48	4.53	8.41	22.1	7.93	7.88	6.36	27.7	11.0	8.51	16.1	6.60	4.26	13.1	2.86	7.77
Proposed algorithm	18.9	7.95	25.0	41.3	12.2	12.0	11.2	26.2	20.0	24.3	38.8	19.3	7.82	14.8	13.6	26.7
SGBM1 [26]	18.9	21.1	17.8	38.7	11.0	36.4	11.6	40.0	13.6	25.4	20.0	8.74	5.97	17.6	10.7	18.3
MANE [27]	21.3	11.6	22.9	45.9	12.4	12.3	15.1	24.7	22.3	31.1	39.9	17.3	9.67	22.5	12.5	51.0
SPS [24]	22.3	20.1	28.0	56.5	13.8	16.8	13.4	37.3	23.8	30.3	30.8	13.0	9.13	19.0	13.4	23.6
IEBIMst [25]	22.7	14.1	18.2	103	13.2	12.7	11.1	26.4	22.5	20.9	13.9	16.3	16.8	11.5	6.16	26.8
DSGCA [28]	23.5	12.7	28.7	58.7	14.8	14.7	16.0	35.8	24.5	29.4	31.0	20.2	12.1	19.2	14.3	39.3
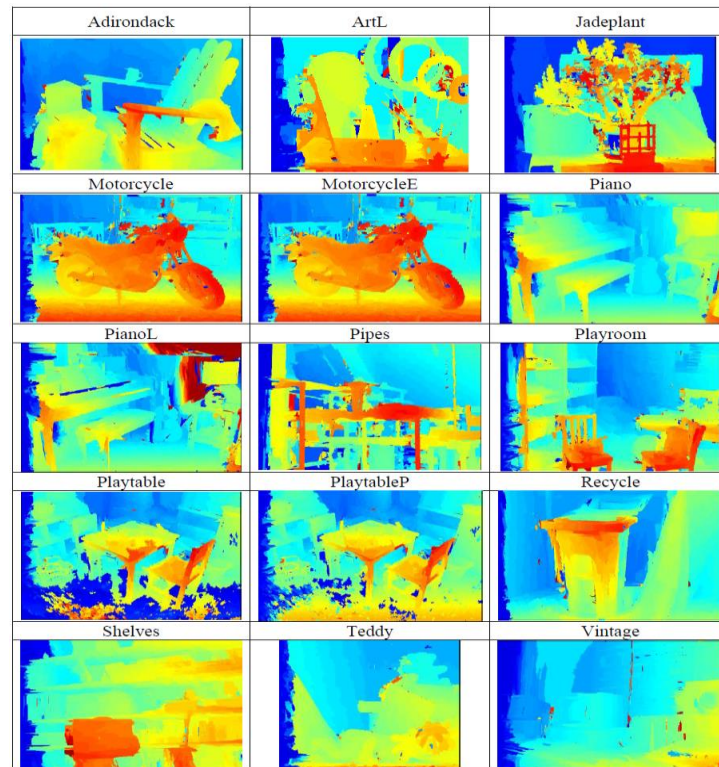
Figure 4. Image results from proposed algorithm was evaluated using Middlebury benchmark dataset

4.   CONCLUSION

This article presents a framework of stereo matching algorithm. The filter used in the proposed framework allows it to achieve the desired accuracy to coincide with the results produced and able to remove noise. Frameworks that start with cost computation that use census transform are able to increase the effectiveness on the disparity map. While the cost aggregation that uses segment-tree is able to reduce noise after process before and it preserves the preliminary disparity map's object boundaries. The WTA strategy proposed in the optimization section further strengthens the framework by normalizing the floating point numbers in accordance with the disparity values. The last framework known as refinement disparity map is to use weighted median filtering to reduce residual noise and improved the final disparity map's efficiency. The entire framework is able to compete with other works. Based on the results released from Middlebury benchmarks, it is able to obtain second low average errors at 9.68% for nonocc errors and 18.9% for all errors. The overall results of the findings are improved. In future work, a more extensive investigation should be conducted by extending our method to include more ways necessary to further reduce the inaccuracies in the existing results. Additionally, a long-term possibility that should be investigated is improving skills for optimum implementation on graphics processing unit (GPU) architecture in order to improve the method and speed of cost computation.

