Pixel Based Cost Computation Using Weighted Distance Information for Cross-Scale Stereo Matching

Yong-Jun Chang, Yo-Sung Ho; Gwangju Institute of Science and Technology (GIST); Gwangju, Republic of Korea

Abstract

Depth information is one of the most important elements in generating three-dimensional (3D) content. Stereo matching methods estimate depth information using the binocular characteristic. The estimated depth information is typically represented by a disparity value. Therefore, two slightly different viewpoints are used to find the disparity value. However, in the homogeneous region, corresponding point finding is problematic since the area is textureless. In order to solve this problem, we propose a pixel based cost computation method using weighted distance information for cross-scale stereo matching. The proposed method uses a hierarchical structure to accurately estimate disparity values in the homogeneous region. We also employ the distance information to complement the pixel based cost function. The experiment results show that the proposed method exceeds the conventional cross-scale stereo matching in terms of produces accurate disparity values.

1. Introduction

In these days, there are a lot of 3D content in our lives. This 3D content gives realistic 3D images to people. The 3D content is usually made by various images such as stereo images and multi-view images. Stereo images give a 3D effect using the binocular disparity. Two images which are captured by a stereo camera have different viewpoints for same objects. If one of the objects is located near the camera, then the disparity of this object will have the large value. On the contrary, if the object is far from the camera, then the object will have the small disparity value. Therefore, human eyes can perceive depth information of objects using this characteristic.

Multi-view images are composed of many images which have different viewpoints. Hence, it is possible to see different views for the same scene depending on the viewer's position. Multi-view images are captured by multi-array cameras. However, using multi-array cameras sometimes causes inconvenient situation because of the huge camera array system. In order to overcome this negative point, view synthesis method using a few captured images was proposed [1].

The view synthesis generates an image which has a virtual viewpoint using captured images. The depth information of captured images is needed to generate the virtual viewpoint image. There are many ways to acquire the depth information of target objects. Stereo matching methods are generally used to obtain depth values from captured images. These methods use the same characteristic with human eyes. These methods check the binocular disparity between stereo images.

Stereo matching methods are composed of two different algorithms. First one is a global method and the other one is a local method. The global method considers whole pixels in the image to estimate depth values. Therefore, it estimates quite accurate depth values. However, this method has a high complexity problem. On the contrary, the local method considers specific pixels in the image to find depth values. Therefore, it is usually faster than the global method. However, this method usually has less accurate depth values in the result image.

Stereo matching methods find disparity values between two corresponding points in stereo images. Therefore, it is very important to search accurate corresponding points between two images. However, searching for corresponding points in some regions is sometimes difficult. These regions cause disparity errors in a disparity map which is the result of stereo matching methods. The homogeneous area is one of the regions to cause disparity errors in the result image. This region does not have any textures. For this reason, it is difficult to find corresponding points in this region. In this paper, we introduce the conventional method to estimate accurate disparity values in the homogeneous region. After that, we propose an improved method to enhance the conventional method.

2. Stereo Matching Method Using Cross-Scale Cost Aggregation

2.1 Hierarchical Structure for Cost Aggregation

A cross-scale cost aggregation is one of the methods to estimate accurate disparity values in homogeneous regions [2]. It uses the local stereo matching method to find the optimal corresponding point between stereo images. This method uses a hierarchical structure to aggregate matching costs of each scale image. Fig. 1 shows the hierarchical structure of stereo images.

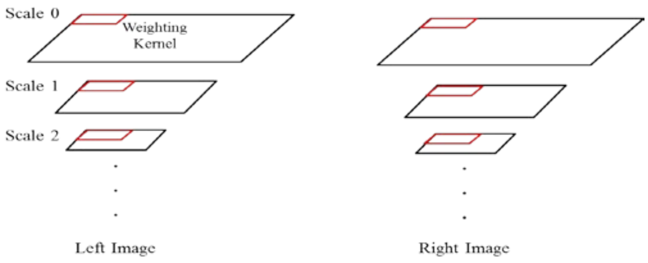


Figure 1. Hierarchical Structure of Stereo Images

The initial matching cost at each scale image is calculated by pixel based cost function [3]. This cost function is defined by Eq. 1.

$$C(i,d) = (1-\alpha)\cdot min(\|I(x_i,y_i) - I'(x_i-d,y_i)\|,\tau_1) + \alpha\cdot min(\|\nabla_x I(x_i,y_i) - \nabla_x I'(x_i-d,y_i)\|,\tau_2) \quad (1)$$

The cost function C is a weighted sum of both color information and gradient information. In Eq. 1, where i means the position of current pixel. Its pixel coordinate can be represented like (x_i, y_i). I is the pixel value of the color image. I' represents the color value of pixel which is the corresponding point of pixel i. The corresponding point can be defined by the disparity value d. Therefore, I' is the color value at $(x_i - d, y_i)$. $\nabla_x I$ and $\nabla_x I'$ are gradient values. τ_1 and τ_2 are truncation values which set upper boundaries of color differences and gradient differences.

