https://doi.org/10.2352/ISSN.2470-1173.2019.11.IPAS-257

Enhanced Guided Image Filter Using Trilateral Kernel for Disparity Error Correction

Yong-Jun Chang and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdangwagi-ro, Buk-gu, Gwangju, 61005, South Korea

Abstract

Stereo matching methods estimate the depth information from stereoscopic images using the characteristics of binocular disparity. We try to find corresponding points from the left and right viewpoint images to obtain the disparity map. Once we get the disparity value between the two correspondences, we can calculate the depth information of the object. If there is no texture in the image, the stereo matching operation may not find the accurate disparity value. Especially, it is quite difficult to estimate correct disparity values in occluded regions. In this paper, we propose a new method to detect disparity errors in the disparity map and correct those errors using the enhanced guided image filter (GIF). An error correction using the GIF causes disparity-smoothing near the occluded region because of a coefficient-smoothing process in the GIF. Thus, our method employs a trilateral kernel for the coefficient smoothing process to alleviate the problem.

1. Introduction

As three-dimensional (3D) video contents become popular, various image processing techniques for 3D video contents generation have been developed. Generally, the depth information is used to generate 3D video contents that can provide users with realistic and immersive experience. The quality of the 3D video contents depends on the accuracy of the depth information. Therefore, it is very important to acquire a high-quality depth information from the object. One way to obtain the depth information from the captured objects is to use stereo matching methods that can estimate the disparity map from corresponding points in the stereoscopic images. Then, the depth information is calculated from the disparity value based on the binocular cue.

Stereo matching methods generally produce good matching results in the textured region since the disparity value is calculated from the correct corresponding points. However, we may have many disparity errors in both textureless and occluded regions because there are no features and no correspondences for the disparity estimation. Disparity errors in those regions lead to a poor-quality depth map and even make the quality of 3D video contents worse. On the other hand, the depth estimation using a depth camera can improve the quality of depth information in those regions. A time-of-flight (ToF) based depth camera uses an infrared ray for the depth estimation. This camera calculates the distance between the depth camera and the object using ToF of the infrared ray that is reflected from the object. However, the depth camera generally shows poor performances in outdoor environments because of sunlight.

One method to improve disparity errors in both textureless and occluded regions is applying a post-processing to stereo matching results. This method improves the quality of the disparity value in those regions without using the depth camera. Therefore, we propose a new idea to improve the quality of the disparity map using the guide image filter (GIF) with a trilateral kernel.

2. Disparity Error Detection and Correction

2.1 Disparity Error Detection

A disparity correction refines disparity errors in the disparity map. In order to apply the disparity correction only to error pixels, a disparity error detection should be performed first. In the proposed method, a cross-checking method is used for detecting initial disparity errors [1]. To use this method, both left and right viewpoint disparity maps are obtained first. Once we have two viewpoints' disparity maps, we can find a corresponding point of a reference pixel based on the disparity value. After that, we can compare disparity values between corresponding points. If those values are not same, the reference pixel is regarded as an error pixel. Fig. 1 shows how the cross-checking method detects disparity errors.

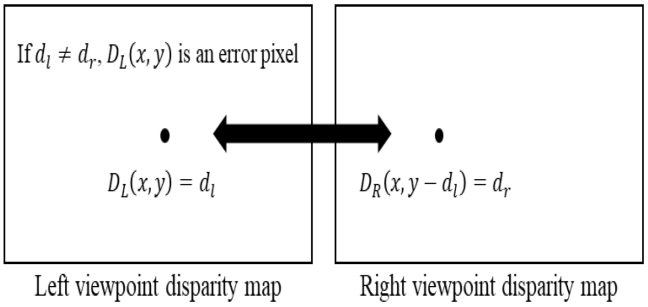
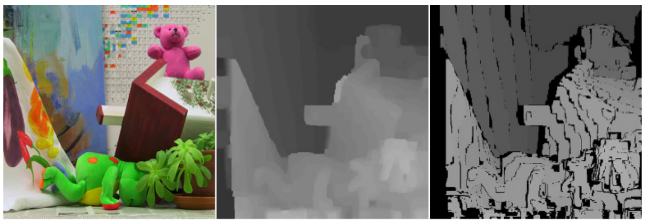


Figure 1. Cross-checking method for initial error detection

In Fig. 1, D_L and D_R are pixel values in the left and the right viewpoint disparity maps and D_L is a reference pixel. Each pixel value corresponds to a disparity value of d_l and a disparity value of d_r, respectively. If both disparity values are not same, D_L is regarded as an error pixel. A result of initial error detection is represented in Fig 2.

