https://doi.org/10.2352/ISSN.2470-1173.2018.13.IPAS-464

Error Correction for Time-of-Flight Images Using Validity Classification

Yunseok Song and Yo-Sung Ho; Gwangju Institute of Science and Technology; Gwangju, Republic of Korea

Abstract

In this paper, we devise a method that reduces distance errors in time-of-flight (ToF) images. Errors are exhibited at boundaries and surfaces that are not capable of reflecting the infrared ray. For the proposed method, at least two ToF cameras are required in the camera setup. ToF distance error region is estimated by comparing the captured ToF image with warped ToF image from the neighboring ToF camera. The distance values in the error region are replaced. A number of methods are examined to select the optimum replacement value. After distance error reduction, this method is inserted into the aforementioned depth map generation framework. The performance is analyzed by evaluating a synthetic image which is generated by the depth map result.

Introduction

In general, 3D video adds depth perception to 2D video, providing a realistic feel. The market for 3D video has grown extensively since the successes of numerous 3D commercial films in the late 2000s. Now with more advanced technologies, content providers attract customers with marketable 3D contents and 3D displays such as stereoscopic or auto-stereoscopic displays reconstruct satisfactory 3D video. Fig. 1 shows 3D services using a variety of displays.

Generally, the process of left and right eyes seeing slightly different scenes achieves 3D experience [1]. In other words, 3D perception is derived from two separate views. 3D displays such as 3DTV on the market provide stereoscopic images to users. In the near future, users will most likely be able to experience 3D perceptional depth freely.

The 3D video system transmits compressed N views of color and depth video data. At the receiver end, M views are generated based on the N decoded views and synthesized views. Hence, the number of output views is always greater than that of input views [8]. For view synthesis, decoded color and depth video data are used as input in depth image based rendering. Thus, the quality of decoded color and depth data directly affects the quality of rendered images.

Depth images present the distance between the camera and the object. Generally, depth images are produced by depth cameras or estimated by stereo matching. Depth cameras allow fast data acquisition but cost can be an issue. In addition, interference must be checked which is caused by frequency overlaps. Stereo matching does not have limitations of depth cameras [9], [10]; however, it can be time-consuming, thus, not suitable for practical applications.

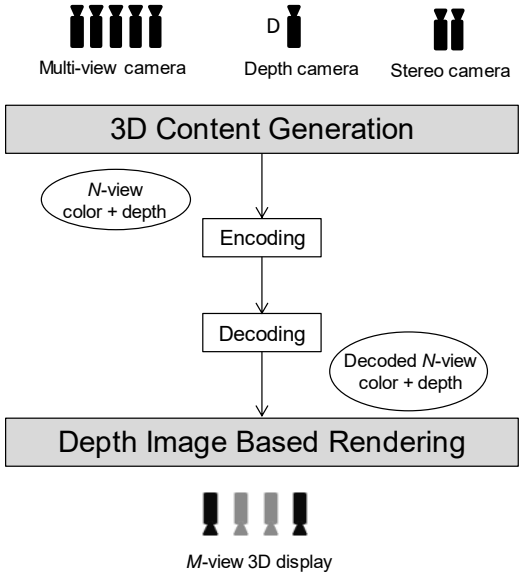


Figure 1. Framework of a 3D video system including 3D content production and depth image based rendering

Depth Map Generation Using ToF Images

ToF cameras are useful for obtaining object distances in a scene [11, 12]. Yet, they can produce low resolution images only. In order to match the resolution of depth images and color images, ToF-to-color view 3D warping is applied using ToF and color camera parameters. Intrinsic and extrinsic camera parameters are necessary for this process [13]. Intrinsic camera parameters include focal length, principal point, and skew coefficients. In addition, extrinsic camera parameters represent rotation and translation characteristics of the camera.

In the ToF image, each pixel is projected to a 3D point, i.e., world coordinate, then this 3D point is projected to the destination image. For ToF-to-color 3D warping, the ToF camera and the color camera are source and destination, respectively. The 2D image coordinate at the destination image is acquired by the 2D image point at the source image and its intrinsic and extrinsic camera parameters.

