VOL. 11, NO. 10, MAY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

ANALYSIS ON SEGMENT-BASED DOUBLE STAGE FILTER ALGORITHM FOR STEREO MATCHING

Teo Chee Huat, Nurulfajar bin Abd Manap and Wira Hidayat Mohd Saad
Department of Electronic Engineering, Faculty of Electronics and Computer Engineering
Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
E-Mail: nicteo@student.utem.edu.my

ABSTRACT

The focus of the stereo matching algorithms development is commonly on eliminating unwanted aspects such as noises, occlusions and unwanted regions. In this paper, a segment-based stereo matching algorithm is introduced, Double Stage Filter (DSF). DSF is a hybrid algorithm which involves the basic approaches such as block matching, dynamic programming and median filter. The main feature of DSF is concern on segmentation and merging process. These processes are to remove the noises and horizontal stripes of raw disparity depth map. Segmentation process of DSF algorithm is to segment the optimized raw disparity depth map into several parts according to the pixel colours. Merging process is to combine the segmented segments into a disparity map. The datasets used for the DSF algorithm are from Middlebury Stereo Vision page and captured by LNC IP stereo camera. Results obtained are analyzed by various methods and functions such as system of Middlebury, Similarity Structural Index Metric (SSIM), Peak to Signal Noise Ratio (PSNR) and Mean Square Errors (MSE).

Keywords: stereo matching, segment-based, double stage filter, segmentation, merging.

INTRODUCTION

Stereo matching is a wide topic in the field of computer vision and attracts attention among researchers. There are various stereo matching algorithms has been developed, which can be found in the Middlebury Stereo Vision Website by Scharstein and Szelinski [1]. This website page provides dataset of stereo images and evaluation system for researchers to make comparison on the results obtained between their proposed stereo matching algorithms with others. Basically, most stereo matching algorithms consist of four common steps ; matching cost computation, cost aggregation, disparity computation optimization and disparity refinement [1]. However, the four steps are not compulsory to be applied in a stereo matching algorithm, it is depends on the design of the algorithm.

In the development of stereo matching algorithm, the step which attracts the most interest among the researchers is on the disparity refinement. Disparity refinement is the step to improve the quality of a raw disparity map for stereo vision applications. A raw disparity map may contain errors such as noises, occlusions and mismatches. Thus, there are approaches to overcome the errors; median filter, cross-checking, sub-pixel estimation and surface fitting [2]–[5]. Median filter is used as part of the research with combination of developed segmentation and merging on disparity refinement.

The main concern of this research is to design a stereo matching algorithm with less complexity in computation and capable to minimize the errors of a raw disparity depth map. Basic approaches are combined and used in the research such as Basic Block Matching (BBM) and Dynamic Programming (DP). BBM is used to determine the matching corresponding by using Sum of Absolute Difference (SAD) [9]. DP operates as optimization to compute the minimum cost of matching pixels between the corresponding scanlines. Unfortunately, there is flaw with this approach on the inter-scanline of the raw disparity depth map obtained by dynamic programming [2, 7, 10, 11, 12]. Horizontal streaks appear on the raw disparity depth map attained due to the scanline operation of dynamic programming. In order to reduce the streaks on the raw disparity depth map, segmentation is applied to extract the raw disparity depth into segments according to pixel colors of the content in the map. The segments will be smoothed out homogeneously by basic median filter. The smoothed segments will be merged up and go through second stage of median filter as final disparity map. As there are two stages of median filter occupied, the developed segment-based algorithm is called as Double Stage Filter (DSF).

Datasets used for DSF algorithm and other two algorithms, BBM and DP algorithm are from Middlebury webpage and captured by LNC IP stereo camera which will discussed further in the following sections. The results obtained by the algorithms are evaluated by objective and subjective evaluation. Objective evaluation is using evaluation functions such as Similarity Structural Index Metric (SSIM) [6], Peak to Signal Noise Ratio (PSNR) [7] and Mean Square Errors (MSE) [8]. Subjective evaluation is based on human eye's perception upon the results obtained. Meanwhile, results are also evaluated using the system provided by Middlebury which using Root-Mean-Squared (RMS) errors and the percentage of bad matching pixels.

