DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-449

Occlusion and Error Detection for Stereo Matching and Hole-Filling Using Dynamic Programming

Eu-Tteum Baek and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdan-gwagiro, Buk-gu, Gwangju 61005, Republic of Korea
Email: {eutteum, hoyo}@gist.ac.kr

Abstract

Occlusion is the key and challenging problem in stereo matching, because the results from depth maps are significantly influenced by occlusion regions. In this paper, we propose a method for occlusion and error regions detection and for efficient hole-filling based on an energy minimization. First, we implement conventional global stereo matching algorithms to estimate depth information. Exploiting the result from a stereo matching method, we segments the depth map occlusion and error regions into non-occlusion regions. To detect occlusion and error regions, we model an energy function with three constraints such as ordering, uniqueness, and color similarity constraints. After labeling the occlusion and error regions, we optimize an energy function based MRF via dynamic programing. In order to evaluate the performance of our proposed method, we measure the percentages of mismatching pixels (BPR). And we subjectively compare the results of our proposed method with conventional methods. Consequently, the proposed method increases the accuracy of depth estimation, and experimental results show that the proposed method generates more stable depth maps compared to the conventional methods.

Keywords: occlusion, disparity estimation, stereo matching, hole-filling, dynamic programming

1. Introduction

Stereo depth estimation is a widely researched topic in computer vision, and it is related to many applications such as 3D movie, 3D printing, object detection, and 3D reconstruction. Depth information represents distance information between a camera and objects in a captured scene. In general, depth information can be obtained by some methods such as active depth cameras, passive depth cameras, and hybrid depth cameras. Active depth sensor acquires depth information with a physical sensor, which emit their own light onto the scene, and derive its depth information [1]. Usually, the active depth cameras are more effective and efficient in generating high quality depth data indoors than the passive sensors. Passive depth cameras measure correlation of scenes captured from two or more cameras [2]. Hybrid depth cameras integrate the active and passive methods to generate more accurate depth data and to cover their weaknesses [3].

Computer visual system adopted the basic principles of the depth estimation from human visual system model. Depth perception arises from a variety of depth cues, for example, monocular cues, binocular cues, differences in brightness, and focus [4]. In the computer visual system, the Binocular cues is the most important source of depth perception. Generally, human percepts depth using the distance of a same object between the viewpoints. Therefore, most of stereo matching approaches exploit the binocular cue.

However, depth data acquisition with the binocular cue suffer from occlusion problem, which is the important problem in stereo matching. Occlusion means that occluded pixel is apparent in the source image, but there is no corresponding pixel in the target image. Figure 1 represents the occlusion and the non-occlusion. Because an object is obscured by the view of some objects or regions, occluded pixels are only visible in the reference image, but in the target image. Figure 2 shows the stereo images which the reference images have yellow regions, which denote occlusion regions.

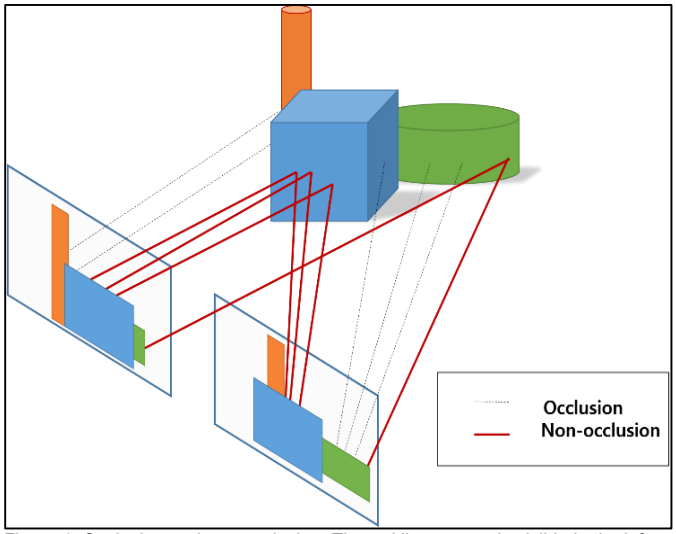


Figure 1. Occlusion and non-occlusion: The red lines are only visible in the left and right images

