**Indonesian Journal of Electrical Engineering and Computer Science**
Vol. 25, No. 1, January 2022, pp. 223~231
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v25.i1.pp223-231 

# A new function of stereo matching algorithm based on hybrid convolutional neural network

**Mohd Saad Hamid[1,2], Nurulfajar Abd Manap[1], Rostam Affendi Hamzah[2], Ahmad Fauzan Kadmin[1,2], Shamsul Fakhar Abd Gani[2], Adi Irwan Herman[3]**
[1]Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[2]Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[3]Product & Test Engineering, Texas Instruments, Batu Berendam, Melaka, Malaysia

**Article Info**

***Article history:***

Received Apr 23, 2021
Revised Nov 23, 2021
Accepted Nov 30, 2021

***Keywords:***

Convolutional neural network
Directional intensity
Stereo matching algorithm
Stereo vision

**ABSTRACT**

This paper proposes a new hybrid method between the learning-based and handcrafted methods for a stereo matching algorithm. The main purpose of the stereo matching algorithm is to produce a disparity map. This map is essential for many applications, including three-dimensional (3D) reconstruction. The raw disparity map computed by a convolutional neural network (CNN) is still prone to errors in the low texture region. The algorithm is set to improve the matching cost computation stage with hybrid CNN-based combined with truncated directional intensity computation. The difference in truncated directional intensity value is employed to decrease radiometric errors. The proposed method's raw matching cost went through the cost aggregation step using the bilateral filter (BF) to improve accuracy. The winner-take-all (WTA) optimization uses the aggregated cost volume to produce an initial disparity map. Finally, a series of refinement processes enhance the initial disparity map for a more accurate final disparity map. This paper verified the performance of the algorithm using the Middlebury online stereo benchmarking system. The proposed algorithm achieves the objective of generating a more accurate and smooth disparity map with different depths at low texture regions through better matching cost quality.



***Corresponding Author:***

Mohd Saad Hamid
Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik, Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
Email: mohdsaad@utem.edu.my

## 1. INTRODUCTION

Recent years have seen a rise in the number of significant advancements in the stereo vision area. As one of the active topics under stereo vision, the stereo matching problem still appears to be actively discussed among researchers around the globe today [1]. One of the objectives in stereo matching algorithm development is to produce a disparity or depth map. The map contains the disparity values between pixel locations of matching features in the left and right images. The map can provide the depth information between the camera and object. This information is valuable for applications such as three-dimensional (3D) reconstruction, obstacle avoidance, robotics, and navigation. The stereo vision approach also has been used to generate the 3D facial image for facial recognition [2].

Besides the stereo-based approach, as mentioned by [3], the single view or monocular-based method can also produce depth-related information. This approach can generate a disparity map using a single image. However, as [3] and [4] pointed out, the monocular depth method was less accurate than the stereo-based approach. It is because the multi-view contains a broader amount of information compared to the single-view approach. Another merit for the stereo-based method, as mentioned by [4], is the capability to outperform the monocular method in recovering the objects of interest.

Another method for generating depth-related information is light detection and ranging (LiDAR). LiDAR uses the laser approach to sense the depth and perform the depth measurement. LiDAR method can provide accurate 3D points [4]. However, this method is not very cost-effective and time-consuming [5]. Due to the highlighted shortcomings in monocular and LiDAR methods, we are inspired to continue working on the stereo-based method and propose our algorithm.

Figure 1 illustrates the crucial steps for the stereo vision algorithm. The formalization of the algorithm steps is related to the previous literature [6]-[9]. The first step is matching cost computation. Some researchers utilize the directional intensity difference to compute matching costs [10]. The calculation of absolute and squared differences is also commonly used because it requires low computational complexity. The next step, cost aggregation, can decrease the errors in the initial matching cost. The edge-preserving filters such as bilateral filter (BF) [11] and guided filter (GF) [12] can achieve the purpose. These filters maintain a good edge while smoothing the input. Thus it provides better results in the aggregation step than the low pass filters (Gaussian and box filter). The third step is in charge of assigning a value to the disparity map. Winner-take-all (WTA) optimization is the most popular method for this step for the local approach. In WTA, the smallest cost value disparity will be selected for every pixel location [7] to generate an initial disparity map. The third step's initial disparity may still contain errors caused by occlusions, low textures, and invalid matches [13]. So, the final step may include multiple post-processing steps to refine the map. The authors of [12] used the left-to-right consistency (LRC) check process to identify the void matched pixels. The median filter is also commonly used for local refinement [14]-[16]. It is another type of non-linear filter for denoising purposes. Due to its complexity, this final step can add additional time to the overall process.