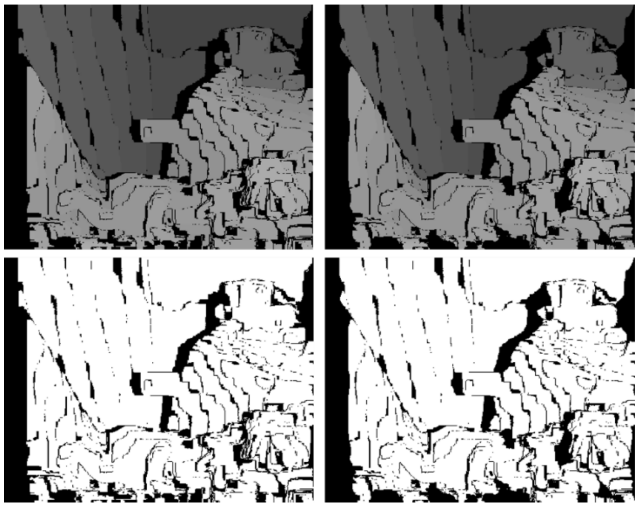


(a) Color image *(b) HBP [2]* *(c) Initial error map*

Figure 2. Result of initial error detection

In Fig. 2(b), HBP means a stereo matching result using a hierarchical belief propagation [2]. Therefore, Fig. 2(b) shows an original stereo matching result that does not use any post-processing methods. Fig. 2(c) represents the result of initial error detection. Black pixels in Fig. 2(c) mean pixels having disparity errors.

Initial disparity errors are detected by the cross-checking method. However, there may still be outliers in the initial error map. To remove those outliers, a median filter is applied to an error mask of the initial detected map [3]. The error mask is a binary image representing both error and non-error regions. Fig. 3 shows the error mask and a filtered error map. Fig. 3(a) represents the initial error mask and the initial error map. If the median filter is applied to the error mask in Fig. 3(a), filtered error mask and error map are obtained depicted in Fig. 3(b). The filtered error map is acquired from the filtered error map.



(a) Initial error detection (b) Filtered error detection

Figure 3. Outlier elimination using median filter

2.2 Disparity Error Correction

Disparity errors detected by the cross-checking method and the median filter are represented as zero values in the error map in Fig. 3(b). To correct those errors, a guided image filter (GIF) is applied to the filtered error map [3, 4]. The GIF defines a linear model for an output estimation as defined in (1).

$$q_i = a_k I_i + b_k, \forall i \in \omega_k \tag{1}$$

In (1), where q is the output value, I is the pixel value of guidance image, and both a_k and b_k are linear coefficients in the kernel ω_k that is centered on pixel k. It is assumed that the linear coefficients of each pixel in the kernel have a constant value. As the center pixel moves, the output values of the pixels in the kernel are overlapped. Therefore, the final output value is calculated as an average value of the output values of each pixel in the kernel. This process change the linear model in (1) to (2).

$$q_i = \bar{a}_i I_i + \bar{b}_i \tag{2}$$

In (2), where $\bar{a}_i$ is the average value of coefficients a_k, and $\bar{b}_i$ is also the average value of coefficients b_k. Therefore, each of average coefficients in (2) mean average values of a_k and b_k when the centered pixel is i in the kernel, respectively. Both coefficients a_k and b_k are calculated by (3) and (4).

$$a_k = \frac{COV(I,p)}{var(I)+\varepsilon} \tag{3}$$

$$b_k = \bar{p}_k - a_k \mu_k \tag{4}$$

In (3), where $COV(I,p)$ is a covariance value between the kernel of guidance image I and that of input image p, $var(I)$ is a variance value of I, and ε is a regularization parameter. In (4), where $\bar{p}_k$ means an average value of the input kernel, and μ_k is also an average value of the guidance kernel. All calculation in both (3) and (4) are performed on a kernel basis.

The GIF in [4] introduces an extension to color filtering. If the guidance image is multichannel and the input image has the gray scale, (2) is redefined in (5). Equations of coefficients in (3) and (4) are also changed to (6) and (7).

$$q_i = \bar{\mathbf{a}}_i^T \mathbf{I}_i + \bar{b}_i, \forall i \in \omega_k \tag{5}$$

$$\mathbf{a}_k = (\mathbf{COV}_k + \varepsilon \mathbf{U})^{-1} \left(\frac{1}{|\omega|} \sum_{i \in \omega_k} \mathbf{I}_i p_i - \mu_k \bar{p}_k \right) \tag{6}$$

$$b_k = \bar{p}_k - \mathbf{a}_k^T \mu_k \tag{7}$$

In (7), where $\bar{\mathbf{a}}_i^T$ is a transpose of the matrix $\bar{\mathbf{a}}_i$, and $\mathbf{I}_i$ is the matrix from the guidance image. Both $\bar{\mathbf{a}}_i$ and $\mathbf{I}_i$ are 3×1 sized matrices. In (6), where $\mathbf{COV}_k$ means a covariance matrix of $\mathbf{I}$ in the kernel, and $\mathbf{U}$ represents an identity matrix. Both matrices have a 3×3 array.

