International Journal of Computers and Applications

ISSN: 1206-212X (Print) 1925-7074 (Online) Journal homepage: http://www.tandfonline.com/loi/tjca20

Extracting depth information from stereo images using a fast correlation matching algorithm

Mozammel Chowdhury, Junbin Gao & Rafiqul Islam

To cite this article: Mozammel Chowdhury, Junbin Gao & Rafiqul Islam (2018): Extracting depth information from stereo images using a fast correlation matching algorithm, International Journal of Computers and Applications, DOI: 10.1080/1206212X.2018.1494895

To link to this article: https://doi.org/10.1080/1206212X.2018.1494895

Published online: 11 Jul 2018.

INTERNATIONAL JOURNAL OF COMPUTERS AND APPLICATIONS
https://doi.org/10.1080/1206212X.2018.1494895

Extracting depth information from stereo images using a fast correlation matching algorithm

Mozammel Chowdhury[a], Junbin Gao [b] and Rafiqul Islam[a]

[a]School of Computing and Mathematics, Charles Sturt University, Sydney, Australia; [b]Busines Analytics Discipline, School of Business, The University of Sydney, Sydney, Australia

ABSTRACT
Stereo matching algorithms are essential for recovering depth information of objects in many computer vision applications including 3D reconstruction, robot navigation, autonomous driving and so on. Most of the stereo algorithms generally rely on two types of matching technique: global and local matching. The state-of-the-art stereo algorithms that measure disparity or depth with high accuracy are generally based on global methods. However, they are not suitable for real-time applications because of high computational costs. The local algorithms, on the other hand, are very fast but they provide less computational accuracy compared to the global methods. To make a tradeoff between computation speed and accuracy, this paper proposes an efficient local correlation approach for depth estimation using a pruning proposal. This paper also evaluates the performance of different matching cost functions/algorithms for disparity or dense estimation. Experimental evaluation confirms that our proposed pruning method for point correspondence is able to achieve a significant accuracy with high computational speed that can be very useful for real-time environments.

ARTICLE HISTORY
Received 27 September 2017
Accepted 21 June 2018

KEYWORDS
Depth extraction; disparity; window cost function; stereo imaging

1. Introduction

Stereo algorithms have attracted great importance in the field of computer vision and robotics in order to acquire depth information of objects from a pair of stereo images, called left and right image sequence [1]. Stereo algorithms deal with two major problems: one is the correspondence matching, in which features corresponding to the same entities are to be matched across the image sequences and another is the reconstruction problem, in which stereo or 3D depth information is to be reconstructed from the correspondences. Fast and accurate correspondence matching is the key issue in depth estimation since many factors such as noise, distortion, reflection of light, lack of texture and occlusion may affect the correspondence matching [2].

Stereo algorithms are generally classified into two broad categories: local and global algorithms. The local algorithms [3–6] also referred to as window-based or area-based algorithms, determine disparity or dense depth by matching the similarity between the pixels within two local windows in the stereo image pair. These algorithms make cost aggregation and consider implicit smoothness assumption. Cost aggregation is conventionally done locally by summing matching cost within windows with constant disparity. On the other hand, global methods [7–11] perform disparity optimization on a global energy function defined over entire image pixels by simultaneously imposing a smoothness constraint.

Local methods are, typically, considered to be more straightforward and simple, and hence adequately fast for real-time applications, but of lower accuracy. However, global methods use elaborate models to describe the matching process which results in stereo disparity of high accuracy, but at the cost of a computational load which leads to be restrictive for real-time tasks or large data manipulation. Reducing high computation time in correspondence matching is still a vital issue for stereo algorithms. To overcome the limitations of the stereo algorithms, this paper proposes a fast and robust local method for dense depth measurement. Our algorithm employs a correlation measure characterized by local intensities for finding the disparity or depth by computing the similarity or matching costs between the pixels within two windows in the stereo image pair. The focal contributions or novelty of the proposed algorithm are summarized below:

- We integrate a pruning technique with the conventional SSD (sum of square differences) algorithm for estimating window costs or correspondence matching

CONTACT Mozammel Chowdhury mochowdhury@csu.edu.au

costs. Based on the pruning proposal, the proposed stereo matching algorithm computes the window cost for candidate pixels in the right image only whose intensities are different within a certain threshold value δ to the intensity of the pixel in the left image. This pruning technique in disparity or depth calculation reduces the computation load significantly.