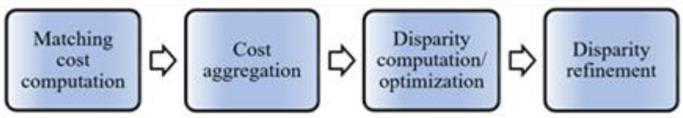



Figure 1. Stereo vision algorithm steps

Deep learning has become the driving force behind advancements in the computer vision field [17]. Deep learning utilizes artificial neural networks to learn a large amount of data [18] mentioned that conventional stereo vision performance will be improved by incorporating deep learning for recognition tasks. Implementation of deep learning on stereo vision can be divided into two approaches. The first approach is the deep learning approach combined with a traditional handcraft algorithm matching cost with a convolutional neural network (MC-CNN-acrt) [14]. MC-CNN-acrt architecture outperforms other stereo matching conventional methods on Middlebury [19] benchmarking systems [14] construct and train their convolutional neural network (CNN) based network using binary classification to solve matching cost computation steps. Another researcher in [20] proposed CNN-based similarity learning in Euclidean space using a dot product. The method computes faster than MC-CNN, but the result is less accurate. CNN is also used in stereo matching because of its feature extraction capabilities and vigorous radiometric difference [15]. The second approach is the pure deep learning end-to-end network. This approach does not require any handcrafted algorithm. The end-to-end style deep learning network performs all stereo matching stages in one combination network. Geometry and context network (GC-Net) by [21] uses 2D CNN to form cost volumes and uses soft-argmin layer to regress the disparity values [22] introduce PSMNet, a faster and more accurate disparity map than GC-Net.

The hybrid method between learning and handcraft algorithm was also introduced by [23] recently. Their deep learning network produces a disparity map and predictions of uncertainties. They also propose a handcrafted method to enhance their disparity map using a modified SGBM algorithm (SGBMP) by [23], [24] introduces a hybrid model between CNN and the learning-based conditional random field (CRF) model to form an end-to-end learning network. The performance of the hybrid and pure end-to-end methods will be discussed further in section 3.

The author of [25] concluded, the end-to-end method still has some demerits in the ill-posed region and is computationally expensive. Additionally, [14] also pointed out that the raw disparity map generated by a CNN is prone to errors in the low texture and the occluded region. The problem with the low texture region is also mentioned by [13], affecting the similarity measurement. Other researchers [26]-[28] also stated that the low texture region is one of the challenging areas to cater to in stereo matching methods. So, in the following section of this paper, we will present our proposed stereo matching algorithm. This paper's main contribution is the hybrid converged classification CNN fused with directional intensity information for the matching cost computation stage. The main objective is to produce a better disparity map by focusing on the low texture region errors.

## 2. THE PROPOSED METHOD

As mentioned in the earlier section, we present our proposed algorithm as categorized in [6]. A summary view of our algorithm is illustrated in Figure 2. Our proposed algorithm is comprised of four stages,

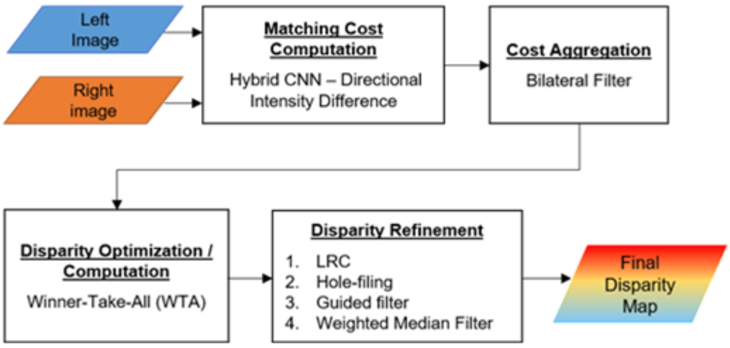



Figure 2. The proposed stereo vision algorithm steps

### 2.1. Matching cost computation

Our CNN-based model is inspired by the MC-CNN-acrt [14] architecture. However, we did improve the architecture, as discussed in [29]. In this paper, we extend the improvement through our new findings. As discussed in our previous work, the converged classification method in our Siamese-based CNN model [29] has been used for the matching cost computation stage. We then fused it with additional directional intensity differences to improve the accuracy of matching costs computed in this stage. We maintained the eight layers of our CNN architecture as in our previous work. The network will produce a similarity score that provides a binary classification of good or bad matching. Based on (1), $C_{CNN}$ reflects the cost value for all disparities d at each pixel position $p$.

$$C_{CNN}(p,d) = -s\left(P_N^L(p), P_N^R(p-d)\right) \tag{1}$$

Input patches of $NxN$ size from left and right image, $P_N^L$ and $P_N^R$ respectively supplied to the CNN. The minus sign would translate the score of similarity to the initial matching cost. The matching cost values obtained by (1) are also discussed in [29]. Another crucial part of this cost computation stage is the directional intensity difference. The second part of the cost function is defined in (2).