The GIF is applied only to holes representing error pixels in the filtered error map. This process is performed iteratively until all the error pixels are corrected. Fig. 4 shows results of the error correction using the GIF. In Fig. 4, disparity maps in the first row are experimental results using the guidance image changed to the gray scale and results in the second row show disparity maps from the color guidance image.

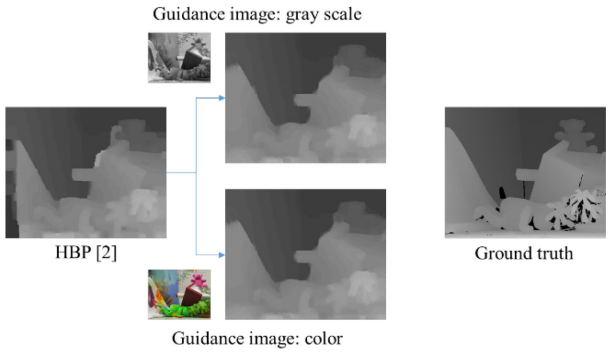


Figure 4. Results of disparity error correction

Results of disparity error correction in Fig. 4 shows better quality than the result of stereo matching method [2]. However, we can check that there are some blurred pixels near the boundary region in both results. Fig. 5 represents enlarged images of the results. In Fig. 5, regions that are cropped by boxes are blurred.

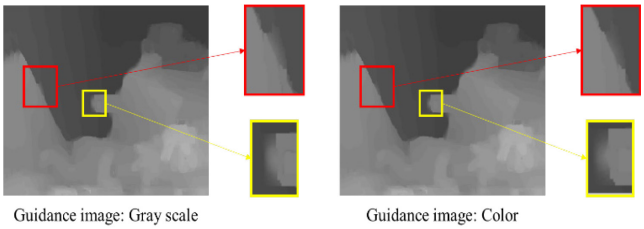


Figure 5. Enlarged images

3. Error Correction with Enhanced Filter

3.1 Analysis of Conventional Method

The GIF has the coefficient-smoothing process for the output estimation. In this process, the average filter is used for smoothing coefficients. Since the basic average filter does not consider the color difference between the foreground and the background, it may mix the coefficient information of different objects. If we use the gray scale image as the guidance image, the output may show much more blurred pixels near the edge region than that of using the color image as depicted in Fig. 5.

The GIF in [4] also introduces the coefficient smoothing process using the Gaussian filter. It is called the Gaussian guided filter (GGF). Since the Gaussian filter considers the pixel distance between the current pixel and the neighboring pixels in the kernel, it gives a large weighting value to pixels near the current pixel. Therefore, it can alleviate the problem of coefficient smoothing process better than the average filter theoretically. Fig. 6 shows comparison results between the GIF and the GGF.

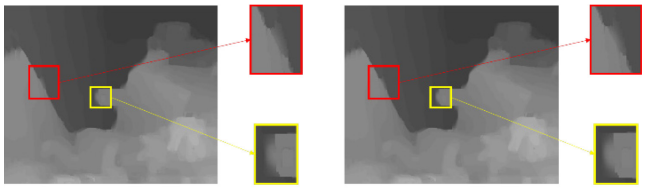


Figure 6. Comparison results between GIF and GGF

In Fig. 6, both results use the gray scale image as the guidance image and there are no big differences between the result of GIF and GGF. However, the bad pixel rate (BPR) in Table 1 quantitatively indicates that the GGF has better results than the GIF. The BPR is a percentage value representing the ratio of error pixels among all the pixels [5]. The error pixel means a pixel that have different disparity value compared with the ground truth.

Table 1. Quantitative comparisons using BPR

	All	Non-occ.	Occ.	Textured	Textureless	Disc.
HBP [2]	11.62	5.1	86.62	5.03	6.08	17.5
GIF [3, 4]	11.16	5.91	68.39	5.87	6.16	20.63
GGF [4]	11.02	5.79	68.33	5.73	6.06	20.41

In Table 1, 'All' is the BPR of all regions in the result, 'Non-occ.' is the non-occluded region, 'Occ.' is the occluded region, 'Textured' is the textured region, 'Textureless' is the textureless region, and 'Disc.' is the discontinuity region. Since the BPR means an error rate, the lower value represents the better result.

3.2 Enhanced Guided Image Filter

The disparity error correction with the GGF shows better BPR in the discontinuity region than that of the GIF. However, it has blurred pixels near the edge region and the BPR in the discontinuity region is still higher than that of HBP in Table 1. One method to improve this problem is to use a concept of the joint bilateral upsampling (JBU) that uses the bilateral filter (BF) for the depth map upsampling [6].