Most approaches exploit the ordering constraint or uniqueness constraint to optimize the problem via dynamic programing [5, 6]. Dynamic programming can independently yield a global minimum for each scanline in a polynomial time. The simplest method based on the uniqueness constraint apply cross-checking algorithm to detect occlusion [7]. Unfortunately, the ordering and uniqueness constraints have limitations. In the region of narrow holes or thin objects, the ordering constraint is violated. In addition, the uniqueness constraint is not appropriate for scenes containing

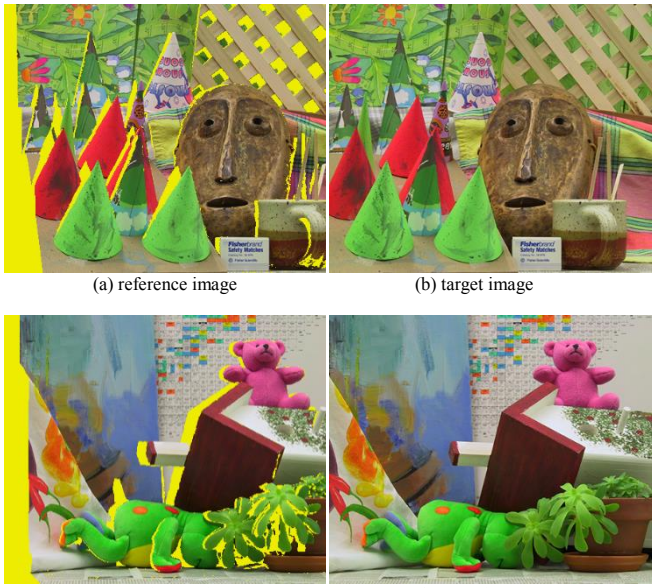
©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-449

horizontally slanted surfaces due to a discrete representation of disparity.

Kolmogorov et al. introduced an MRF based energy function for penalizing occluded pixels [8]. The defect of the algorithm is that the penalty of the occlusion term depends on only the uniqueness constraint. Using a two-step local method, Liu et al. presented an occlusion handling method [9]. To compute an initial matching cost, they use contrast contest histogram descriptors. Then, disparity estimation is performed via two-pass weighted cost aggregation considering segmentation based adaptive support weights. Jang et al. proposed a method for occlusion detection and refinement [10]. The algorithm optimizes an energy function considering warping, cross check, and luminance difference constraints via graph cuts. Even though reasonable occluded pixels are obtained, occlusion refinement has a drawback. In other words, Jang's algorithm does not consider smoothness constraint to generate smooth surface. Therefore, the occlusion refinement part should be improved.

The first goal of our work is to detect accurate occlusion regions using an energy function containing three constraints and optimize its energy function via an expectation maximization (EM) algorithm. The second goal of our approach is to refine the occlusion and error regions using a dynamic programing.

The rest of the paper is arranged as follows. Section II described stereo depth estimation, Section III occlusion detection and occlusion refinement in detail. In Section IV, experiment results and discussion is presented. Finally, the conclusion is described in Section V.



(a) reference image (b) target image

(c) reference image (d) target image

Figure 2. Occlusions: (a) and (b) occlusion in the cones stereo image, and (c) and (d) occlusion in the teddy stereo image. Yellow areas in the reference image are occluded from the right camera

2. Stereo depth estimation

Stereo matching algorithms estimate the distance of objects using stereo image. These methods can be categorized into local and global methods. Local method measure dissimilarity using local support window. Conventional local costs functions include the sum of absolute differences (SAD), the sum of squared differences (SSD), normalized cross correlation (NCC), and the census transform [11]. In contrast to local methods, global methods consider stereo disparity estimation as a labeling problem where the pixels of the reference image are nodes and the estimated disparities are labels. Global methods minimize an energy function via optimization techniques such as dynamic programming [12], graph cuts [13], belief propagation [14], and semi-global matching [15].

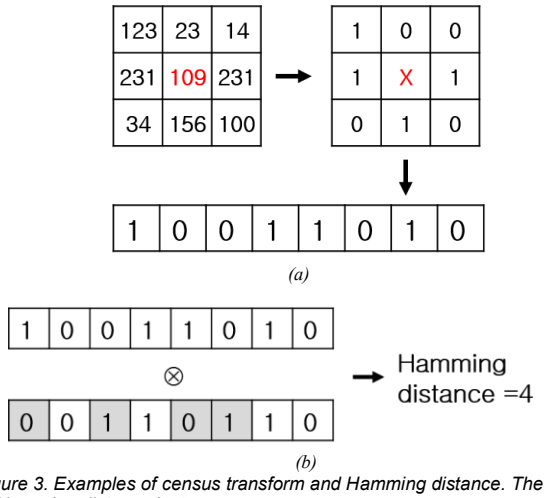


Figure 3. Examples of census transform and Hamming distance. The example of Hamming distance is 4