From depth image warping, the depth image corresponding to the color camera view is acquired. These images contain empty pixels due to the resolution difference between ToF and color cameras. Joint bilateral filter (JBF) is used to fill such areas [15]. The filter is applied to the object region only, assuming the approximate depth value range of the object is known. JBF is an extension of the bilateral filter [14], which is widely used for edge preservation.

$$D(x,y)=\frac{\sum_u\sum_v W(u,v)\cdot D_i(x,y)}{\sum_u\sum_v W(u,v)} \quad (1)$$

In (1), Di(x, y) and D(x, y) denote the value at (x, y) coordinate in the warped depth image and the final depth image which is to be filled, respectively. Di(x, y) is available as a result of ToF-to-color warping. (u, v) is the neighbor coordinate of (x, y). r represents the kernel size. In the experiments, the kernel size r is 11. W represents the weight, which is zero if the pixel value in the warped depth image is zero. Otherwise, the weight is a multiple of spatial weight f(u, v) and range weight g(u, v). This is represented in (2).

$$W(u,v)=\begin{cases}0 & ,\text{if } D_i(x,y)=0\\ f(u,v)\cdot g(u,v) & ,\text{otherwise}\end{cases} \quad (2)$$

The spatial weight is based on the intensity difference. When computing the intensity difference, JBF uses the color image while the bilateral filter uses the depth image itself. JBF produces more reliable spatial weights since color data difference can be more specific. The range weight is the same in both JBF and the bilateral filter. These weights are explained by (3) and (4). σf and σg are sigma values for the spatial and range weights which are Gaussian values; their values are 2 and 8, respectively, in the experiments.

$$f(u,v)=\exp\left\{-\frac{|I(x,y)-I(u,v)|^2}{2\sigma_f{}^2}\right\} \quad (3)$$

$$g(u,v)=\exp\left\{-\frac{(x-u)^2+(y-v)^2}{2\sigma_g{}^2}\right\} \quad (4)$$

Camera System Overview

The devised camera system consists of two ToF cameras positioned above three color cameras. Fig. 1 exhibits the devised camera system. The image resolution is 176×144 and 1280×720 for ToF and color, respectively. The distance between centers of color cameras is 6.5 cm and there is no space between the pair of ToF cameras. The objective is to minimize errors in T0 and T1. We carry out ToF distance correction using a ToF image pair prior to depth map generation process explained in the previous section.

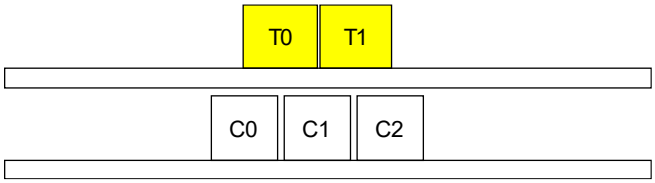


Figure 2. Camera system: two ToF cameras (T0 and T1) positioned above three color cameras (C0, C1, and C2)

Error Region Classification

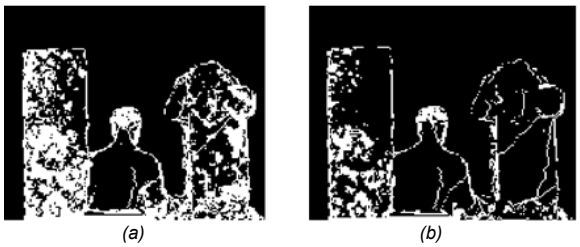
Distance correction is applied to samples that show inconsistency between the ToF images. We estimate erroneous regions by comparing the ToF image with an image warped from the other ToF. If the absolute difference between these images is greater than a threshold, this sample is treated as an erroneous distance value. This is described by (2) where ToF_diff is the binary difference image. In the equation, ToF0-to-ToF1 warped image is compared with ToF1. This can be executed vice versa. Fig. 3 shows captured ToF images and a warped ToF image. Since ToF0 has inconsistent distance data in the head region, the warped image exhibits holes. Subsequently, a large portion of the head becomes the error region in the binary difference image.