The depth map upsampling means the process of converting a low-resolution depth map to a high-resolution. In this process, holes are generated due to the resolution difference. One of the methods to fill those holes uses the JBU [7]. The JBU uses the high-resolution image as the guidance image for the depth map upsampling. It applies the BF to the guidance image for the hole filling process. The BF is composed of the spatial and range kernels as depicted in (8) [8].

$$W_{BF} = e^{-\frac{\|c_i - c_j\|^2}{2\sigma_r^2}} \cdot e^{-\frac{\|x_i - x_j\|^2}{2\sigma_s^2}} \tag{8}$$

In (8), where c is a pixel value, x is a pixel position, σ_r is a parameter for the range kernel, σ_s is that of the spatial kernel, i is an index of the current pixel, and j is also an index of the neighboring pixel. The spatial kernel and the range kernel are multiplied each other. Finally, the kernel of BF is represented as W_{BF}.

The process of depth map upsampling is very similar to the disparity error correction. For this reason, the method introduced in [6] uses the concept of JBU for the coefficient smoothing process. This method is called the modified guided image filter (MGIF). Since the JBU considers both the difference of pixel values and the difference of pixel positions, it gives large weighting values to pixels that are close to the current pixel position and have similar pixel value to that of the current pixel. As a result, the MGIF prevents the mixture of the coefficient information in the GIF. The comparison result of disparity error correction using the JBU is represented in Fig. 7. In Fig. 7, the result of MGIF shows better results than other results.

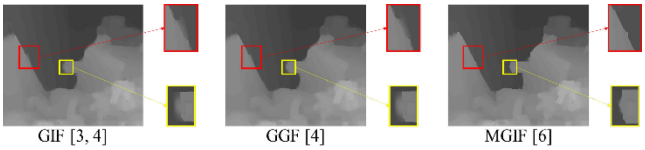


Figure 7. Comparison results among GIF, GGF, and MGIF

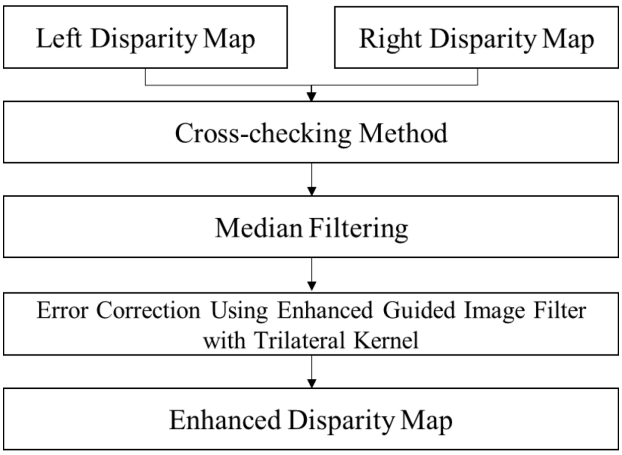


Figure 8. Flowchart of proposed method

The disparity error correction with the MGIF alleviates blurred pixels near the edge regions in the disparity map as depicted in Fig.

7. However, there are still some outliers in the result. The proposed method enhances the GIF using a trilateral kernel.

Fig. 8 shows a flowchart of the proposed method. In Fig. 8, we detect the disparity errors from the initial disparity map. The way to detect the disparity errors is same with the method introduced in Section 2.1. Holes in the error detected disparity map are filled with the enhanced guided image filter with the trilateral kernel.

The coefficient smoothing using the JBU prevents the blurred pixels near the edge region. However, the BF used in JBU considers only the pixel value difference and the distance difference between the current pixel and the neighboring pixel in the filtering process. Therefore, if two pixels have the same pixel value although they are on different objects, the BF may show the wrong result and it can cause outliers in the disparity map.

The proposed method checks a local similarity to prevent this problem. The local similarity detects the majority pixels in the BF kernel of the disparity map. Therefore, the purpose of measurement of the local similarity is to give an additional information to the BF used for the coefficient smoothing. The meaning of the additional information is the coefficient information of the majority pixels in the kernel. We use this information as the third kernel of the BF with spatial and range kernels. We call this a trilateral kernel. It is because the probability that the current pixel has the similar information to that of the majority pixels in the kernel is high. Fig. 9 shows how to calculate the local similarity value in the kernel.

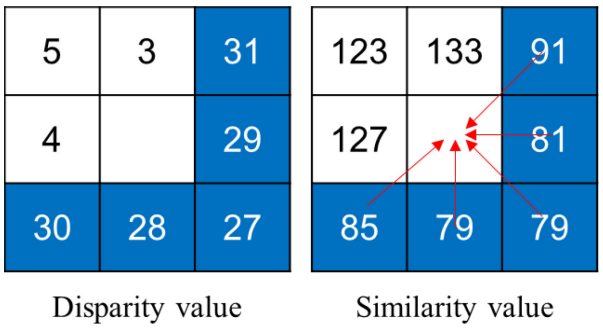


Figure 9. Local similarity

In Fig. 9, there are kernels for the calculation of local similarity. The size of kernel is same with the filter for the coefficient smoothing. In this example, we use the 3×3 sized kernel. In order to calculate the local similarity, we calculate the disparity differences between the reference pixel and the target pixels. If the reference pixel belongs to the majority pixels, the sum of the disparity differences can be small. In the opposite case, the sum of the disparity differences can be large. Colored pixels in Fig. 9 represents the majority pixels. An equation of the local similarity for the example in Fig. 9 is defined in (9).

$$s_i = \sum_j |d_i - d_j| \tag{9}$$

In (9), where s_i is the similarity value at the pixel i, d is the disparity value of the pixel, and j is an index of neighboring pixels which are not holes.

