Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 3, June 2024, pp. 1685~1693
ISSN: 2302-9285, DOI: 10.11591/eei.v13i3.5708

Stereo matching algorithm using deep learning and edge-preserving filter for machine vision

Shamsul Fakhar Abd Gani[1,2], Muhammad Fahmi Miskon[1], Rostam Affendi Hamzah[2], Mohd Saad Hamid[2], Ahmad Fauzan Kadmin[2], Adi Irwan Herman[3]
[1]Fakulti Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka, Malaysia
[2]Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik, Universiti Teknikal Malaysia Melaka, Malaysia
[3]Product and Test Engineering, Texas Instruments (Malaysia), Melaka, Malaysia

Article Info

Article history:

Received Jan 6, 2023
Revised Sep 29, 2023
Accepted Dec 19, 2023

Keywords:

Bilateral filter
Deep learning
Disparity map
Edge preserving filter
Stereo matching
Stereo vision

ABSTRACT

Machine vision research began with a single-camera system, but these systems had various limitations from having just one point-of-view of the environment and no depth information, therefore stereo cameras were invented. This paper proposes a hybrid method of a stereo matching algorithm with the goal of generating an accurate disparity map critical for applications such as 3D surface reconstruction and robot navigation to name a few. Convolutional neural network (CNN) is utilised to generate the matching cost, which is then input into cost aggregation to increase accuracy with the help of a bilateral filter (BF). Winner-take-all (WTA) is used to generate the preliminary disparity map. An edge-preserving filter (EPF) is applied to that output based on a transform that defines an isometry between curves on the 2D image manifold in 5D and the real line to eliminate these artefacts. The transform warps the input signal adaptively to allow linear 1D filtering. Due to the filter's resistance to high contrast and brightness, it is effective in refining and removing noise from the output image. Based on experimental research employing a Middlebury standard validation benchmark, this approach gives high accuracy with an average non-occluded error of 6.71% comparable to other published methods.

Corresponding Author:

Shamsul Fakhar Abd Gani
Faculty of Electrical and Electronic Engineering Technology
Universiti Teknikal Malaysia Melaka (UTeM)
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
Email: shamsulfakhar@utem.edu.my

1. INTRODUCTION

Machine vision research started with one camera. Due to their single point-of-view, it had many limitations, hence stereo cameras were developed to rebuild a 3D environment and fix defects produced by a single camera without depth information. Vision or LiDAR are the main approaches to get depth data for terrain or surface reconstruction. LIDAR scans the world and creates 3D distinct surfaces. Despite its accuracy and vast field views, it is hard to provide the hardware and interfaces and uses a lot of power, hence stereo vision is used to overcome it. The overlap between the visual fields of the left and right eyes enables humans to analyse and extract depth from seeing objects using two eyes. Stereo vision algorithms replicate these complex visuals by modelling them using a number of mathematical approaches. Research on stereo vision focuses mostly on stereo matching, which is sometimes referred to as the stereo correspondence problem [1]; where incorrect matches between images of several sensors obtruse the actual matches. Stereo matching uses arithmetic to find pixels in 2D stereoscopic images that match a 3D scene. Normalised

epipolar geometry simplifies finding correspondences on the same epipolar line producing a disparity map as the output. Stereo vision allows robots to estimate distance, plan collision-free paths, grasp objects with precision, and operate independently in dynamic environments [2]-[5]. Self-driving cars can determine the relative position of lane markings and make accurate decisions for trajectory planning, ensuring smooth and safe navigation [6]-[9]. Other applications are 3D face recognition [10], surface regeneration [11]-[14], robotic surgery [15]-[19], virtual reality [20]-[23], and augmented reality [24]. Numerous research publications have been circulated in this topic, and substantial progress has been achieved. The number of journal papers and books published by ScienceDirect in the domain of stereo vision between 2000 and 2022 is presented in Figure 1. Each year, new methods are derived, with a focus on i) accuracy and ii) time consumption. When deciding on a stereo method, these criteria should be carefully evaluated.

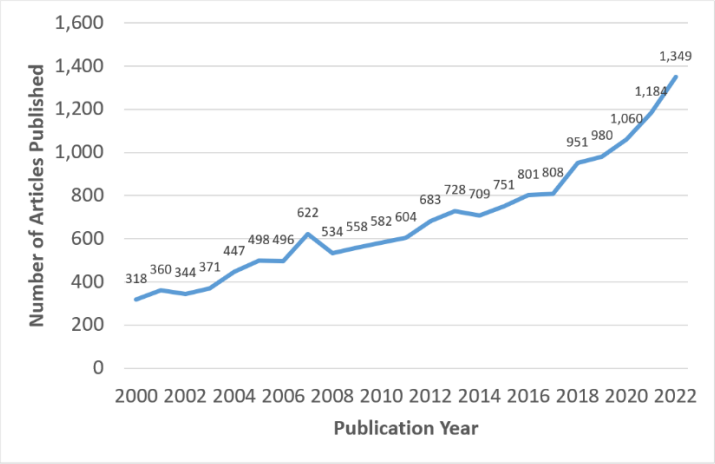


Figure 1. The increasing number of articles in journals or books published by ScienceDirect corresponding to the search term "stereo vision" from 2000 until 2022

