CONVERGED CLASSIFICATION NETWORK FOR MATCHING COST COMPUTATION

[1]Mohd Saad Hamid, [2]Nurulfajar Abd Manap, [3]RostamAffendi Hamzah, [4]Ahmad FauzanKadmin

[1]Faculty of Electrical and Electronic Engineering Technology, UniversitiTeknikal Malaysia Melaka,Durian Tunggal, Malaysia. Email: mohdsaad@utem.edu.my
[2]Faculty of Electronics and Computer Engineering, UniversitiTeknikal Malaysia Melaka,Durian Tunggal, Malaysia. Email: nurulfajar@utem.edu.my
[3]Faculty of Electrical and Electronic Engineering Technology, UniversitiTeknikal Malaysia Melaka,Durian Tunggal, Malaysia. Email: rostamaffendi@utem.edu.my
[4]Faculty of Electrical and Electronic Engineering Technology, UniversitiTeknikal Malaysia Melaka,Durian Tunggal, Malaysia. Email: fauzan@utem.edu.my

***Abstract*:**
Stereoscopic vision lets us identify the world around us in 3D by incorporating data from depth signals into a clear visual model of the world. The stereo matching algorithm capable of producing the disparity or depth map in computer. This map is crucial for many applications such as 3D reconstruction, robotics and autonomous driving.The disparity map also prone to errors such as noises in the region which contains object occlusions, reflective regions, and repetitive patterns.So we propose this stereo matching algorithm to produce a disparity map and to reduce the errors by incorporating a deep learning approach. This paper focused on matching cost computation step as an initial step to produce the disparity or depth map. The proposed convolutional neural network designed with the output neurons in the classification part scaled-downin converging style. The raw cost generated aggregated by the normalized box filter. Then the disparity map computed using Winner Take All approach. The final disparity map refined using Weighted Median Filter. Overall quantitative results for the proposed work performed competitively compared to other established stereo matching algorithm based on the Middlebury standard benchmark online system.

Keywords* :convolutional neural network; matching cost computation, stereo matching*

I. INTRODUCTION

The study on stereo vision mainly focused on stereo matching[1]. Stereo matching remains as a challenging area in computer vision to this day ([2]–[4]). A stereo matching algorithm capable of producing disparity or depth map. In the development of the stereo matching algorithm, the main objective is to find the disparity value calculated based on the object in left and right image pairs. Disparity value obtained based on the differences in the pixel location of corresponding features recorded in the left and right image. The distance between camera and object revealed through the depth map [5]. The importance of depth perception by stereo matching also highlighted in [6].According to[6], the depth information enables us to use for multiple application such as scene reconstruction, virtual and augmented reality, obstacle avoidance and several other applications. The authors of[7] perform stereo matching for 3D face reconstruction. They used the spatial-temporal integral image (STII) for faster matching cost computation in stereo matching for the reconstruction process. According to [8] and [9], the information from the disparity map will provide a further study on 3D projective transformation. However, image noises and repetitive textures influenced the quality of the disparity map and led to inaccurate disparity map produced[10].

Another camera-based method to obtain depth-related information includes the monocular based method([11]–[13]). One of the great work on monocular depth estimation is done by [11]. The authors implement unsupervised learning in their monocular vision work. Their work also highlighted and compared in [6]. However, based on the results obtained by [6], the monocular depth method found to be less accurate as compared to their stereo-based method. Their stereo method capable of recovering more objects of interest as compared to the monocular vision method by [11]. The main caused highlighted by [6] was the differences in the size of the information. The multi-view processing can provide more information than a single view method.

Another related method, Light Detection and Ranging (LiDAR) is one of the common methods for outdoor range sensing. It is a laser-based measurement technique and different from the camera-based technique (monocular and

ISSN: 2005-4238 IJAST

stereo) for depth sensing. As highlighted by [14] LiDAR capable of producing 3D points accurately, but it is not very cost-effective and time-consuming. Moreover, this issue also mentioned in [15]. LiDAR also commonly used to produce training dataset, as mentioned in [16]. The demerits on the monocular and LiDAR methods motivate us to proceed with the work on stereo vision.

As mentioned in [17], both local and global approach are the main categories in stereo vision algorithm. In the local method, disparity computation at any point in the image determined by the intensity values within a predefined support window. Because of this behaviour, the local method able to run faster with low computational complexity [17]–[20],[19],[20]. The global method is another interesting topic on stereo matching. According to [21], the global method produces disparity based on energy minimisation process, commonly based on Markov Random Field (MRF). The global method provides better accuracy for the disparity output. However, it will incur more computational complexity [22]. Energy minimisation was the critical function in a global method.The energy minimisation in global method focuses on data term and smoothness term [23].