The average value of the disparity differences represents the local similarity well. However, we enhance the equation of the local similarity using the distance weighting function as defined in (10).

$$s_i = \sum_j w_s \cdot |d_i - d_j| \tag{10}$$

We use the weighted sum of the disparity differences in (10). In (10), where w_s is same with the spatial kernel in (8). Therefore, the trilateral kernel is defined in (11) and this kernel is used for the coefficient smoothing process instead of the BF.

$$W_{TRI} = e^{-\frac{\|c_i - c_j\|^2}{2\sigma_r^2}} \cdot e^{-\frac{\|x_i - x_j\|^2}{2\sigma_s^2}} \cdot e^{-\frac{s_j}{2\sigma_l^2}} \tag{11}$$

4. Experimental Results

We tested our method using four different images: *Teddy*, *Cones*, *Tsukuba*, and *Venus* [9]. Both gray scale and color images are used for the experiments. In terms of the gray scale image, we transformed the color dataset to the gray scale dataset. Fig. 10 shows experimental results compared with those of conventional methods. All the results in Fig. 10 were generated by the color datasets. Test images represents *Teddy*, *Cones*, *Tsukuba*, and *Venus* from top to bottom. In Fig. 10, there are some blurred pixels near the edge region for each of results using conventional methods. However, in terms of our method, we alleviate the blurred pixels near the edge region.

On the other hand, all the results in Fig. 11 were generated by the gray scale datasets except for the stereo matching results. The enlarged images of *Teddy* and *Venus* in Fig. 10 are shown in Fig. 12 for further comparison. In Fig. 12, we can check that the results from MGIF and our method looks better than other methods. However, it is difficult to compare the results of the MGIF and the proposed method in detail. Therefore, some enlarged images from results of the MGIF and the proposed method are shown in Fig. 13. In Fig. 13, the upper image represents the result of the MGIF and the bottom one is that of our method.

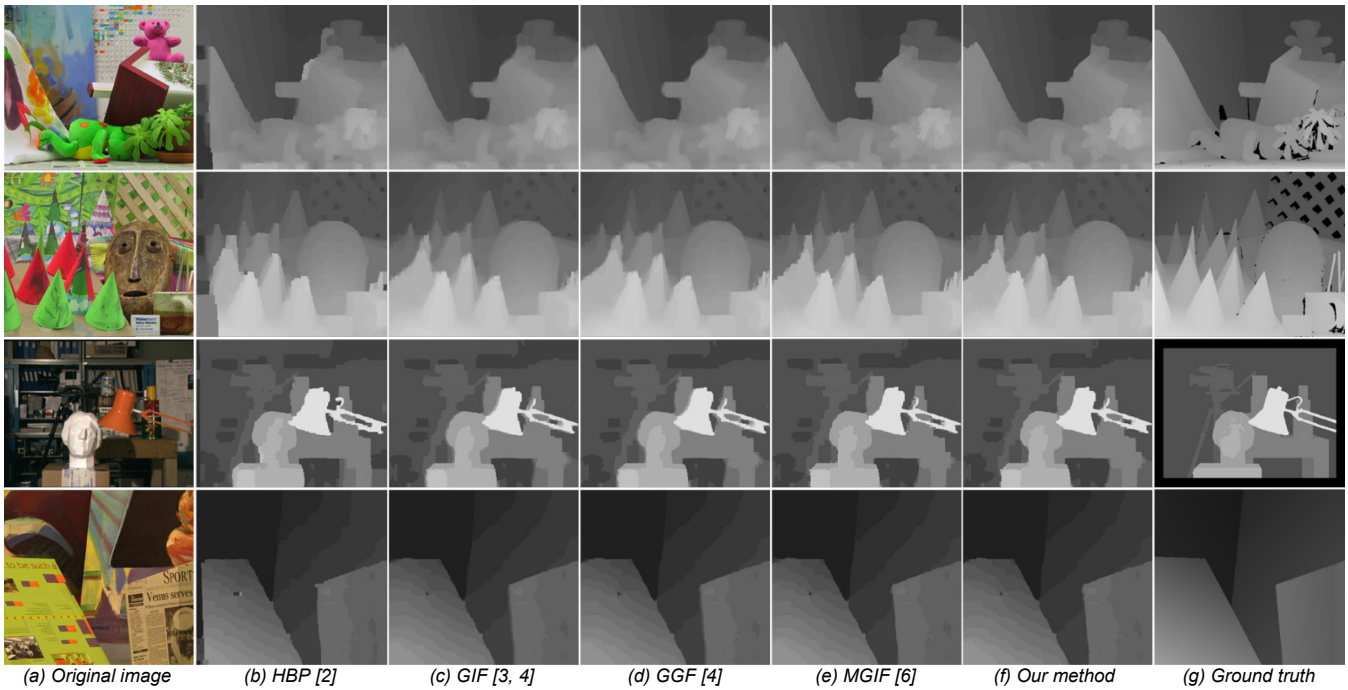
For the quantitative comparison, we measured the BPR [5]. Table 2 shows the error rates of experimental results using the gray scale datasets. Table 3 shows BPR results using the color datasets.

