INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2018/ M42934
July 2018, Ljubljana,

Source	**Gwangju Institute of Science and Technology (GIST)**
Status	**Report**
Title	**[MPEG-I-Visual]: Disparity Error Refinement Using Disparity Information of Verified Pixels**
Author	**Yong-Jun Chang (GIST, yjchang@gist.ac.kr)** **Yo-Sung Ho (GIST, hoyo@gist.ac.kr)**

Abstract

The stereo matching method generally shows a good matching result near textured regions that have many features. However, the stereo matching method sometimes does not find accurate disparity values in some regions that do not have enough features like homogeneous regions. Therefore, there are some disparity errors in the disparity map. This report proposes a post-processing method for refinement of disparity errors in the disparity map.

1 Introduction

Depth information is a very important data for immersive media content. One of the methods for depth estimation is a stereo matching. The stereo matching methods uses a characteristic of binocular disparity for making a disparity map. It searches corresponding points between left and right viewpoint images. After that, a disparity value between two corresponding points is calculated. The stereo matching method estimates depth information depending on color information of features. Therefore, the disparity map may have disparity errors because of homogeneous regions. Those errors cause a falling-off in quality of 3D video content.

2 Cross-checking Method for Disparity Error Detection

Stereo matching results have left and right disparity maps. Based on both disparity maps, we can verify the disparity accuracy as depicted in Fig. 1. In Fig. 1, I_L and I_R are disparity values in both left and right disparity maps. Since we already know the disparity value of both I_L and I_R, a cross-checking method can be used for the verification of disparity values [1]. If disparity values of both pixels are not same, those pixels may have wrong disparity values.

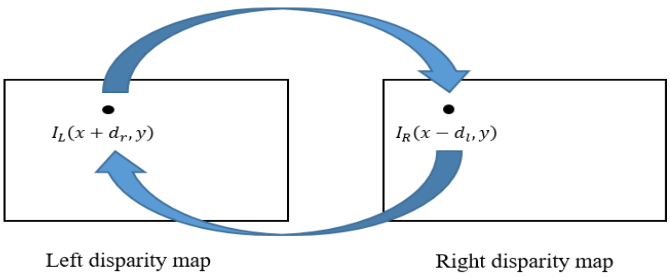
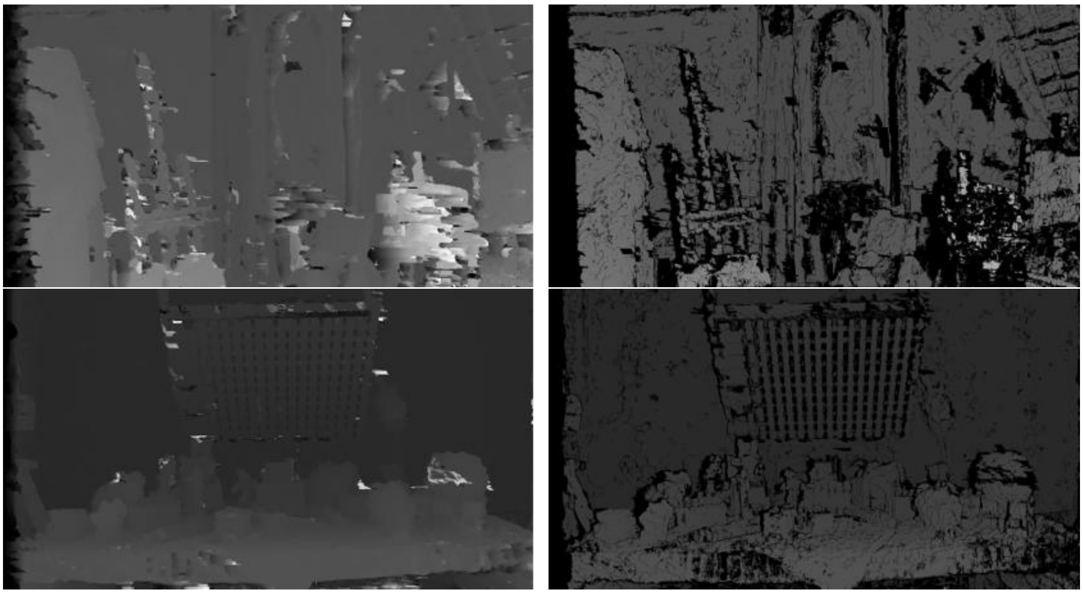


