The 2012 International Workshop on Advanced Image Technology

Depth Map Boundary Enhancement Using Random Walk

Gyo-Yoon Lee and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
261 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Korea
E-mail: {gyoyoon, hoyo}@gist.ac.kr

***Abstract*—** Depth information is essential for depth image-based rendering (DIBR), which is one of the rendering processes for the virtual view using a color image and its corresponding depth map. Although there are several depth estimation methods, more accurate depth estimation is still required. Since inaccurate depth information along object boundaries causes serious rendering errors, depth boundary information is very important. In this paper, we propose a new depth map filtering algorithm to solve the boundary mismatch problem. In order to preserve depth boundaries, we smooth depth values of the non-boundary region using the random walk probability. After calculating a similarity measure between the current pixel and its neighboring pixels using the random walk probability model, we have applied a weighted averaging filter based on the random walk probability. Finally, we have obtained a depth map with higher quality compared to other methods. Experimental results showed enhancement of depth boundaries.

Keywords: depth map, random walk probability, bilateral filter

I. INTRODUCTIONS

As the three-dimensional (3D) video becomes attractive in a variety of 3D multimedia applications, the multi-view video with corresponding depth map is essential: they are often called as multi-view video-plus-depth (MVD) data. In near future, consumers will be able to experience 3D depth impression and choose their own viewpoints in the immersive visual scenes created by 3D videos. Recently, the ISO/IEC Moving Picture Experts Group (MPEG) has investigated for the standardization on 3D and multi-view video coding [1].

It is important to estimate accurate depth information from real-world scenes for the high quality multi-view video. Although various depth estimation methods have been studied in the field of computer vision, accurate measurement of depth information from natural scenes is still an unsolved problem.

In general, depth estimation methods can be classified into two categories: passive depth sensing and active depth sensing. The active depth sensing method usually uses physical sensors, such as laser, infrared ray (IR), or light pattern, to obtain depth information from natural scenes directly. Structured light patterns [2] and depth cameras [3] [4] are good examples of these approaches. Nevertheless, these direct depth estimation tools and systems are quite expensive for consumers. Besides, the depth camera captures only a low-resolution depth map. Thus, we need to up-sample the low-resolution depth map to obtain a high-quality 3D video. Recently, time-of-flight (TOF) depth cameras of low price and small size have been introduced and applied for 3D home video games and other multimedia services.

On the other hand, the passive depth sensing method calculates depth information indirectly from 2D images captured by two or more cameras. Typical examples include shape from focus [5] and stereo matching [6]. The advantage of indirect depth estimation is low price because we can create depth maps using ordinary color cameras. However, accuracy of the depth map depends on color image characters. And it cannot obtain depth information in occlusions and textureless regions. The accurate depth map affects the 3D video quality. Especially, inaccurate depth values along the object boundary cause the serious 3D rendering error. Thus, it is needed to match the depth map and the color image to improve the depth map and 3D rendering quality.

In this paper, we propose the depth filter to reduce mismatch depth values along object boundaries. We design the smoothing filter which consists of the color-based weighting factor. The weighting value is random walk color similarity between the current pixel and neighboring pixels of the defined block. Experimental results show that our method based on the random walk probability can improve boundary matching accuracy. Besides, the proposed method efficiently handles depth values of the occlusion region. The enhanced depth map also helps improve the 3D rendering quality.

II. MULTI-VIEW DEPTH ESTIMATION

A. Disparity and Depth

Fig. 1 illustrates the relationship between disparity and depth. Suppose that a certain 3D point is projected onto the right image plane and it is located at (x_r, y). This 3D point is also projected onto the left image plane and it is located at (x_l, y). Then, the relationship between disparity d and depth Z can be defined by

$$Z = \frac{Bf}{d} = \frac{Bf}{x_l - x_r} \tag{1}$$

This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2011.

where B represents the camera distance and f represents the focal length of each camera. Eq. (1) proves that we can find the depth if we estimate the disparity by using the correspondence of multi-view videos.

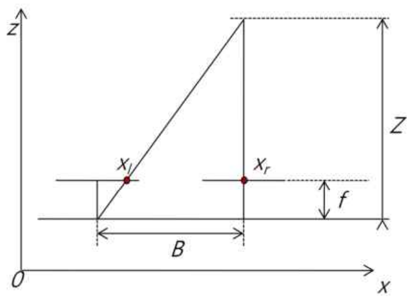


Figure 1. Relationship between disparity and depth

B. Disparity Computation

The task of stereo matching is the computation of 3D data from 2D input images. It is exactly what the human visual system is doing when we perceive depths. Since two images captured by our eyes are obtained from slightly different perspectives, the position of a scene point in one view is horizontally displaced in the other view. The amount of the displacement allows reasoning about the depth of the scene point. As shown in Figure 2, we determine pairs of points that correspond to same scene point in binocular images in stereo matching. The length of the horizontal displacement vector is commonly called as disparity. Basically, the disparity of a pixel is inversely proportional to the distance of the pixel from cameras.