The epipolar line is the result of projecting a line from the focus point of the left image onto the image plane of the right image [25]. This line denotes probable equivalent locations in the right image for a pixel in the left image. The locations and orientations of the cameras, together with other characteristics like focal length and field of vision are utilized to decide the orientations and positions of the epipolar lines. Rerectifying images along epipolar lines simplifies disparity candidates along a single x-axis [26], [27]. Common image rectification techniques may be broken down into three distinct transformation categories: projective, affine, and shearing. Stereo correspondence mechanism is then utilized to identify the pixels in the stereo images that are identical. Using the triangulation principle as shown in (1), the depth information is finally retrieved:

$$z = \frac{f.b}{x_L - x_R} = \frac{f.b}{d} \quad (1)$$

where z is the depth, f is the camera focal length, b is the baseline space in the middle of cameras' optical centre, with d as disparity. In a given stereo arrangement with constant f and b, the disparity scope limits the depth range into $[d_{min}, d_{max}]$. Erroneous pixels and other ambiguities in the input images will present an explicit influence on the quality of the output map. As illustrated in Figure 2, Scharstein and Szeliski [1] first suggested a common outline for stereo vision processes, which was executed by utilizing a sequenced multi-stage system. The framework receives the input image pair from a stereo camera that act as the stereo sensor. This framework relies on the assumption that the image pair input has been rectified. In the first stage, the cost function is computed to determine the degree of similarity among patches in the input image pair. Secondly, cost aggregation was conducted, followed by the use of filters to remove noise. The process was then replicated for the rest of the pixels in the left image. After that, disparity improvement is achieved by applying optical low-pass filter to refine it. Winner-take-all (WTA) strategy is then implemented where only the disparity with the lowest cost stays active while all other disparity candidates are shut down.

Stereo matching research can be categorized as a global or local optimization strategy based on how the disparity is computed. Local methods utilise disparity depending on the connection between pixel intensities (grayscale, RGB colours, texture patterns) inside a particular local support window. Sum of absolute differences (SAD), sum of squared differences (SSD), and normalised cross-correlation (NCC) are

some example of local methods. It estimates the disparity by comparing the surroundings of a pixel *p* in the left image to the surroundings of a pixel *q* in the right image, where *q* has been translated across a candidate disparity *p* as illustrated in Figure 3. Computation of this method is usually fast. However, the quality suffers, especially in the depth discontinuity area [28], [29].

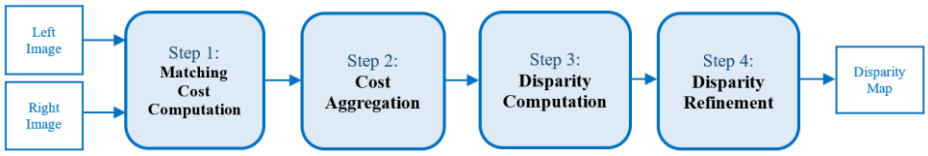


Figure 2. A sequential multistage stereo vision model [1]

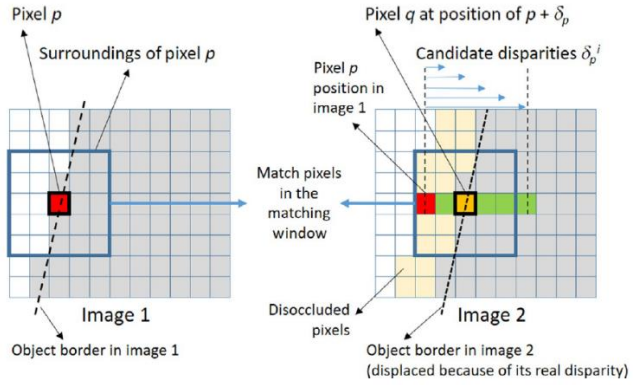


Figure 3. Matching windows in matching cost computation [30]

Global optimization strategies approach the disparity question as a process of reducing a preset global energy function. A number of solutions have been created by means of a markov random field (MRF). These methods may either be categorised as graph cut (GC) or belief propagation (BP) methods. The BP method lowers the energy function by continuously sending signals from the current node to nearby nodes in the MRF network [31], [32]. Since [33]-[35]'s introduction of convolutional neural network (CNN) trained on tiny image patch pairs with known actual disparity, attention to a deep learning-based stereo vision system has grown significantly. CNN outperforms traditional approaches in terms of error rate and processing time, but it remains challenging to identify optimal corresponding spots in fundamentally ill-posed areas, such as areas with repetitive shapes, areas that are obscured, reflective planes, and texture-less areas. This study discusses a stereo corresponding method that utilises a deep learning-based hybrid approach to generate stereo image features for computing the matching cost and an EPF to construct the resulting disparity map. CNN is used to extract features from the image pairs dataset and compute the corresponding cost, whereas BF and WTA aggregate costs and refine discrepancies.