The energy function used in stereo matching usually consists of a correspondence data term and a smoothness term. Data term measures how well the observations are matched. Smoothness term assumes that pixels that are adjacent to each other may have a similar disparity.

$$E(d) = \sum_{s} D_s(d_s) + \lambda \cdot \sum_{s,t \in N(s)} S_{s,t}(d_s, d_t) \quad (1)$$

where d_s is the disparity map for a reference image I(x, y) and λ is a weight parameter that adjusts smoothness of the result. $D_s(\ )$ is the data term and $S_{s,t}(\ )$ is the smoothness term. We exploit data term as Hamming distance defined by

$$D_s(d_s) = Hamming(I_c(p), \bar{I}_c(\bar{p}_d)) \quad (2)$$

where $I_c(p)$ and $\bar{I}_c(\bar{p}_d)$ are transformed vectors using census transform, which is a non-parametric local transform method. Hamming distance is the number of differences between two vectors as shown in Fig. 3(a). Let $I_c(p)$ denotes census transform of one point p. The center pixel's intensity value is replaced by the bit string composed of set of boolean comparisons such that in a square window and $I_c(p)$ is defined as

$$I_c(p) = \bigotimes_{q \in N_p} \xi(I(p), I(q)) \quad (3)$$

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-449

where $\otimes$ denotes concatenation, N_p is neighboring pixels in a window, and ξ denotes transform represented as

$$\xi(I(p), I(q)) = \begin{cases} 0, if\ I(p) < I(q) \\ 1,\ otherwise \end{cases} \quad (4)$$

Census transform converts relative intensity difference to 0 or 1 in 1 dimensional vector form. Figure 3(b) represents an example of the census transform of a window with respect to the center pixel.

Alejo Concha et al. evaluated several cost functions [16]. The results of [16] show that truncated L1 and L2, Tukey and Geman-MacClure have the best performance. Therefore, we use smoothness term as truncated L2-norm defined by

$$S_{s,t}(d_s, d_t) = \min(\lambda \mid d_s - d_t \mid, T_s) \quad (5)$$

where T_s is the truncation value to constrain the high cost increase. Figure 4 shows the graph of truncated L2-norm. In order to optimize the energy functions, we employ the multi label algorithm which calculate optimal solution, is called the "alpha-expansion" algorithm [17].

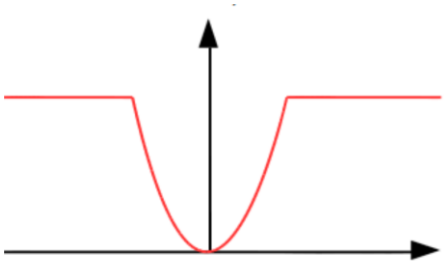


Figure 4. Graph of truncated L2-norm.

3. Occlusion handling

3.1 Occlusion detection

Occlusions are a principal challenge for the accurate computation of visual correspondence. Occluded pixels are visible in only reference image shown in Fig. 1.

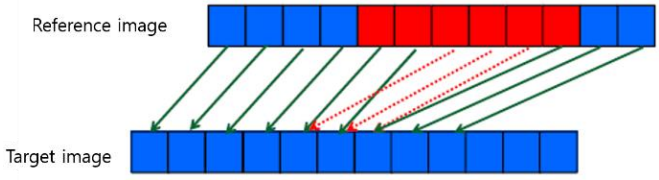


Figure 5. Warping constraint

We exploit the ordering constraint, uniqueness constraint, and color similarity constraint to formulate an energy function defined by

$$E_T(D^R) = \lambda_1 E_g(D^R) + \lambda_2 E_o(D^R, D^T) + \lambda_3 E_c(D^R) \quad (6)$$

where $E_g(D^R)$, $E_o(D^R, D^T)$, and $E_c(D^R)$ are ordering constraint term, uniqueness constraint term, and color similarity term. λ_1, λ_2, and λ_3 are weights. D^R and D^T are the reference and target disparity maps respectively. Ordering constraint predicts candidates of occluded pixels. Pixels in the reference image are projected to the target image. In the case of many-to-one mapping, a pixel which possesses the largest disparity value is selected as the visible pixel, but the rest of the matching pixels become occluded pixels shown Fig. 5. The red-colored pixels are regarded as candidates of occluded pixels. Ordering constraint is defined as