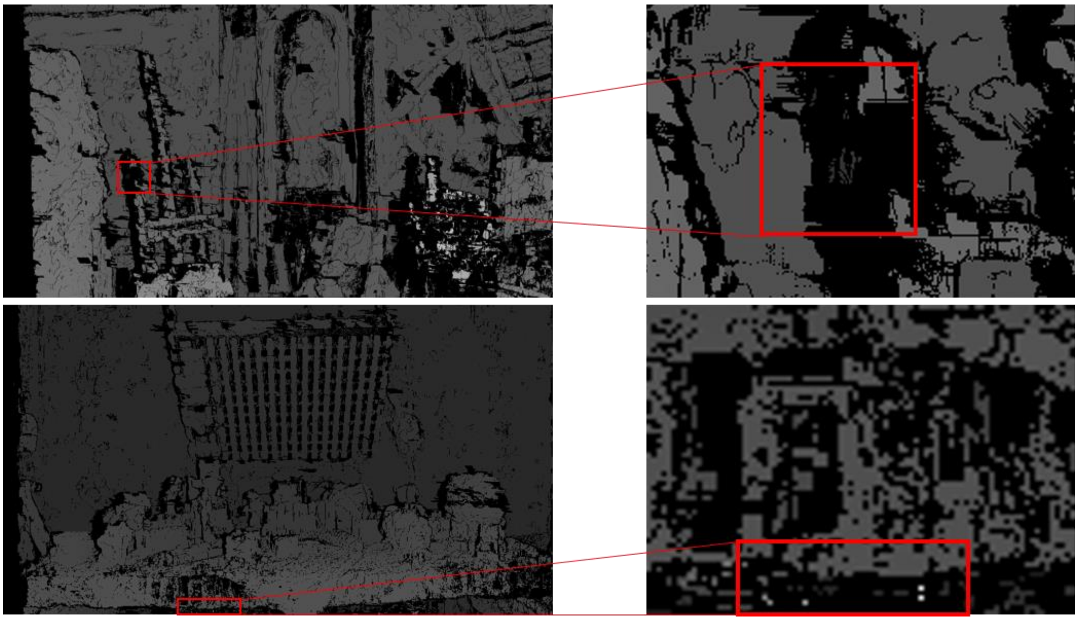
Figure 1. Cross-checking method

Fig. 2 shows results of the cross-checking method for the detection of disparity errors. In Fig. 2, pixels values that are detected as error pixels are replaced to zero values. The other pixels that are not regarded as error pixels keep their own pixel values. The initial disparity map can be estimated by using a three step search (TSS) algorithm [2, 3].