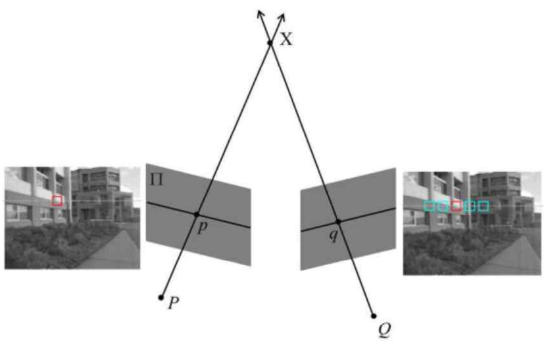


Figure 2. Principle of stereo matching

According to the way to find out correspondence points, we can classify stereo matching algorithms into two categories: local and global correspondence methods. The local correspondence method relies on local information to determine the corresponding point. On the other hand, the global correspondence method depends on information of a whole scanline or the entire image to compute disparities. While the local correspondence method is sensitive to ambiguous regions, such as uniform zones or occlusion regions, the global method is less sensitive to regions. However, we need more complicated computation to estimate disparities in the global approaches. In general, global correspondence methods consider smoothness constraint to generate the depth map. Most global approaches are formulated by energy minimization. When we define the energy $E(\varepsilon)$ at disparity ε, $E(\varepsilon)$ is calculated by

$$E(\varepsilon) = E_{data}(\varepsilon) + \lambda E_{reg}(\varepsilon) \tag{2}$$

where $E_{data}(\varepsilon)$ is the cost of matching left and right images at a disparity ε, and $E_{reg}(\varepsilon)$ is a regularization term to preserve the discontinuity of disparities at a disparity ε. In the local correspondence method, we assume that tiny image patches have similar intensity patterns across views. In order to find out the point of the maximum correspondence, the local approach usually makes a window move on the scan line on the other view image. The final disparity is obtained by selecting the point of the highest matching score. Since the matching score of a pixel in the local method is not influenced on disparities of neighboring pixels, we have only to concentrate on the pixel locally. In contrast to the local approaches, the global correspondence method generates their smoothness models using neighboring pixels.

C. Depth Enhancemetn Filter

Since the passive depth estimation quality depends on the color image character, many post processing methods are proposed. Joint bilateral filter is most used to obtain clear object boundaries [7].

$$D_p = \frac{1}{k_p} \sum_{q \in \Omega} I_q f(\|p-q\|) g(\|I_p - I_q\|) \tag{3}$$

where D_p is the depth values of the current pixel p. $f(\|p-q\|)$ means the distance between p and q. $g(\|I_p - I_q\|)$ is a Gaussian color distribution between pixels and k_p is a normalization factor. By making the depth values of similar color pixels smooth, Eq. (3) generates improved boundary information.

Yang *et. al* proposed the iterative depth filter which has the cost function [8]. The cost function includes the joint bilateral filter. By using the cost function, the smallest cost depth value of depth candidates (current, left, right, top and bottom pixels) becomes the depth value of the current pixel.

$$C_{(i)}(y,x,d) = \min(\eta * L, (d - D_{(i)}(y,x)^2)$$

$$D(x,y) = \arg\min_{d \in d_p} \frac{\sum_{u \in u_p} \sum_{v \in v_p} W(u,v) C(u,v,d)}{\sum_{u \in u_p} \sum_{v \in v_p} W(u,v)} \tag{4}$$

It is affected by neighboring pixel which has already modified depth values.

Lee *et. al* proposed the depth enhancement method which used temporal information [9]. It defines the temporal weighting function by using motion estimation. The temporal weighting function also enhances coding efficiency.

III. RANDOM WALK DEPTH FILTER

We define the weighting function which is determined by color similarity. First, we define the block of the depth map. The block size depends on the depth map resolution. In case of the high-resolution depth map, the large block size is appropriate. We find similarity between the center pixel and border pixels of the define block. We calculate the color similarity which represents the color variation along paths from border pixels to the center of the block.

We design the weighted smoothing filter. Since similar color pixels have similar depth values in the block, we use color similarity and discontinuity for the depth boundary. The weighting value is defined as color similarity between the center pixel and border pixels of the current block. Figure 3 shows the color similarity between the center and borders.

