OCCLUSION HANDLED BLOCK-BASED STEREO MATCHING WITH IMAGE SEGMENTATION

Jisu Kim, Cheolhyeong Park, Ju O Kim and Deokwoo Lee

Department of Computer Engineering, Keimyung University, Daegu 42601, Republic of Korea

ABSTRACT

This paper chiefly deals with techniques of stereo vision, particularly focuses on the procedure of stereo matching. In addition, the proposed approach deals with detection of the regions of occlusion. Prior to carrying out stereo matching, image segmentation is conducted in order to achieve precise matching results. In practice, in stereo vision, matching algorithm sometimes suffers from insufficient accuracy if occlusion is inherent with the scene of interest. Searching the matching regions is conducted based on cross correlation and based on finding a region of the minimum mean square error of the difference between the areas of interest defined in matching window. Middlebury dataset is used for experiments, comparison with the existed results, and the proposed algorithm shows better performance than the existed matching algorithms. To evaluate the proposed algorithm, we compare the result of disparity to the existed ones.

KEYWORDS

Occlusion, Stereo vision, Segmentation, Matching.

1. INTRODUCTION

One can understand, recognize and analyze real world three dimensional scenes using parallax generated from both eyes [1]. Computer vision, especially stereo vision, utilizes the parallax principle in order to make a machine achieve understanding real world scenes. The technique, called computer vision, has brought tremendous attention in diverse areas such as recognition, machine learning, deep learning, etc. Stereo matching is one of the procedures for depth estimation from multiple-view camera systems (usually stereo camera system). To achieve accurate depth values, disparity should be calculated with high accuracy. Accurate stereo matching can be achieved if one can provide sufficiently accurate rectification result that is from camera calibration. Disparity is a difference of horizontal pixel locations (vertical location is identical if rectification is successfully done) between a pair of images each of which is projection of 3D real world scene of target [2]. The results of stereo matching are reliable if the vertical pixel locations are identical between a pair of images. Unfortunately, in practice, it is difficult to find the perfect algorithm that can be applied to all of the scenes due to the existence of ambiguity inherent with images, e.g., discontinuities in boundaries or depth, occlusion or texture-less regions. Therefore, numerous methods for stereo matching have been proposed so that various cases stated above can be solved [3,4]. In general, the most popularly employed techniques for the stereo matching can be categorized into two methods, the one is global matching and the other one is local one [5]. Global matching method usually deals with the cases of existence of texture-less region or occlusions. Global method provides accurate disparities

whereas it has a limitation of higher computational complexity compared to the local one. Graph-cut, belief propagation or dynamic programming based stereo matching are the widely used techniques of the global matching method. Graph-cut theoretic method borrows the concept of computer network, and disparity values are assigned based on using max-flow / min-cut algorithm that has shown superior performances of disparity estimation [6]. Belief propagation also has shown the competitive result of matching accuracy, however, it consumes high computation time to achieve convergence of the optimal solution [7]. To alleviate the limitation of computational complexity, dynamic programming(DP) based stereo matching and disparity estimation has been proposed. DP based disparity estimation finds the optimal path for disparity that is defined in disparity space image (DSI) generated by the relationship between a reference and a test image (or a pair of images) [8]. Local method for stereo matching, contrary to the global one, is simple to implement, shows lower computational complexity. However, it is rather vulnerable to have errors in the cases of occlusion, texture-less scenes, and to provide low accuracy of matching in boundary regions. Thus local method is popularly used in the areas that require real-time processing and high speed applications [9]. Local method can be categorized into two techniques, the one is feature based method and the other one is area-based method. The former one fully exploits edge, corner or cave information instead of only using pixel values of an image [10]. SIFT(Scale invariant feature transform) and SURF(Speed-up robust feature) are well known algorithms of the feature based technique [11,12]. As well known, SIFT is invariant to light, noise, view-points, rotation, scale, etc. SURF, based on multi-scale space theory, can alleviate the limitation of computational complexity of SIFT algorithm. SURF carries out feature extraction using Hessian matrix [12]. Block based matching algorithm is one of the local based matching techniques. It searches the matched areas by calculating cross correlation and by searching the region that results in the minimum difference of pixel values between the regions each of which belongs to the reference and the test image, respectively. Block based matching algorithm shows competitiveness in that it is fast and simple to be implemented.

