Region Based Stereo Matching Method with Gradient and Distance Information

Yong-Jun Chang and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdangwagi-ro, Buk-gu, Gwangju, 61005, Republic of Korea
E-mail: {yjchang, hoyo}@gist.ac.kr

***Abstract*— Stereo matching methods estimate depth information of captured images. One way to estimate accurate depth values is to use the distance information. This method enhances the disparity map by preserving the edge region. In order to preserve the depth discontinuity near the edge region, it uses the distance information as a new weighting value for the matching cost function. However, this method has a high complexity problem. To overcome this problem, we propose region based stereo matching method with gradient and distance information. Since the distance transform calculates the pixel distance from the edge region, we can classify whether the pixel is near the edge region or not. In other words, some regions near the edge have small distance transformed values. For this reason, our method divides regions depending on the value of distance transformed pixel. After that, different cost functions are applied to each region for improving the computation efficiency.**

I. INTRODUCTION

Depth information of the captured scene is one of the important elements to generate a three-dimensional (3D) content. There are several methods to get depth information of the object. First, camera based methods can give depth information. A depth camera measures the depth from the object using the infrared rays or the structured light. Both of them estimate the depth information directly. Thus, we can acquire the depth map quickly. However, this camera has some limitations on shooting environment. Since the infrared ray is easily influenced by the sunlight, a lot of depth errors can be generated in the depth map. In addition, the depth camera cannot measure the depth of the object that is located far from the camera. For these reasons, the depth camera is not suitable for an outdoor environment.

A stereo matching method is the second way to get depth information of the object. This method uses stereo images that are captured by a stereo camera to estimate the depth value. Therefore, unlike the depth measurement using the depth camera, it is not influenced by the outdoor environment. The stereo matching method estimates depth information using the characteristic of binocular disparity. There are corresponding points between two stereo images. The disparity value is the result of the subtraction between corresponding points. This value has a large number if the object that includes those points is located far from the camera. On the other hand, the disparity value is very small if the object is located near the camera. Thus, the depth value is represented as the disparity value using this characteristic. We call the result image of the stereo matching method a disparity map.

In general, the 3D content is generated by depth information. Therefore, the accurate disparity map is very important in the 3D content generation. Although the stereo matching method estimates the depth value without the limitation of shooting environment, there is a matching ambiguity problem in the textureless region and the edge region. In terms of the textureless region, a cross-scale cost aggregation method was proposed to improve the disparity accuracy [1]. The stereo matching method using the distance transform was also proposed to preserve the depth discontinuity in the edge region [2-4]. In this paper, we propose an adaptive stereo matching method using various information to preserve the depth discontinuity with lower complexity than that of the conventional method [3].

II. DEPTH PRESERVING IN EDGE REGION

A. Distance Transformation

The pixel distance from the edge region is calculated by the distance transform [2, 3]. This transformation uses a 3×3 window kernel to change the pixel value to the distance value. The window kernel is defined in (1).

$$r_{i,j}^{k} = \min \begin{bmatrix} r_{i-1,j-1}^{k-1} + \beta & r_{i,j-1}^{k-1} + \alpha & r_{i+1,j-1}^{k-1} + \beta \\ r_{i-1,j}^{k-1} + \alpha & r_{i,j}^{k-1} & r_{i+1,j}^{k-1} + \alpha \\ r_{i-1,j+1}^{k-1} + \beta & r_{i,j+1}^{k-1} + \alpha & r_{i+1,j+1}^{k-1} + \beta \end{bmatrix} \quad (1)$$

(a) Original image (b) DT map

Fig. 1 Result of distance transformation

In (1), where $r_{i,j}^{k}$ represents the kth transformed value of the current pixel at (i, j). When the number of k is zero, it means

the initial edge information. α and β regulate the strength of the transformation. This kernel is applied to the edge image that has the pixel value of zero in the edge region and the value of one in the other region like the binary image. The distance value of the current pixel is determined by the neighboring pixel that has the minimum distance transformed value in the kernel. Therefore, the pixel has the large distance value if it is located far from the edge region.

Fig. 1 shows the result of the distance transformation. In Fig. 1, the original color image is represented in Fig. 1(a). A distance transformed (DT) map is represented in Fig. 1(b). In Fig. 1(b), pixels near the edge region have lower pixel values than other regions.