<Original disparity map> <Error detected disparity map>

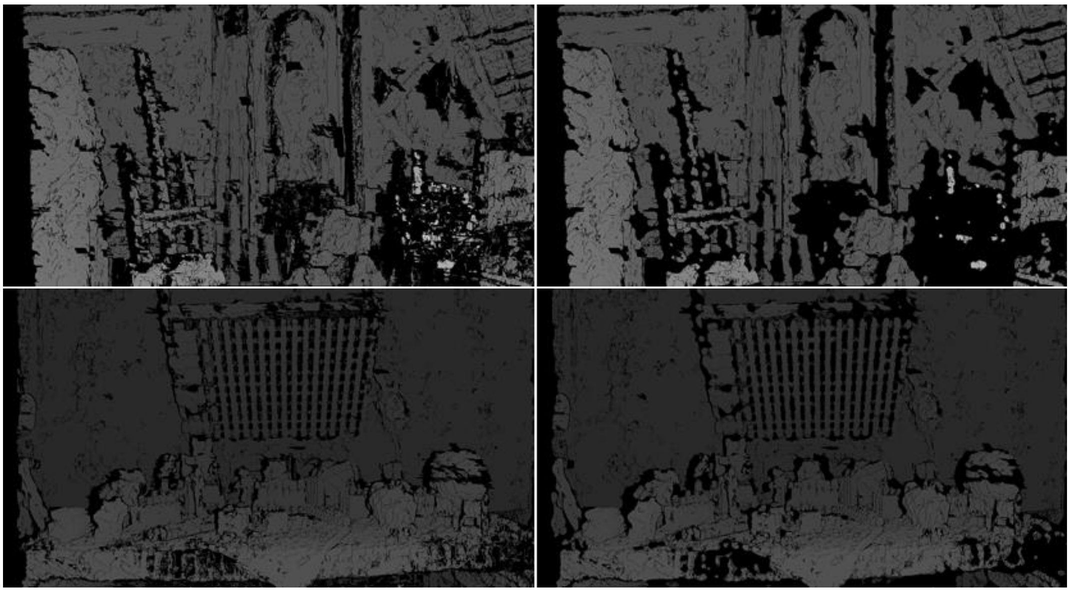
Figure 2. Results of disparity error detection



<Error detected image> <Outliers>

Figure 3. Outliers in error detected images

The cross-checking method detects most errors in the disparity map. However, some outliers are still not detected as depicted in Fig. 3. Outliers are usually located in hole regions and there are many zero pixels around the outlier. Therefore, we remove the outliers by applying the median filter to those regions. After that, a filtered error map can be acquired by the filtered mask. Fig. 4 represents the result of median filtering. As shown in Fig. 4, some outliers are detected by the median filter. After the median filtering, there are many holes in the filtered image. To fill those holes, a guided image filter is applied to the filtered image [4, 5].



<Error detected disparity map> <Filtered disparity map>

Figure 4. Results of disparity error detection

3 Refinement of Disparity Errors

The guided image filter refines a pixel value considering color differences in a guidance image. Therefore, this filtering method are used for preserving disparity values near the object boundary. In this document, the guided image is the color image and the input image is the disparity map. The guidance image filter is quite similar to a joint bilateral filter. The joint bilateral filter can also preserves disparity values of edge region in the input disparity map. However, this filtering method has to calculate weight values of every pixels in the kernel and this process is applied to all the pixels in the image.

On the contrary, the guided image filter estimates output image by using a local linear model [4]. This model is defined in (1).

$$q_i = \bar{a}_i I_i + \bar{b}_i \quad (1)$$

In (1), where i is one of the pixels in the kernel, q is the output image, I_i is the input image, and both $\bar{a}_i$ and $\bar{b}_i$ are average coefficients of a_i and b_i for the linear model. The guided image filter calculates both coefficients using covariance, variance, and mean values of input and guidance images. Hence, those values of each pixel in input and guidance images should be calculated first. After that, both a_i and b_i are calculated based on (2) and (3).