$$\begin{cases} E_g(D^R) = 1,\ if\ D^R = occlusion\ candidate \\ E_g(D^R) = \alpha,\ elseif\ D^R = largest\ value\ of\ candidate \\ E_g(D^R) = 0,\ ortherwise \end{cases} \quad (7)$$

where α is a small positive value. Uniqueness constraint Evaluate the mutual consistency from both disparity maps and both color images. Uniqueness constraint is defined as

$$\begin{cases} E_o(D^R, D^T) = 0,\ if\ D^R(x_z) = D^T(x_z - D^R(x_z)) \\ E_o(D^R, D^T) = 1,\ otherwise \end{cases} \quad (8)$$

where x_z is a pixel in a reference image. If a particular pixel in the image is not an occluded pixel, the disparity values from the left and the right disparity maps should be consistent as shown in the Fig. 6.

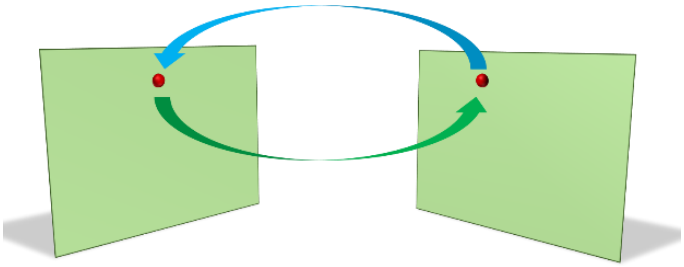


Figure 6. Uniqueness constraint

Color similarity in the color cube (RGB) is measured by the Euclidean distance. Color similarity measurement is represented as

$$E_c(C^R, C^T) = \frac{1}{W_C} \sqrt{\sum_{i=1} C_i^R(x_z) - C_i^T(x_z - D^R(x_z))} \quad (9)$$

where C^R and C^T are the reference and target color images respectively. W_C is a normalization factor.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-449

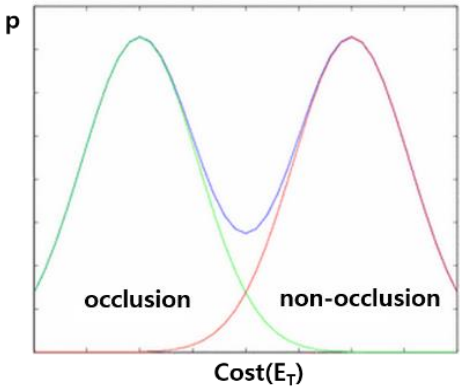


Figure 7. Gaussian mixture model (GMM): parameters (k=2)

In order to estimate occlusion regions, we exploit expectation maximization (EM) algorithm for Gaussian mixture model (GMM). EM is an iterative method for finding maximum likelihood and is a parametric optimization algorithm. Therefore, two GMM parameters are estimated by using the occlusion energy function from (6). Figure 7 shows an illustrative example of estimating a one dimensional Gaussian.

3.2 Occlusion refinement

After occlusion detection, the reasonable disparity value should be filled to the occluded pixel. In [2], they separate occlusion regions into two parts. Figure 8 shows the reference image and the corresponding occlusion map.

(a) (b)

Figure 8. Two kinds of occlusion.

Let a left image is a reference image. There are the left-side occlusion and the general occlusion. Left-side occlusion occurs because of non-existence of left-side region in the leftmost of a right image. The part in the orange rectangle in Fig. 8 (b) indicates the left-side occlusion, and the rest of the occlusion is the general part. General occlusion obscures an object or regions on target plane from a reference image. In the case of left-side occlusion, we fill an occlusion from right to left, but we assign disparity value in general occlusion from right to left.