B. Depth Estimation Using Distance Transformation

The distance transformation calculates the pixel distance from the edge region. Therefore, not only the pixel in the edge region but also pixels near that region are represented by specific distance values after the distance transformation. Thus, the accuracy of stereo matching near the edge region is improved using the distance transformation. Jang et al. proposed the depth preserving method in the edge region using this transformation [3]. This method obtains the DT map first. After that, weighting values are calculated based on the map. In addition to those weighting values, the matching cost is computed by the cost function that is defined as follows

$$D_s(d_s) = \frac{\sum_{t\in N(s)} W_{s,t}(dt_t)|I_L(x_t,y_t)-I_R(x_t+d_s,y_t)|}{\sum_{t\in N(s)} W_{s,t}(dt_t)} \quad (2)$$

where $D_s(d_s)$ is the matching cost depending on the disparity candidate d_s. This cost function is applied to the energy function as a data term. Where s represents the position of the current pixel, t means one of the neighboring pixels in the matching window $N(s)$, and $W_{s,t}(dt_t)$ is the weighting function that is used for the similarity measurement between two corresponding pixels. It is composed of two weighting functions. Those are represented in (3) and (4).

$$W_{s,t}(dt_t) = f(dt_t)\cdot g\left(\left|I_{L,s} - I_{L,t}\right|\right) \quad (3)$$

$$f(dt_t) = 1 - e^{-\frac{dt_t^2}{2\sigma_f}}, g\left(\left|I_{L,s} - I_{L,t}\right|\right) = e^{-\frac{|I_{L,s}-I_{L,t}|^2}{2\sigma_g}} \quad (4)$$

In (3), dt_t means the distance transformed value at pixel position t. In (4), each function shows the weighting function that composes the total weighting function $W_{s,t}(dt_t)$. $f(dt_t)$ calculates the weighting values using the distance transformed value. $g\left(\left|I_{L,s} - I_{L,t}\right|\right)$ is the weighting function of color differences between the current pixel and the neighboring pixels in the kernel. The data term is used for the energy function as follows

$$E(d) = \sum_s D_s(d_s) + \sum_{s,t\in N(s)} S_{s,t}(d_s, d_t) \quad (5)$$

where $S_{s,t}(d_s, d_t)$ is the smoothness term that checks the disparity continuity between the current pixel and the neighboring pixel. d_s and d_t represent disparity candidates of pixel s and t, respectively. The total energy function $E(d)$ is formulated by the data term and the smoothness term.

III. Adaptive Stereo Matching Method Using Various Information in Stereo Images

A. Problem of Distance Transformation

The distance transformation calculates the distance value using the 8bit gray level. Therefore, the pixel value in the DT map has one of the numbers from 0 to 255. In Fig. 1(b), there are white regions in the DT map. Pixels in those regions have 255 values. It means that all pixels in those regions are very far from the edge region. The conventional method which is proposed by Jang et al. applies (2) to all pixels in the DT map [3]. However, if all the pixels have same values, the weighting function $f(dt_t)$ in (4) also calculates same weighting values. This problem increases the unnecessary computation. As a result, the conventional method has a high complexity problem.

B. Distance Transformation Based Region Division

The division of image region makes it possible to avoid the unnecessary computation of weighting value by applying different cost functions to each region [4]. In order to divide the image region, the distance transformation is used. Fig. 2 shows a result of region division using the DT map.

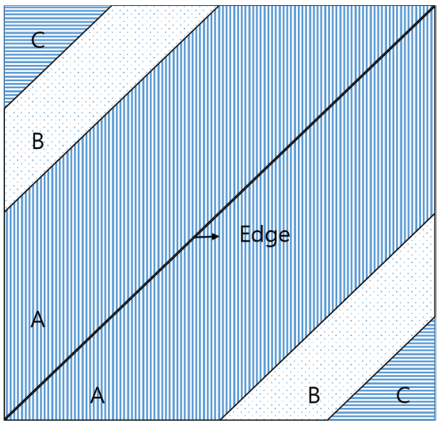


Fig. 2 Region division using distance transformation

In Fig. 2, the region A is the group of pixels which are located near the edge region. A threshold τ is defined to separate the region A and other regions such as the region B and the region C. If the current pixel in the DT map has the value that is equal or larger than 0 and smaller than τ, then it is regarded as the pixel in the region A. In terms of the region B, if the distance value of the current pixel is equal or larger than τ and smaller than 255, we regard that the pixel is located in the region B. Lastly, the region C is the group of pixels that have 255 distance values in the DT map.