The matching cost using the pixel based cost function usually has cost noise because the pixel based cost computation does not consider the cost consistency between the current pixel and neighboring pixels. Fig. 2 represents the cost noise which is caused by the pixel based cost computation.

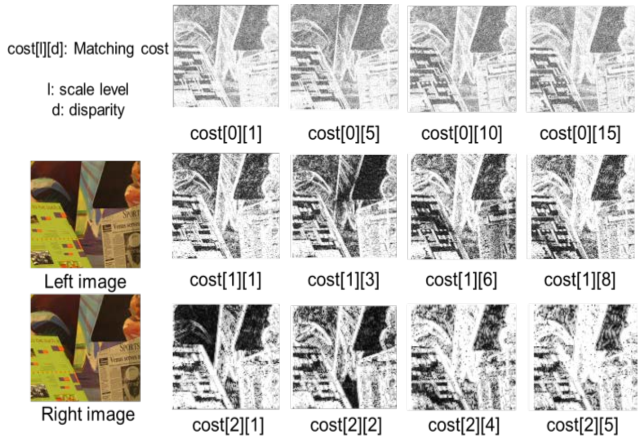


Figure 2. Matching Cost Noise

In Fig.2, each row represents the matching cost of different scale images. Each column shows the cost result depending on different disparity values. As you can see in these results, there are some cost noises among black regions. Actually, these regions should have similar matching costs. In order to reduce these noises, the intra scale cost aggregation is proposed [2].

2.2 Intra Scale Cost Aggregation

The intra scale cost aggregation refines the matching cost to reduce the cost noise. The weighted least square optimization is used to define the refined cost function. This function is defined as follows.

$$\tilde{C}(i,d) = arg \min_{s} \frac{1}{Z_i} \sum_{j \in N_i} K(i,j) \| s - C(j,d) \|^2 \qquad (2)$$

In Eq.2, where $\tilde{C}$ is the refined matching cost. K is the weighting kernel to reduce the cost noise. Eq. 2 finds the optimal matching cost for the pixel i using the weighted least square optimization. It calculates the squared difference between the cost of current pixel i and that of neighboring pixel j. Z_i is a sum of weighting values in the kernel K. Eq. 2 can be solved by a partial differential. As a result, the equation of intra scale cost aggregation is defined in Eq. 3.

$$\tilde{C}(i,d) = \frac{1}{Z_i} \sum_{j \in N_i} K(i,j) C(j,d) \qquad (3)$$

The noise of matching cost is refined by using Eq. 3. The refined matching costs are depicted in Fig. 3.

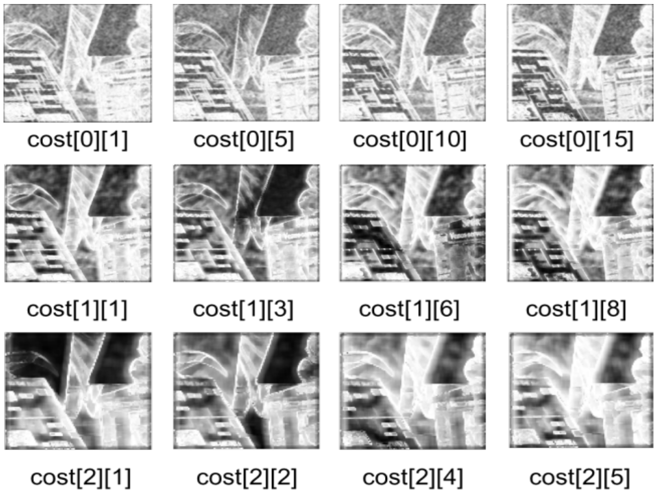


Figure 3. Refined Matching Cost

2.3 Inter Scale Cost Aggregation

The intra scale cost aggregation refines the matching cost using the weighted least square optimization. It compares the current pixel cost with neighboring pixel cost to reduce the cost noise. The initial cost at each scale image can be refined by this aggregation method. However, the intra scale cost aggregation does not consider the relationship of matching cost among different scale images.