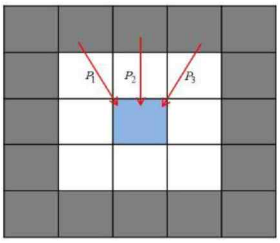


Figure 3. Color similarity between the center and borders of the block

In Fig. 3, $P_1, P_{2,...}$ means the probability which is color similarity. Eq. (5) represents the depth filter using probability values.

$$D_n(i) = \sum P(i,j) D_{n-1}(j) \tag{5}$$

where *i, j* are current pixel and border pixels, respectively. *n* means the iterative number. Since proposed depth filter is the smoothing filter using the probability, it does not rapidly change the depth values. Thus, by performing the depth filter several times we gradually change the depth values. D represents the depth map.

If we use only the color difference of near pixels, the unclear color edge disturbs matching the depth map. So, we define the color similarity as the random walk probability. Since the random walk probability includes not only color values but also color variation between pixels, it includes color information of the border and neighboring pixels. Because it considers color information of all pixels in the block, proposed depth filter is efficient for ambiguous color region along the object boundary.

The random walk probability is based on graph theory [10]. A graph consists of a pair $G = (V, E)$ with vertexes $v \in V$ and edges $e \in E \subseteq V \times V$. There is the cost between vertices which are neighboring. Eq. (6) means the cost between pixel *i* and *j*.

$$w_{ij} = \exp\left(-\frac{\left|z_i - z_j\right|^2}{\sigma}\right) \tag{6}$$

The random walk probability is appropriate to find clear color boundary. Thus it is efficient for the depth map boundary. The random walk probability is the sum of cost along the path from border pixels to the center pixel in the block. It is same as color variation from far pixel to the current pixel. The proposed method finds the obvious boundary using color variation. However, there are many paths to the current pixel. Since each path has the sum of cost, we need to select one path to define the random walk probability. Figure 4 shows the probability corresponding path.

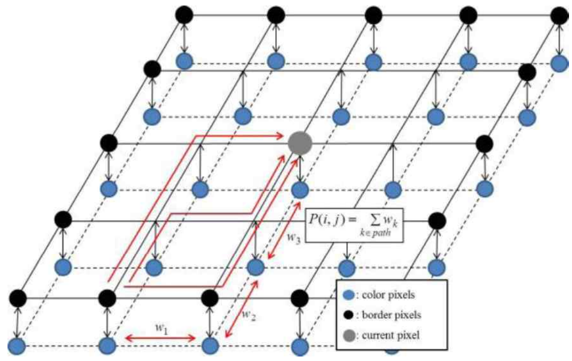


Figure 4. Sum of cost along paths

The random walk probability is defined as minimum sum of path cost. We solve the minimization problem using the Dirichlet problem.

$$D[x] = \frac{1}{2} \sum_{i,j \in E} w_{ij}(x_i - x_j)^2 = \frac{1}{2} x^T L x \tag{7}$$

where L is the Laplacian matrix. The Laplacian matrix is

$$L_{ij} = \begin{cases} d_i & \text{if } i = j, \\ -w_{ij} & \text{if } v_i \text{ and } v_j \text{ are adjancent,} \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Eq. (7) is composed of border pixels and other pixels. Therefore, we decomposed Eq. (5) into

$$D[x] = \frac{1}{2}[x_{border}^T \quad x_{non-border}^T] \begin{bmatrix} L_{border} & B \\ B & L_{non-border} \end{bmatrix} \begin{bmatrix} x_{border} \\ x_{non-boder} \end{bmatrix} \tag{9}$$

To minimize Eq. (7), we find critical point.

$$L_{non-border} X = -B^T M_{borders} \tag{10}$$

It is easy to solve Eq. (10) which is composed of simple matrixes calculation. The random walk probability which is obtained by Eq. (10) is applied to Eq. (5). Although complexity of process for the random walk probability of each pixel is high, we solve the complexity problem using parallel processing. Since calculation of each pixel does not affect operation of other pixels, it is appropriate for parallel processing. So we reduce operation time using general-purpose computing on graphics processing units (GPGPU).

IV. EXPERIMENTAL RESULT

In our experiments, we used various sequences and depth maps which are obtained by using various depth estimation methods, such as belief propagation (BP), depth estimation reference software (DERS), depth cameras etc. Table I explains the sequences for experiment. For experiments, the 7×7 block size is used.

Figure 5 and Fig. 6 show enhanced depth maps and Fig. 7 and Table II represent the synthesized view quality.