C. Adaptive Stereo Matching with Region Division

The region A in Fig. 2 is composed of pixels near the edge region including edge pixels. Therefore, it is important to

estimate exact corresponding points in this region for the depth preserving problem. Even though the equation (2) performs well in the edge region for the depth preserving problem, it has a high complexity problem. Thus, we apply a pixel based cost computation using color and gradient information to this region [4, 5]. Pixel values in the edge region have very large differentials. The gradient value of the original image shows those differentials in the gradient image. Accordingly, corresponding points in the edge region can be estimated well by using gradient and color images. The cost function is defined as follows

$$D_s(d_s) = k \cdot |I_L(x_s, y_s) - I_R(x_s + d_s, y_s)| + (1-k) \cdot |\nabla I_L(x_s, y_s) - \nabla I_R(x_s + d_s, y_s)|, if\ 0 \le dt_s < \tau \quad (6)$$

where I and ∇I represent color and gradient values, respectively. Those two types of information are combined to form the cost function D_s. The strength of each information is regulated by the weighting factor k. This equation is applied to pixels which have one of the values from 0 to τ in the DT map.

In terms of the region B, pixels in this region have specific distance values unlike pixels in the region C which have the distance value of 255. In other words, all pixels in the region B have one of the values from τ to 255. However, those pixels are located further from the edge region than pixels in the region A. Since this region does not include the edge region, there are no pixels which have the large differential of pixel values. Therefore, it is difficult to estimate accurate corresponding points using color and gradient images. In order to improve this problem, we add the gradient weighting term to the cost function that is defined in (2). A new cost function is represented as follows

$$D_s(d_s) = \frac{\sum_{t \in N(s)} G_{s,t} W_{s,t}(dt_t) |I_L(x_t, y_t) - I_R(x_t + d_s, y_t)|}{\sum_{t \in N(s)} G_{s,t} W_{s,t}(dt_t)} \quad (7)$$

where $G_{s,t}$ is the weighting function using the gradient image. This weighting function calculates differences of gradient values between the current pixel and the neighboring pixels in the kernel. The weighting function is formulated in (8).

$$G_{s,t}\left(\left|I_{gL,s} - I_{gL,t}\right|\right) = e^{-\frac{\left|\nabla I_{L,s} - \nabla I_{L,t}\right|^2}{2\sigma_d}} \quad (8)$$

In (8), where $\nabla I_{L,s}$ is the gradient value of the current pixel in the window kernel. $\nabla I_{L,t}$ means that of the neighboring pixel in the kernel. This function is computed using the gradient image.

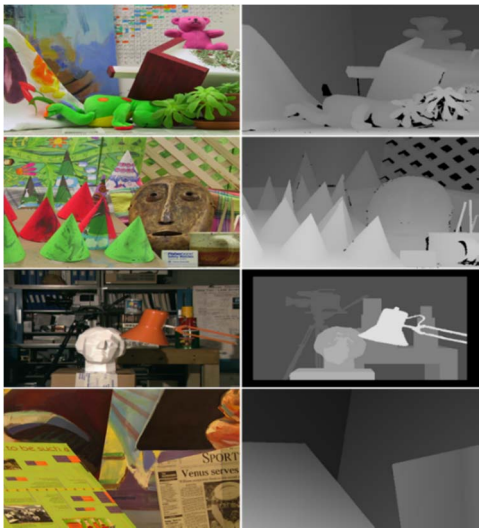
The region C has pixels which have the distance value of 255. For this reason, the weighting value using the DT map does not have any significant effect on the stereo matching. In addition, this region has a small variation of intensity value in the color image. Therefore, we apply a different cost function as follows

$$D_s(d_s) = \frac{\sum_{t \in N(s)} k \cdot \left|I_{gL}(x_s, y_s) - I_{gR}(x_s + d_s, y_s)\right| + (1-k) \cdot \left|\nabla I_L(x_s, y_s) - \nabla I_R(x_s + d_s, y_s)\right|}{9}, \ if\ dt_s = 255 \quad (9)$$

where I_g is the pixel value of the gray scale image. The 3×3 window is used to calculate the cost function D_s.