$$C_{DI}(p,d) = \alpha \cdot min(\parallel I_L(p) - I_R(p) \parallel, \tau_1) + (1-\alpha) \cdot min\left(\parallel \underset{x}{\nabla} I_L(p) - \underset{x}{\nabla} I_R(p) \parallel, \tau_2\right) \tag{2}$$

Here $I(p)$ represent the colour vector of a pixel at position $p$. $\nabla x$ is the directional intensity difference in the $x$-direction. $A$ balances the directional intensity difference values, while $\tau_1$ and $\tau_2$ are truncation values. This paper improves the matching cost step from our previous work [29] by combining it with the directional intensity difference-based cost volume, $C_{DI}$, as in (3).

$$C_{IM}(p,d) = C_{DI}(p,d) + \lambda \cdot C_{CNN}(p,d) \tag{3}$$

The parameter λ will balance the cost volumes. The raw matching cost volume cost from this stage is defined as $C_{IM}$.

### 2.2. Cost aggregation

In the second stage, we refine the raw matching costs to create a more accurate disparity map. It is because the initial cost volume generated from the previous stage is prone to noises. Therefore, we aggregate cost volumes in this stage using a bilateral filter (BF). The BF is responsible for performing smoothing operations while keeping the edges sharp on the matching costs. The BF [11] defined in (4) will aggregate the raw matching cost volume, $C_{IM}$.

$$BF[I]p = \frac{1}{W_p}\sum_{q\epsilon S} G_{\sigma_s}(\parallel p - q \parallel) G_{\sigma_r}(\mid I_p - I_q \mid) \tag{4}$$

Where $Wp$ is the normalization factor. $I_p$ and $I_q$ represent the intensity at pixel position $p$ and $q$, respectively. $p$ represents the $(x, y)$ coordinates pixel of interest while the neighbouring coordinate $q$ is within support region $S$. The aggregated cost is denoted as $C_{AGG}$ in (5) at the end of this step.

$$C_{AGG}(p, d) = BF[I]C_{IM}(p, d) \tag{5}$$

### 2.3. Disparity computation / optimization

Using winner-take-all (WTA) optimization, we computed the disparity map for this disparity computation stage. We employ WTA optimization as defined in (6) to produce an initial disparity map, $d(p)$.

$$d(p) = \underset{d\epsilon d_r}{\operatorname{argmin}}\left(C_{AGG}(p, d)\right) \tag{6}$$

The sets of disparity values of $d_r$ obtained from the ground truth. The $d(p)$ is the chosen disparity value for position $d(p)$, which represents 2D coordinates $(x, y)$.

### 2.4. Disparity refinement

This refinement stage consists of several steps. As discussed by [12], left-right consistency (LRC) check and interpolation were also used to deal with occlusion and mismatches. Firstly, we employ LRC to identify invalid pixels. Then the detected invalid pixel is replaced with a valid pixel using the hole-filling process. After that, we implemented a guided image filter (GIF) due to its good edge-preserving capabilities. The filter kernel for GIF is implemented in this paper as defined in (7).

$$G_{p,q}(I) = \frac{1}{|w|^2}\sum_{q\epsilon W_k}\left(1 + \frac{(I_p - \mu_k)(I_q - \mu_k)}{\sigma_k^2 + \varepsilon}\right) \tag{7}$$

Where $I$ the guidance image and $p$ represent the *(x; y)* coordinates pixel of interest. Another pixel position, $q$, denote the neighbouring pixel in the support region wk. Then, $\sigma$ and $\mu$ are the variance and mean of the intensity values, respectively. The parameter, $\varepsilon$ is used as a control element for the smoothness term. The refined disparity map using GIF is defined as $d_{GF}$ in (8).

$$d_{GF}(p) = G_{p,q}(I)d(p) \tag{8}$$

The weighted median filter (WMF) was implemented to remove any existing outliers in the disparity map. We use the following cosine similarity weight function as defined in (9) for the weighted median filtering.

$$W_p^{cos} = \frac{I_p \cdot I_q}{\parallel I_p \parallel \cdot \parallel I_q \parallel} \tag{9}$$

$$d_f(p) = W_p^{cos} h(p) d_{GF}(p) \tag{10}$$

After the WMF is implemented, the final disparity map generated from the whole proposed algorithm represented by $d_f(p)$ in (10). The following section will provide an overview of the proposed method's performance quantitatively and qualitatively,