TABLE I. SEQUENCES, RESOLUTION AND DEPTH ESTIMATION METHODS

Sequence	Resolution	Depth estimation
Cafe	1920×1080	Hierarchical BP
Newspaper	1024×768	DERS
Office	640×480	Kinect camera

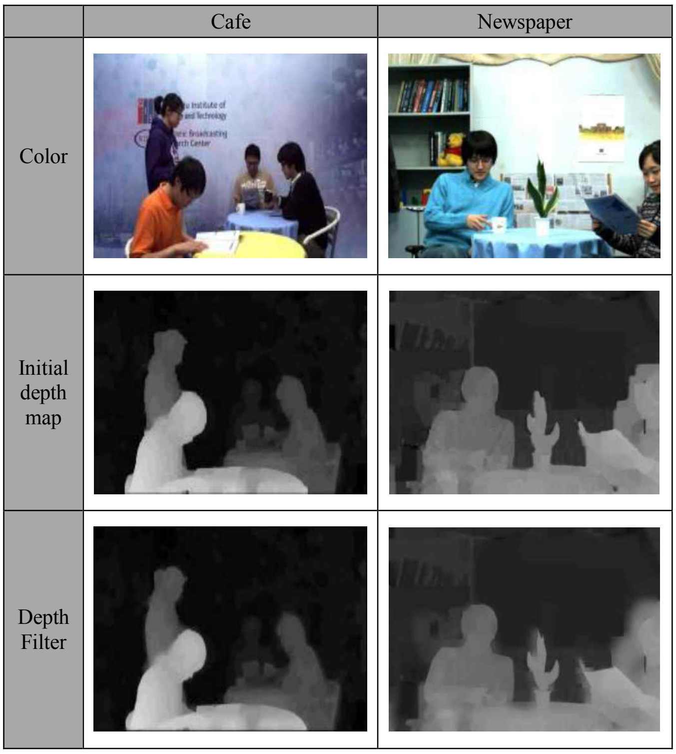
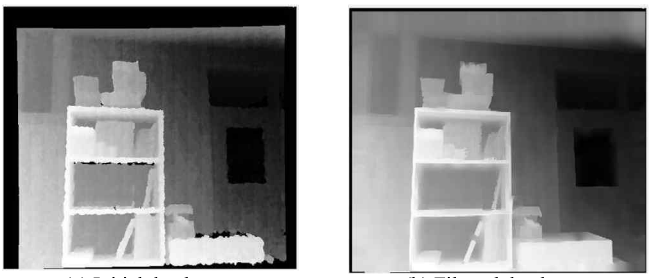


Figure 5. Comparisons of depth maps



(a) Initial depth map (b) Filtered depth map

Figure 6. Kinect camera depth map

(a) Original depth map (b) Filtered depth map

Figure 7. Synthesized right view

TABLE II. PSNR OF SYNTHESIZED VIEWS

	Original depth (dB)	Filtered depth map (dB)
Café	32.1729	32.6592
Newspaper	29.1415	29.1720

V. CONCLUSION

In this paper, we have proposed a depth filter to enhance depth boundary information. Considering color variation in the block, we improve the confidence of depth information. To design the filter including color information, we find the random walk probability from border pixels to the center pixel in the block. We also perform a smoothing filter using the random walk probability for weighting values. Experimental results show that proposed depth filter enhances clear boundary information. Besides, the enhanced depth map improves quality of the synthesized view.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 M8595, "FTV-Free Viewpoint Television," 2002.

[2] D. Scharstein, and R. Szeliski, "High-accuracy stereo depth maps using structured light," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 195-202, June 2003.

[3] Y. Netzer, "Depth sensing camera systems and methods," ed: Google Patents, 1999.

[4] G. Iddan and G. Yahav, "3D imaging in the studio and elsewhese," Proc. of SPIE Vidometrics and Optical Methods for 3D Shape Measurements, pp. 48–55, Jan. 2001.

[5] S.K. Nayar, Y. Nakagawa, "Shape from focus," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16(8), pp. 824-831, 1994.

[6] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame," Intel Journal Computer Vision, vol. 47, no. 1, pp. 7-42, 2002.

[7] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," ACM Transactions on Graphics, vol. 26, p. 96, 2007.

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

[9] S.B. Lee, Y.S. Ho, "Temporally Consistent Depth Map Estimation Using Motion Estimation for 3DTV," International Workshop on Advanced Image Technology (IWAIT), pp. 149(1-6), 2010.

[10] L. Grady, "Random walk for image segmentation," IEEE transactions on pattern analysis and machine intelligence, pp. 1768-1783, 2006.