In order to handle the problem, we formulate an energy function for assigning occlusion based on MRF-MAP model defined as

$$E_{occ}(d) = E_d(d) + \lambda_{occ} \cdot E_s(d) \tag{10}$$

where $E_d(d)$ is a data term, and $E_s(d)$ is a smoothness term. We exploit the data term as

$$E_d(s,d) = \frac{1}{dist(s,t)} \exp\left(-\frac{dif_{s,t}}{\sigma^2}\right) \; s.t. \;\; t \in non-occ$$
$$where dif_{s,t} = \sum_{c\in\{R,G,B\}} |I_c(s) - I_c(t)| \tag{11}$$

where dist(s, t) is Euclidean distance from s to t, and $dif_{s,t}$ denotes color dissimilarity. We use the smoothness term as truncated l2-norm from (5). We minimize the energy function for occlusion refinement via dynamic programming, which is an efficient algorithm for solving sequential decision problems. Dynamic programming divide large problem into a small sub problem. It stores all results of sub problem. And calculates sub problem only one time. In hole filling algorithm, it is similar to shortest path algorithm shown in Fig. 9.

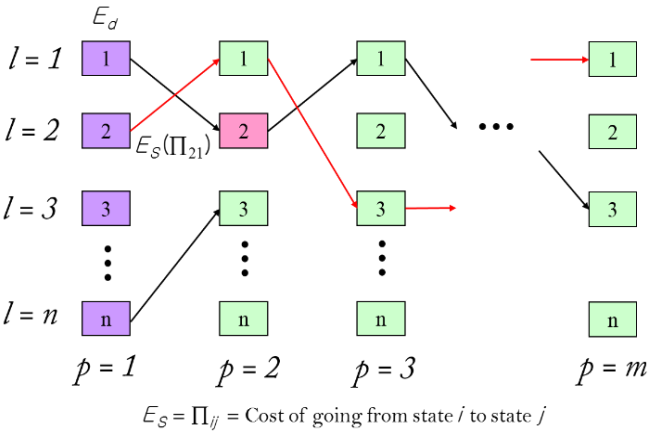


Figure 9. Dynamic programming for hole filling

In order to enhance disparity, we use guided image filtering, which is a kind of edge preserving noise removal filter [18]. The filter weights $W_{p,q}$ are expressed as

$$W_{s,t} = \frac{1}{|w|^2} \sum_{k:(s,t)\in wk} (1 + (I_s - \mu_k)(\sum_k + \varepsilon U)^{-1}(I_t - \mu_k)) \tag{12}$$

where |w| is the total number of pixels in a window w_k centered at pixel k, and ε is a smoothness parameter. $\sum_k$ and μ_k are the covariance and mean of pixel intensities within w_k. I_s, I_t and μ_k are 3 × 1 vectors, while $\sum_k$ and the unary matrix U are of size 3 × 3.

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-449

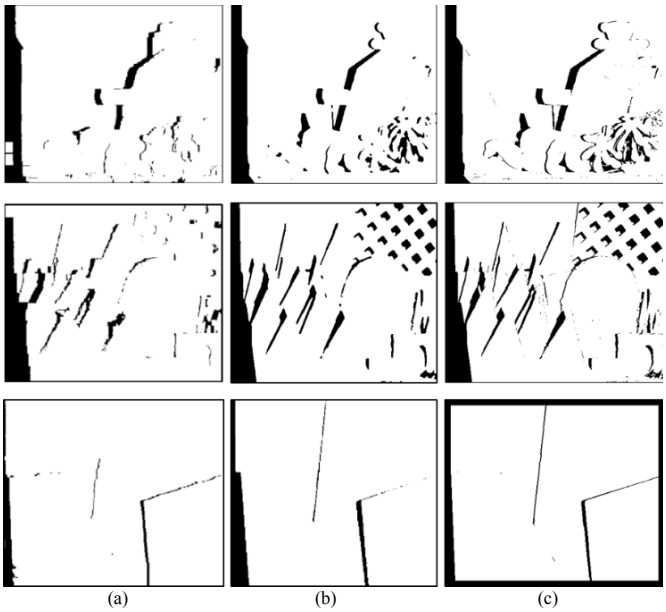
4. Experimental results

In order to evaluate objectively the performance of our method, we exploit the percentages of mismatching pixels (BPR), whose absolute difference is greater than 1, was used. First, we evaluate the occlusion detection using ground truth, provided by [21-22]. Figure 10 shows the visual comparison of occlusion detection compared with Jang's method. Table I shows the percentage of the bad matching pixels between the results of the proposed method and ground truths of the occlusion map. The results show that the proposed method is a high performance method compared to conventional method.