The previous study on stereo vision algorithm by [23], [17], [10], [24], [19]mentioned there are four main steps to produce disparity map from the stereo based algorithm. The matching step is the most crucial in stereo vision[25]. The summary of the general steps in stereo vision algorithm mentioned in [23]and [17] illustrated in Fig.1.

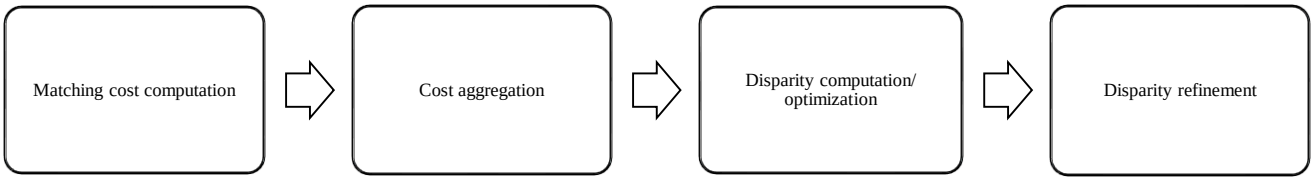


Fig.1.Stereo Vision Algorithm Steps

Szeliski and Scharstein (2002) in[23] also described that different algorithms might employ different step sequence combination. The aggregation step was often skipped in global approach due to the redundant purpose of global smoothness constraint when it performs the optimization step after the disparity computation step[23]. In general, the main steps for stereo vision algorithm as follows:

- Matching Cost Computation:

This step involves the calculation of the cost of assigning a special disparity to each pixel[25]. The examplesof traditional handcraft method for this step are Sum of Absolute Difference algorithm (SAD), Sum of Squared Difference (SSD), Normalized Cross-Correlation (NCC), Zero Mean Normalized Cross-Correlation (ZNCC), rank transform and census transform (CT) as explained in [23] and [17].

- Cost Aggregation

Mostly related to the local approach, the cost aggregation step involves summing or averaging over a support region in the disparity space image (DSI)[26]. The mainobjective of this step is to reduce noise in the matching cost. [3] aggregate the cost using low pass filters such as box filter and Gaussian filter. Another type of filter used for the purpose is the edge-preserving filters such as bilateral filter (BF), guided filter (GF), which preserve good edge and better results in the aggregation process ([3], [17], [26]). Other variations of the GF based filter used also mentioned in [27] and [28].

- Disparity Computation/Optimization

This step responsible to assign the disparity map value ([17], [19], [29],[30],[20]). The commonly used method used for this step is Winner Take All (WTA) optimization ([31]–[38]). In WTA, the disparity associated with the lowest cost value chosen at each pixel location ([17], [39]). Other examples for the optimization stage are dynamic programming (DP), simulated annealing (SA), scanline optimization (SO) and graph cut (GC) as mentioned by Scharstein and Szeliski (2002) in [23].

- Disparity Refinement

After the third step, the generated disparity or depth map may contain noises, errors such as invalid matches, and occlusion. Some of the methods implement slanted plane smoothing for the occlusion problem ([40], [41]). Left-to-right-consistency check (LRC) used to detect invalid pixels ([3], [17]). At this step, multiple filtering techniques used to refine the output before producing the final disparity map. Median filter techniques are commonly used for local refinement as implemented in [34]–[37], [42]–[45]. This last step also might introduce extra timing to the overall process due to its complexity.

As mentioned in [46], artificial intelligence AI has become a concentrated topic in exaggerated publicity by the mass media. One of the interesting area in artificial intelligence area these days is machine learning. The beauty of the machine learning algorithm is the capability to instruct the computer to react or deciding on particular condition without having to program the computer explicitly. Due to the flexibility of the machine learning algorithm to learn via self-train, analysis, observation and experience. Machine learning algorithm capable of adapting new situation through pattern and trend detection for better results.

In recent years,deep learning as a subset of machine learning has become the catalyst to evolve the stereo vision area. Deep learning implementation has boosted the performance of stereo vision application as described by [47]. Furthermore, as mentioned in [48] and [49], the traditional stereo vison cannot match the human performance for recognition tasks. However, with the assimilation of deep learning into their algorithm, they can match human performance. Since then, researchers around the globe have been working to refine and implement deep learning in real-world stereo visionapplication. Concerning the stereo matching works, deep learning also has been applied in two ways. Some researcher mix convolutional neural network (CNN) with traditional handcraft algorithm and some also proposed a new end to end CNN based networks ([6], [15], [50]–[56]).