(a) (b) (c)

Figure 3. Captured ToF images and warped ToF image : (a) ToF 0; (b) ToF1; (c) ToF0-to-ToF1 warped image

In the experiments, we compare the results while varying the threshold: 30, 50, 70, and 90. Threshold 30 would mean the margin for distance difference that can be accepted is 30 mm. The larger the difference threshold, less samples would be classified as erroneous samples. In addition, if the coordinate in the ToF-to-ToF warped image is a hole, this coordinate is regarded as erroneous distance region, as described in (5). Fig. 4 displays binary difference images with different threshold values.

$$\text{ToF\_diff}(x,y)=\begin{cases}255, & |\text{ToF0\_to\_ToFl}(x,y)-\text{ToFl}(x,y)|>\text{threshold}\\ & \text{or ToF0\_to\_ToFl}(x,y)=0\\ 0, & \text{otherwise}\end{cases} \quad (5)$$



(a) (b)

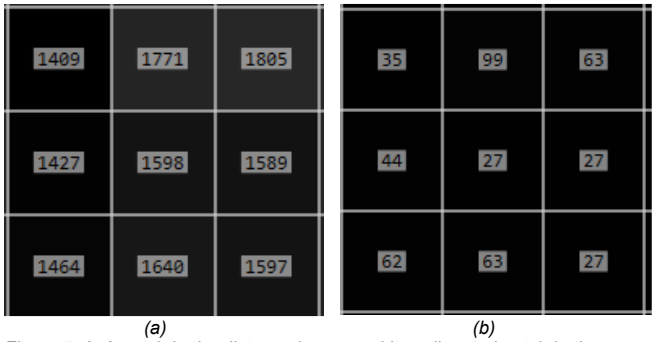
Figure 4. Binary difference image with varying threshold: (a) Threshold: 30; (b) Threshold: 70

Distance Correction Using Valid Neighbors

After determining which regions to apply distance correction, 3×3 filtering is applied to them. Since the image resolution is small, i.e., 176×144, only a small window works. We test four filtering methods: median, averaging, averaging without minimum and maximum, selection of sample with maximum amplitude. When capturing using the ToF camera, amplitude image is available as well. Hence, we explored the relativity in respect to

the distance image. Higher amplitude can mean higher reflectivity within the surface, i.e., more reliable distance sensing. Fig. 5 shows an example of 3×3 patch in the distance image and its collocated patch in the amplitude image.

Neighbor samples that are classified as erroneous samples are disregarded in the filtering process. In other words, erroneous samples are affected by reliable samples only. If reliable samples does not exist among the neighbor samples, the center sample cannot be modified. Thus, iterative execution is followed. Once the erroneous sample is modified, it is classified as a reliable sample and this is reflected in the following iteration.



(a) (b)

Figure 5. 3×3 patch in the distance image and its collocated patch in the amplitude image: (a) Distance; (b) Amplitude

In Fig. 5, the center sample with distance value 1598 needs to be modified. Although it is not shown, neighbor sample 1805 is classified as erroneous; thus, this value does not influence the outcome.

1) Median: median{1409, 1771, 1427, 1589, 1464, 1640, 1597} = 1589

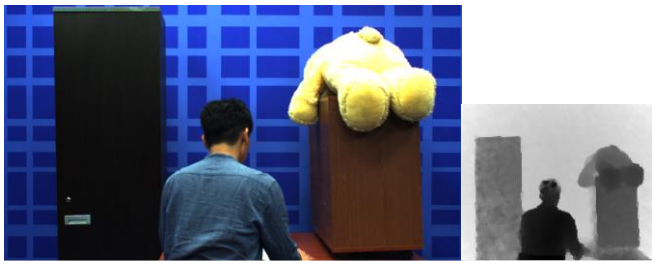
2) Averaging: avg{1409, 1771, 1427, 1589, 1464, 1640, 1597} = 1557

3) Averaging without minimum and maximum: avg{1427, 1589, 1464, 1640, 1597} = 1543

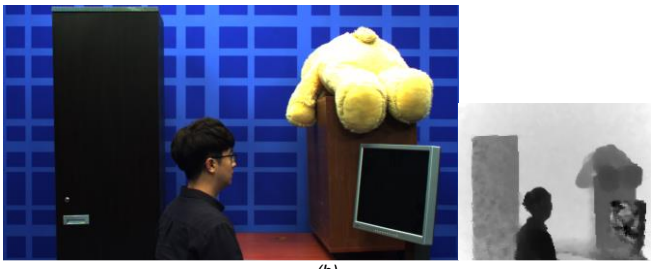
4) Maximum amplitude: Neighbor sample of distance 1771 possesses the largest amplitude (99).