The inter scale cost aggregation considers the cost consistency among different scale images [2]. In order to check the cost consistency, a new term is added to the equation of intra cost aggregation. It is defined in Eq. 4.

$$\hat{v} = arg \min_{\{s^k\}_{k=0}^{K}} \left(\sum_{k=0}^{K} \frac{1}{Z_{i^k}^{k}} \sum_{j^k \in N_{i^k}} K(i^k, j^k) \left\| s^k - C^k(j^k, d^k) \right\|^2 + \lambda \sum_{k=1}^{K} \left\| s^k - s^{k-1} \right\|^2 \right) \qquad (4)$$

In Eq. 4, where $\hat{v}$ represents the vector set of inter scale costs. k means the scale level. The first term shows the intra scale cost aggregation of each scale level. The second term of this equation checks the cost consistency between the current scale and the previous scale. λ is a regularization parameter.

From Eq. 2, we can induce the relationship between the intra scale cost and the inter scale cost. This relationship is acquired by applying partial differential to Eq. 4. As a result, the relationship between the intra scale cost and the inter scale cost is defined as follows.

$$A\hat{v} = \tilde{v},$$
$$\hat{C}^0(i^0, d^0) = \sum_{k=0}^{K} A^{-1}(0,k)\tilde{C}^k(i^k, d^k) \tag{5}$$

In Eq. 5, where $\tilde{v}$ means the set of intra scale costs. A is a tridiagonal matrix which is induced by partial differential of Eq. 4. Hence, the intra scale cost can be defined using the inverse matrix of A. $\hat{C}^0$ is the inter scale cost in the finest scale image.

2.4 Problem of Cross-Scale Cost Aggregation

The conventional cross-scale cost aggregation algorithm uses the pixel based cost function to calculate the initial cost of each scale image. The pixel based cost function uses color information and gradient information [3]. These two types of information are robust to the stereo matching near edge regions in stereo images and they are also robust to the stereo matching using the image which is under the various illumination. However, this cost function still has a matching ambiguity in homogeneous regions. Fig. 4 describes the matching ambiguity problem of the conventional cross-scale cost aggregation in textureless regions.

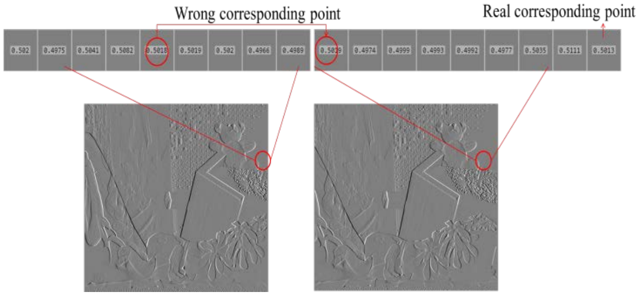
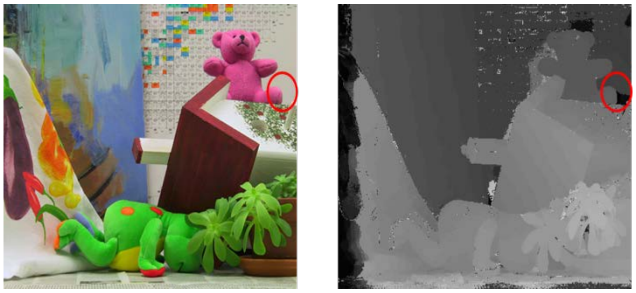


Figure 4. Matching Ambiguity Problem

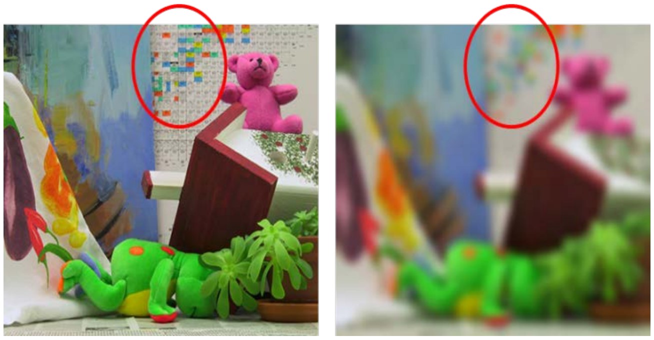


(a) Original Image *(b) Disparity Map*

Figure 5. Disparity Errors in Homogeneous Regions

In Fig. 4, two images are gradient images and circle areas represent textureless regions in stereo images. Not only the color image but also the gradient image does not have specific pixel values in textureless regions. For this reason, the pixels in these regions can have the lowest cost values at wrong corresponding points. This problem causes disparity errors in homogeneous regions like Fig. 5.