We propose a new CNN based method matching cost computation. Inspired by [31], we use the base architecture of MC-CNN-acrt as in [31] where it has been used by other researchers to calculate matching cost computation for their stereo matching algorithm ([10], [33], [36], [37], [45], [57]).

As published by the authors of [42], their work inspired many other researchers on the usage of deep learning for matching cost computation and other stereo matching steps. Their MC-CNN-acrt architecture outperforms other stereo matching conventional methods. The MC-CNN-acrt contains several layers combined as a Siamese network. As illustrated in **Fig. 2**, L1 is the convolutional layer (with 32 kernels of 5x5) and L2-L8 made of fully connected(FC) layers. L2 and L3, the FC layers contain 200 neurons each. Left and right feature output from L3 concatenated together to form a single 400-dimensional vector. L4-L7 contains 300 neurons each. Zbontar&LeCunmould the matching cost computation as a binary classification problem at L8 with two neurons produces good and bad matches. They verified the effectiveness of using CNN for matching cost computation by replacing it with other methods such as SAD, CT, and NCC for Middlebury and KITTI (2012 & 2015) benchmark. Their method outperforms different pure handcraft algorithm.

Chen et al.[58]proposed another method of matching cost computation. In contrast to [42], the authors compute similarity in the Euclidean space using a dot product. The process enables calculation time faster than MC-CNN. However, the accuracy obtained less than the MC-CNN. The employed inner product of which indicate matching score that will be large in case of a similar patch and vice versa.

Chen & Yuan in[32]highlighted the issues of using fix size patches and equal weight. According to them, using a larger patch, however, will lead to better accuracy in the texture-less region with ahigher computational cost. Following [42], they modified the network by adding several CNN subnetwork for cost computation. They found that having a multi-scale approach in their network produce better output accuracy and also maintain efficiency in term of test time. Based on their output also, it is proven that a larger input patch producesa more accurate result as compared to a smaller patch.

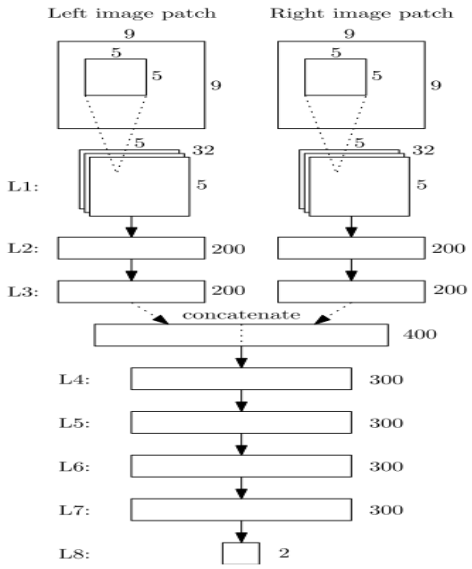


Fig. 2.MC-CNN-acrt architecture [31]

Wen in[36] stated that the main reason to implemented CNN is because of the feature extraction capabilities available in CNN. They apply Siamese network as described in [59] and also performed in [31], [42]. It contains two branches that share the same architecture. The image patches from left and rightsent to the network in two different branches. Left and right image patches supplied to the three convolution layers per branch where for each layer, the ReLUactivation function equipped for each convolution process. They cast the problem as a binary classification problem where the final output of the overall network is binary output 0 or 1 to indicate the similarity between input left and right.

This paper focuses on the effect of converging the classification part of the architecture to calculate the matching cost. This paper arranged as follows. The next section will describe the proposed method. Section III will explain the results and the discussions based on Middlebury version 3 [60] stereo benchmarking evaluation system. Section IV will conclude the paper with the overall conclusions of the work.

II. THE PROPOSED ALGORITHM

The four stages of our proposed algorithm as follows:

A. Matching Cost Computation

Our approach on the matching cost computation inspired by the original works of [31], [42]. We use architecture mentioned in the MC-CNN-acrt, as illustrated in Fig. 2as the baseline for this work. Our main focus is on the classification part of this Siamese base network (refer **Fig. 2**). We redesigned the classification part of the baseline architecture by scaling down the output neurons in a fully connected layer in a converging manner, as illustrated in**Fig. 3**. The output neuron for layer 5 (FC5) reduced to 300 neurons. The next layer FC6 set to 250 neurons, and the FC7 will produce 100 neurons. A similar approach to the original baseline, the output neurons for last layer 8 (FC8) set to 2 for binary similarity classification. All layers except the input layer and final layer equipped with Rectified Linear Unit (ReLU) activation function.