## 3. RESULTS AND DISCUSSION

Several tests were executed on a personal computer (PC) platform to examine the proposed method's performance using Middlebury v3 datasets [19]. The main code runs on the Python platform, utilizing Keras and Tensorflow library to perform the CNN architecture's inference. The hardware used in this test is a PC with Intel Core i5 3.0 GHz with 16 GB DDR3 RAM and the Nvidia GTX1060 GPU. We maintained the hyperparameters of our CNN model, similar to our previous work. The image datasets for the tests were taken from Middlebury online benchmarking system [19]. The results of the proposed algorithm are also compared with other methods such as MC-CNN-acrt and MC-CNN-fst by [14], pyramid stereo matching network

(PSMNet_ROB) by [22], hybrid CNN+CRF models by [24] (labelled as JMR), SGBMP by [23], MC-CNN-WS by [30] and line segment based efficient large scale stereo matching (LS_ELAS) by [31]. Table 1 shows the results obtained for *All* errors (error of invalid disparity values in all pixels). Another metric used is shown in Table 2, which illustrates the *NonOcc* error percentage (error of invalid disparity values in non-occluded pixel).

Based on the *All* errors results in Table 1, the proposed algorithm framework outperforms the original MC-CNN-acrt and other published methods. The proposed method can reduce the average errors down to 9.29%. It can also perform better than another recently published deep learning method denoted as SGBMP [23] and JMR [24]. However, there is a considerable difference between the SGBMP method's results and ours, especially on the Jadeplant, PianoL, Vintage and PlayTable images.

Based on the quantitative comparison of *NonOcc* error percentage in Table 2, the proposed method performed moderately. However, the proposed method still performed better than the end-to-end network-based method, PSMNet [22]. The proposed method generates output with 6.05% of errors, while PSMNet produces 9.60%. Thus, except for the Vintage and Piano images, our proposed method outperforms PSMNet in almost every image.

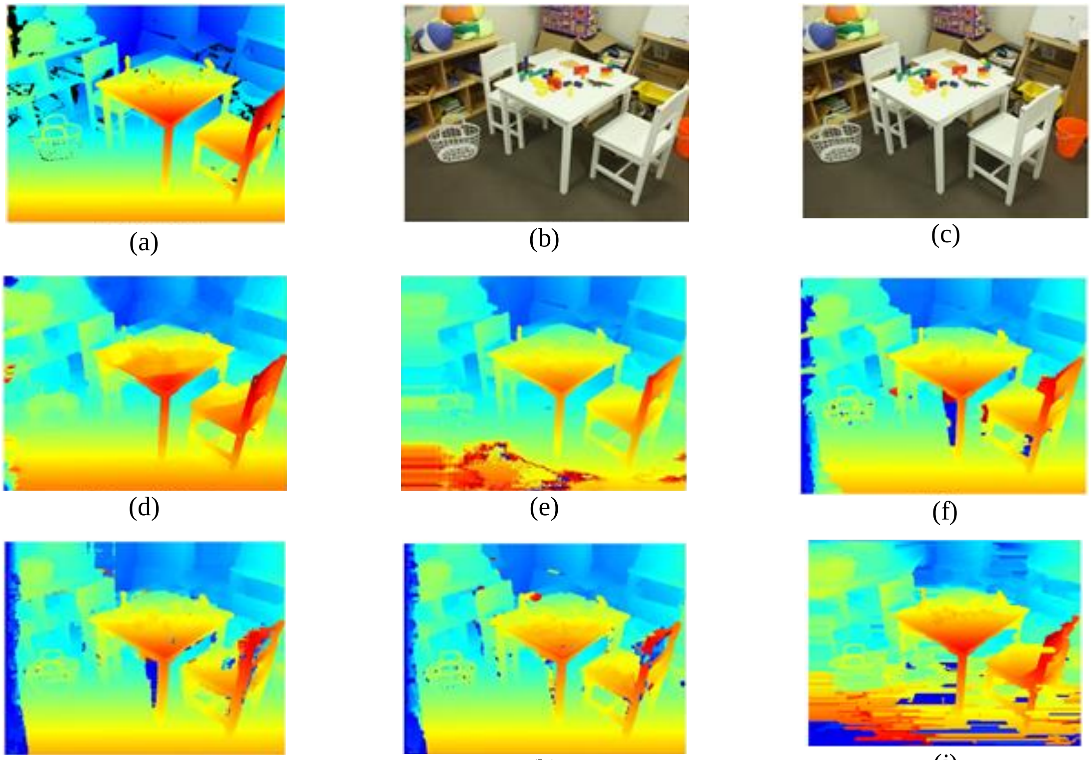
Table 1. Results from Middlebury benchmark - All error