- A fuzzy rule-based disparity refinement technique is incorporated, which is able to reduce false correspondence matching.
- We formulate the accurate dense depth equations from the camera geometry model.
- We evaluate the accuracy of different correlation matching cost functions.

The rest of this paper is organized as follows. Section 1 demonstrates disparity and depth estimation process from stereo images. Window cost or matching cost functions are reported in Section 2. The proposed algorithm for disparity or depth estimation is described in Section 3 while Section 4 presents the disparity refinement technique. The experimental results are reported in Section 5. Finally, Section 6 concludes the paper and provides directions for future work.

2. Dense disparity and depth estimation

The disparity or depth information can be obtained by computing the similarity or matching costs between the pixels within two windows in the stereo image pair. In stereo imaging system, two images of the same scene are taken from slightly different viewpoints using two cameras: left and right camera, which are placed in the same lateral plane. For most pixels in the left image, there is a corresponding pixel in the right image in the same horizontal line. The difference in the coordinates of the corresponding pixels is known as the disparity, as shown in Figure 1, can be expressed by the following equation.

$$d = x_L - x_R. \tag{1}$$

The depth information of the object can be extracted from the stereo imaging system using triangulation geometry of the camera settings, and the estimated disparity [12,13]. In this work, the pinhole camera model is used for computation of dense depth information. The process is illustrated in Figure 2.

Let, L and R are to be two pinhole cameras with parallel optical axes; O_L and O_R are two center points of the left and right camera respectively with same focal length f. The baseline, which is the line connecting the two lens centers of the cameras, is perpendicular to the optical axes. Let b to be the baseline distance and x_L is the x-coordinate of the projected 3D point onto the left camera

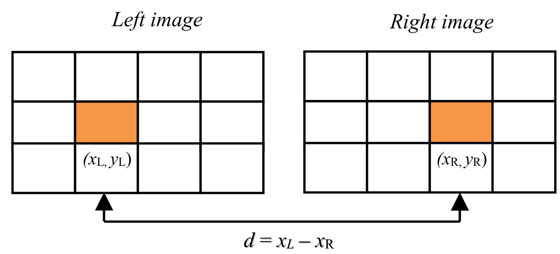


Figure 1. The pixel x_R in the right image is the corresponding pixel of x_L in the left image.

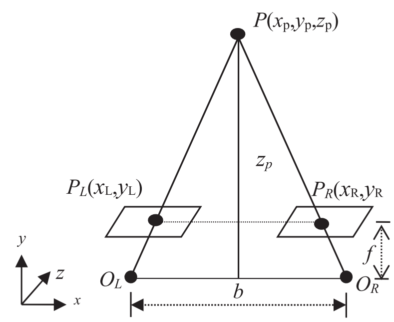


Figure 2. Dense depth estimation using triangulation.

image plane and x_R is the x-coordinate of the projection onto the right image plane.

For a world 3D point $P(x_p, y_p, z_p)$, we can derive the equations for stereo dense depth estimation from the camera geometry model as follows:

$$\frac{z_p}{b} = \frac{z_p - f}{b - (x_L - x_R)}. \tag{2}$$

$$\text{Thus the depth, } z_p = \frac{bf}{x_L - x_R} = \frac{bf}{d}. \tag{3}$$

Our aim is to recover the 3D point P from its projections P_L and P_R. Therefore, we can get:

$$x_p = \frac{bx_L}{x_L - x_R} = \frac{bx_L}{d}. \tag{4}$$

$$y_p = \frac{by_L}{x_L - x_R} = \frac{by_L}{d}. \tag{5}$$

Since the depth, z_p indicates a distance value (i.e. in mm or cm), we need to modify Equation (3) for its uniformity because, the parameters (b, f, d) in the equation possess different units. Otherwise, it would provide erroneous result during depth or distance measurement between the stereo cameras and the object. Accordingly, we reform Equation (3) through converting the unit of the disparity value (d) by dividing it with the pixel size (normally in mm/pixel) of the camera. Thus the accurate depth information is given by,

$$z_p = \frac{bf}{ds}, \tag{6}$$

where s is the size of a pixel of the stereo camera. Therefore, once we estimate the disparity of a reference pixel, we can easily extract the depth information of that pixel.