The output of the network will provide a similarity score, which provides binary classification. Softmax function has been used as the activation function in the last layer to provide good and bad matches. It will return the value of good and bad matching.

$$\mathrm{C_{CNN}}(\mathrm{p}, d) \;=\; -\mathrm{s}(\, P_N^L(\mathrm{p}), P_N^R(\mathrm{p} \;-\; d)\,) \qquad (1)$$

The matching cost value $\mathrm{C_{CNN}}(\mathrm{p}, d)$ represent the matching cost at each position p for all disparities, d. It comes from the output of the CNN network where the inputs coming from $N$by$N$ left image input patch, P_N^L and N *by* N right image input patch with, P_N^R. The minus sign will convert the similarity score to the matching cost.

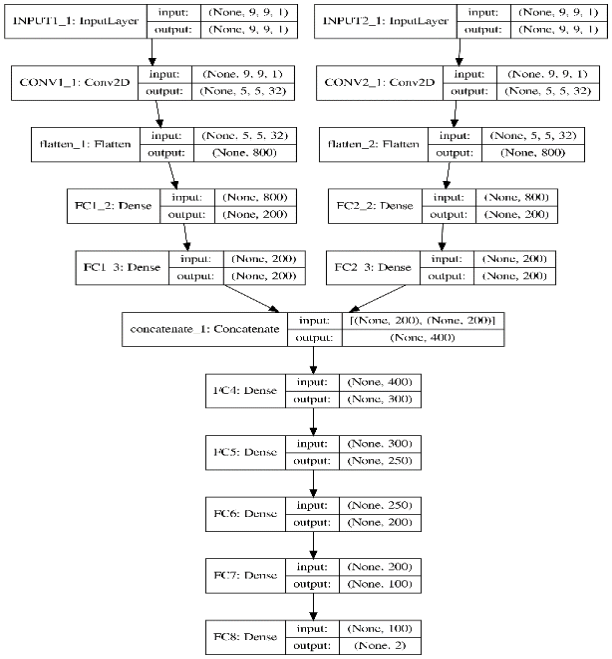


Fig. 3.The Proposed Architecture

B. Cost Aggregation

In order to produce a more accurate disparity map, we need to refine the raw matching cost. Inspired by [61]we implement cost aggregation steps using normalised box filter from the OpenCV library[62]. The output of the cost aggregation step labeled as $C_{AGG}(p,d)$. C_{AGG}(p,d) is the cost volume after aggregation,

$$C_{AGG} = \sum_{k,l} C_{CNN}(p,d)h(k,l) \qquad (2)$$

The kernel, $h(k,l)$is the normalised box filter kernel of 5x5 size applied to the initial matching cost computed, C_{CNN}.

C. Disparity Computation / Optimization

For the disparity computation step, we computed the final disparity map by using WTA optimization. We employ WTA optimization as per the following equation.

$$d = arg\min_{d \in d_r}(C_{AGG}(p,d)) \qquad (3)$$

The maximum disparity value of $d_r$obtained from the ground truth of data. As mentioned earlier, in the WTA approach, disparity associated with the lowest cost value chosen at each pixel location

D. Disparity Refinement

For the disparity refinement step, multiple processing steps involved. Inspired by the application of the Weighted Median(WM) filter in [61], we apply the WM filter to remove outliers while maintaining the edges. Firstly the initial disparitymap applied with Weighted Median Filter (WMF) from [62]. Next, we used LRC to detect invalid pixels. After that,we applied WMF again to help refine the final disparity map. We did not implement the exact post-processing method implemented in [31] for both baseline (in [31]) and the proposed method.

III. EXPERIMENTAL RESULTS

The implementation of the architecture to produce a disparity map done using Python, Keras, TensorFlow, and OpenCV library running on top of Ubuntu Linux 18.04. We executed the code on a hardware platform of consists of a personal computer with Intel Core i5 3.0 GHz processor and equipped with 16GB DDR3 RAM and GPU NVidia GTX1060 3GB. The inference process of the CNN network performed using GPU for matching cost calculation while the other stage 2 until 4 of the algorithm run using CPU. We train the network using the Adam optimization algorithm, with the learning rate set to 0.0001. Batch is size set to 128 and trained for 50 epochs.