Table 2. BPR of experimental results using gray scale datasets

	All	Non-occ.	Occ.	Textured	Textureless	Disc.
HBP [2]	11.62	5.1	86.62	5.03	6.08	17.5
GIF [3, 4]	11.16	5.91	68.39	5.87	6.16	20.63
GGF [4]	11.02	5.79	68.33	5.73	6.06	20.41
MGIF [6]	9.73	5.06	55.12	4.97	5.58	16.7
Our method	**9.62**	**4.98**	**54.28**	**4.89**	**5.39**	**16.38**

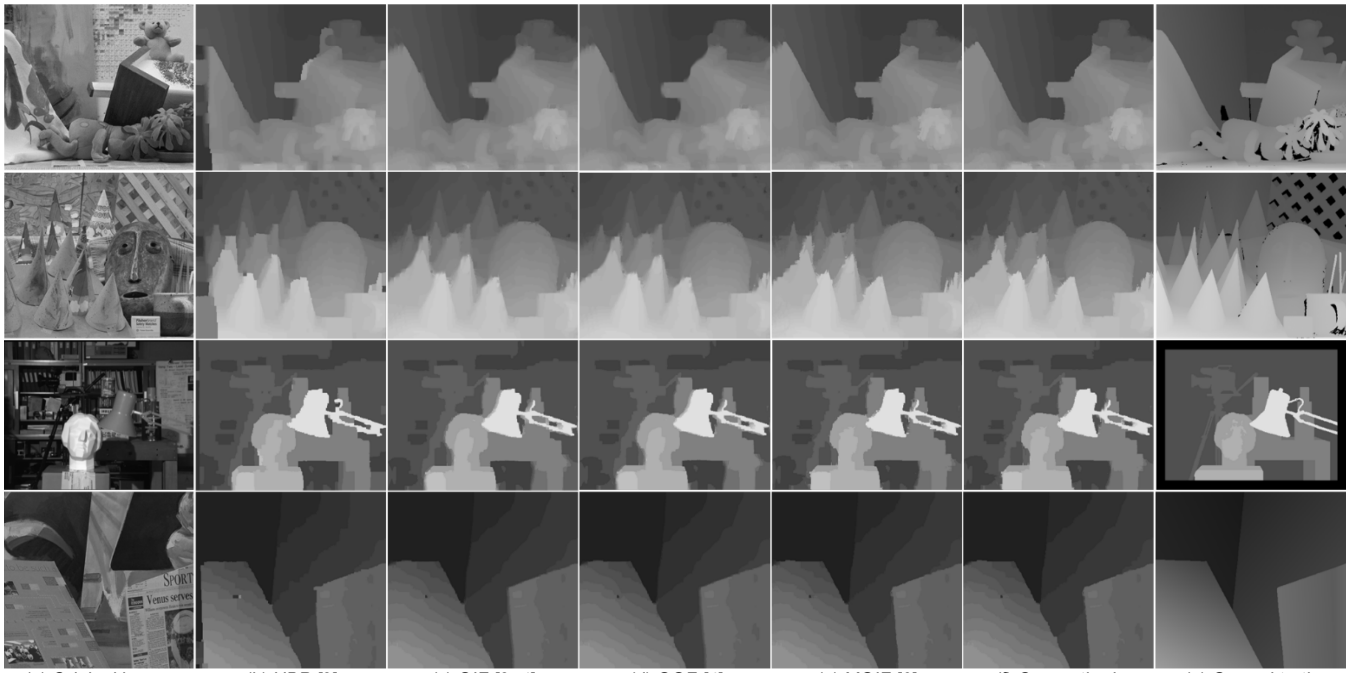
Table 3. BPR of experimental results using color datasets

	All	Non-occ.	Occ.	Textured	Textureless	Disc.
HBP [2]	11.62	5.1	86.62	5.03	6.08	17.5
GIF [3, 4]	10.82	5.65	65.93	5.6	5.91	19.72
GGF [4]	10.73	5.56	66.29	5.5	5.88	19.58
MGIF [6]	9.32	4.85	**50.93**	4.74	5.52	15.99
Our method	**9.29**	**4.83**	51.11	**4.73**	**5.39**	**15.89**