The cross-scale cost aggregation method also has a problem in some textured regions. Since this method uses the hierarchical structure to acquire more accurate disparity values in homogeneous regions, low scale images can lose textural edges in some regions. Fig. 6 shows the problem of hierarchical structure.



(a) 450 x 375 *(b) 57 x 47*

Figure 6. Problem in Low Scale Images

In Fig. 6, Fig. 6(a) describes the original scale image and circle regions have textural edges. On the contrary, same regions in Fig. 6(b) lose detail things of textural edges. Therefore, this problem also causes disparity errors in these regions.

Fig. 7 shows disparity errors in textural edges. As you can see in this picture, there are some disparity noises in box regions.

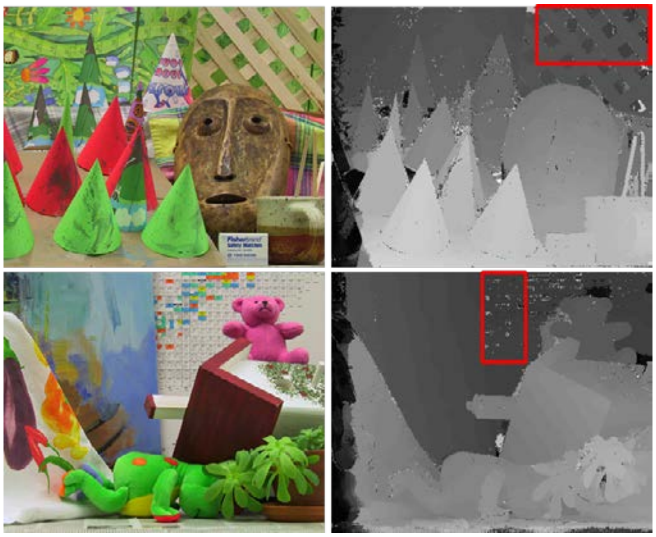


Figure 7. Disparity Errors in Textured Regions

In order to improve the disparity accuracy in homogeneous regions and some textured regions, we propose the modified initial cost function using weighted distance information.

3. Cost Function Using Distance Information

3.1 Distance Transform

The distance transform was used by Jang et al. to preserve discontinuity depth values in the disparity map [4]. The distance transform calculates the distance value from edge regions to the current pixel position. Eq. 6 is the kernel of distance transform.

$$r_{i,j}^{k} = min\begin{bmatrix} r_{i-1,j-1}^{k-1} + m & r_{i,j-1}^{k-1} + n & r_{i+1,j-1}^{k-1} + m \\ r_{i-1,j}^{k-1} + n & r_{i,j}^{k-1} & r_{i+1,j}^{k-1} + n \\ r_{i-1,j+1}^{k-1} + m & r_{i,j+1}^{k-1} + n & r_{i+1,j+1}^{k-1} + m \end{bmatrix} \quad (6)$$

In Eq. 6, where $r_{i,j}^{k}$ is the current distance value at k^{th} iteration. i, j represent the pixel coordinates. m and n are weighting factors. Hence, the distance value of current pixel at k^{th} iteration is replaced by the pixel which has the smallest distance value in previous iteration step. Fig. 8 shows how to do the distance transform.

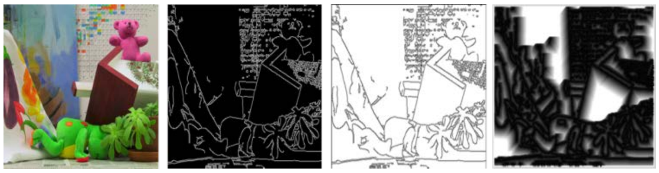

(a) Color (b) Edge (c) Reversal (d) Distance Map

Figure 8. Process of Distance Transform

Fig. 8(a) shows the original color image. Fig. 8(b) is an edge image of Fig. 8(a). The edge image is obtained by using Canny edge detection [5]. Fig. 8(c) is the reversal image of Fig. 8(b) and Fig. 8(d) is the distance map which is the result of distance transform. The distance map can be acquired using this process. In Fig. 8(d), we can check that the pixel which is located far from edge regions has high distance value. On the contrary, the pixel which is near edge regions has low distance value.