The inference process performed to calculate matching costs using the stereoimages from Middlebury online benchmarking system.The Middlebury v3 [60] dataset contains 30 sets of stereo images (15 training, images, and 15 testing images). These training images were developed to determine theperformance of an algorithm and can be uploaded several times. However, the testing imagesare only for the final evaluation.

We evaluated the proposed method using Middlebury v3 stereo benchmark system to test the accuracy of the disparity map generated using proposed architecture. We compared the proposed architecture with the original baseline architecture published in [31]. Both architectures have been applied with the same process to show the effectiveness of the architecture for matching costs. Here we provide qualitative as well as the quantitative result for the comparison. Refer toTable Iand

Table II, which contains the result from the Middlebury benchmark system [60] to compare the effectiveness of the proposed architecture. The table contains results obtained for *NonOcc*error (error of invalid disparity values in non-occluded pixel) and *All* error (error of invalid disparity values in all pixels).

Based on the quantitative result obtained in Table I and

Table II, the proposed architecture outperform the original baseline architecture. For the error of invalid disparity values in all pixels, the proposed architecture performed better than the baseline for all 15 images. For the error of invalid disparity values in non-occluded pixels, there is a significant improvement shown for 11 out of 15 images as compared to the baseline architecture.

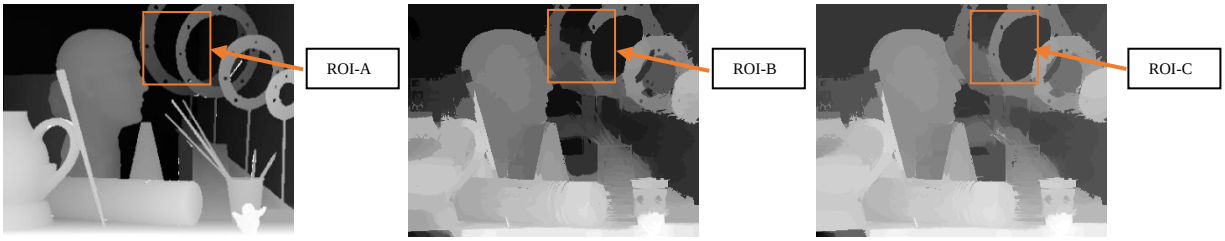
ISSN: 2005-4238 IJAST

Table I Quantitative Comparison Between Proposed and Baseline – Middlebury Benchmark – NonOcc Error

Images	Adirondack	ArtL	Jadeplant	Motorcycle	MotorcycleE	Piano	PianoL	Pipes	Playroom	Playtable	PlaytableP	Recycle	Shelves	Teddy	Vintage	Weight
Proposed	5.8	8.88	30.9	7.04	6.64	7.32	6.81	10.6	8.27	5.31	5.27	5.33	10.2	4.36	11.1	9.04
Baseline	5.78	8.96	33.5	7.77	7.5	7.16	7.14	10.3	8.32	5.31	5.27	4.98	11.2	4.38	24.5	9.91
MC-CNN-acrt[42]	0.76	2.49	16.3	1.27	1.27	1.83	5.07	2.29	2.27	3.11	3.03	2.48	4.41	1.07	14.8	3.81
MC-CNN-fst[42]	1.21	2.84	10	1.62	1.61	3.17	13.2	3.2	3.13	5.78	2.97	1.95	6.26	1.12	9.16	3.87
PSMNet_ROB[52]	7.32	9.69	44.5	5.55	6.12	5.01	9.82	9.86	7.33	4.4	4.43	3.73	11.1	3.44	8.07	9.6
MC-CNN-WS[63]	1.66	4.27	12.8	2.26	2.18	3.21	11.7	4.27	3.49	3.78	3.31	1.83	7.02	2	14.3	4.63
LS_ELAS [64]	8.46	3.83	41.1	5.12	5.8	5.54	8.97	7.44	8.76	22.4	3.47	6.93	8.26	2.29	13.1	9.66

Table IIQuantitative Comparison Between Proposed and Baseline – Middlebury Benchmark – All Error