$$a = \frac{COV(I,p)}{\sigma_I^2 + \epsilon} \quad (2)$$

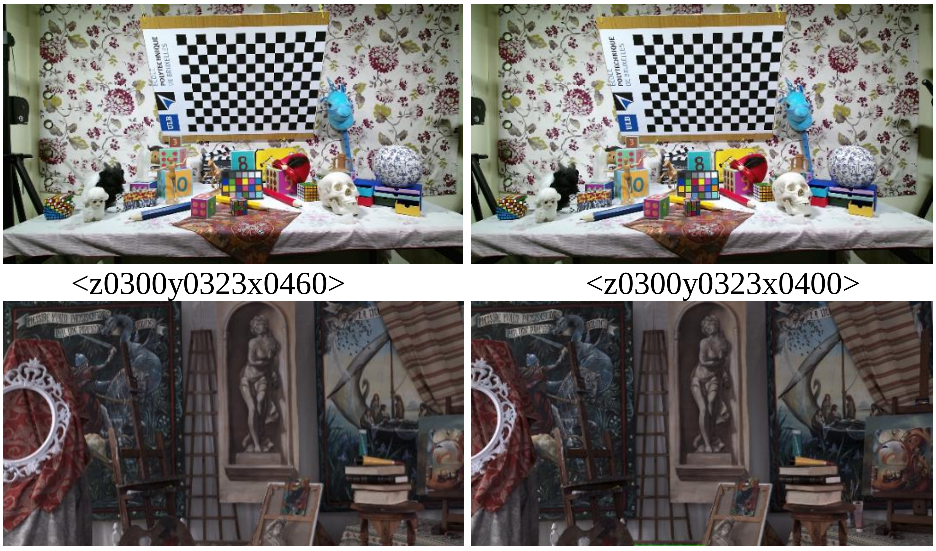
$$b = \mu_p - a\mu_I \quad (3)$$

In (2), where $COV(I,p)$ is the covariance value between the kernel of input image and that of guidance image, σ_I^2 is a variance value of the kernel of input image, and ϵ is a regularization parameter. After calculating the coefficient a, the other coefficient b can be calculated as depicted in (3). In (3), where μ_p and μ_I are the average pixel value of the guidance image and the input image, respectively.

The guidance image filter is only applied to holes in the filtered disparity map. Therefore, it uses verified neighboring pixels for the hole filling. If there are no verified pixels for the filtering process, then the current pixel is kept at zero value. The filtering process up to this point is repeated until zero pixels in the filtered disparity map disappear [5].

4 Experimental Results

We test the disparity error detection and refinement using ULB dataset version 2 [6] and TechnicolorPainter [7]. In terms of ULB dataset version 2, we used both z0300y0323x0400 and z0300y0323x0460 images. Those are stereo images and the spacing between stereo images is 6cm. The other sequence TechnicolorPainter from Technicolor has 7cm spacing between two adjacent cameras. In our experiment, we use TechnicolorPainter_pr_000_#04 and TechnicolorPainter_pr_000_#06. The distance between two images is 14cm. Fig. 5 shows test sequences that we used for the experiment. In Fig. 5, each image in the left and the right columns represents the left and the right viewpoints, respectively.