Jang's method uses distance information as weighting values in the matching cost function to preserve edge regions in the disparity map [4]. In this paper, we use distance information as one of the cost terms.

3.2 Modified Distance Transform

The conventional distance transform calculates the distance value from edge regions considering eight different directions [4]. However, stereo matching methods use rectified images to search corresponding points. For this reason, the corresponding point is always searched in the same scan line. Hence, it is enough to calculate the distance value in the x direction. We modify the kernel of distance transform as follows.

$$r_{i,j}^{k} = min\left[r_{i-1,j}^{k-1} + n \quad r_{i,j}^{k-1} \quad r_{i+1,j}^{k-1} + n\right] \quad (7)$$

Eq. 7 represents the kernel of modified distance transform. In Eq. 7, where n is the weighting factor. The process of modified distance transform is same with that of conventional distance transform. Fig. 9 describes new distance maps.

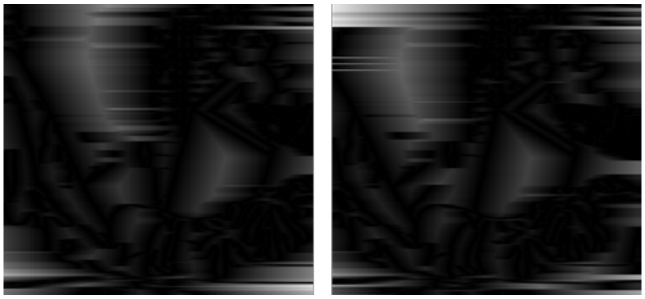

(a) Left Image (b) Right Image

Figure 9. Distance Transform Using Modified Kernel

3.3 Modified Cost Function

The initial cost function which is defined in Eq. 1 uses two types of information. First one is color information and the other one is gradient information. Our proposed method adds the distance information term to the equation of conventional matching cost. Eq. 8 represents the modified cost function.

$$\begin{aligned} C(i,d) &= \alpha' \cdot min(\|I(x_i,y_i) - I'(x_i - d, y_i)\|, \tau_1) \\ &+\beta' \cdot min(\|\nabla_x I(x_i,y_i) - \nabla_x I'(x_i - d, y_i)\|, \tau_2) \\ &+\gamma' \cdot \|dt(x_i,y_i) - dt'(x_i - d, y_i)\| \end{aligned} \quad (8)$$

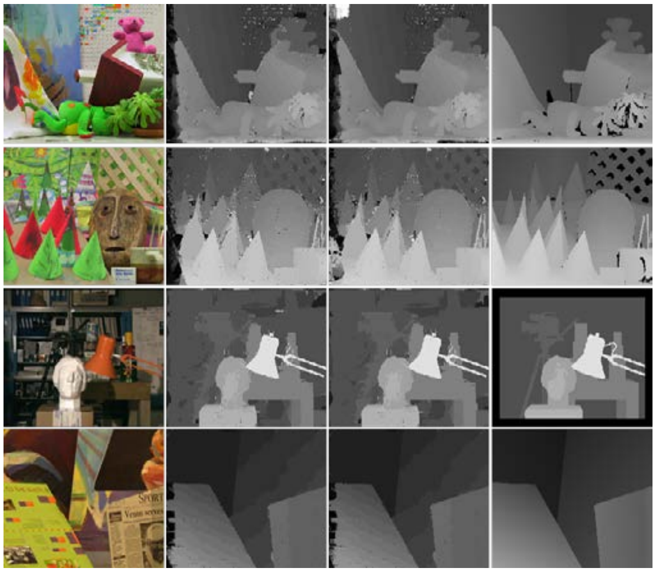
In Eq. 8, the last term represents distance information. dt is the distance value in the distance map. α', β' and γ' are weighting values. We add the weighted distance information to the conventional cost function because the distance value causes matching errors in occlusion regions.