Images	Adirondack	ArtL	Jadeplant	Motorcycle	MotorcycleE	Piano	PianoL	Pipes	Playroom	Playtable	PlaytableP	Recycle	Shelves	Teddy	Vintage	Weight
Proposed	8.18	11.8	50.1	9.66	9.43	8.39	7.47	15.7	11	6.48	6.19	5.93	11.6	5.33	13.2	12.4
Baseline	8.25	12	53.8	11.5	11.3	8.64	7.92	15.5	13.2	7.21	7	5.62	12.7	6.77	24.8	13.9
MC-CNN-acrt[42]	4.24	18.7	34.1	7.21	7.22	6	9.35	13.5	18.3	9.71	9.37	4.64	6.62	9.31	21.6	11.8
MC-CNN-fst[42]	5.32	19.2	32.6	8.75	8.83	8.12	17.2	15.5	18.6	13.6	9.75	5	8.91	10.4	15.8	12.8
PSMNet_ROB[52]	8.83	13.9	68.4	8.26	9.16	5.89	10.5	14.4	9.38	5.54	5.52	4.98	11.6	3.87	9.66	13.3
MC-CNN-WS[63]	5.73	20.5	36.3	9.39	9.37	8.13	16.1	16.7	18.7	11.5	10.1	5.05	9.83	11	20.8	13.7
LS_ELAS[64]	9.31	5.9	64.5	7.24	7.65	6.25	9.69	12.8	10.1	23.9	4.27	7.39	8.48	2.98	14	12.9



(a)Ground Truth Image *(b) Output from Baseline* *(c) Output from Proposed*

Fig. 4.Comparison Disparity Output with the Ground Truth from Middlebury- ArtL Image

For a qualitative comparison between the proposed and baseline architecture, we present in the following Fig. 4. The figure contains a comparison of the disparity output between ground truth, baseline, and the proposed method. The qualitative comparison focused on the region of interest (ROI), as highlighted in the orange line box. The table also highlighted the error in baseline method output, as highlighted in ROI-B. The proposed architecture capable of reducing the error in ArtL image in the highlighted ROI-C.

Both Table I and II also contain a performance comparison with other algorithms. The overall performance in terms of *NonOcc*error, the proposed architecture performed better thanLS_ELAS[64] and PSMNet_ROB[52]. As for *All* error, the proposed method produces more accurate results than LS_ELAS[64], MC-CNN-fst[42], PSMNet_ROB[52], and MC-CNN-WS [63]. It shows that the proposed architecture is competitive with other stereo matching algorithms.

IV. CONCLUSIONS

In this work, we focus on improving the architecture of original work by [31]. The result shows that the proposed method which based on converged classification on CNN for matching cost computation steps and equipped with other post-processing steps capable of producing comparable results to other established methods. Even though it performs lower than the results of the final MC-CNN-acrt algorithm in the Middlebury benchmark standard, but it proves that the proposed method at the classification part improves overall results. Furthermore, it performs better than the original baseline with simple stereo matching post-processing steps applied. Overall results of the proposed method show competitive results compared to other stereo matching algorithms. We hope to further advance the architecture with different dilation to the CNN layer and equip it with better stereo matching steps. We also want to enhance the cost aggregation and disparity refinement steps using a new image enhancement method in the future.

V. ACKNOWLEDGEMENTS

This work supported by the Ministry of Education (MOE), Malaysia, and sponsored byUniversitiTeknikal Malaysia Melaka under grant number: JURNAL/2018/FTK/Q00008.

REFERENCES

1. S. Damjanović, F. van der Heijden, L. J. Spreeuwers, F. Van Der Heijden, and S. Group, “Local Stereo Matching Using Adaptive Local Segmentation,” ISRN Mach. Vis., vol. 2012, pp. 1–11, 2012.
2. T. Xue, A. Owens, D. Scharstein, M. Goesele, and R. Szeliski, “Multi-frame stereo matching with edges, planes, and superpixels,” Image Vis. Comput., 2019.
3. 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.
4. J. Pang, W. Sun, J. S. J. Ren, C. Yang, and Q. Yan, “Cascade Residual Learning: A Two-Stage Convolutional Neural Network for Stereo Matching,” in Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017, 2018, vol. 2018-Janua, pp. 878–886.
5. J. Fu and J. Liang, “Virtual view generation based on 3D-dense-attentive GAN networks,” Sensors (Switzerland), vol. 19, no. 2, 2019.
6. N. Smolyanskiy, A. Kamenev, and S. Birchfield, “On the importance of stereo for accurate depth estimation: An efficient semi-supervised deep neural network approach,” IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2018-June, pp. 1120–1128, 2018.
7. K. Fu, Y. Xie, H. Jing, and J. Zhu, “Fast spatial–temporal stereo matching for 3D face reconstruction under speckle pattern projection,” Image Vis. Comput., vol. 85, pp. 36–45, 2019.
8. N. A. Manap, G. Di Caterina, J. Soraghan, V. Sidharth, and H. Yao, “Smart surveillance system based on stereo matching algorithms with IP and PTZ cameras,” 3DTV-CON 2010 True Vis. - Capture, Transm. Disp. 3D Video, pp. 4–7, 2010.
9. X. Ma, S. Wang, W. Liu, F. Ma, A. Wang, Y. Sheng, Y. Li, and H. Ming, “Optimized stereo matching algorithm for