Experiment Results

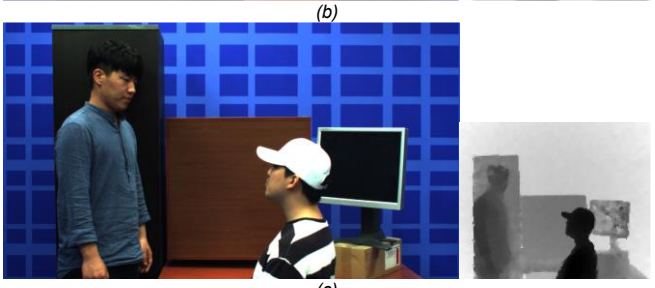
We conducted experiments on four test images with flat background. The arranged camera setup is covered in Section 4.1. Since the proposed method relies on ToF camera setup, public dataset images were not used. Background subtraction and color correction were applied to color images prior to depth map generation process. Fig. 6 represents the test images captured by color and ToF test images.



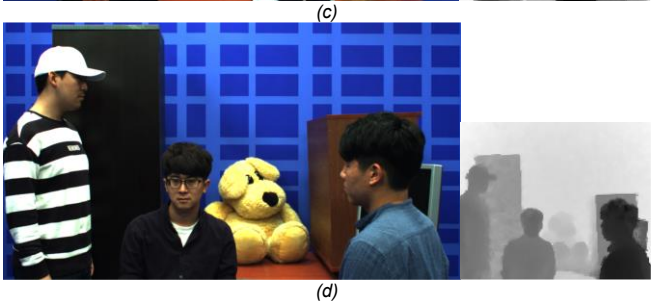
(a)



(b)



(c)



(d)

Figure 6. Four test images captured by color and ToF cameras: (a) Image 1; (b) Image 2; (c) Image 3; (d) Image 4

The acquired depth maps are used to warp the source color view to the adjacent target color view. The color image captured at the target view is used as an anchor for PSNR calculation. In real applications, color views would be warped to arbitrary viewpoints. Holes exist in the warped color image since the result depth maps are not ground truth data. Thus, image in-painting was employed to fill such holes. Table 1, Table 2, Table 3, and Table 4 list the comparison of captured C2 and virtual C2. This is for the evaluation of generated D1 while varying the difference threshold. PSNR values are calculated on foreground values only, discarding background. Fig. 8 represent the depth map and virtual color image results for Image 4.

Table 1: Comparison of captured C2 and virtual C2, difference threshold: 90

	Original ToF		Median		Average		Average w/o min, max		Max amplitude	
	PSNR	Holes	PSNR	Holes	PSNR	Holes	PSNR	Holes	PSNR	Holes
Image 1	25.11	3064	26.50	1605	26.45	1602	26.52	1578	26.66	1621
Image 2	22.27	2686	25.04	1110	25.11	1313	25.15	1315	25.07	1213
Image 3	17.94	2476	18.67	946	18.73	956	18.71	979	18.89	1258
Image 4	19.00	1385	20.83	689	21.01	691	20.99	701	21.09	839
Average	21.08	2403	22.76	1088	22.83	1141	22.84	1143	22.93	1233

Table 2: Comparison of captured C2 and virtual C2, difference threshold: 70

	Median		Average		Average w/o min, max		Max amplitude	
	PSNR	Holes	PSNR	Holes	PSNR	Holes	PSNR	Holes
Image 1	26.50	1605	26.45	1602	26.52	1578	26.66	1621
Image 2	25.04	1110	25.11	1313	25.15	1315	25.07	1213
Image 3	18.67	946	18.73	956	18.71	979	18.89	1258
Image 4	20.83	689	21.01	691	20.99	701	21.09	839
Average	22.76	1088	22.83	1141	22.84	1143	22.93	1233

Table 3: Comparison of captured C2 and virtual C2, difference threshold: 50

	Median		Average		Average w/o min, max		Max amplitude	
	PSNR	Holes	PSNR	Holes	PSNR	Holes	PSNR	Holes
Image 1	26.51	1607	26.46	1612	26.61	1615	26.63	1807
Image 2	25.01	1089	25.04	1307	25.08	1301	25.1	1198
Image 3	18.60	924	18.74	912	18.71	900	18.97	1253
Image 4	20.84	689	21.05	695	21.00	682	21.24	841
Average	22.74	1077	22.82	1132	22.85	1124.5	22.99	1275