In this paper, we propose the approach to stereo matching based on block based techniques and segmentation of sub-regions of images. The proposed method improves the existed block based matching method. This work maintains the performance of block-based method while increasing the accuracy of matching result by adding segmentation technique. To substantiate the propose approach, experimental results are provided, and the results outperform the existed result of block based matching. The rest of this paper is organized as follows. Section 2 briefly introduces the current block based matching followed by section 3 that introduces the proposed approach in this work. Section 4 provides the experimental results followed by concluding remark in section 5.

2. BLOCK BASED STEREO MATCHING

Block based stereo matching is the area-based matching technique. Prior to searching task, searching window is established followed by cross correlation is calculated. Correlation quantifies the similarity between sub-regions belonging to a pair of images, and the most similar areas result in the least square difference. Determination of size of searching window is also important because the size affects to performance of matching task, e.g., accuracy of matching, error sensitivity and computational complexity.
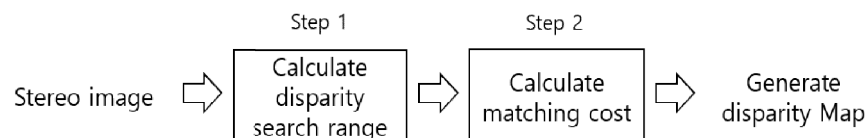

Figure 1. Overall flow of stereo matching and disparity map estimation

Figure 1 shows the overall flow of stereo matching. Matching is carried out within specific range, called search range. In practice, search range also needs to be optimized in order to achieve system efficiency, but this is not discussed in this paper (Figure 2).
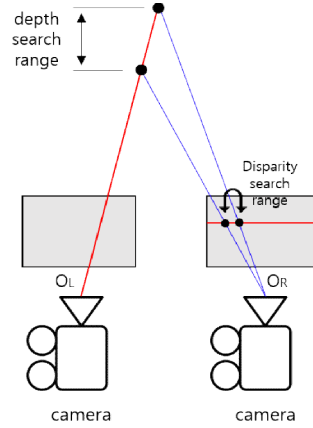

Figure 2. Search range for stereo matching

The similarity is also quantified by calculating mean square error between a pair of sub-regions contained in the matching windows, and the error is called matching cost. Ultimately, stereo matching is carried out so that the algorithm finds the point that generates the minimum matching cost. Numerous method is used to calculate matching cost, and one of the most widely used methods is SAD (sum of absolute difference). The point of the least SAD is decided to be the matching area, and the disparity is defined at this point.

3. PROPOSED APPROACH TO STEREO MATCHING

This section is the most contribution of this paper and details the proposed method. Stereo matching is carried out based on the block based techniques. In addition, image segmentation is combined with the block based method in order to increase matching accuracy. This work also deals with the occlusion. Figure 3 shows overall flow diagram of the proposed approach in this paper.
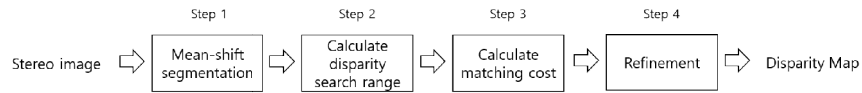

Figure 3. Overall flow of the proposed algorithm for stereo matching

Image segmentation is performed using mean-shift algorithm, and the clustered sub-regions are used for stereo matching. Mean-shift algorithm, one of the clustering methods, is based on the greedy algorithm, and the clustering is performed by searching the area of highest density of the data points [13]. Searching range of mean shift algorithm, called searching radius, needs to be established prior to carrying out clustering. Eq. 1 is mathematical modeling of mean-shift algorithm that updates the mean area (or the area of the highest density).