| Method | Proposed | JMR | SGBMP | MC-CNN-acrt | MC-CNN-fst | PSMNet_ROB | MC-CNN-WS | LS_ELAS |
|---|---|---|---|---|---|---|---|---|
| Adirondack | 5.37 | 2.17 | 6.50 | 4.24 | 5.32 | 8.83 | 5.73 | 9.31 |
| ArtL | 8.36 | 18.00 | 9.33 | 18.70 | 19.20 | 13.90 | 20.50 | 5.90 |
| Jadeplant | 37.60 | 24.70 | 56.80 | 34.10 | 32.60 | 68.40 | 36.30 | 64.50 |
| Motorcycle | 7.24 | 5.98 | 5.04 | 7.21 | 8.75 | 8.26 | 9.39 | 7.24 |
| MotorcycleE | 7.29 | 6.90 | 5.43 | 7.22 | 8.83 | 9.16 | 9.37 | 7.65 |
| Piano | 6.63 | 6.14 | 4.77 | 6.00 | 8.12 | 5.89 | 8.13 | 6.25 |
| PianoL | 7.88 | 7.27 | 14.80 | 9.35 | 17.20 | 10.50 | 16.10 | 9.69 |
| Pipes | 11.40 | 11.00 | 7.85 | 13.50 | 15.50 | 14.40 | 16.70 | 12.80 |
| Playroom | 9.78 | 17.50 | 7.62 | 18.30 | 18.60 | 9.38 | 18.70 | 10.10 |
| Playtable | 4.72 | 8.18 | 10.60 | 9.71 | 13.60 | 5.54 | 11.50 | 23.90 |
| PlaytableP | 4.48 | 7.44 | 3.78 | 9.37 | 9.75 | 5.52 | 10.10 | 4.27 |
| Recycle | 3.79 | 2.96 | 3.19 | 4.64 | 5.00 | 4.98 | 5.05 | 7.39 |
| Shelves | 8.44 | 7.81 | 5.00 | 6.62 | 8.91 | 11.60 | 9.83 | 8.48 |
| Teddy | 3.64 | 8.98 | 3.35 | 9.31 | 10.40 | 3.87 | 11.00 | 2.98 |
| Vintage | 9.96 | 10.30 | 30.00 | 21.60 | 15.80 | 9.66 | 20.80 | 14.00 |
| **Average** | **9.29** | **9.57** | **11.20** | **11.80** | **12.80** | **13.30** | **13.70** | **12.90** |

Table 2. Results from Middlebury benchmark - NonOcc error

| Method | JMR | MC-CNN-acrt | MC-CNN-fst | MC-CNN-WS | Proposed | SGBMP | PSMNet_ROB | LS_ELAS |
|---|---|---|---|---|---|---|---|---|
| Adirondack | 0.92 | 0.76 | 1.21 | 1.66 | 3.23 | 3.87 | 7.32 | 8.46 |
| ArtL | 2.18 | 2.49 | 2.84 | 4.27 | 5.79 | 4.96 | 9.69 | 3.83 |
| Jadeplant | 6.01 | 16.30 | 10.00 | 12.80 | 19.00 | 29.30 | 44.50 | 41.10 |
| Motorcycle | 1.26 | 1.27 | 1.62 | 2.26 | 4.59 | 3.45 | 5.55 | 5.12 |
| MotorcycleE | 1.27 | 1.27 | 1.61 | 2.18 | 4.50 | 3.89 | 6.12 | 5.80 |
| Piano | 2.21 | 1.83 | 3.17 | 3.21 | 5.88 | 3.82 | 5.01 | 5.54 |
| PianoL | 4.03 | 5.07 | 13.20 | 11.70 | 7.31 | 14.40 | 9.82 | 8.97 |
| Pipes | 2.12 | 2.29 | 3.20 | 4.27 | 7.05 | 3.94 | 9.86 | 7.44 |
| Playroom | 1.94 | 2.27 | 3.13 | 3.49 | 6.75 | 5.09 | 7.33 | 8.76 |
| Playtable | 2.20 | 3.11 | 5.78 | 3.78 | 3.58 | 9.74 | 4.40 | 22.40 |
| PlaytableP | 1.65 | 3.03 | 2.97 | 3.31 | 3.51 | 2.70 | 4.43 | 3.47 |
| Recycle | 1.30 | 2.48 | 1.95 | 1.83 | 2.97 | 2.91 | 3.73 | 6.93 |
| Shelves | 5.51 | 4.41 | 6.26 | 7.02 | 7.39 | 4.64 | 11.10 | 8.26 |
| Teddy | 1.15 | 1.07 | 1.12 | 2.00 | 2.51 | 1.80 | 3.44 | 2.29 |
| Vintage | 3.73 | 14.80 | 9.16 | 14.30 | 8.08 | 26.10 | 8.07 | 13.10 |
| **Average** | **2.30** | **3.81** | **3.87** | **4.63** | **6.05** | **7.25** | **9.60** | **9.66** |