2. PROPOSED METHOD

Figure 4 is a representation of the suggested approach for the experiment. CNN is first employed to obtain image pair information and determine the likeness measures. In cost aggregation, a BF that preserves edges while eliminating raw noise is used. WTA technique is then used to compute disparity by substituting minimal disparity value for minimum cost aggregation. Left and right image checking is performed to identify acceptable and unacceptable pixels in the image. The process generates obstruction zones and erroneous pixels, particularly in low-texture regions. By filling in, inaccurate disparity map image element are replaced with acceptable values. In the last step, an EPF [36] is performed to remove any residual noise created from the previous filling in procedure.

2.1. Matching cost computation

The convolution layer is a crucial component of CNN, comprising of a sequence of square-shaped kernels. Despite their small size, these filters will accommodate the whole volume depth. As the convolution layer is constructed, the quantity of layer depth will correspond to the number of filters employed in the layer

immediately before it. Each kernel reads the input region, total up the dot product, and saves the output in an activation layer. Next, the feature map layer is integrated to construct the input volume for the subsequent network layer. Rectified linear unit (ReLU), a non-linear activation function, is added after each convolution layer. ReLU keeps the dispensation required to operate a neural network from growing exponentially. The computation cost of introducing new ReLU layers to a growing CNN increases linearly. Pooling layer gradually decreases the dimension of the input, decreasing system parameters and processing, and aids in preventing fitting problem. The cost of matching may therefore be easily determined based on the CNN output.

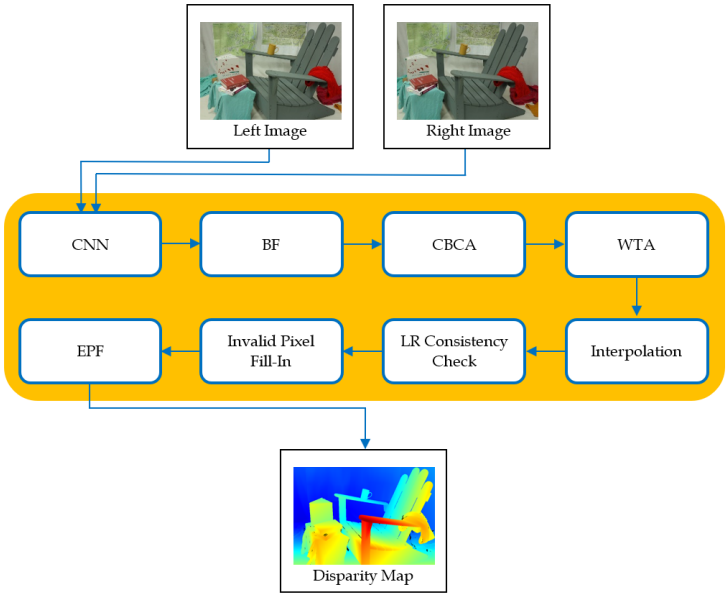


Figure 4. Block diagram of the proposed method

2.2. Cost aggregation

Cost aggregation is designed to lower corresponding ambiguity by applying a filter to smooth out the high noise at the preliminary raw matching cost. It is required because the information collected for a single pixel when computing the matching cost is insufficient for accurate matching. Bitwise operations on binary strings were utilised to construct the cost aggregate volume in [37]. This approach is quick since it requires little computing, but its efficiency is low. It treats binary numbers that are comparable in different regions of interest. Another method is by doing segmentation which was projected in [38] where segment tree (ST) was utilised in which pixels were sorted by reference colour and intensity into distinct segments. This approach yielded precise answers for the textured sections, but poor precision for the plain colour and non-textured parts. BF is used to boost the precision at the object's edges and minimise the noise on the insides of the edges. The BF kernel is represented by (2):

$$BF[I]_p = \frac{1}{W_p}\sum_{q\in s} G_{\sigma s}(|p-q|)G_{\sigma r}(I_p - I_q)I_q \quad (2)$$

where *Gσs* is 'space' parameter which determines the positive effect of faint pixels, *Gσr* is 'ranmge' parameter which determines the impact of pixel q having a concentration amount varying from I_p.