$$m_h(x) = \frac{\sum_{i=1}^{n} x_i k\left[\left\|\frac{x-x_i}{h}\right\|^2\right]}{\sum_{i=1}^{n} k\left[\left\|\frac{x-x_i}{h}\right\|^2\right]} - x, (1)$$

where $m_h(x)$ is the updated mean value at the location of data x (x is also the center of the searching window). n is a number of the points in searching window, i.e., data points are represented as x_1, x_2, ……… x_n. h is a radius of a searching window, and $k(\cdot)$ represents kernel function. According to Eq. 1, $m_h(x)$ is updated to the location of higher data density, and it converges to the highest density point. The result of segmentation using mean shift algorithm is shown in Figure 4

.

Figure 4. Original image (left) and segmentation and clustering result (right)

As shown in Figure 4, mean-shift algorithm provides the clustering result that preserves edge information while normalizing the pixel values in the same cluster. In this paper, to evaluate matching cost (or dissimilarity) C_{SAD} is employed, and C_{SAD} is written as

$$C_{SAD}(x_L, x_R) = \sum_{m=-W}^{+W} \sum_{n=-W}^{+W} |I_R(x_L + m, s + n) - I_R(x_R + m, s + n)| \quad (2)$$

where x_L, x_R are real values, horizontal coordinates of the pixels that are contained in the left and the right matching windows of images, respectively. Size of the matching window is W X W. I_R is a sub-region of an image and s represents the vertical coordinate of a pixel (since a pair of images is already rectified, vertical coordinate of the matched pixels are identical). By using SAD, matching cost is calculated, and the sample result is provided in Figure 5.
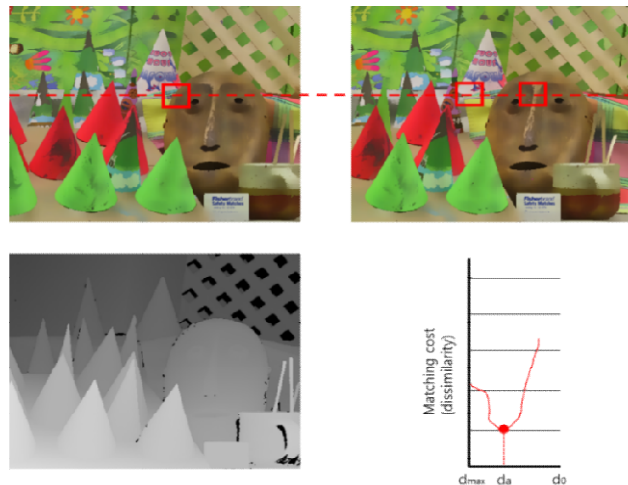

Figure 5. Matching cost calculated on the scanline. The minimum matching cost corresponds to the matched area.

Stereo matching sometimes shows weakness in the case of occlusion. To avoid error due to occlusion, prior to matching, detection of the occlusion area is required. This paper also deals with detecting regions of occlusion. By carrying out occlusion detection, one can provide high quality disparity map. In this work, to detect occlusion, gradient of an image is calculated. The