The qualitative comparison of the methods is illustrated in Figure 3. The ground truth image of PlayTable image is shown in Figure 3(a). The PlayTable image pair (illustrated as left and right image in Figure 3(b) and Figure 3(c) respectively) contains a low texture region in the floor area of images. Based on Figure 3(d), the proposed method produces a smooth disparity map with different depths in the region with low texture and plain colour. Qualitatively, the map produced is more accurate than the disparity map generated by other methods such as SGBMP, JMR, MC-CNN-WS, MC-CNN-acrt and LS_ELAS (illustrated in Figure 3(e), Figure 3(f), Figure 3(g), Figure 3(h) and Figure 3(i) respectively). This proposed method achieved the objective of this work, to reduce the error in the low texture region, as mentioned in the earlier section. The *All* error recorded for the PlayTable image is the best among other methods, as illustrated in Table 1.



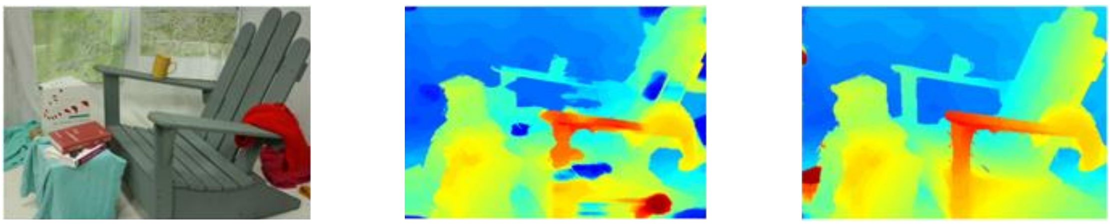
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 3. Middlebury - PlayTable image comparison, (a) ground truth, (b) left, (c) right, (d) proposed method, (e) SGBMP, (f) JMR, (g) MC-CNN-WS, (h) MC-CNN-acrt, (i) LS_ELAS

The effectiveness of combining the cost volumes $C_{DI}$ and $C_{CNN}$ into $C_{IM}$ in step 1 as discussed in section 2.1, is proven. The combined cost volumes, $C_{IM}$ helps reduce the error from 16.2% to 9.29% for All errors and from 14.4% to 6.05% for NonOcc error in the Middlebury online stereo benchmarking system. For qualitative comparison, Figure 4(a) depicts the left image of the Adirondack image. Figure 4(b) illustrated the disparity map for the Adirondack image when $C_{DI}$ was used without $C_{CNN}$. In addition, Figure 4(c) demonstrates the effect of combining both cost volumes ($C_{IM}$). Thus, this combination contributes to better accuracy generated through the proposed matching cost computation steps.



(a) (b) (c)

Figure 4. Middlebury-Adirondack image (a) left image, (b) without proposed method, and (c) proposed method

## 4. CONCLUSION

In conclusion, we demonstrate the hybrid learning-based method combined with a handcrafted method to perform matching cost computation. The initial cost volume created may contain noises, so the cost aggregation step was implemented to reduce errors. BF has been used as an edge-preserving filter to maintain the sharp edge while smoothing the input. WTA was employed to compute the initial disparity map. The final disparity map generated through post-processing steps includes LRC, hole filling, GIF, and WMF. Based on the final disparity map, we conclude that the proposed algorithm enhances the disparity map's accuracy in the low texture region while maintaining the object edge. Furthermore, the proposed algorithm can perform competitively compared to other published methods based on the Middlebury stereo benchmarking system.

## ACKNOWLEDGEMENTS

This work is supported by the Centre for Research and Innovation Management (CRIM), Universiti Teknikal Malaysia Melaka (UTeM).