REFERENCES

[1]   S. S. N. Bhuiyan and O. O. Khalifa, "Efficient 3D stereo vision stabilization for multi-camera viewpoints," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 3, pp. 882–889, 2019, doi: 10.11591/eei.v8i3.1518.
[2]   W. Budiharto, A. Santoso, D. Purwanto, and A. Jazidie, "Multiple moving obstacles avoidance of service robot using stereo vision," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 9, no. 3, pp. 433-444, 2011, doi: 10.12928/telkomnika.v9i3.733.
[3]   E. Winarno, A. Harjoko, A. M. Arymurthy, and E. Winarko, "Face recognition based on symmetrical half-join method using stereo vision camera," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, no. 6, pp. 2818-2827, 2016, doi: 10.11591/ijece.v6i6.pp2818-2827.
[4]   R. A. Hamzah, H. Ibrahim, and A. H. A. Hassan, "Stereo matching algorithm for 3D surface reconstruction based on triangulation principle," in *2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering*

(ICITISEE), 2016, pp. 119–124, doi: 10.1109/ICITISEE.2016.7803059.

[5] I. Vedamurthy *et al.*, "Recovering stereo vision by squashing virtual bugs in a virtual reality environment," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1697, 2016, doi: 10.1098/rstb.2015.0264.

[6] H. Xi and W. Cui, "Wide baseline matching using support vector regression," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 11, no. 3, pp. 597-602, 2013, doi: 10.12928/telkomnika.v11i3.1144.

[7] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*, 2001, pp. 131–140, doi: 10.1109/SMBV.2001.988771.

[8] Q. Yang, "A non-local cost aggregation method for stereo matching," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1402–1409, doi: 10.1109/CVPR.2012.6247827.

[9] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *CVPR 2011*, 2011, pp. 3017–3024, doi: 10.1109/CVPR.2011.5995372.

[10] R. A. Setyawan, R. Soenoko, M. A. Choiron, and P. Mudjirahardjo, "Matching algorithm performance analysis for autocalibration method of stereo vision," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 1105-112, 2020, doi: 10.12928/telkomnika.v18i2.14842.

[11] R. A. Hamzah, H. N. Rosly, and S. Hamid, "An obstacle detection and avoidance of a mobile robot with stereo vision camera," in *2011 International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, 2011, pp. 104–108, doi: 10.1109/ICEDSA.2011.5959032.

[12] S. Ahmed, M. Hansard, and A. Cavallaro, "Constrained optimization for plane-based stereo," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3870–3882, 2018, doi: 10.1109/TIP.2018.2823543.

[13] Y. Lee and C. -M. Kyung, "A memory- and accuracy-aware gaussian parameter-based stereo matching using confidence measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 1845–1858, 2021, doi: 10.1109/TPAMI.2019.2959613.

[14] L. Kong, J. Zhu, and S. Ying, "Local stereo matching using adaptive cross-region-based guided image filtering with orthogonal weights," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–20, 2021, doi: 10.1155/2021/5556990.

[15] J. Žbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1592–1599, doi: 10.1109/CVPR.2015.7298767.

[16] H. Hirschmüller, P. R. Innocent, and J. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors," *International Journal of Computer Vision*, vol. 47, pp. 229–246, 2002, doi: 10.1023/A:1014554110407.

[17] N. Ma, Y. Men, C. Men, and X. Li, "Accurate dense stereo matching based on image segmentation using an adaptive multi-cost approach," *Symmetry*, vol. 8, no. 12, 2016, doi: 10.3390/sym8120159.

[18] W. Yuan, C. Meng, X. Tong, and Z. Li, "Efficient local stereo matching algorithm based on fast gradient domain guided image filtering," *Signal Processing: Image Communication*, vol. 95, 2021, doi: 10.1016/j.image.2021.116280.

[19] R. A. Hamzah, M. G. Y. Wei, and N. S. N. Anwar, "Stereo matching based on absolute differences for multiple objects detection," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 17, no. 1, pp. 261-267, 2019, doi: 10.12928/telkomnika.v17i1.9185.

[20] 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)*, 2016, pp. 236–241, doi: 10.1109/SiPS.2016.49.

[21] S. Fakhar A. G, M. Saad H, A. Fauzan K., R. Affendi H., and M. Aidil A., "Development of portable automatic number plate recognition (ANPR) system on Raspberry Pi," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 3, pp. 1805-1813, 2019, doi: 10.11591/ijece.v9i3.pp1805-1813.