3. Window cost or matching cost calculation

The local correlation-based matching algorithms conventionally find dense disparity based on matching windows of pixels through the statistical measures such as the sum of absolute differences (SAD) [3], the sum of square differences (SSD) [5] or normalized cross correlation (NCC) [6]. To determine the correspondence of a pixel in the left image, the matching costs or window costs (SAD/SSD/NCC) are computed for all candidate pixels within the windows in the right image within a search range. The pixel in the right image that gives the best window cost (i.e. the minimum SAD/SSD value or the maximum NCC value) indicates the corresponding pixel of the candidate pixel in the left image. The conventional window cost functions or algorithms calculate the correspondence or disparity using the following measures:

$$W_c^{\mathrm{SAD}}(x,y,d) = \sum_{i=1}^{W_x}\sum_{j=1}^{W_y} |f_L(x+i,y+j) - f_R(x+i+d,y+j)|, \tag{7}$$

$$W_c^{\mathrm{SSD}}(x,y,d) = \sum_{i=1}^{W_x}\sum_{j=1}^{W_y} \{f_L(x+i,y+j) - f_R(x+i+d,y+j)\}^2, \tag{8}$$

$$W_c^{\mathrm{NCC}}(x,y,d) = \frac{\sum_{i=1}^{W_x}\sum_{j=1}^{W_y} |f_L(x+i,y+j)\times f_R(x+i+d,y+j)|}{\sqrt{\sum_{i=1}^{W_x}\sum_{j=1}^{W_y} |f_L^2(x+i,y+j)| \sum_{i=1}^{W_x}\sum_{j=1}^{W_y} |f_R^2(x+i+d,y+j)|}}, \tag{9}$$

where $f_L(x,y)$ and $f_R(x,y)$ are the intensities of the pixels at a position(x,y) in the left and right images, respectively, $W_C(x,y,d)$ is the window cost of a pixel at position(x,y) in the left image with disparity d, w_x and w_y are the window width and height, respectively.

4. Proposed algorithm for disparity/dense estimation

To estimate dense disparity or depth, we search the corresponding pixel in the right image, for every reference pixel in the left image. The pixel in the right image that gives the best window cost indicates the corresponding pixel of the reference pixel in the left image. For corresponding matching, it requires computing the window costs for all candidate pixels within the search range, $-d_{\max}$ to $+d_{\max}$.

The computation time for stereo correspondence matching depends on window size and disparity search range. With a view to achieve a substantial gain in matching accuracy with less expense of computation time, we employ a pruning technique in this proposed method. To determine the correspondence of a pixel in the left image based on the pruning proposal, we just compute the window cost for candidate pixels in the right image whose intensities are different within a certain threshold value δ to the intensity of the pixel in the left image. We use the modified version of SSD algorithm integrating the pruning proposal. Our proposed algorithm for depth estimation is illustrated in the following steps:

(1) For each pixel (x, y) in the left image, search the corresponding pixel in the right image within a search range:

for $d' = -d_{\max}$ to $+d_{\max}$ **do**

if $|f_L(x,y) - f_R(x+d',y)| <$ threshold **then**

(2) Calculate $W_C(x,y,d')$

(3) Find the disparity d such that, $d = \arg\min W_C(x,y,d')$

(4) Calculate the depth, $z = \frac{bf}{ds}$

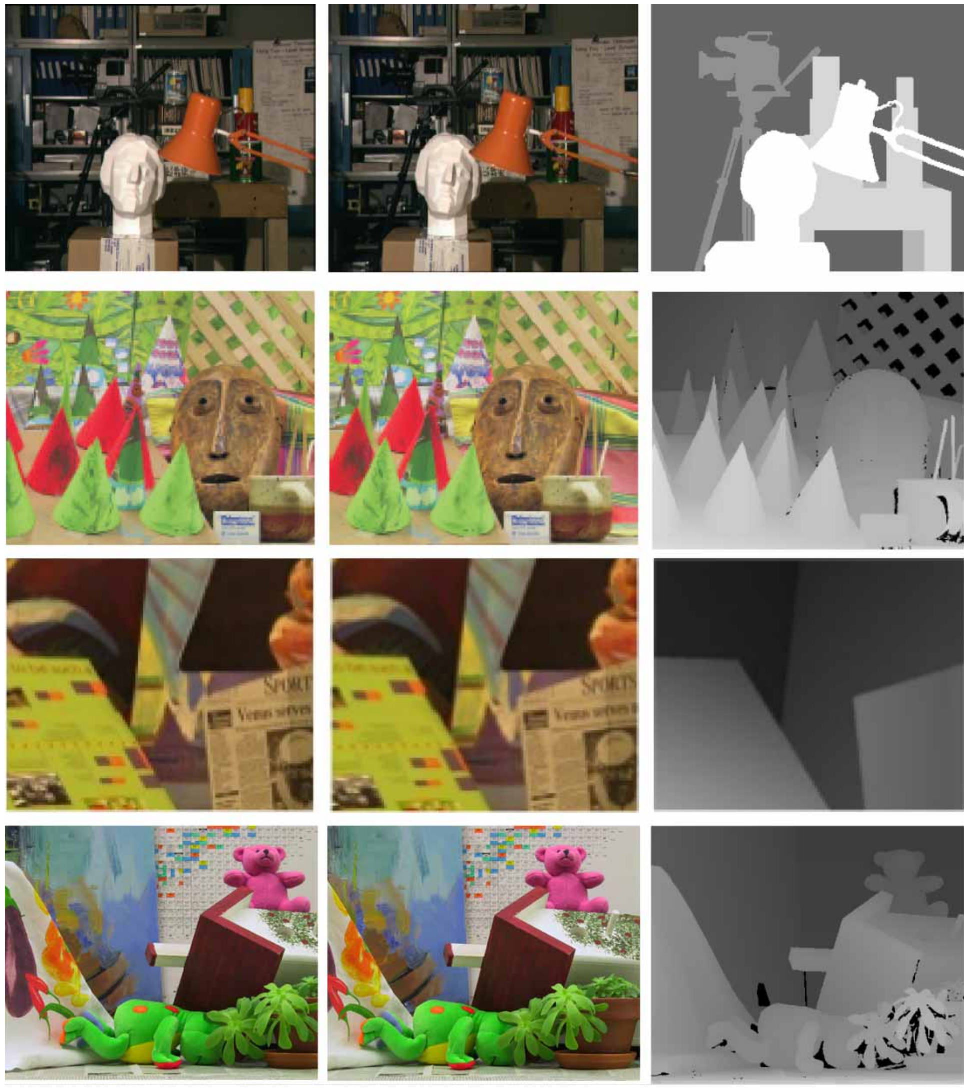
(5) Repeat steps 1 and 3 to calculate disparities and corresponding depth for all pixels in the left image.

5. Disparity refinement

In order to refine the disparity, we employ a fuzzy filtering algorithm [14]. This refinement technique propagates disparities from textured foregrounds to un-textured backgrounds. It allows us to enforce local consistency in the disparity values which accordingly reduces false matching rate and improves the accuracy.