2.3. Disparity computation

Disparity computation determines the specific combination of disparities for which normalisation of disparity values happens. In this stage, WTA is implemented which can be illustrated as in (3):

$$d_p = \arg\min_{d\in D} C(p,d) \quad (3)$$

where *dp* is the disparity with the least expensive cost, *C(p,d)* is the cumulative cost calculated in the second stage, *D* is the set of all acceptable individual disparity, with the highest amount selected depending on the true disparity map. Because local approaches accumulate support areas by adding or making an even distribution of them, their precision is susceptible to noise and ambiguity areas.

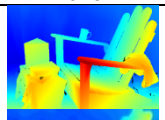
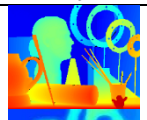
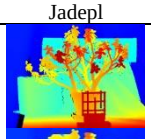
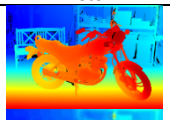
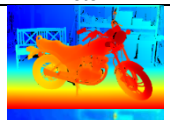
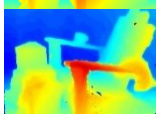
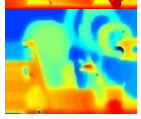
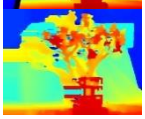
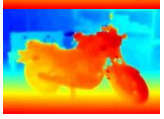
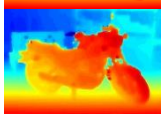
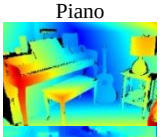
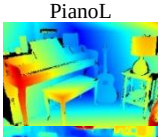
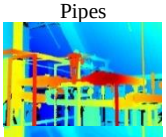
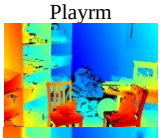
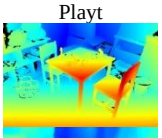
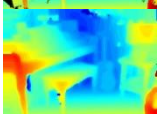
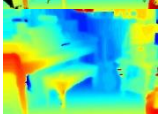
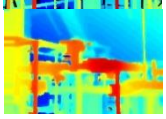
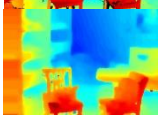
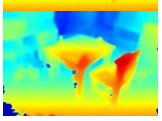
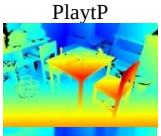
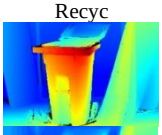
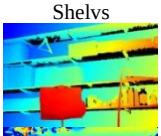
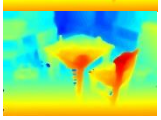
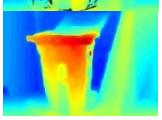
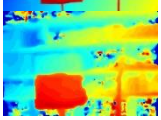
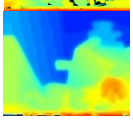
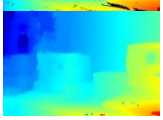
2.4. Disparity refinement

In the last stage, postprocessing and disparity refinement standard techniques were executed. Peng *et al.* [39] implemented the integration of mean shift and superpixel with segmentation method (SEG). This approach clusters the disparity map based on colour. The weighted median filter (WMF) is used in [40]. WMF combined box aggregation with a weighted median. This approach effectively removes noise from outliers while preserving the edges. The median filter (MF) was utilised in [41], [42]. The precision was good around the edges, but MF generates a significant amount of inaccuracy in places with poor texture. BF was utilised to improve the edge qualities, although its processing time was longer in [43]. This proposed technique finishes with several continuous processes, starting with occlusion management, incorrect pixel management, and noise reduction. To remove artifacts, the kernel is applied with an EPF based on a transform that specifies an isometry between curves on the 2D image manifold in 5D and the actual line, which then performs high-quality edge-preserving filtering on images. The transform preserves the geodesic distance between points on these curves, warping the input signal adaptively so that 1D filtering can be carried out in linear time and efficiently.

3. RESULTS AND DISCUSSION

This experiment is done using a Windows 10 with Intel Xeon 2.80GHz CPU, an Nvidia Quadro P1000 GPU with 8GB DDR4 RAM. The accuracy is assessed utilising Middlebury v3 benchmarking system where training images are assessed on the percentage of erroneous pixels in both occluded and non-occluded (NON-OCC) pixels. The deep learning component retains the parameters as in the experiment by [34]. We explored different approaches, including bitwise operations on binary strings and segmentation-based methods, such as ST to smoothen the matching cost and enhance precision. By considering support zones and utilizing techniques like BF [44], we improved the matching results, especially at object edges and textured regions. We employed cumulative cost calculations and selected the disparity with the lowest cost for each pixel. While local approaches is susceptible to noise and ambiguity, our method accounted for these limitations by considering neighboring pixel information and achieving substantial improvements in the final disparity map accuracy. Table 1 illustrates the ground truth from Middlebury compared with the output disparity map from the proposed system along with the error rate.