integral imaging microscopy and its potential use in precise 3-D optical manipulation," Opt. Commun., vol. 430, no. May 2018, pp. 374–379, 2019.
10. G. Song, H. Zheng, Q. Wang, and Z. Su, "Training a convolutional neural network for disparity optimization in stereo matching," in 2017 International Conference on Computational Biology and Bioinformatics, ICCBB 2017, 2017, pp. 48–52.
11. C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017.
12. D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in Advances in Neural Information Processing Systems, 2014.
13. R. P. Padhy, S. Verma, S. Ahmad, S. K. Choudhury, and P. K. Sa, "Deep Neural Network for Autonomous UAV Navigation in Indoor Corridor Environments," Procedia Comput. Sci., vol. 133, pp. 643–650, 2018.
14. C. Strecha, W. Von Hansen, L. Van 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.
15. G. Yang, J. Manela, M. Happold, and D. Ramanan, "Hierarchical Deep Stereo Matching on High-resolution Images," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
16. Y. Kuznietsov, J. Stückler, and B. Leibe, "Semi-supervised deep learning for monocular depth map prediction," in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017.
17. R. A. Hamzah and H. Ibrahim, "Literature survey on stereo vision disparity map algorithms," J. Sensors, vol. 2016, 2016.
18. G. Popovi, A. Hadviger, I. Markovi, and I. Petrovi, "Computationally efficient dense moving object detection based on reduced space disparity estimation," in International Federation of Automatic Control, 2018, vol. 22, pp. 360–365.
19. N. Manap, S. Hussin, … A. D.-J. of, and U. 2018, "Performance Analysis on Stereo Matching Algorithms Based on Local and Global Methods for 3D Images Application," Journal.Utem.Edu.My, vol. 10, no. 2, pp. 23–28, 2018.
20. O. Zeglazi, M. Rziza, A. Amine, and C. Demonceaux, "A hierarchical stereo matching algorithm based on adaptive support region aggregation method," Pattern Recognit. Lett., vol. 112, pp. 205–211, 2018.
21. Z. Wang, S. Zhu, Y. Li, and Z. Cui, "Convolutional neural network based deep conditional random fields for stereo matching," J. Vis. Commun. Image Represent., vol. 40, no. Part B, pp. 739–750, 2016.
22. C. S. Panchal and A. B. Upadhyay, "Depth Estimation Analysis Using Sum of Absolute Difference Algorithm," Int. J. Adv. Res. Electr., vol. 3, no. 1, pp. 6761–6767, 2014.
23. 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.
24. X. Huang and Y. J. Zhang, "An O(1) disparity refinement method for stereo matching," Pattern Recognit., vol. 55, pp. 198–206, 2016.
25. B. Salehian, A. M. Fotouhi, and A. A. Raie, "Dynamic programming-based dense stereo matching improvement using an efficient search space reduction technique," Optik (Stuttg)., vol. 160, pp. 1–12, 2018.
26. Y. Xu, Y. Zhao, and M. Ji, "Local stereo matching with adaptive shape support window based cost aggregation," Appl. Opt., vol. 53, no. 29, p. 6885, 2014.
27. G. S. Hong and B. G. Kim, "A local stereo matching algorithm based on weighted guided image filtering for improving the generation of depth range images," Displays, vol. 49, pp. 80–87, 2017.
28. Williem and I. K. Park, "Cost aggregation benchmark for light field depth estimation," J. Vis. Commun. Image Represent., vol. 56, pp. 38–51, 2018.
29. L. F. S. Cambuim, J. P. F. Barbosa, and E. N. S. Barros, "Hardware module for low-resource and real-time stereo vision engine using semi-global matching approach," Proc. 30th Symp. Integr. Circuits Syst. Des. Chip Sands - SBCCI '17, pp. 53–58, 2017.
30. A. J. Malekabadi, M. Khojastehpour, and B. Emadi, "Comparison of block-based stereo and semi-global algorithm and effects of pre-processing and imaging parameters on tree disparity map," Sci. Hortic. (Amsterdam)., vol. 247, no. May 2018, pp. 264–274, 2019.
31. J. ˇZbontar and Y. LeCun, "Computing the Stereo Matching Cost with a Convolutional Neural Network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015), 2015, pp. 1592–1599.
32. J. Chen and C. Yuan, "Convolutional neural network using multi-scale information for stereo matching cost computation," Proc. - Int. Conf. Image Process. ICIP, vol. 2016-Augus, pp. 3424–3428, 2016.
33. A. Seki and M. Pollefeys, "Patch based confidence prediction for dense disparity map," Br. Mach. Vis. Conf. 2016, BMVC 2016, vol. 2016-Septe, no. c, pp. 23.1-23.13, 2016.
34. A. Seki and M. Pollefeys, "SGM-Nets: Semi-global matching with neural networks," in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, pp. 6640–6649.
35. M. Yang and X. Lv, "Learning both matching cost and smoothness constraint for stereo matching," Neurocomputing, vol. 314, pp. 234–241, 2018.
36. 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.
37. 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, 2018, vol. 2018-Janua, no. 1, pp. 1–6.
38. Z. Li and L. Yu, "Compare stereo patches using atrous convolutional neural networks," ICMR 2018 - Proc. 2018 ACM Int. Conf. Multimed. Retr., pp. 473–480, 2018.
39. R. Szeliski, Computer Vision : Algorithms and Applications. London: Springer London, 2011.