calculation is performed in a pair of images, especially at the same region of both images. If the edge is detected at the same region (i.e., gradient is identical at the same region), the region is considered no occlusion. If not, in other words, gradient is radically increased in only one image, it is considered the occlusion is existed. If occlusion is existed, disparity value or gradient has high standard variation. Based on the standard variation, the areas of false matching are detected. The area of false matching is filled with the neighboring values and the disparity map is refined. Figure 6 shows detected occluded regions and the refined disparity map.
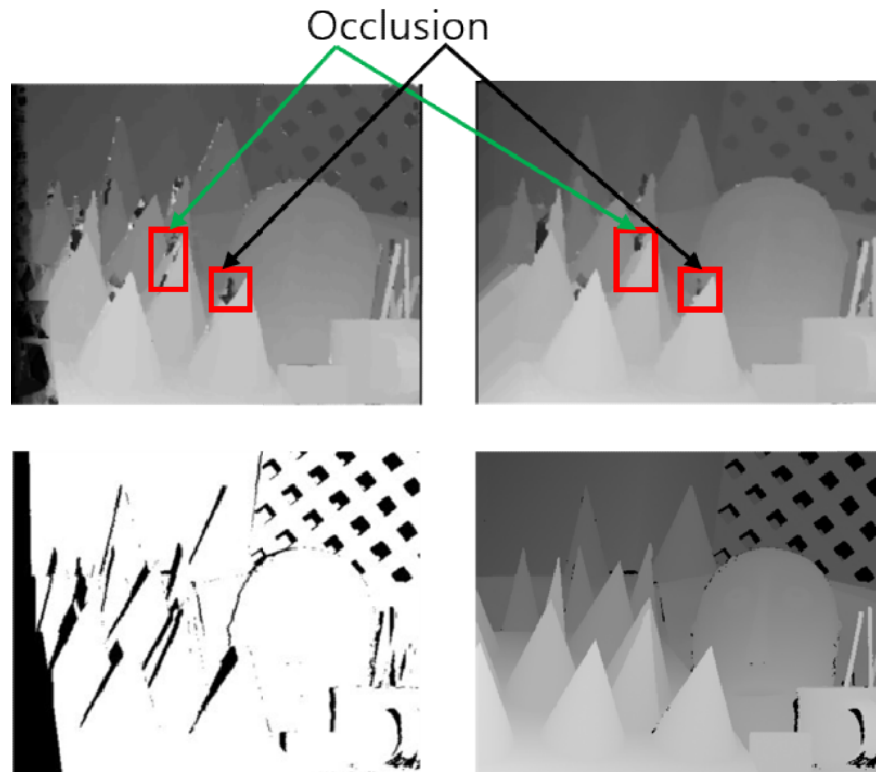

Figure 6. Detection of occlusion based on gradient of image (left-top and right-top), Occluded regions (left-bottom) and refined disparity map(right-bottom)

4. EXPERIMENTS

In the experiments, middlebury dataset is used. Comparison and evaluation are also carried out using the results provided in the website of middlebury stereo vision [14]. To evaluate the proposed approach in this paper, MSE(mean square error) of the disparity map results between the methods and the ground-truth is calculated [15]. Ground truth of disparity value is provided *a priori* in the page of middlebury stereo dataset. The comparison is performed between the proposed method and block-based matching technique. Although only the results from two images are presented, the experiments have been conducted using all of the images in *2003 dataset* contained in middlebury dataset.

As shown in Table 1, the performance is affected by the size of matching window. However, in overall, the proposed method outperforms the conventional block based matching algorithm.

Table 1. Mean square error of the disparity values

Method	Image 1				Image 2			
	3 X 3 (window size)		9 X 9		3 X 3		9 X 9	
	Left	Right	Left	Right	Left	Right	Left	Right
Block based matching	11516	11259	11770	1000	7832	7442	8082	7573
Proposed	11130	10396	11456	10722	7728	7380	8065	7477
Improvement(%)	3.35	7.66	2.66	2.52	1.32	0.83	0.02	1.26

5. CONCLUSION

This paper presents the stereo matching algorithm that combines block-based matching and image segmentation while occlusion is handled. Prior to performing stereo matching, image segmentation is carried out so that the feature is accurately extracted and boundary information is preserved. While preserving strength of the conventional matching algorithm the proposed method could increase the accuracy of matching and the result of disparity estimation. However, there still exists matching errors due to the various environments of the images. In the future work, learning based feature extraction and matching will be used in order to achieve adaptive matching can be performed.

REFERENCES

[1]   Hartely, Richard. & Zisserman, Andrew (2003) *Multiple View Geometry in Computer Vision*, Computer graphics, image processing and robotics, Cambridge University Press.