Table 1. The ground truth of Middlebury v3 dataset compared to our algorithm output

Image	Adiron	ArtL	Jadepl	Motor	MotorE
Ground truth					
Ours					
NON-OCC	4.04%	8.05%	22.10%	4.43%	4.43%
Image	Piano	PianoL	Pipes	Playrm	Playt
Ground truth					
Ours					
NON-OCC	4.22%	6.23%	8.14%	5.53%	5.02%
Image	PlaytP	Recyc	Shelvs	Teddy	Vintge
Ground truth					
Ours					
NON-OCC	4.49%	2.93%	12.10%	3.16%	6.88%

The quantitative evaluation results are shown in Table 2 for NON-OCC error and Table 3 for all error. The proposed method achieves the lowest error for NON-OCC error in Table 2 for the recycle bin and teddy bear photos, with 2.93% and 3.16% errors, respectively. Complicated images, such as jadeplant and shelves exhibit the worst accuracy at 22.10% and 12.10% owing to the system's incapacity to identify repetitive shapes, such as the tiny leaf bundles in the jadeplant image. Other areas that are difficult to match include textureless objects and shadows, such as ArtL and recycle images. These regions contain similar pixel values, and the likelihood of a mismatch is high. The technique proposed in this article, on the other hand, pinpointed the exact location of the difference. However, the proposed technique can reconstruct a nearly exact disparity map with distinctive discontinuities. The disparity level is applied exactly when the distance contours are clearly distinguished. The suggested approach reduces salt-and-pepper noise while preserving the dividing lines on the margins. Tables 2 and 3 present several published techniques to demonstrate the performance of the proposed work. The Middlebury evaluation shows that the proposed stereo corresponding approach can produce accurate results with an average NON-OCC error of 6.71%. It demonstrates that the suggested approach is comparable to recently published techniques and can be implemented as a comprehensive algorithm.

Table 2. Comparative results for NON-OCC error using Middlebury v3 evaluation platform

Method	Avg Err	Adiron	ArtL	Jadepl	Motor	Piano	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
SGM [45]	5.29	1.85	4.25	14.60	2.76	3.61	5.63	4.24	15.70	2.67	2.31	7.78	1.50	13.90
Ours	6.71	4.04	8.05	22.10	4.43	4.22	8.14	5.53	5.02	4.49	2.93	12.10	3.16	6.88
ELAS [46]	7.65	5.66	3.15	29.10	3.15	4.41	6.07	6.37	16.90	2.70	5.82	10.70	2.24	16.70
PSMNet_ROB [47]	9.60	7.32	9.69	44.50	5.55	5.01	9.86	7.33	4.40	4.43	3.73	11.10	3.44	8.07
BSM [37]	13.4	7.27	11.40	30.50	6.67	10.80	10.50	12.50	24.40	12.80	7.42	16.40	4.88	32.80

Table 3. Comparative results for ALL error using Middlebury v3 evaluation platform

Method	Avg Err	Adiron	ArtL	Jadepl	Motor	Piano	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
SGM [45]	8.51	2.46	7.83	32.1	5.17	5.12	11.3	6.15	18.5	3.59	2.84	8.35	2.53	15.5
ELAS [46]	10.9	6.68	5.3	51.2	5.36	4.99	11.1	8.97	18.7	3.56	6.43	11.1	2.99	17.1
Ours	12.5	8.08	13.10	53.40	9.00	5.56	16.20	8.92	9.50	7.47	5.76	13.30	4.79	10.00
PSMNet_ROB [47]	13.3	8.83	13.9	68.4	8.26	5.89	14.4	9.38	5.54	5.52	4.98	11.6	3.87	9.66
BSM [37]	23.5	12.7	28.7	58.7	14.8	16	24.5	29.4	31	20.2	12.1	19.2	14.3	39.3

4. CONCLUSION

Our research in stereo vision has led to the development of a comprehensive method that incorporates matching cost computation using CNN, cost aggregation to reduce ambiguity, disparity computation through WTA optimization, and an EPF as a post-processing technique for refinement. Our method achieved accurate disparity estimation by effectively addressing noise, occlusion, and texture-related challenges. Comparative evaluations demonstrated its competitive performance, with an average NON-OCC error of 6.71% on the Middlebury benchmark. Future research efforts should focus on addressing the limitations associated with repetitive shapes, textureless regions, and improving the robustness of disparity estimation in challenging scenarios. By continuing to refine the proposed method, we can contribute to the ongoing progress in stereo vision and its applications, ultimately enhancing our understanding and utilization of 3D perception.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to Ministry of Higher Education (MoHE) Malaysia and Universiti Teknikal Malaysia Melaka (UTeM) for providing the funding necessary to accomplish this study via Fundamental Research Grant Scheme (FRGS) No: FRGS/1/2020/TKO/UTeM/02/12.

REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, 2002, doi: 10.1023/A:1014573219977.