40. P. Brandao, E. Mazomenos, and D. Stoyanov, "Widening siamese architectures for stereo matching," Pattern Recognit. Lett., vol. 120, pp. 75–81, Apr. 2019.
41. W. Luo, A. G. Schwing, and R. Urtasun, "Efficient Deep Learning for Stereo Matching," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 5695–5703.
42. 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.
43. M. Yang, Y. Liu, and Z. You, "The Euclidean embedding learning based on convolutional neural network for stereo matching," Neurocomputing, vol. 267, pp. 195–200, 2017.
44. A. Shaked and L. Wolf, "Improved stereo matching with constant highway networks and reflective confidence learning," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, no. v, pp. 6901–6910, 2017.
45. S. Joung, S. Kim, B. Ham, and K. Sohn, "Unsupervised stereo matching using correspondence consistency," in 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 2518–2522.
46. F. Chollet, Deep Learning with Python, 1st ed. Greenwich, CT, USA: Manning Publications Co., 2017.
47. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012, no. December, pp. 1097–1105.
48. D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3642–3649.
49. D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," Neural Networks, vol. 32, pp. 333–338, 2012.
50. N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox, "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation Nikolaus," in 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4040–4048.
51. A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, R. Kennedy, A. Bachrach, and A. Bry, "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.
52. J.-R. Chang and Y.-S. Chen, "Pyramid Stereo Matching Network," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
53. G. Yang, H. Zhao, J. Shi, Z. Deng, and J. Jia, "SegStereo: Exploiting Semantic Information for Disparity Estimation," in European Conference on Computer Vision(ECCV), 2018, pp. 1–16.
54. X. Song, X. Zhao, H. Hu, and L. Fang, "EdgeStereo: A Context Integrated Residual Pyramid Network for Stereo Matching," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11365 LNCS, pp. 20–35, 2019.
55. K. Swami, K. Raghavan, N. Pelluri, R. Sarkar, and P. Bajpai, "DISCO: Depth Inference from Stereo using Context," in 2019 IEEE International Conference on Multimedia and Expo (ICME), 2019, pp. 502–507.
56. T. P. Nguyen and J. W. Jeon, "Wide context learning network for stereo matching," Signal Process. Image Commun., vol. 78, no. January, pp. 263–273, 2019.
57. S. Jeong, S. Kim, B. Ham, and K. Sohn, "Convolutional Cost Aggregation For Robust Stereo Matching," in 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 2523–2527.
58. 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, no. d, pp. 1–9.
59. S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, vol. 07-12-June, no. i, pp. 4353–4361.
60. D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8753, no. 1, pp. 31–42, 2014.
61. 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, 2013, pp. 49–56.
62. G. Bradski, "The OpenCV Library," Dr. Dobb's J. Softw. Tools, 2000.
63. 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, 2017.
64. R. A. Jellal, M. Lange, B. Wassermann, A. Schilling, and A. Zell, "LS-ELAS: Line segment based efficient large scale stereo matching," in Proceedings - IEEE International Conference on Robotics and Automation, 2017.

ISSN: 2005-4238 IJAST