STRUCTURAL DESIGN OF DOUBLE STAGE FILTER (DSF)

The overview design structure of Double Stage Filter (DSF) algorithm is shown in Figure-1. DSF algorithm consists of two phases: Phase 1 which consist of basic approaches in stereo matching and Phase 2 which is the new development of double stage filter. In Phase 1, it

VOL. 11, NO. 10, MAY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

consists of three basic components which needed in a stereo matching algorithm according to [1] such as matching cost computation, cost aggregation and disparity computation or optimization. Meanwhile, for the Phase 2 is the refinement part which involves new developed methods of segmentation, image merging process and organized basic median filter. Phase 1 is to obtain the raw disparity map by applying the three stated basic component, while in Phase 2; the raw disparity map will be refined by using the new developed methods of DSF algorithm. The samples of raw disparity map and the refined disparity map are shown in Figure-2 (a) and Figure-2 (b).

In this new developed algorithm, Sum of Absolute Differences (SAD) is used as the area-based matching cost function to compare the image regions between each block with the region of interest. This cost function is chosen as part of this algorithm due to its simple similarity measures. However, the matching cost may cause noises due to each pixels select its disparity of all other pixels independently. The limitation can be reduced in the following steps of the algorithm such as aggregation and optimization.

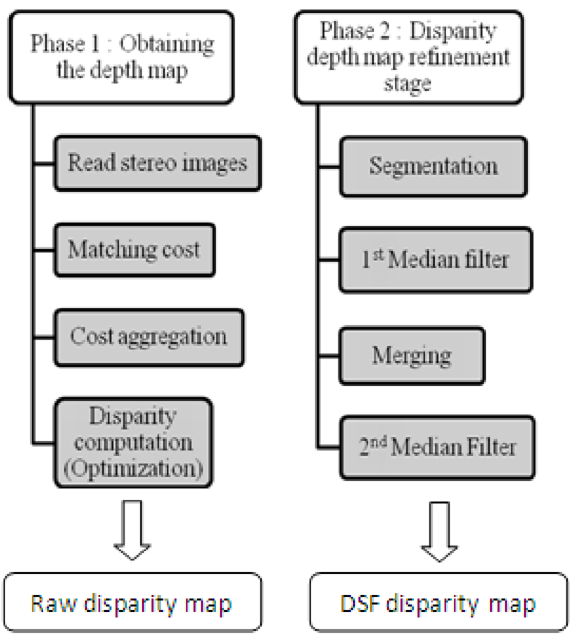
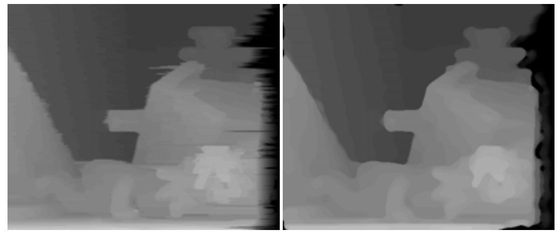


Figure-1. General design of block diagram for DSF



(a) (b)

Figure-2. (a) Raw disparity map; (b) Refined disparity map.

Cost aggregation is depends on window-based to sum up or take the average over the cost values from Disparity Space Image (DSI). DSI can be defined as a three-dimensional data structure to determine the matching cost of individual pixel for all possible shifting of pixels. DSI can be applied at fixed window sizes by additional technique so that the computation is independent on the window size [9]. Therefore, fixed square window of cost aggregation is applied in this algorithm as it is commonly used algorithm for applications which concern on real-time output due to fast and easy implementation. The size of the window need to be chosen appropriately where a small window does not provide good results on less-textured regions, while the bigger size of window do not have the ability to capture on tiny and thin objects.