[2] D. Murray and C. Jennings, "Stereo vision based mapping and navigation for mobile robots," *Proceedings of International Conference on Robotics and Automation*, Albuquerque, NM, USA, 1997, pp. 1694-1699 vol. 2, doi: 10.1109/ROBOT.1997.614387.

[3] R. A. Hamzah, S. F. Abd Ghani, A. F. Kadmin, and K. A. A. Aziz, "A practical method for camera calibration in stereo vision mobile robot navigation," *2012 IEEE Student Conference on Research and Development (SCOReD)*, Pulau Pinang, Malaysia, 2012, pp. 104-108, doi: 10.1109/SCOReD.2012.6518620.

[4] 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, p. 433, Dec. 2011, doi: 10.12928/telkomnika.v9i3.733.

[5] M. N. K. M. S. M. Aras, S. S. Abdullah, S. Y. Binti Othman, M. Sulaiman, M. F. Basar, and M. K. M. Zambri, "Fuzzy logic controller for depth control of underwater remotely operated vehicle," *Journal of Theoretical and Applied Information Technology*, vol. 91, no. 2, pp. 275–288, 2016.

[6] Y. Shi, Y. Guo, Z. Mi, and X. Li, "Stereo CenterNet-based 3D object detection for autonomous driving," *Neurocomputing*, vol. 471, pp. 219–229, Jan. 2022, doi: 10.1016/j.neucom.2021.11.048.

[7] C. Tao, H. He, F. Xu, and J. Cao, "Stereo priori RCNN based car detection on point level for autonomous driving," *Knowledge-Based Systems*, vol. 229, p. 107346, Oct. 2021, doi: 10.1016/j.knosys.2021.107346.

[8] M. Z. R. Z. Ahmadi, A. Jidin, K. B. Jaffar, M. N. Othman, R. N. P. Nagarajan, and M. H. Jopri, "Minimization of torque ripple utilizing by 3-L CHMI in DTC," in *Proceedings of the 2013 IEEE 7th International Power Engineering and Optimization Conference, PEOCO 2013*, Jun. 2013, pp. 636–640, doi: 10.1109/PEOCO.2013.6564625.

[9] F. A. Phang *et al.*, "Integrating Drone Technology in Service Learning for Engineering Students," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 15, p. 78, Aug. 2021, doi: 10.3991/ijet.v16i15.23673.

[10] 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, p. 2818, Dec. 2016, doi: 10.11591/ijece.v6i6.12970.

[11] S. S. M. Allan *et al.*, "Stereo Correspondence and Reconstruction of Endoscopic Data Challenge," *arXiv*, 2021, doi: 10.48550/arXiv.2101.01133.

[12] R. A. Hamzah, A. F. Kadmin, M. S. Hamid, S. F. A. Ghani, and H. Ibrahim, "Improvement of stereo matching algorithm for 3D surface reconstruction," *Signal Processing: Image Communication*, vol. 65, pp. 165–172, Jul. 2018, doi: 10.1016/j.image.2018.04.001.

[13] A. K. M. Z. Hossain, N. Bin, J. Abedalrahim, M. Kamrul, and M. Rafiqul, "A Tree-profile Shape Ultra Wide Band Antenna for Chipless RFID Tags," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, 2021, doi: 10.14569/ijacsa.2021.0120469.

[14] A. W. Y. K. J. A. J. Alsayaydeh *et al.*, "DEVELOPMENT OF PROGRAMMABLE HOME SECURITY USING GSM SYSTEM FOR EARLY PREVENTION," *ARPN Journal of Engineering and Applied Sciences*, vol. 16, no. 1, pp. 88–97, 2021.

[15] M. Ye, E. Johns, A. Handa, L. Zhang, P. Pratt, and G. Z. Yang, "Self-Supervised Siamese Learning on Stereo Image Pairs for Depth Estimation in Robotic Surgery," *The Hamlyn Symposium*, Jun. 2017, doi: 10.31256/hsmr2017.14.

[16] P. Pratt, D. Stoyanov, M. Visentini-Scarzanella, and G.-Z. Yang, "Dynamic Guidance for Robotic Surgery Using Image-Constrained Biomechanical Models," in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2010, pp. 77–85, doi: 10.1007/978-3-642-15705-9_10.

[17] M. H. Jopri, A. R. Abdullah, M. Manap, T. Sutikno, and M. R. Ab Ghani, "An Identification of Multiple Harmonic Sources in a Distribution System by Using Spectrogram," *Bulletin of Electrical Engineering and Informatics*, vol. 7, no. 2, pp. 244–256, Jun. 2018, doi: 10.11591/eei.v7i2.1188.

[18] A. R. Abdullah, N. Norddin, N. Q. Z. Abidin, A. Aman, and M. H. Jopri, "Leakage current analysis on polymeric and non-polymeric insulating materials using time-frequency distribution," 2012 IEEE International Conference on Power and Energy (PECon), Kota Kinabalu, Malaysia, 2012, pp. 979-984, doi: 10.1109/PECon.2012.6450360.