6. Experimental evaluation

In this section, we evaluate the performance of our proposed algorithm and compare with other stereo methods in terms of accuracy and execution time. In order to



(a) Left image (b) Right image (c) Ground truth

Figure 3. Test stereo image pairs: Tsukuba head (top), cones (second row), venus (third row) and teddy (bottom).

demonstrate the effectiveness of our proposed algorithm, we present the processing results using standard stereo image pairs taken from Middlebury stereo datasets with their ground truth values [15]. Experiments are carried out on a computer with 2.8 GHz Intel Core i7 processor. The algorithm has been implemented using Visual C++. Figure 3 shows the standard stereo image pairs with their ground truth. The ground truth images are histogram equalized for visualization purpose. The size of the left and right view of input image pairs is 384×288 (width $\times$ height) pixels and that of ground truths is 348×252 pixels.

The estimated disparity maps with different algorithms for using the test image pairs are shown in Figure 4. The quantitative results of disparity estimation for different algorithms are reported in Table 1. The accuracies shown in this table represent the percent of correct disparities, i.e. same value as that of ground truth. The table demonstrates that the highest matching accuracy is obtained in case of the Tsukuba stereo image pair using a window of size 11×11 which is about 95%. The disparities are computed with a search range of -5 to $+5$ pixels for a threshold level of 20. Empirical analysis suggests that these parameter values provide optimized results.

We compare our algorithm with some advanced related methods [7–11] in terms of accuracy and execution time. The results are reported in Table 2. Experimental results reveal that our proposed method achieves a substantial gain in matching accuracy with less expense of computation load. Although, there are several methods with high accuracy, they require high computation time comparable to our method.

Experimental evaluation shows that the matching accuracy boosts up with the increase in the size of