[22] *Middlebury Stereo Evaluation - Version 3*, Middlebury stereo evaluation, 2021. [Online]. Available: https://vision.middlebury.edu/stereo/eval3/

[23] N. Einecke and J. Eggert, "Anisotropic median filtering for stereo disparity map refinement," in *Proceedings of the International Conference on Computer Vision Theory and Applications*, 2013, vol. 2, pp. 189–198, doi: 10.5220/0004200401890198.

[24] R. A. Hamzah, R. A. Rahim, and H. N. Rosly, "Depth evaluation in selected region of disparity mapping for navigation of stereo vision mobile robot," in *2010 IEEE Symposium on Industrial Electronics and Applications (ISIEA)*, 2010, pp. 551–555, doi: 10.1109/ISIEA.2010.5679404.

[25] K. Zhang, J. Li, Y. Li, W. Hu, L. Sun, and S. Yang, "Binary stereo matching," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 356–359. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6460145

[26] M. Kitagawa, I. Shimizu, and R. Sara, "High accuracy local stereo matching using DoG scale map," in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, 2017, pp. 258–261, doi: 10.23919/MVA.2017.7986850.

[27] W. Mao and M. Gong, "Disparity filtering with 3D convolutional neural networks," in *2018 15th Conference on Computer and Robot Vision (CRV)*, 2018, pp. 246–253, doi: 10.1109/CRV.2018.00042.

[28] Y. Li and S. Fang, "Removal-based multi-view stereo using a window-based matching method," *Optik*, vol. 178, pp. 1318–1336, 2019, doi: 10.1016/j.ijleo.2018.10.126.

BIOGRAPHIES OF AUTHORS

Muhammad Nazmi Zainal Azali 🆔 🔗 SC ⓒ currently pursuing the M.Sc. degree in Electronic Engineering from Universiti Teknikal Malaysia Melaka. His current research interests focusing on stereo vision and digital image processing. He is also interested in electronic soldering and circuit. He can be contacted at email: nazmi_z@icloud.com.

Rostam Affendi Hamzah 🆔 Ⓖ SC Ⓒ graduated from Universiti Teknologi Malaysia where he received his B.Eng majoring in Electronic Engineering. Then he received his M. Sc. majoring in Electronic System Design Engineering and PhD majoring in Electronic Imaging from the Universiti Sains Malaysia. Currently he is a lecturer in the Universiti Teknikal Malaysia Melaka teaching digital electronics, digital image processing and embedded system. He can be contacted at email: rostamaffendi@utem.edu.my.

Zarina Mohd Noh 🆔 Ⓖ SC Ⓒ received the Ph.D. degree from Universiti Putra Malaysia. She is currently a senior lecturer also M.Sc. Co-Supervisor at Universiti Teknikal Malaysia Melaka.Her research interest includes image processing and computer embedded system engineering. She can be contacted at email: zarina.noh@utem.edu.my.

Izwan Zainal Abidin 🆔 Ⓖ SC Ⓒ is currently the Managing Director and Chief Executive Officer of Terra Drone Technology Malaysia Sdn Bhd, a Joint Venture company between himself and Terra Drone Corporation of Japan, the number one Remote Sensing Drone Service Provider in the world in 2019 and 2020. He can be contacted at email: izwan@terradrone.com.my and izwan@terra-drone.co.jp.

Tg Mohd Faisal Tengku Wook 🆔 Ⓖ SC Ⓒ graduated from Universiti Sains Malaysia (USM), in 2000 and finished his Master of Business Administration (Advanced Operations Management) from Universiti Teknikal Malaysia Melaka (UTeM), in 2011. He is currently a senior Teaching Engineer attached to Electronic and Computer Engineering Technology Department, Faculty of Electrical and Electronic Engineering Technology, UTeM. He started his career as Production Engineer in Soshin Electronics (M) Sdn. Bhd. During in Soshin, Tengku plays an active role in Yield improvement projects and transfer new product from Japan. After 7 years he leave Soshin for Konica Minolta. In Konica Minolta he is person in charge for Western Digital project. After 5 years he decide to join UTeM as an academician. He can be contacted at email: tgfaisal@utem.edu.my.