[19] M. Manap, M. H. Jopri, A. R. Abdullah, R. Karim, M. R. Yusoff, and A. H. Azahar, "A verification of periodogram technique for harmonic source diagnostic analytic by using logistic regression," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 17, no. 1, p. 497, Feb. 2019, doi: 10.12928/telkomnika.v17i1.10390.

[20] C. B. L. Xiaoxiao *et al.*, "Recent progress in 3D vision," *Journal of Image and Graphics*, vol. 26, no. 1, pp. 1389–1428, 2021, doi: 10.11834/JIG.210043.

[21] J. Zhao and R. S. Allison, "The Role of Binocular Vision in Avoiding Virtual Obstacles While Walking," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 7, pp. 3277–3288, Jul. 2021, doi: 10.1109/tvcg.2020.2969181.

[22] C. Uz-Bilgin, M. Thompson, and E. Klopfer, "Stereoscopic Views Improve Spatial Presence but Not Spatial Learning in VR Games," *PRESENCE: Virtual and Augmented Reality*, vol. 28, pp. 227–245, Jan. 2019, doi: 10.1162/pres_a_00349.

[23] J. Tong, L. M. Wilcox, and R. S. Allison, "The impacts of lens and stereo camera separation on perceived slant in Virtual Reality head-mounted displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 11, pp. 3759–3766, Nov. 2022, doi: 10.1109/tvcg.2022.3203098.

[24] M. von Atzigen *et al.*, "Marker-free surgical navigation of rod bending using a stereo neural network and augmented reality in spinal fusion," *Medical Image Analysis*, vol. 77, p. 102365, Apr. 2022, doi: 10.1016/j.media.2022.102365.

[25] R. A. Hamzah *et al.*, "A study of edge preserving filters in image matching," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 111–117, Feb. 2021, doi: 10.11591/eei.v10i1.1947.

[26] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Machine Vision and Applications*, vol. 12, no. 1, pp. 16–22, Jul. 2000, doi: 10.1007/s001380050120.

[27] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, no. 1–2, pp. 87–119, Oct. 1995, doi: 10.1016/0004-3702(95)00022-4.

[28] 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, Sep. 2019, doi: 10.11591/eei.v8i3.1518.

[29] A. Z. Jidin, R. Hussin, L. W. Fook, and M. S. Mispan, "An Automation Program for March Algorithm Fault Detection Analysis," in *2021 IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2021 and 2021 IEEE Conference on Postgraduate Research in Microelectronics and Electronics, PRIMEASIA 2021*, IEEE, Nov. 2021, pp. 149–152, doi:

10.1109/APCCAS51387.2021.9687806.
[30] O. Stankiewicz, G. Lafruit, and M. Domański, "Multiview video: Acquisition, processing, compression, and virtual view rendering," in *Academic Press Library in Signal Processing*, vol. 6, pp. 3–74, 2018, doi: 10.1016/b978-0-12-811889-4.00001-4.
[31] Y. Liu and J. K. Aggarwal, "Local and Global Stereo Methods," in *Handbook of Image and Video Processing*, Elsevier, 2005, pp. 297–308, doi: 10.1016/b978-012119792-6/50081-4.
[32] A. Z. Jidin *et al.*, "Generation of New Low-Complexity March Algorithms for Optimum Faults Detection in SRAM," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 8, pp. 2738–2751, Aug. 2023, doi: 10.1109/tcad.2022.3229281.
[33] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, IEEE, May 2010, doi: 10.1109/iscas.2010.5537907.
[34] J. Žbontar and Y. Lecun, "Stereo matching by training a convolutional neural network to compare image patches," *Journal of Machine Learning Research*, vol. 17, pp. 1–32, 2016.
[35] N. M. A. Brahin, H. M. Nasir, A. Z. Jidin, M. F. Zulkifli, and T. Sutikno, "Development of vocabulary learning application by using machine learning technique," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 1, pp. 362–369, Feb. 2020, doi: 10.11591/eei.v9i1.1616.
[36] E. S. L. Gastal and M. M. Oliveira, "Domain Transform for Edge-Aware Image and Video Processing," *ACM Transactions on Graphics*, vol. 30, no. 4, pp. 1–12, Jul. 2011, doi: 10.1145/2010324.1964964.
[37] K. Z. J. L. Y. L. W. H. L. S. S. Yang, "Binary stereo matching," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 356–359.
[38] X. Mei, X. Sun, W. Dong, H. Wang, and X. Zhang, "Segment-Tree Based Cost Aggregation for Stereo Matching," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2013, doi: 10.1109/cvpr.2013.47.
[39] Y. Peng, G. Li, R. Wang, and W. Wang, "Stereo matching with space-constrained cost aggregation and segmentation-based disparity refinement," in *Three-Dimensional Image Processing, Measurement (3DIPM), and Applications 2015*, R. Sitnik and W. Puech, Eds., Mar. 2015, p. 939309, doi: 10.1117/12.2083741.
[40] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, "Constant time weighted median filtering for stereo matching and beyond," in *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, Dec. 2013, pp. 49–56, doi: 10.1109/ICCV.2013.13.
[41] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," in *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, Nov. 2011, pp. 467–474, doi: 10.1109/ICCVW.2011.6130280.
[42] S. F. A. Gani, R. A. Hamzah, R. Latip, S. Salam, F. Noraqillah, and A. I. Herman, "Image compression using singular value decomposition by extracting red, green, and blue channel colors," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 168–175, Feb. 2022, doi: 10.11591/eei.v11i1.2602.
[43] Q. Yang, "Recursive Approximation of the Bilateral Filter," *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 1919–1927, Jun. 2015, doi: 10.1109/TIP.2015.2403238.
[44] S. F. A. Gani, M. F. Miskon, and R. A. Hamzah, "Depth Map Information from Stereo Image Pairs using Deep Learning and Bilateral Filter for Machine Vision Application," in *2022 IEEE 5th International Symposium in Robotics and Manufacturing Automation, ROMA 2022*, IEEE, Aug. 2022, pp. 1–6, doi: 10.1109/ROMA55875.2022.9915680.
[45] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, Feb. 2008, doi: 10.1109/tpami.2007.1166.
[46] A. Geiger, M. Roser, and R. Urtasun, "Efficient Large-Scale Stereo Matching," in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2011, pp. 25–38, doi: 10.1007/978-3-642-19315-6_3.
[47] J. R. Chang and Y. S. Chen, "Pyramid Stereo Matching Network," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 5410–5418, doi: 10.1109/CVPR.2018.00567.