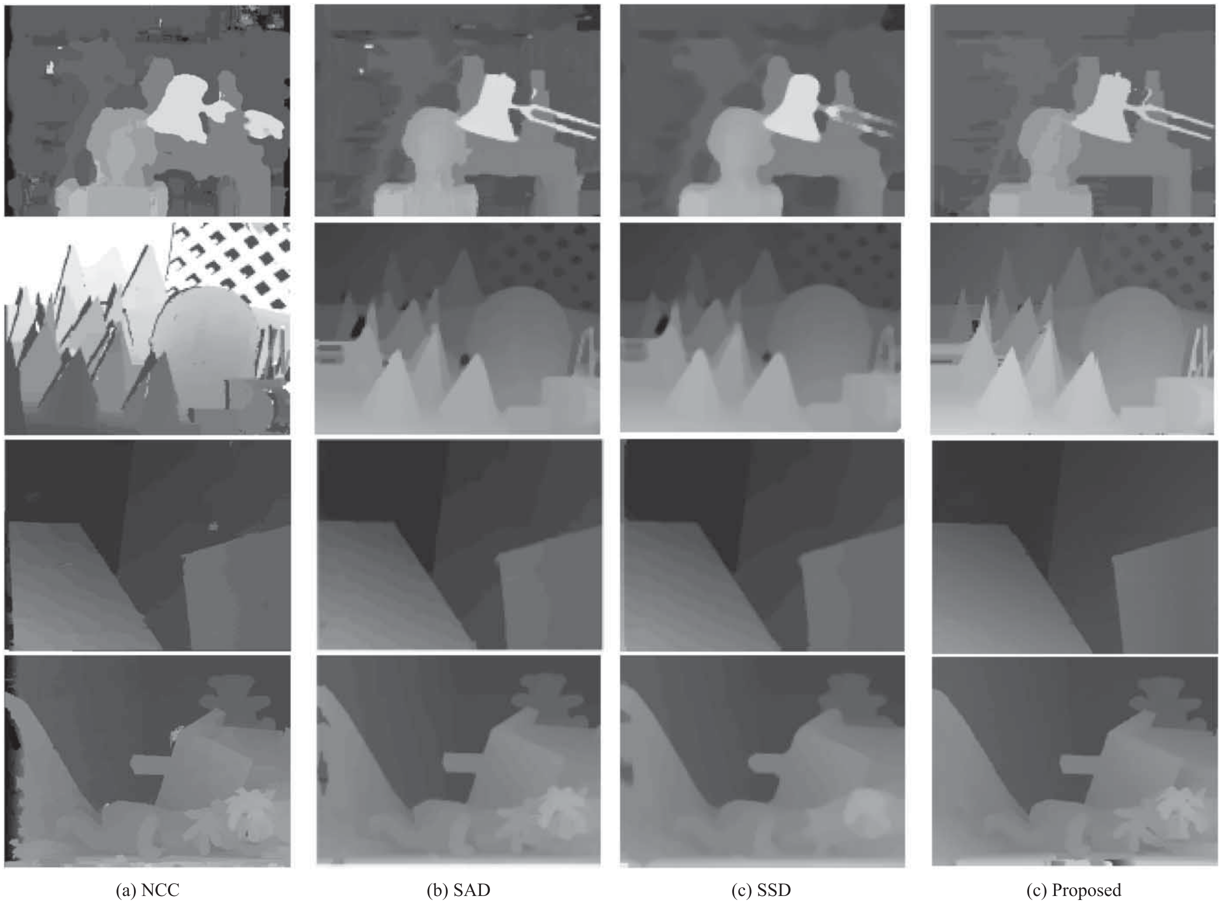


Figure 4. Estimated disparity map for different algorithm using standard benchmarking stereo image pairs: Tsukuba head (top), Cones (second row), venus (third row) and teddy (bottom).

Table 1. Disparity matching accuracy (in %) for different algorithms using standard stereo image pairs.

	Tsukuba head				Cones				Venus				Teddy			
Window size	NCC	SAD	SSD	Proposed	NCC	SAD	SSD	Proposed	NCC	SAD	SSD	Proposed	NCC	SAD	SSD	Proposed
3×3	79.2	81.8	85.6	92.7	78.4	81.2	84.7	91.6	78.6	81.9	85.1	88.7	78.8	81.4	84.9	89.8
7×7	82.8	84.9	87.2	93.8	80.7	83.8	86.8	92.9	80.8	84.1	87.3	92.5	80.4	84.7	86.8	91.9
11×11	85.1	86.6	90.4	95.6	84.8	86.1	89.3	93.8	84.2	86.9	89.6	94.1	83.9	85.9	88.6	93.5

Table 2. Comparisons of matching accuracy and computational time.

Method	Matching accuracy (%)	Computation time (second)
Convolution NN [7]	97.4	100
Graph segmentation [10]	97.46	48
Disparity gradient [11]	96.3	3.04
Linear matching [8]	96.37	15
Large-scale [9]	98	0.96
NCC	79.2	0.82
SAD	81.8	0.47
SSD	85.6	0.65
Proposed method	92.7	0.28

the window; however, computation time significantly increases. But, our intension is to reduce the computation time while keeping the accuracy to a practical stage. Empirically we find that a window of size 3×3 pixels is a good choice on the basis of both computational speed and matching accuracy.

7. Conclusion

In this paper, we propose a fast disparity estimation algorithm based on correlation matching with a pruning proposal. The effectiveness of this algorithm has been justified using standard benchmark stereo image datasets. Experimental results confirm that we can reduce the corresponding matching load significantly using the pruning technique while achieving a substantial matching accuracy. We investigate the effects of different window cost functions for different stereo image pairs. We believe that this proposed method will be useful for many applications where a very fast estimation of dense disparities is required. Our next target is to improve the algorithm to achieve more matching accuracy.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Mozammel Chowdhury is a Doctoral Researcher at the School of Computing and Mathematics, Charles Sturt University, Australia. Prior to this, Mr Chowdhury was a Faculty Member of the Department of Computer Science and Engineering at Jahangirnagar University, Bangladesh. He has expertise in Computer Vision, Machine Intelligence, Image Analysis, Cyber-Security, eHealth, E-learning and E-governance. Mr Chowdhury received his B.Sc. (Honors) and MS (Research) degree in Computer Science and Engineering from Jahangirnagar University. He has published around 55 research articles in reputed journals and conference proceedings. He is serving as an editorial board member/reviewer of numerous international journals and conferences.