[2]   Mühlmann, Karsten & Maier, Dennis & Hesser, Jürgen & Männer, Reinhard, (2002) "Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation", *International Journal of Computer Vision*, Vol. 47, No. 1, pp.79-88.

[3]   Xu, Jintao & Yang, Qingxiong & Feng, Zuren, (2016) "Occlusion-Aware Stereo Matching", *International Journal of Computer Vision*, Vol. 120, No. 3, pp.256-271.

[4]   Kim Kyung Rae & Kim Chang Su, (2016) "Adaptive smoothness constraints for efficient stereo matching using texture and edge information*", 2016 IEEE International Conference on Image Processing (ICIP)*, pp.3429-3433.

[5]   Brown, Myron Z & Burschka, Darius & Hager, Gregory D, (2003) "Advances in computational stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 25, No. 8, pp.993-1008.

[6]   Huang, Xiaoshui & Yuan, Chun & Zhang Jian, (2015) "Graph Cuts Stereo Matching Based on Patch-Match and Ground Control Points Constraint", Advances in Multimedia Information Processing – PCM, Vol. 9315, pp14-23.

[7]　Mozerov, Mikhail G & Weijer, Joost van de, (2015) "Accurate Stereo Matching by Two-Step Energy Minimization", *IEEE Transactions on Image Processing,* Vol. 24*,* No. 3, pp.1153-163.

[8]　Salehian, Behzad & Fotouhi, Ali M & Raie, Abolghasem A, (2018) "Dynamic programming-based dense stereo matching improvement using an efficient search space reduction technique", *Optik*, Vol. 160, pp.1-12.

[9]　Zhu, Shiping & Yan, Lina, (2017) "Local stereo matching algorithm with efficient matching cost and adaptive guided image filter". *The Visual Computer,* Vol. 33, No. 9, pp. 1087-1102.

[10]　Kang, C & Kim, J & Lee, S & Nam, K, (1997) "Stereo Matching Using Dynamic Programming with Region Partition". *Journal of the Institute of Electronics and Information Engineers,* Vol. 20, No. 1, pp.479-482.

[11]　Lowe, David G, (1999) "Object recognition from local scale-invariant features", *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp.1-8.

[12]　Bay, Herbert & Tuytelaars, Tinne & Gool, Luc V, (2008) "Speeded-Up Robust Features (SURF)", *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp.345-359.

[13]　Lee, K-M. and Lin, C-H, (2017) "And Image Segmentation and Merge Hierarchical Region using Mean-Shift Tracking Algorithm", *Proceedings of Annual Conference of IEIE*, pp.704-706.

[14]　Scharstein, D & Szeliski, R, (2002) "A taxonomy and evaluation of dense two-frame stereo corresponding algorithms", *International Journal of Computer Vision,* Vol. 47, No. 1, pp.7-42.

[15]　Scharstein, D & Szeliski, R, (2003) "High-Accuracy Stereo Depth Maps Using Structured Light", *IIEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.195-202.

Author

Jisu Kim is in department of computer engineering, Keimyung University, Daegu, Republic of Korea. He is currently working on image processing, computer vision, signal processing and machine learning. He is currently pursuing his M.S degree in computer engineering.

Cheolhyeong Park is in department of computer engineering, Keimyung University, Daegu, Republic of Korea. He is currently working on geometric image analysis, computer vision, computer graphics and machine learning. He is in the course of integrated B.S and M.S degree in computer engineering.

Ju O Kim is in department of computer engineering, Keimyung University, Daegu, Republic of Korea. He is currently working on image analysis and Processing. He is pursuing B.S degree in computer engineering.

Dr. Deokwoo Lee is an Assistant Professor in the department of computer engineering at Keimyung University. Dr. Lee has received B.S degree in electrical engineering from Kyungpook National University, Daegu, Republic of Korea, and M.S and Ph.D degree from North Carolina State University, Raleigh, NC, USA, respectively. He has been working on the areas of computer vision, image processing, signal processing and machine learning. In particular, he has been conducting camera calibration, bio-signal analysis and image denoising.