## REFERENCES

[1] T. Xue, A. Owens, D. Scharstein, M. Goesele, and R. Szeliski, "Multi-frame stereo matching with edges, planes, and superpixels," *Image Vis. Comput.*, 2019, doi: 10.1016/j.imavis.2019.05.006.
[2] E. Winarno, I. H. Al Amin, and W. Hadikurniawati, "Asymmetrical half-join method on dual vision face recognition," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 6, pp. 3411-3420, 2017, doi: 10.11591/ijece.v7i6.pp3411-3420.
[3] H. Ham, J. Wesley, and Hendra, "Computer vision based 3D reconstruction : A review," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 4, pp. 2394-2402, 2019, doi: 10.11591/ijece.v9i4.pp2394-2402.
[4] C. Strecha, W. V. Hansen, L. V. Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008, doi: 10.1109/CVPR.2008.4587706.
[5] G. Yang, J. Manela, M. Happold, and D. Ramanan, "Hierarchical Deep Stereo Matching on High-Resolution Images," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5510-5519, doi: 10.1109/CVPR.2019.00566.
[6] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo," *Int. J. Comput. Vis.*, vol. 47, no. 1, pp. 7-42, 2002, doi: 10.1109/SMBV.2001.988771.
[7] R. A. Hamzah and H. Ibrahim, "Literature survey on stereo vision disparity map algorithms," *J. Sensors*, vol. 2016, 2016, doi: 10.1155/2016/8742920.
[8] M. S. Hamid, N. A. Manap, R. A. Hamzah, and A. F. Kadmin, "Stereo matching algorithm based on deep learning: A survey," *J. King Saud Univ. - Comput. Inf. Sci.*, 2020, doi: 10.1016/j.jksuci.2020.08.011.
[9] N. A. Manap, S. F. Hussin, A. M. Darsono, and M. M. Ibrahim, "Performance Analysis on Stereo Matching Algorithms Based on Local and Global Methods for 3D Images Application," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 2, pp. 23-28, 2018.
[10] K. Zhang, Y. Fang, D. Min, L. Sun, S. Yang, and S. Yan, "Cross-Scale Cost Aggregation for Stereo Matching," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 5, pp. 965-976, 2017, doi: 10.1109/TCSVT.2015.2513663.
[11] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications," *Found. Trends Comput. Graph. Vis.*, vol. 4, no. 1, pp. 1-73, 2009, doi: 10.1561/0600000020.
[12] S. Zhu, Z. Wang, X. Zhang, and Y. Li, "Edge-preserving guided filtering based cost aggregation for stereo matching," *J. Vis. Commun. Image Represent.*, vol. 39, pp. 107-119, 2016, doi: 10.1016/j.jvcir.2016.05.012.
[13] P. Brandao, E. Mazomenos, and D. Stoyanov, "Widening siamese architectures for stereo matching," *Pattern Recognit. Lett.*, vol. 120, pp. 75-81, Apr. 2019, doi: 10.1016/j.patrec.2018.12.002.
[14] J. ˇZbontar and Y. LeCun, "Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches," *J. Mach. Learn. Res.*, vol. 17, pp. 1-32, 2016, doi: 10.1186/s13568-015-0106-7.
[15] S. Wen, "Convolutional neural network and adaptive guided image filter based stereo matching," in *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2017, no. 1, pp. 1-6, doi: 10.1109/IST.2017.8261530.
[16] H. Xu, "Stereo matching and depth map collection algorithm based on deep learning," in *IST 2017 - IEEE International Conference on Imaging Systems and Techniques, Proceedings*, vol. 2018-Janua, no. 1, 2018, pp. 1-6, doi: 10.1109/IST.2017.8261504.
[17] T. Q. Vinh, L. H. Duy, and N. T. Nhan, "Vietnamese handwritten character recognition using convolutional neural network," *Indonesian Journal of Applied Informatics (IJAI)*, vol. 9, no. 2, pp. 276-283, 2020, doi: 10.11591/ijai.v9.i2.pp276-283.
[18] D. Ciresan, U. Meier, J. Schmidhuber, D. Cires, and U. Meier, "Multi-column Deep Neural Networks for Image Classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, no. February, pp. 3642-3649, doi: 10.1109/CVPR.2012.6248110.
[19] D. Scharstein *et al.*, "High-resolution stereo datasets with subpixel-accurate ground truth," in *German Conference on Pattern Recognition*, 2014, doi: 10.1007/978-3-319-11752-2_3.
[20] Z. Chen, X. Sun, L. Wang, Y. Yu, and C. Huang, "A Deep Visual Correspondence Embedding Model," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1-9, doi: 10.1109/ICCV.2015.117.
[21] A. Kendall *et al.*, "End-to-End Learning of Geometry and Context for Deep Stereo Regression," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-Octob, pp. 66-75, doi: 10.1109/ICCV.2017.17.
[22] J.-R. Chang and Y.-S. Chen, "Pyramid Stereo Matching Network," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, doi: 10.1109/CVPR.2017.730.
[23] Y. Hu, W. Zhen, and S. Scherer, "Deep-Learning Assisted High-Resolution Binocular Stereo Depth Reconstruction," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 8637-8643, doi: 10.1109/ICRA40945.2020.9196655.

[24] P. Knöbelreiter, C. Reinbacher, A. Shekhovtsov, and T. Pock, "End-to-end training of hybrid CNN-CRF models for stereo," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, 2017, pp. 1456-1465, doi: 10.1109/CVPR.2017.159.

[25] X. Song, X. Zhao, H. Hu, and L. Fang, "EdgeStereo: A Context Integrated Residual Pyramid Network for Stereo Matching," *Lect. Notes Comput. Sci.*, vol. 11365, pp. 20-35, 2019, doi: 10.1007/978-3-030-20873-8_2.