Table 4: Comparison of captured C2 and virtual C2, difference threshold: 30

	Median		Average		Average w/o min, max		Max amplitude	
	PSNR	Holes	PSNR	Holes	PSNR	Holes	PSNR	Holes
Image 1	26.55	1512	26.55	1622	26.6	1659	26.81	1859
Image 2	25.02	1042	25.04	1299	25.10	1268	25.22	1199
Image 3	18.58	990	18.75	945	18.73	1003	19.08	1270
Image 4	20.89	725	21.05	689	21.02	678	21.3	856
Average	22.76	1067	22.85	1139	22.86	1152	23.10	1296

Figure 7. Depth map results and virtual warped color images for Image 4: (a) Using original ToF; (b) Using median filter; (c) Using averaging; (d) Using averaging without minimum and maximum

In terms of PSNR and number of holes, averaging without minimum and maximum method showed the best result as the distance modification method. In addition, we concluded difference threshold 50 is the optimum choice. While median and averaging methods still produced reliable results, maximum amplitude selection method was not sufficient. Compared to the case when using the original ToF, the proposed method led to enhanced synthetic images. Fig. 8, Fig. 9, and Fig. 10 demonstrate the depth map and warped color image results for Image 1, Image 2, and Image 3, respectively.

Conclusion

In this paper, we devised a method that reduces distance errors in ToF images. The target errors are exhibited at boundaries and surfaces that are not capable of reflecting the infrared ray. For the proposed method, at least two ToF cameras are required in the camera setup. ToF distance error region is estimated by comparing the captured ToF image with a warped image from the neighboring ToF camera. The distance values in the error region are replaced. A number of methods were examined to select the optimum replacement value. Averaging without minimum and maximum method showed the best performance. After distance error reduction, this method is inserted into the aforementioned depth map generation framework. The performance was analyzed by evaluating synthetic images, which are generated by the depth map results. In regards to the distance modification method, averaging without minimum and maximum method showed the best result.

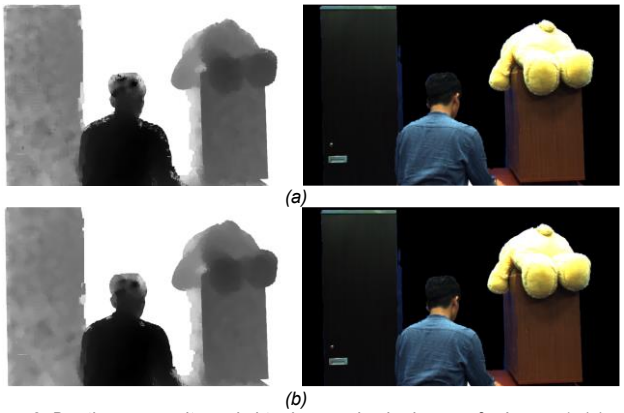


Figure 8. Depth map results and virtual warped color images for Image 1: (a) Using original ToF; (b) Using proposed method

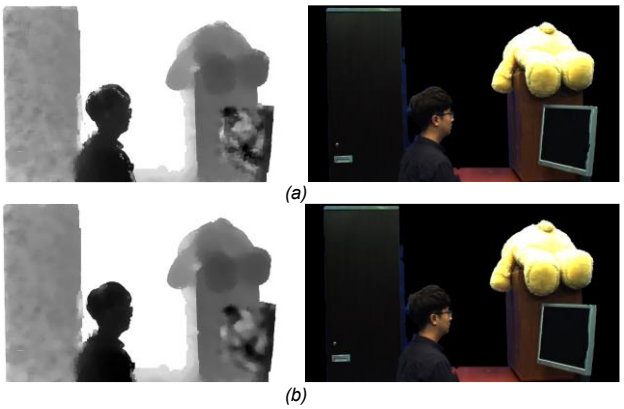


Figure 9. Depth map results and virtual warped color images for Image 2: (a) Using original ToF; (b) Using proposed method

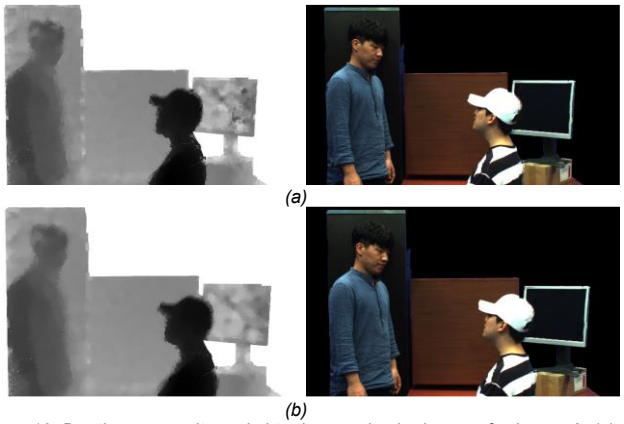


Figure 10. Depth map results and virtual warped color images for Image 3: (a) Using original ToF; (b) Using proposed method

Acknowledgment

This work was supported by 'The Cross-Ministry Giga KOREA Project' grant funded by the Korea government(MSIT) (GK17C0100, Development of Interactive and Realistic Massive Giga- Content Technology).