For the step of disparity computation or optimization, dynamic programming approach is applied in DSF algorithm to reduce the noises on disparity image that obtained from the steps of matching cost and aggregation which used SAD with fixed square window [10]. This approach has its own characteristic to smooth the edges of the objects boundaries which found in the noisy disparity image from the previous step of aggregation [11], [12]. It will refine the disparity image between rows in horizontal direction as scanline. The direction of dynamic programming in refining the noises of a disparity image between rows may cause lines or known as "streaks" resulting on the raw disparity depth map obtained.

In Phase 2 of DSF algorithm, the major concern is to remove the streaks and reduce the remaining noises such as occlusion, depth discontinuities and less texture regions of the raw disparity depth map that obtained by dynamic programming. Segmentation is used in this algorithm to extract the raw disparity depth map into several parts according to pixel colours of the content in the map. Selection on the particular range of pixel colour only will be segmented and filtered by median filtering process. Median filtering is a non-linear operation, which most frequently used in image processing in reducing noises while preserving edges of an image [13]. Several segmented parts will go through the first stage of median filtering with suitable window size in part-by-part to remove noises on every part of the disparity map. Step of merging will add up the filtered parts into a new disparity map. However, there will be remaining outliers between the several filtered parts after the merging step. In the last

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

step, second stage of median filtering is applied to remove the outliers and to reduce the remaining noises of the new disparity map.

OBJECTIVE EVALUATION

This section presents the results of disparity depth maps obtained from the existing stereo matching algorithms; Basic block Matching (BBM), Dynamic Programming (DP) and the proposed algorithm; Double Stage Filter (DSF). The input stereo pair images used for implementation on the stereo matching algorithms are Tsukuba, Venus, Teddy and Cones which taken from the Middlebury Stereo Vision Page. The results obtained are shown in Figure-3. The first column images (a) are the ground truths. The second column images, (b) show the results obtained from BBM algorithm. The third column images (c) show the results obtained from DP algorithm. The fourth column images (d) show the results from the proposed algorithm, DSF.

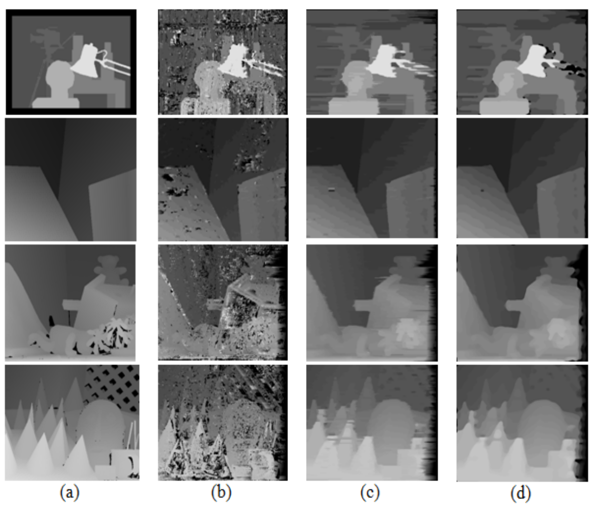


Figure-3. Results of three stereo matching algorithms by applying the Middlebury benchmark datasets.

The results in the second column in Figure-3 used the square window SAD of 7×7 for ('Venus'), and 3×3 for ('Tsukuba', 'Teddy' and 'Cones'). Meanwhile, there are two mask sizes applied for each stereo pair in DSF algorithm due to two stages of median filtering on Phase 2 of the algorithm. For the first stage of median filtering, the optimum mask size of 7×7 for ('Tsukuba' and 'Venus') and 11×11 for ('Teddy' and 'Cones') is used. In the second stage of median filtering, the optimum mask size of 7×7 is used for ('Tsukuba') and 11×11 for ('Teddy', 'Cones' and 'Venus'). These parameters of window sizes and mask sizes are chosen as optimum sizes in obtaining the best result for disparity maps of the algorithms. In the second column, it shows the disparity maps which obtained from BBM algorithm that contain noises all around the map. The third column indicates that DP algorithm successfully removed the random noises of BBM, unfortunately there are still horizontal 'steaks' appeared in the disparity maps. Meanwhile, in the fourth column shows that the proposed algorithm, DSF completely removed the horizontal 'streaks' of DP algorithm while preserving the edges of the disparity map.