4. Experiment Results

The proposed method uses the initial matching cost with weighted distance information. The proposed cost function has three different information terms. First term is about color information, second term is about gradient information and the last term is about distance information. Each term has the weighting value. In Eq. 8, where α' is set to 0.1, β' is set to 0.89 and γ' is set to 0.01. The weighting value α which is used in Eq. 1 is set to 0.89. Initial matching costs of each scale image are aggregated by the method of cross-scale cost aggregation [2]. We used a bilateral to refine the matching cost [6]. λ in Eq. 4 is set to 0.3.

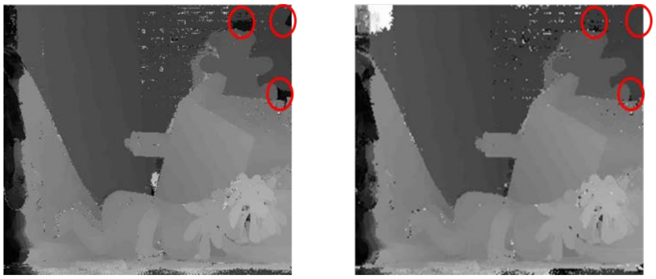
In order to implement the proposed cost function, four test images are used: *Teddy*, *Cones*, *Tsukuba* and *Venus*. We compared the error rate of proposed method with that of the conventional method which used the initial matching cost with color information and gradient information. The error rate checks error pixels in the disparity map. If the disparity difference between the pixel of the result image and that of the ground truth image is larger than 1, then that pixel is considered as an error pixel [7].

Fig. 10 shows result images. The first row image is *Teddy*, the second row image is *Cones*, the third row image is *Tsukuba* and the last row image is *Venus*. Pictures in Fig. 10(a) are original color images. Fig. 10(b) represents result images using the CA method. The CA method means the conventional matching cost function. Fig. 10(c) describes disparity maps using the proposed method and pictures in Fig. 10(d) are ground truth images.



(a) Original (b) CA Method (c) Our Method (d) Ground Truth
Figure 10. Comparison Images

As you can see in Fig. 10, our proposed method has better disparity quality in homogeneous regions than the conventional method. Fig. 11 shows enlarged images of Fig. 10.

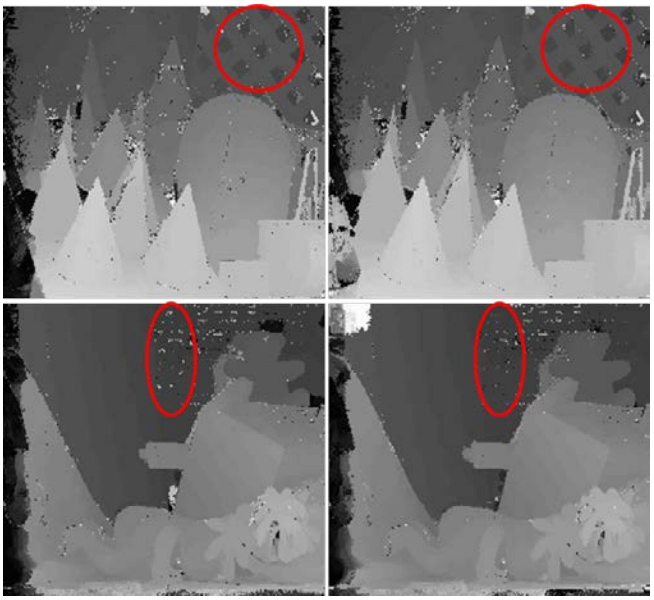


(a) CA Method (b) Proposed Method
Figure 11. Enlarged Images

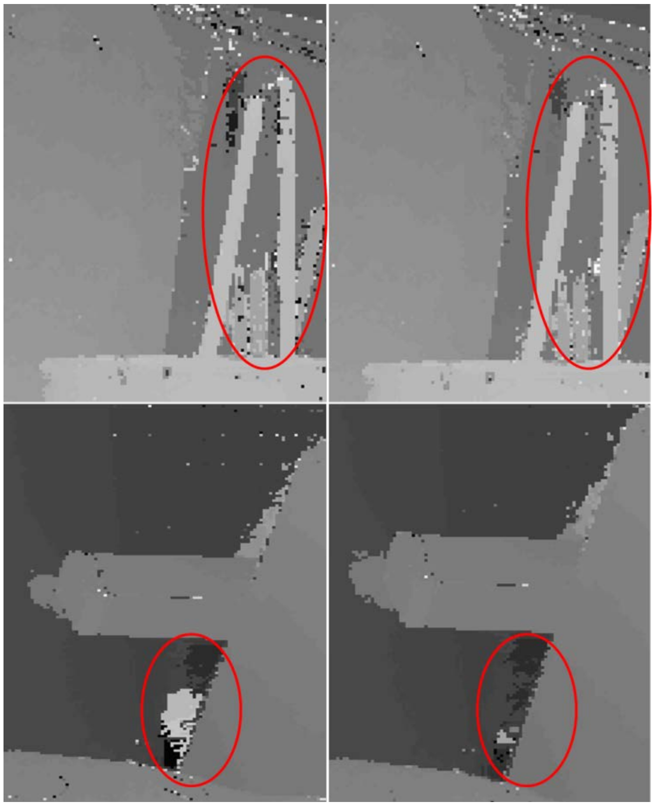
In Fig. 11, we can check that the disparity result using the proposed method finds more accurate disparity values in textureless regions than the result of conventional method.