BIOGRAPHIES OF AUTHORS

Shamsul Fakhar Abd Gani completed his Bachelor of Engineering (Computer Engineering) with honours from Universiti Malaysia Perlis and his Master of Technology (Internet and Web Computing) from Royal Melbourne Institute of Technology, Australia. Currently, he is a senior lecturer at Universiti Teknikal Malaysia Melaka conducting primarily computer vision-related research. He can be contacted at email: shamsulfakhar@utem.edu.my.

Muhammad Fahmi Miskon is an Associate Professor at Universiti Teknikal Malaysia Melaka's Department of Mechatronics Engineering. He holds a Bachelor of Electrical Engineering (Mechatronics) from Universiti Teknologi Malaysia and a Master of Science in Mechatronics from the University of Newcastle Upon Tyne, United Kingdom. He holds a Ph.D. in Electrical and Computer Engineering from Australia's Monash University. Novelty detection and robotics are among his research interests. He can be contacted at email: fahmimiskon@utem.edu.my.

Rostam Affendi Hamzah graduated from Universiti Teknologi Malaysia with a Bachelor of Engineering in Electronic Engineering. In 2010, he graduated from Universiti Sains Malaysia with a Master of Science in Electronic System Design Engineering. In 2017, he received Ph.D majoring in Electronic Imaging from Universiti Sains Malaysia. His research interests consist of computer vision, pattern recognition, and digital image processing. He can be contacted at email: rostamaffendi@utem.edu.my.

Mohd Saad Hamid received his Bachelor of Engineering with a major in Computer from Multimedia University. In 2014, he received his Master of Engineering in Computer and Communication from Universiti Kebangsaan Malaysia. He was formerly employed by Telekom Malaysia Berhad and Continental Automotive Components (M) Sdn Bhd and is currently pursing a doctoral degree at Universiti Teknikal Malaysia Melaka. Computer vision, deep learning, image processing, and embedded systems are his research interests. He can be contacted at email: mohdsaad@utem.edu.my.

Ahmad Fauzan Kadmin is a Chartered Engineer and Professional Technologist affiliated with UTeM as a senior lecturer with over 16 years of experience in the field of electronic and computer engineering, computer vision, and medical electronics. He graduated with a Bachelor's Degree in Electronics Engineering from Universiti Sains Malaysia and a Master's Degree in Computer & Communication Engineering from Universiti Kebangsaan Malaysia. He has published a number of technical and engineering papers in the fields of computer vision and medical image processing. He can be contacted at email: fauzan@utem.edu.my.

Adi Irwan Herman graduated with a Bachelor's Degree in Computer Engineering Technology (Computer Systems) with honours from Universiti Teknikal Malaysia Melaka. He has served with Texas Instruments Melaka for over 7 years, and his current research interest involves computer engineering-related disciplines. He can be contacted at email: adiirwanherman@gmail.com.