Table-1. Results of MSE value for Tsukuba.

Window size	BBM	DP	DSF
3X3	353.77	213.49	223.33
5X5	295.59	219.57	219.43
7X7	271.48	222.03	223.53
9X9	259.63	224.85	219.35
11X11	251.08	226.92	220.25

Table-2. Results of PSNR value for Tsukuba.

Window size	BBM	DP	DSF
3X3	12.64	14.84	14.64
5X5	13.42	14.72	14.72
7X7	13.79	14.67	14.64
9X9	13.99	14.61	14.72
11X11	14.13	14.57	14.70

Table-3. Results of SSIM value for Tsukuba.

Window size	BBM	DP	DSF
3X3	0.14	0.52	0.56
5X5	0.28	0.53	0.55
7X7	0.38	0.53	0.56
9X9	0.44	0.53	0.56
11X11	0.47	0.53	0.56

The MSE values in Table-1 show the results of computed disparity map directly proportional to the gradually increasing window sizes among three stereo matching algorithms. For BBM algorithm with diamond shape on line is shows that the MSE values for Tsukuba stereo pair are gradually decreasing and this indicates that as window size increasing, more errors are reduced for BBM algorithm. On the other hand, DP algorithm is vice versa to BBM algorithm as the MSE values for Tsukuba stereo pair that generated by DP algorithm are raised with the window size extending respectively. The results of DP algorithm indicates that it performs well when using smaller size of window rather than the bigger window size due to its scanline optimization on every row of pixels. Smaller size of window may scan on content of image precisely and more errors can be reduced compared to bigger window size. Meanwhile, bigger size of window may missed scanning on tiny objects in the content of disparity map.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Based on the results in Table-2, it is clearly shows that the PSNR values computed from BBM algorithm is gradually increasing with increasing of window sizes. This represent that more noises are able to be reduced with larger window size for BBM algorithm. For DP algorithm, the PSNR values generated are gradually decreasing, which means less noise is removed and this shows that DP algorithm can only operate efficiently on reducing errors on image with smaller window sizes compared to BBM algorithm. Meanwhile, for DSF algorithm, the window sizes are not much affect on the performance for the whole algorithm, the graph shows that the PSNR values obtained from DSF algorithm are remain stable.

Based on the results of BBM algorithm, Table-3 indicates that the SSIM value is gradually raising and this shows that BBM algorithm require bigger window size to obtain a better quality of disparity map. However, the SSIM values of disparity map for each window size of BBM algorithm are lower than the SSIM values of disparity map for DP algorithm and DSF algorithm. For DP algorithm, the SSIM values obtained are remaining steady while the SSIM values for DSF algorithm are not much different among the window sizes.

SUBJECTIVE EVALUATION

For the part of subjective evaluation, the datasets used for evaluation are consist of stereo pair images that captured from LNC IP camera and the stereo pair images from Middlebury website page [14], [15]. The LNC IP camera used is only able to capture out left and right images without generating the ground truth of images. Therefore, the results obtained can only be evaluated subjectively by human's eyes observation on the disparity depth maps. The analysis of performance on the stereo matching algorithms from human's eyes perception will be further described in section.

The stereo pair images used for computation in the research are the images captured from LNC IP camera as shown in Figure-4. The LNC IP camera used is designed with various sensors for supporting media streaming, face detection, motion detection and image enhancement. In the research, LNC IP camera is used to capture stereo pair images, however, this camera is unable to generate the ground truth images. Therefore, the results obtained can only be evaluated subjectively by human's eyes observation on the disparity depth maps. The specifications of LNC IP camera as shown in Table-4.

Table-4. Specifications of LNC IP camera.

Specifications	
Sensor	**HD CMOS**
Day/Night	Auto Switch
Pixels	Approximate 200M
e-shoot	1/2 to 1/10000s
Image enhancement	Auto gain, Auto exposure, Auto white balance
Lens	Fix focus, IRIS support, C/CS mount lens support
Resolution	1920 × 1080
Format	H.264, MPEG-4, JPEG
Frame rate	H264/ MPEG-4 : 30f/s

Figure-4. LNC IP camera.