Fig. 12 also shows the enlarged images of Fig. 10. In Fig. 10, disparity errors in textured regions are reduced comparing with results of the conventional method. Result images using the proposed method also have better disparity values in discontinuity regions and occlusion regions than the conventional method. The term of distance information detects edge regions quite well. For this reason, disparity errors near edge regions are decreased.

Fig. 13 describes the error reduction near edge regions. As you can see in Fig. 13, the proposed method estimates more accurate disparity values in edge regions than the cross-scale cost aggregation method using the conventional matching cost function.



(a) CA Method (b) Proposed Method
Figure 12. Comparison Images in Textured Regions



(a) CA Method (b) Proposed Method
Figure 13. Error Reduction in Edge Regions

We also prepare comparison tables to compare experiment results more objectively. Comparison tables show the error rate of each algorithm. These tables are described in Table 1 and Table 2.

Table 1 shows the result of the conventional method. In Table 1, we checked error rate of disparity maps in all regions, textureless regions and discontinuity regions. Table 2 represents the result of the proposed method. Table 2 shows same types of error rate with Table 1. As you can see in Table 1 and Table 2, average error rates of the proposed method are lower than those of the conventional method. Especially, the error rate in textureless regions is the most error reduced regions comparing with other regions. It is reduced by 0.41%.

Table 1: Error Rate of Conventional Method

	Error Rate (%)		
	All	Textureless	Discontinuity
Teddy	13.80	4.38	17.04
Cones	13.94	1.14	14.62
Tsukuba	2.67	1.57	9.71
Venus	1.94	1.14	4.22
Average	8.09	2.06	11.40

Table 2: Error Rate of Proposed Method

	Error Rate (%)		
	All	Textureless	Discontinuity
Teddy	13.65	3.77	16.41
Cones	13.61	0.89	14.35
Tsukuba	2.52	1.11	10.11
Venus	1.75	0.84	4.32
Average	7.88	1.65	11.30

5. Conclusion

The conventional cross-scale cost aggregation method uses a pixel based cost function in which color information and gradient information are considered. However, this cost function has a matching ambiguity problem in textureless regions. The proposed method overcomes this problem by adopting distance information which is obtained by a modified distance transform. The term of distance information is added to the conventional pixel based cost function. First, the initial matching cost is calculated by the modified cost function. After that, matching costs of each scale image are aggregated using the cross-scale cost aggregation algorithm. As a result, compared to the conventional method, the proposed method successfully reduces error rates in all regions including, textureless and discontinuity regions.

Acknowledgement

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] C. Lee and Y. S. Ho, "View synthesis using depth map for 3D video," Asia-Pacific Signal and International Processing Association, pp. 350-357, Oct. 2009.

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

[3] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," IEEE Conference on Computer Vision and Pattern Recognition, pp. 3017-3028, June 2011.

[4] 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, Oct. 2014.

[5] J. Canny, "A Computational Approach to Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, Issue 6, pp. 679-698, Nov. 1986.

[6] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," IEEE Conference on Computer Vision, pp. 839-846, Jan. 1998.

[7] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision, vol. 47, Issue 1, pp. 7-42, April 2002.

Author Biography

Yong-Jun Chang received his B.S. in electronic engineering and avionics from the Korea Aerospace University, Gyeonggi-do, Korea (2014). Since then he has studied in the Gwangju Institute of Science and Technology in Gwangju, Korea for master's degree courses. His research interests are stereo matching, video coding, and image processing.

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.