[26] S. Zhu, H. Xu, and L. Yan, "A Stereo Matching and Depth Map Acquisition Algorithm Based on Deep Learning and Improved Winner Takes All-Dynamic Programming," *IEEE Access*, vol. 7, pp. 74625-74639, 2019, doi: 10.1109/ACCESS.2019.2921395.

[27] J. Kang, L. Chen, F. Deng, and C. Heipke, "Context pyramidal network for stereo matching regularized by disparity gradients," *ISPRS J. Photogramm. Remote Sens.*, vol. 157, no. March, pp. 201-215, 2019, doi: 10.1016/j.isprsjprs.2019.09.012.

[28] F. Li, Q. Li, T. Zhang, Y. Niu, and G. Shi, "Depth acquisition with the combination of structured light and deep learning stereo matching," *Signal Process. Image Commun.*, vol. 75, no. April, pp. 111-117, 2019, doi: 10.1016/j.image.2019.04.001.

[29] M. S. Hamid, N. A. Manap, R. A. Hamzah, and A. F. Kadmin, "Converged classification network for matching cost computation," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 6, 2020.

[30] S. Tulyakov, A. Ivanov, and F. Fleuret, "Weakly Supervised Learning of Deep Metrics for Stereo Reconstruction," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, 2017, pp. 1348-1357, doi: 10.1109/ICCV.2017.150.

[31] R. A. Jellal, M. Lange, B. Wassermann, A. Schilling, and A. Zell, "LS-ELAS: Line segment based efficient large scale stereo matching," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 146-152, doi: 10.1109/ICRA.2017.7989019.


## BIOGRAPHIES OF AUTHORS

**Mohd Saad Hamid** received his B. Eng degree majoring in Computer from Multimedia University. He completed his M. Eng majoring in Computer and Communication from Universiti Kebangsaan Malaysia in 2014. Currently, he is a lecturer at Universiti Teknikal Malaysia Melaka. His research interests are computer vision, deep learning, image processing and embedded systems. He can be contacted at email: mohdsaad@utem.edu.my.

**Nurulfajar Abd Manap** earned a Master of Electrical Engineering in image processing (2002, Universiti Teknologi Malaysia) and Bachelor of Electrical Engineering (2000, Universiti Teknologi Malaysia). He obtained his PhD (Image and Video Processing) at the University of Strathclyde, Glasgow in 2012. At present, he teaches various electronics engineering subjects and courses in his affiliation for undergraduate and postgraduate levels. His subjects include Advanced Digital Signal Processing, Computer Vision and Pattern Recognition, Data Structures, Multimedia Technology & Applications and Distributed & High Performance Computing. His research interests are 3D image processing, stereo vision and video processing. He also one of MIET Chartered Engineer and Apple Distinguished Educators (ADE). He can be contacted at email: nurulfajar@utem.edu.my.

**Rostam Affendi Hamzah** 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 from the Universiti Sains Malaysia in 2010. In 2017, he received PhD majoring in Electronic Imaging from Universiti Sains Malaysia. Currently, he is a senior lecturer in the Universiti Teknikal Malaysia Melaka teaching Digital Electronics, Digital Image Processing and Embedded Systems. His research interests are computer vision, pattern recognition and digital image processing. He can be contacted at email: rostamaffendi@utem.edu.my.

**Ahmad Fauzan Kadmin** graduated with a Bachelor Degree in Electronics Engineering from Universiti Sains Malaysia and Master Degree in Computer & Communication Engineering from Universiti Kebangsaan Malaysia. He has over 14 years of experience in the electronic & computer engineering field with technical expert in R&D engineering, computer vision & medical electronics. He is also one of the MIET Chartered Engineers. He can be contacted at email: fauzan@utem.edu.my.

**Shamsul Fakhar Abd Gani** graduated from Universiti Malaysia Perlis (UniMAP) in Bachelor of Engineering (Computer Engineering) with honours in 2006 and later received his Master's degree in Internet & Web Computing in 2015 from Royal Melbourne Institute of Technology (RMIT) Australia. He started his career as an R&D electronic engineer specializing in software design for meter cluster development in Siemens VDO Automotive. Shamsul is now a lecturer in the electronic and computer engineering technology department of FTKEE UTeM. He can be contacted at email: shamsulfakhar@utem.edu.my.

**Adi Irwan Herman** graduated in 2015 with a Bachelor Degree in Computer Engineering Technology (Computer Systems) from the Universiti Teknikal Malaysia Melaka. Currently, he has been working with Texas Instrument for more than five years, with his current research interests being computer engineering-related fields of studies. He can be contacted at email: adiirwanherman@gmail.com.