The evaluation on the stereo pairs for the stereo matching algorithms are using human's eyes perception due to the LNC IP camera which unable to generate ground truth images. The subjective results are presented in Figure-5 and Figure-6.

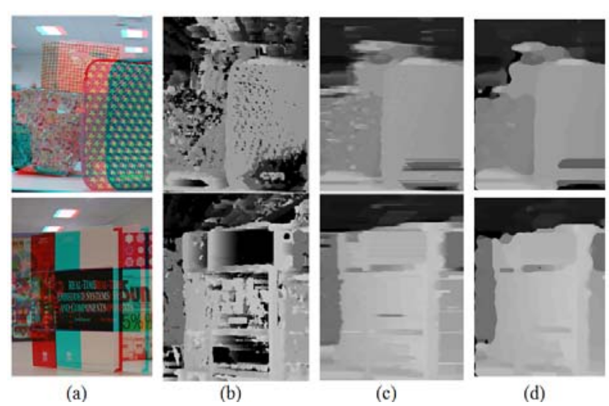


Figure-5. Results from dataset captured using LNC IP camera: (a) Original stereo images; (b) BBM; (c) DP; (d) DSF.

ARPN Journal of Engineering and Applied Sciences

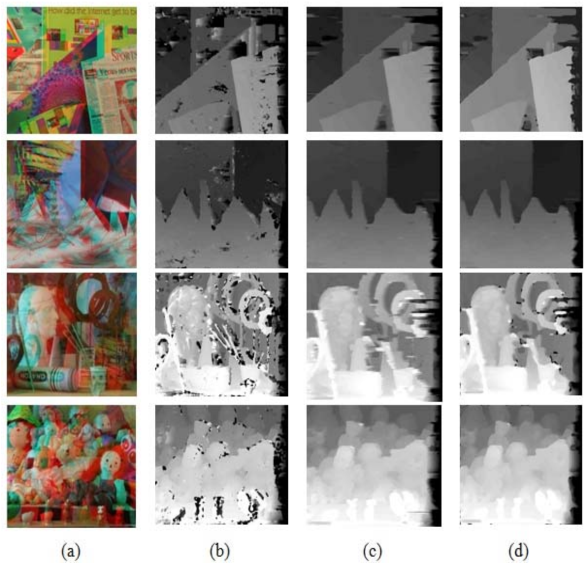


Figure-6. Results of datasets from Middlebury Webpage: (a) Original; (b) BBM; (c) DP; (d) DSF.

In Figure-5, the first and second rows of stereo pair images are captured by LNC IP camera while in Figure-6shows the stereo pair images from the Middlebury Stereo Vision page. The first column of both Figures, Figure-5(a) and Figure-6(a) shows the combination between the original left and right images. Figure-5(b) and Figure-6(b) represent the results obtained by the approach of BBM; the disparity depth maps obtained contain unwanted aspects such as random noises, depth discontinuities and occlusions compared to the results obtained by DP and DSF. Figure-5(c) and Figure-6(c) indicate the disparity depth maps obtained by dynamic programming approach; the results show dynamic programming is able to remove almost all the noises in the results of BBM with scanline optimization. However, dynamic programming approach cause horizontal streaks during the process in removing the unwanted aspects. Figure-5(d) and Figure-6(d) show the disparity depth maps obtained by double stage filter approach where the existing noises and horizontal streaks of disparity depth map in DP are removed. Double stage filter is not only able to remove the horizontal streaks and noises, it also improve the clearness of the objects' shape in the disparity map.