Junbin Gao is a Professor of Big Data Analytics at the University of Sydney Business School, Australia. He received the B.Sc. degree in computational mathematics from the Huazhong University of Science and Technology (HUST), China in 1982, and the Ph.D. degree from the Dalian University of Technology, China, in 1991. He was a Lecturer and Senior Lecturer of Computer Science with the University of New England, Australia, from 2001 to 2005. From 1982 to 2001, he was an Associate Lecturer, a Lecturer, an Associate Professor, and a Professor with the Department of Mathematics at HUST. Dr Gao also served as a Professor of Computing Science with the School of Computing and Mathematics, Charles Sturt University, Australia. His main research interests include machine learning, data mining, Bayesian learning and inference, and image analysis.

Rafiqul Islam is a Senior Lecturer in the School of Computing and Mathematics and Leader of Cyber Security Research Area, Faculty of Business, at Charles Sturt University, Australia. Dr Islam has expertise in Network & Information Security, Cyber-Security, Sensor Network and Machine Learning. He received his PhD and postdoctoral experience from Deakin University, Australia. Dr Islam was a lead researcher on various research projects including several ARC-funded projects. Dr Islam published more than 90 refereed research papers. Dr Islam is also editorial board member for an international journal and General Chair and Technical Program Committee (TPC) member of various international conferences. He is a member of the Institute of Electrical and Electronic Engineers (IEEE) and a founder member of Association of Accounting Technician (ATT).

ORCID

Junbin Gao http://orcid.org/0000-0001-9803-0256

References

[1] Schastein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithm. Int J Comput Vision. 2002;47:7–42.

[2] Lazaros N, Sirakoulis GC, Gasteratos A. Review of stereo vision algorithms: from software to hardware. Int J Optomechatronics. 2008;2:435–462.

[3] Chowdhury M, Gao J, Islam R. An efficient algorithm for stereo correspondence matching. Int J Comput Theory Eng. 2017;9(1):69–72.

[4] De-Maeztu L, Mattoccia S, Villanueva A, Cabeza R. Linear stereo matching. In IEEE international conference on computer vision (ICCV 2011); 2011. p. 1708–1715.

[5] Chowdhury M, Bhuiyan M. Fast window based stereo matching for 3D scene reconstruction. Int Arab J Inf Technol. July 2013;10(4):209–214.

[6] Chowdhury M., Gao J., Islam R. Fast stereo matching with fuzzy correlation. IEEE Conference on Industrial Electronics & Applications (ICIEA 2016). Hefei, China; 2016. p. 678–682.

[7] Yang Q. Stereo matching using tree filtering. IEEE Trans Pattern Anal Mach Intell. 2015;37(4):834–846.

[8] Geiger A, Roser M, Urtasun R. Efficient large-scale stereo matching. In Asian conference on computer vision (ACCV); 2010.

[9] Chowdhury M, Gao J, Islam R. Fuzzy logic based filtering for image de-noising. In IEEE international conference on fuzzy systems (FUZZ-IEEE 2016); Vancouver, Canada. ISBN 978-1-5090-0626-7. p. 2372–2376.

[10] Middlebury stereo datasets.http://vision.middlebury.edu/stereo/data/.

[11] Zbontar, LeCun Y. Computing the stereo matching cost with a convolutional neural network. CVPR; 2015. p. 1592–1599.

[12] Hamzah RA, Ibrahim H, Hassan AHA. Stereo matching algorithm based on per pixel difference adjustment, iterative guided filter and graph segmentation. J Visual Commun Image Represent. 2017;42:145–160.

[13] Park J-M, Song G-Y, Lee J-W. Shape-indifferent stereo disparity based on disparity gradient estimation. Image Vision Comput. 2017;57:102–113.

[14] Chowdhury M, Gao J., Islam R. Distance measurement of objects using stereo vision. 9th Hellenic Conference on Artificial Intelligence (SETN 2016), Thessaloniki, Greece, Article 33; 2016. p. 33.1–33.4. doi:10.1145/2903220.2903247

[15] Chowdhury M, Gao J., Islam R. Islam R. Robust human detection and localization in security applications. J Concurr Comput: Pract Experience. Forthcoming. doi:10.1002/cpe.3977