TABLE 1

Evaluation for occlusion map

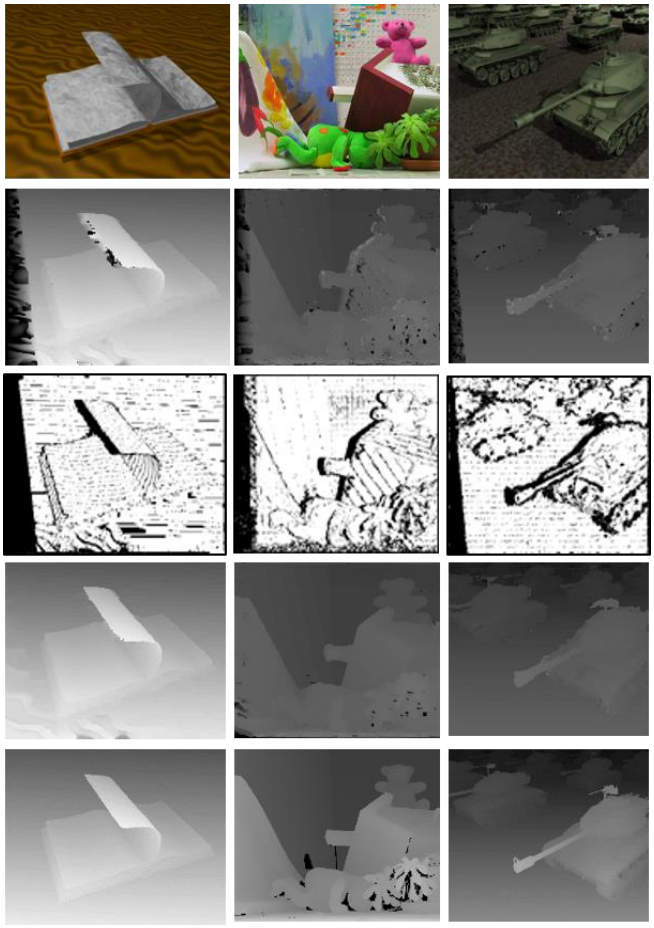
	Teddy	Cone	Venus
Jang's [10]	4.75	6.78	1.16
proposed	1.55	1.70	3.57



(a) (b) (c)

Figure 10. Occlusion detection: (a) Jang's method; (b) our method; (c) ground truth

In the experiment, λ_1, λ_2, and λ_3 weights for occlusion detection are set to 0.7, 0.1, and 0.2 respectively. λ weights for depth estimation is set to 0.3, and λ_{occ} weight for occlusion refinement is set to 0.2 to balance each term of the energy function. Figure 11 represents the results of depth estimation and the results of hole filling. Figure 11 demonstrates that the proposed method improves the quality considerably. Occlusion regions are well refined.



(a) (b) (c)

Figure 11. Dynamic programming for hole filling: (a)book; (b) cone; (c) tank. First row are the color images, second row the results of initial disparity map, third row are the results of occlusion detection, fourth row are Final results, last row are ground truth

TABLE 2

Evaluation for occlusion map

Algorithm		CSBP [20]	Jang's [10]	Proposed	GC+occ [19]
Teddy	nonocc	11.10	6.34	7.81	11.20
	all	20.20	13.62	13.40	17.40
	disc	27.50	17.59	22.75	19.80
Cones	nonocc	5.98	4.96	5.70	5.36
	all	16.50	12.70	12.58	12.40
	disc	16.00	14.44	16.42	13.00

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-449

5. Conclusions

In this paper, we proposed the occlusion detection method for stereo matching and hole-filling method to generate 3D information. The proposed method exploits the MAP-MRF model to define the energy function for generating an initial disparity map. After optimizing the energy function via graph cuts, occlusion was detected by ordering, uniqueness, and color similarity constraints. Further, we assigned reasonable disparity values to occluded pixels using dynamic programming and applied edge preserving noise removal filter. Experimental results show that our method detects more accurate occlusion region compared with a conventional method, and our method produces more accurate disparity maps compared to other methods in terms of bad pixel rate.

Acknowledgment

This research was supported by the 'Cross-Ministry Giga KOREA Project' of the Ministry of Science, ICT and Future Planning, Republic of Korea (ROK). [GK15C0100, Development of Interactive and Realistic Massive Giga- Content Technology]

References

[1] A. Frick , F. Kellner , B. Bartczak and R. Koch, "Generation of 3D-TV LDV-content with time of flight camera," IEEE Int'l Conf. on 3DTV, pp. 45–48, 2009.