CONCLUSIONS

Double Stage Filter (DSF) algorithm is designed with the aim of improving the raw disparity maps in the post-processing stage. DSF consists of segmentation, median filtering and merging process, with less complexity in computation. The disparity maps obtained are evaluated with evaluation functions such as MSE, PSNR and SSIM. DSF algorithm performed well compared to DP and BBM algorithms as the graph of the evaluation functions stated the disparity map obtained by DSF is closer to the ground truth image rather than the other algorithms. In the subjective evaluation, based on evaluation from human eyes perception, the disparity map obtained by DSF is better than BBM and DP algorithm as the disparity map obtained by DSF shows that the noises are reduced and the horizontal streaks on images are completely removed. This indicates that DSF is capable to remove noises, horizontal 'streaks' of DP algorithm and minimize the depth discontinuities of disparity depth maps

ACKNOWLEDGEMENTS

The authors would like to thank to the Machine Learning and Signal Processing (MLSP) research group under Center for Telecommunication Research and Innovation (CeTRI) of Universiti Teknikal Malaysia Melaka (UTeM) and Faculty of Electronics and Computer Engineering (FKEKK), UTeM for sponsoring this work under project PJP/2013/FKEKK (30C)/S01220 and the use of the existing facilities to complete this project.

REFERENCES

[1] D. Scharstein and R. Szeliski. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Proc. IEEE Work. Stereo Multi-Baseline Vis. (SMBV 2001), no. 1, pp. 131-140.

[2] Z. Youlian, H. Cheng, and Z. Lifang. 2013. A Median Image Filtering Algorithm Based on Statistical Histogram. 2013 Fifth Int. Conf. Meas. Technol. Mechatronics Autom. no. 1, pp. 17-20.

[3] N. Sabater, J. Morel and a Almansa. 2010. Block matching reaches theoretical accuracy bounds.

[4] M. Bleyer and M. Gelautz. 2005. A layered stereo matching algorithm using image segmentation and global visibility constraints. ISPRS J. Photogramm. Remote Sens. vol. 59, no. 3, pp. 128-150.

[5] K. Zhang, J. Lu, and G. Lafruit. 2009. Cross-based local stereo matching using orthogonal integral images. IEEE Trans. Circuits Syst. Video Technol., vol. 19, no. 7, pp. 1073–1079.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. vol. 13, no. 4, pp. 600-12.

[7] A. Khaparde, A. Naik, M. Deshpande, S. Khar, K. Pandhari, and M. Shewale. 2013. Performance Analysis of Stereo Matching Using Segmentation Based Disparity Map. no. c, pp. 38-43.

[8] R. Xu, D. Taubman, and A. T. Naman 2014. Nonlinear transform for robust dense block-based

motion estimation. IEEE Trans. Image Process. vol. 23, no. 5, pp. 2222-34.

[9] F. Tombari, L. Di Stefano, S. Mattoccia, A. Mainetti, and D. Arces. 2010. A 3D Reconstruction System Based on Improved Spacetime Stereo. No. December, pp. 7–10.

[10] R. A. Hamzah, K. A. A. Aziz, and A. S. M. Shokri. 2012. A pixel to pixel correspondence and region of interest in stereo vision application. 2012 IEEE Symp. Comput. Informatics, Isc. 2012, pp. 193-197.

[11] O. Veksler. 2005. Stereo Correspondence by Dynamic Programming on a Tree. 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. vol. 2, no. 1, pp. 384-390.

[12] C. S. Park and H. W. Park. 2001. A robust stereo disparity estimation using adaptive window search and dynamic programming search. Pattern Recognit. vol. 34, no. 12, pp. 2573-2576.

[13] J. Huang, T.-Z. Huang, X.-L. Zhao, Z.-B. Xu and X.-G. Lv. 2014. Two soft-thresholding based iterative algorithms for image deblurring. Inf. Sci. (Ny). vol. 271, pp. 179-195.

[14] Z. Lee and T. Q. Nguyen. 2014. Multi-Array Camera Disparity Enhancement. IEEE Trans. Multimed., vol. 16, no. 8, pp. 2168-2177.

[15] G. A. Kordelas, D. S. Alexiadis, P. Daras, and E. Izquierdo. 2015. Enhanced disparity estimation in stereo images. Image Vis. Comput. vol. 35, pp. 31-49.