(a) Original image (b) HBP [2] (c) GIF [3, 4] (d) GGF [4] (e) MGIF [6] (f) Our method (g) Ground truth

Figure 10. Experimental results using color datasets

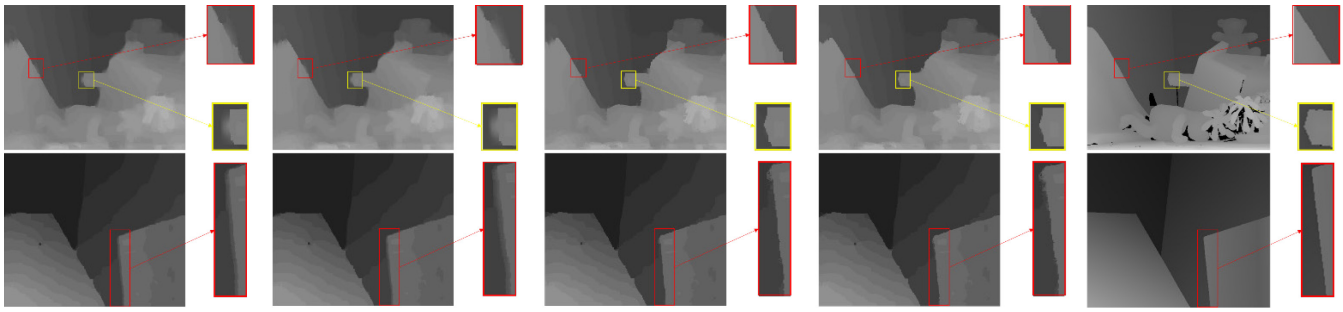


(a) Original image (b) HBP [2] (c) GIF [3, 4] (d) GGF [4] (e) MGIF [6] (f) Our method (g) Ground truth

Figure 11. Experimental results using gray scale datasets

In Table 2, our method shows the best average BPRs compared with other methods in the all regions. On the other hand, our method has worse BPR in the occluded region than the MGIF in Table 3. However, Table 3 also shows that our method shows better BPRs in all the regions except for the occluded region than other methods. When we applied the GIF or the GGF to the error detected disparity map, the average error in 'All' showed lower than that of HBP result. However, in terms of 'Disc.', both GIF and GGF results showed worse error rates than HBP. It is because of the coefficient problem introduced in Section 2.

Our method tried to improve this problem modifying the process of coefficient smoothing using the trilateral kernel. Therefore, in both Table 2 and Table 3, our method reduced error rates not only in all regions but also in the error rate in the discontinuity region compared with HBP. The proposed method showed even better results than the MGIF.



(a) GIF [3, 4] (b) GGF [4] (c) MGIF [6] (e) Our method (g) Ground truth

Figure 12. Enlarged image of 'Teddy' and 'Venus'

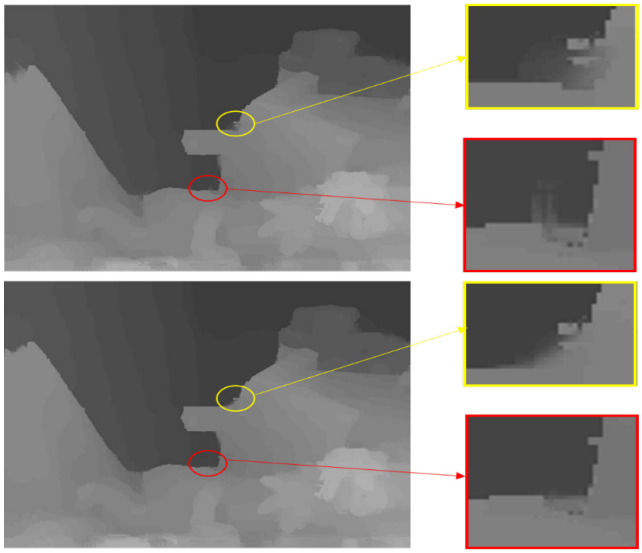


Figure 13. Comparison of enlarged images between MGIF and our method

5. Conclusion

This paper proposed the method for solving the pixel blurring problem near the object boundary which occurs when the disparity map is corrected with the GIF. The conventional GIF uses the average filter for the coefficient smoothing. Since this filter causes the blurring effect near the edge region, we used the concept of the JBU using the trilateral kernel for the coefficient smoothing. As a result, our method alleviates the blurring artifact in the edge region and it also shows a reduction of error rate of about 1.61% in the edge region compared with the results obtained using HBP.

References

[1] G. Egnal and R. Wildes, "Detecting Binocular Half-occlusions: Empirical Comparisons of Five Approaches," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 8, pp. 1127-1133, 2002.

[2] Q. Yang, L. Wang, R. Yang, S. Wang, M. Liao, and D. Nister, "Real-time Global Stereo Matching Using Hierarchical Belief Propagation," British Machine Vision Conference, pp. 989-998, Edinburgh, Scotland, 2006.