References

[1] C. Fehn, “Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3-D TV,” in Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XI, 2004, vol. 5291, no. 2, pp. 93-104.

[2] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, “View generation with 3D warping using depth information for FTV,” Signal Processing: Image Communication, vol. 24, no. 1, pp. 65-72, Jan. 2009.

[3] Y. S. Kang and Y. S. Ho, “An efficient image rectification method for parallel multi-camera arrangement,” IEEE Trans. Consum. Electron., vol. 57, no. 3, pp. 1041-1048, Aug. 2011.

[4] Y. Song, C. Lee, and Y. S. Ho, “Adaptive depth boundary sharpening for effective view synthesis,” in Proc. Picture Coding Symposium, 2012, pp. 73-76.

[5] Y. Song and Y.S. Ho, “Depth map boundary filter for enhanced view synthesis in 3D video,” Journal of Signal Processing Systems, DOI 10.1007/s11265-016-1158-x, pp. 1-9, Aug. 2016.

[6] J. Y. Lee and H. W. Park, “Efficient synthesis-based depth map coding in AVC-compatible 3D video coding,” IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 3, pp. 1107-1116, June 2016.

[7] B. W. Micallef, C. J. Debono, and R. A. Farrugia, “Reducing 3D video coding complexity through more efficient disparity estimation,” IEEE Trans. Consum. Electron., vol. 60, no. 1, pp. 74-82, Feb. 2014.

[8] Y. J. Jung, H. Sohn, S. I. Lee, and Y. M. Ro, “Visual comfort improvement in stereoscopic 3D displays using perceptually plausible assessment metric of visual comfort,” IEEE Trans. Consum. Electron., vol. 60, no. 1, pp. 1-9, Feb. 2014.

[9] W. S. Jang and Y. S. Ho, “Efficient disparity map estimation using occlusion handling for various 3D multimedia applications,” IEEE Trans. Consum. Electron., vol. 57, no. 4, pp. 1937-1943, Nov. 2011.

[10] Y. Zhan, Y. Gu, K. Huang, C. Zhang, and K. Hu, “Accurate image-guided stereo matching with efficient matching cost and disparity refinement,” IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 9, pp. 1632-1645, Sept. 2016.

[11] C. Lee, H. Song, B. Choi, and Y. S. Ho, “3D scene capturing using stereoscopic cameras and a time-of-flight camera,” IEEE Trans. Consum. Electron., vol. 57, no. 3, pp. 1370-1376, Aug. 2011.

[12] Z. Zhang, “Flexible camera calibration by viewing a plane from unknown orientations,” in Proc. IEEE International Conference on Computer Vision, 1999, pp. 666-673.

[13] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, “Joint bilateral upsampling,” ACM Transactions on Graphics., vol. 26, no.3, pp. 1-5, Aug. 2007.

[14] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in Proc. IEEE International Conference on Computer Vision, 1998, pp. 839-846.

[15] D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," IEEE Trans. Image Process., vol. 26, no.3, pp. 1176-1190, March 2012.

Author Biography

Yunseok Song received his B.S. in Electrical Engineering from Illinois Institute of Technology (2008) and M.S. in Electrical Engineering from University of Southern California (2009). He is pursuing a Ph.D. in the School of Electrical Engineering and Computer Science at Gwangju Institute of Science and Technology. His research interests include 3D image processing, video processing, and realistic broadcasting.

Yo-Sung Ho received both his B.S. and M.S. in Electronic Engineering from Seoul National University (1981 and 1983, respectively) and Ph.D. in Electrical and Computer Engineering from University of California at Santa Barbara (1990). Since 1995, he has been with Gwangju Institute of Science and Technology, where he is currently a professor in the School of Electrical Engineering and Computer Science. His research interests include image analysis and image restoration, 3D television, and realistic broadcasting.