[2] W.S. Jang, Y.S. Ho, "Efficient disparity map estimation using occlusion handling for various 3D multimedia applications," IEEE Trans. Consumer Electronics, vol. 57, no. 4, pp. 1937–1943, 2011.

[3] E.K. Lee, Y.S. Ho, "Generation of high-quality depth maps using hybrid camera system for 3-D video," J. Visual Comm. Image Represent, vol. 22 no. 1, pp. 73–84, 2011.

[4] I. P. Howard, Perceiving in depth. New York: Oxford University Press, 2012.

[5] A. F. Bobick and S. S. Intille, "Large occlusion stereo," International Journal of Computer Vision, vol. 33 no. 3 pp. 1–20, 1999.

[6] H. Ishikawa and D. Geiger, "Occlusions, discontinuities, and epipolar lines in stereo," European Conference on Computer Vision, pp. 425–433, 1998.

[7] G. Egnal and R. Wildes, "Detecting binocular halfocclusions: empirical comparisons of five approaches," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24 no. 8, pp. 1127–1133, 2002.

[8] V. Kolmogorov, R. Zabih, "Computing visual correspondence with occlusions using graph cuts," IEEE International Conference on Computer Vision, pp. 508–515, 2001.

[9] T. Liu, P. Zhang, L. Luo, "Dense stereo correspondence with contrast context histogram, segmentation-based two-pass aggregation and occlusion handling," Advances in Image and Video Technology, pp. 449–461, 2009.

[10] W. S.Jang and Y. S. Ho, "Discontinuity preserving disparity estimation with occlusion handling," Journal of Visual Communication and Image Representation, vol. 25 no. 7, pp. 1595-1603, 2014.

[11] R. Zabih and J. Woodfill. "Non-parametric local transforms for computing visual correspondence," European Conf. Computer Vision, pp. 151–158, 1994.

[12] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," Int. J. Computer Vision, vol. 35, no. 3, pp. 269–293, 1999.

[13] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," Int'l Conf. Computer Vision, pp. 508–515, 2001.

[14] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," Exploring Artificial Intelligence in the New Millenium, pp. 239–269, 2003.

[15] H. Hirschmueller, "Stereo vision in structured environments by consistent semi-global matching," Computer Vision and Pattern Recognition, vol. 2, pp. 2386–2393, 2006.

[16] A. Concha, and J. Civera, "n evaluation of robust cost functions for RGB direct mapping, " European Conf. on Mobile Robots, pp. 1-8, 2015.

[17] Y. Boykov, O. Veksler and R. Zabih, "Faster approximate energy minimization via graph cuts", Tran. Pattern Analysis and Machine Intelligence, vol 23, no.11, pp 1222-1239, 2001.

[18] K. He, J. Sun, and X. Tang, "Guided image filtering," European Conf. Computer Vision, pp. 1–14, 2010.

[19] V. kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," In Proc. IEEE International Conference on Computer Vision, pp. 508-515, 2001.

[20] Q. Yang, L. Wang, N. Ahuja, "A constant-space belief propagation algorithm for stereo matching," In Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1458-1465, 2010.

[21] C. Richardt, D. Orr, I. Davies, A. Criminisi, and N. A. Dodgson. "Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid," European Conf. Computer Vision, pp. 6311-6316, 2010.

[22] D. Scharstein and R. Szeliski. Middlebury Data Sets [Online]. Available: http://vision.middlebury.edu/stereo/

Author Biography

Eu-Tteum Baek received his B.S. degree in computer science and engineering from Chonbuk National University, Korea, in 2012 and M.S. degree in Information and Communication Engineering at the Gwangju Institute of Science and Technology (GIST), Korea, in 2015. He is currently working towards his Ph.D. degree in the Department of Information and Communications at GIST, Korea. His research interests are 3D digital image processing, depth estimation, and realistic broadcasting.

Yo-Sung Ho received his B.S. in electronic engineering from the Seoul National University, Seoul, Korea (1981) and his Ph.D. in electrical and computer engineering from the University of California, Santa Barbara (1990). He worked in Philips Laboratories from 1990 to 1993. Since 1995, he has been with the Gwangju Institute of Science and Technology, Gwangju, Korea, where he is currently a professor. His research interests include image analysis, 3D television, and digital video broadcasting.