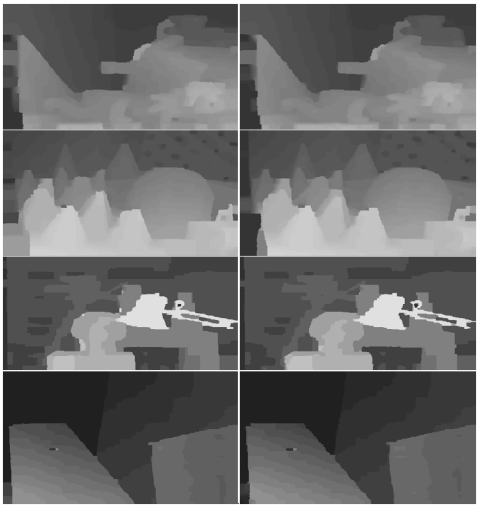
IV. Experiment Results

Four test images were used for implementation: *Teddy*, *Cones*, *Tsukuba* and *Venus*. The resolution of *Teddy* and *Cones* is 450×375. That of *Tsukuba* and *Venus* is 384×288 and 434×383, respevtivley. Those images are shown in Fig. 3. In Fig. 3(a), orders of the test images are *Teddy*, *Cones*, *Tsukuba* and *Venus* from top to bottom. The images represented in Fig. 3(b), they are ground truth images. The energy function that is defined in (2) is optimized by the hierarchical belief propagation [6]. In addition, the threshold τ is set to 200.



(a) Original image (b) Ground truth

Fig. 3 Original and ground truth images



(a) Conventional method (b) Proposed method

Fig. 4 Comparison of result images

Fig. 4 shows the result images of stereo matching. Fig. 4(a) and Fig. 4(b) are result images of the conventional method [3] and those of the proposed method, respectively. In Fig. 4, it is difficult to verify the differences between the conventional method and the proposed method. Thus, the bad pixel rate is also computed for the objective comparison. The error rate comparison is shown in Table Ⅰ.

TABLE Ⅰ
COMPARISON OF ERROR RATE

Algorithm		Conventional method [3]	Proposed method
Teddy	Nonoccluded region	10.68	9.88
	All region	20.61	19.93
	Discontinuity region	27.28	27.08
Cones	Nonoccluded region	6.69	6.22
	All region	17.52	18.24
	Discontinuity region	17.71	17.77
Tsukuba	Nonoccluded region	1.87	1.81
	All region	3.98	3.7
	Discontinuity region	10.26	10.09
Venus	Nonoccluded region	1.23	1.26
	All region	4.16	4.2
	Discontinuity region	15.59	14.61
Average		11.47	11.23

In terms of the error pixel, the pixel is regarded as the error when the difference between the disparity value of the pixel in the result image and that of the pixel in the ground truth is larger than one. All the values in this table are represented as a percentage. The comparison results showed that the average error rate of the proposed method was reduced by 0.24% compared with that of the conventional method [3].

TABLE Ⅱ
COMPARISON OF IMPLEMENTATION TIME

Algorithm	Conventional method [3]	Proposed method
	Time (sec.)	
Teddy	37.86	12.33
Cones	36.95	9.23
Tsukuba	6.28	1.92
Venus	11.47	4.34
Average	23.14	6.96

We also compared the implementation time between the conventional method and the proposed method. Table Ⅱ shows the comparison table of the implementation time. The proposed method performed faster than the conventional method. As a result, the average time of the proposed method was reduced by 16.18 seconds compared with that of the conventional method [3].

V. CONCLUSION

In this paper, we proposed an adaptive stereo matching method using various information such as the distance, the gradient and the color. Our method divides the image region based on the DT map first. In addition, three different cost functions are used depending on the characteristic of the divided image region. As a result, our method relieves the high complexity problem that is one of the problems in the conventional method. The average error rate of the proposed method is reduced by 0.24%. The average implementation time is also reduced by 69.92% compared with that of the conventional method.

ACKNOWLEDGMENT

This work was supported by the 'Civil-Military Technology Cooperation Program' grant funded by the Korea government.

REFERENCES

[1] K. Zheng, Y. Fang, D. Min, L. Sun, S. Yang, S. Yan, and Q. Tian, "Cross-scale cost aggregation for stereo matching," *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1590-1597, June 2014.

[2] G. Borgefors, "Distance transformations in digital images," *Computer Vision*, *Graphics*, *and Image Processing*, vol. 34, issue 3, pp. 344-371, June 1986.

[3] W.S. Jang and Y.S. Ho, "Discontinuity Preserving Disparity Estimation with Occlusion Handling," *J. Visual Communication and Image Representation*, vol. 25, no. 7, pp. 1595-1603, Oct. 2014.

[4] Y.J. Chang and Y.S. Ho, "Adaptive stereo matching method depending on characteristic of regions in stereo images," *Conf. Korean Institute of Smart Media*, pp. 135-138, April 2016.

[5] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," IEEE Conf. Computer Vision and Pattern Analysis, pp. 3017-3024, June 2011.

[6] J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, issue 6, pp. 679-698, Nov. 1986.