[3] Y. J. Chang and Y. S. Ho, "Disparity Error Refinement Using Disparity Information of Verified Pixels," ISO/IEC JTC1/SC29/WG11, pp. 1-6, Ljubljana, Slovenia, 2018.

[4] K. He, J. Sun, and X. Tang, "Guided Image Filtering," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 6, pp. 1397-1409, 2013.

[5] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms," International Journal of Computer Vision, vol. 47, no. 1/2/3, pp. 7-42, 2002.

[6] Y. J. Chang and Y. S. Ho, "Post-processing for Disparity Map Enhancement Using Modified Guided Image Filter," Joint International Workshop on Advanced Image Technology and International Forum on Medical Imaging in Asia, pp. 1-6, Nanyang, Singapore, 2019.

[7] J. Kopf, M. F. Cohen, D. Liscinski, and M. Uyttendaele, "Joint Bilateral Upsampling," ACM Transactions on Graphics, vol. 26, no. 3, pp. 96:1-96:5, 2007.

[8] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," IEEE Conference on Computer Vision, pp. 839-846, Bombay, India, 1998.

[9] http://vision.middlebury.edu/stereo/data/

Acknowledgement

This research was partially supported by the 'Brain Korea 21 Plus Project' of the Ministry of Education & Human Resources Development, Republic of Korea (ROK) [F16SN26T2205], and partially supported by the 'Civil-Military Technology Cooperation Program' grant funded by the Korea government.

Author Biography

Yong-Jun Chang received his B.S. in electronic engineering and avionics from the Korea Aerospace University, Gyeonggi-do, Korea (2014) and his M.S. in electrical engineering and computer science from the Gwangju institute of Science and Technology, Gwangju, Korea (2016). Since then he has studied in the Gwangju Institute of Science and Technology in Gwangju, Korea for Ph.D. courses. His research interests are stereo matching, video coding, and image processing.

Yo-Sung Ho received his B.S. and M.S. degrees in electronic engineering from Seoul National University, Seoul, Korea (1981, 1983) and his Ph.D. in electrical and computer engineering from University of California, Santa Barbara, USA (1990). He worked at ETRI from 1983 to 1995, and Philips Laboratories from 1990 to 1993. Since 1995, he has been with Gwangju Institute of Science and Technology, Gwangju, Korea, where he is currently a professor. His research interests include video coding, 3D image processing, 3DTV, AR/VR, and realistic broadcasting systems.

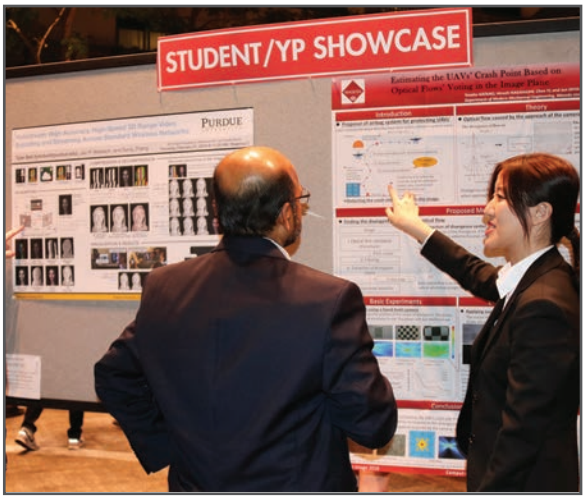
JOIN US AT THE NEXT EI!

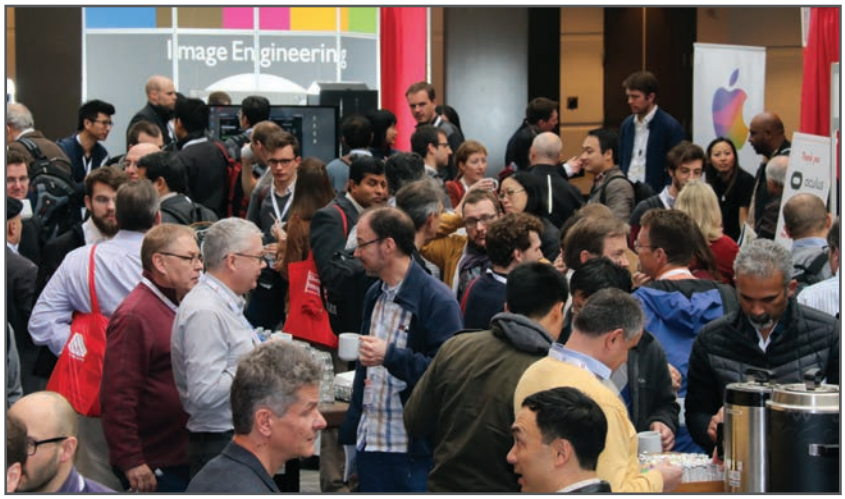
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org