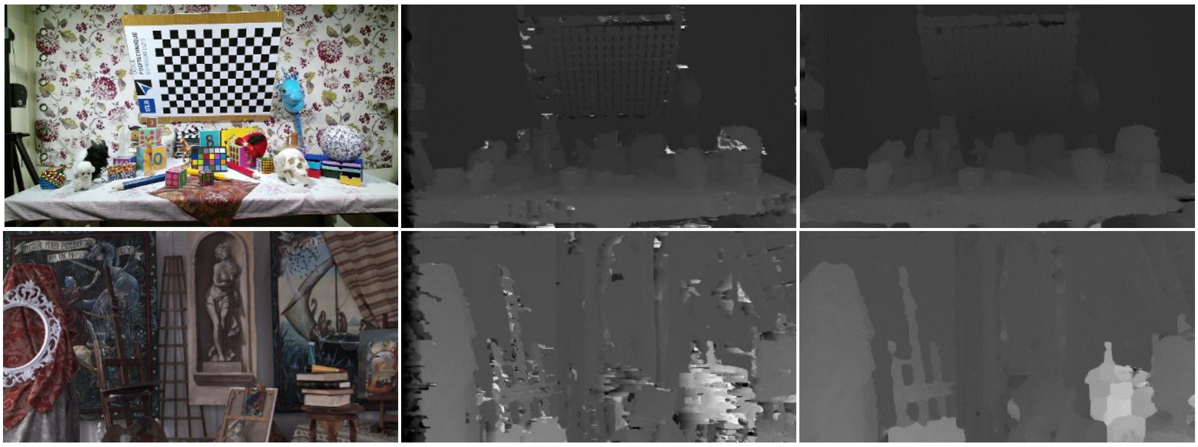
<z0300y0323x0460> <z0300y0323x0400>

<TechnicolorPainter_pr_000_#04> <TechnicolorPainter_pr_000_#06>

Figure 5. Test sequences

Fig. 6 represents stereo matching results using TSS algorithm and results of disparity refinement using the guided image filter. Refinement results show better disparity maps than original results from TSS algorithm. To evaluate those results, an intermediate viewpoint image was synthesized based on output disparity maps. In case of ULB dataset, we made the intermediate viewpoint image between z0300y0323x0460 and z0300y0323x0400. The other intermediate viewpoint image for the Technicolor dataset was also synthesized based on disparity maps from TechnicolorPainter_pr_000_#04 and TechnicolorPainter_pr_000_#06. Synthesized images are represented in Fig. 7.

For the quantitative comparison, we measured PSNR between the original color image and the synthesized image. PSNR results are summarized in Table 1. In Table 1, synthesized images using refined disparity maps have better PSNR value than those of stereo matching results. In addition, the average PSNR is better than that of the stereo matching results.



<Color image> <Stereo matching result> <Disparity refinement>

Figure 6. Experimental results

<Ground truth> <Stereo matching result> <Disparity refinement>

Figure 7. Synthesized images

Table 1. Quantitative Comparison

	Stereo matching result	Disparity refinement
	PSNR (dB)	
TechnicolorPainter	23.06	29.41
ULB dataset ver.2	23.47	27.07
Average	23.27	28.24

<Stereo matching result> <Disparity refinement>

Figure 8. Enlarged images for ULB dataset ver.2

Enlarged images are depicted in Fig. 8 and Fig. 9 for the quality comparison. In both Fig. 8 and Fig. 9, synthesized images using refined disparity maps show better quality in some regions compared with that of stereo matching results.

<Stereo matching result> <Disparity refinement>

Figure 9. Enlarged images for TechnicolorPainter

5 Conclusion

In this document, we proposed the post-processing method for the disparity map. The post-processing method includes a disparity error detection and a refinement. The disparity error detection using the cross-checking method and the median filter should be performed prior to the disparity refinement step. After that, holes that are generated by the error detection are filled by the guided image filter. As a result, the refined disparity map showed better quality than the original one in some regions. However, synthesized images using the refined disparity map had higher average PSNR values than those of stereo matching results.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) Grant funded by the Korean Government(MSIP)(No. 2011-0030079)

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] J. H. Mun, Y. J. Chang, and Y. S. Ho, "Disparity Estimation Using TSS Algorithm," ISO/IEC JTC1/SC29/WG11 MPEG2018/M42432, April 2018.
[3] Y. J. Chang and Y. S. Ho, "Disparity Estimation Using Fast Motion-Search and Local Image Characteristics," Electronic Imaging, pp. 1-6, 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] Y. J. Chang, S. Kim, and Y. S. Ho, "Depth Upsampling Methods for High Resolution Depth Map," International Conference on Electronics, Information, and Communication, pp. 1269-1272, 2018.
[6] D. Bonatto, A. Schenkel, T. Lenertz, Y. Li, and G. Lafruit, "ULB High Density 2D/3D Camera Array Dataset, version 2," ISO/IEC JTC1/SC29/WG11 MPEG2017/M41083, July 2017.
[7] D. Doyen, T. Langlois, B. Vandame, F. Babon, G. Boisson, N. Sabater, R. Gendrot, and A. Schubert, "Light Field Content from 16-Camera Rig," ISO/IEC JTC1/SC29/WG11 MPEG2017/M40010, January 2017.