Depth Image Filter for Mixed and Noisy Pixel Removal in RGB-D Camera Systems

Sung-Yeol Kim, Manbae Kim, and Yo-Sung Ho, *Senior Member, IEEE*

Abstract — *The commercial RGB-D camera produces color images and their depth maps from a scene in real time. However, the active camera creates mixed depth data near the border of different objects, occasionally losing depth information of shiny and dark surfaces in the scene. Furthermore, noise is added to the depth map. In this paper, a new method is presented to resolve such mixed, lost, and noisy pixel problems of the RGB-D camera. In particular, mixed pixel areas are detected using common distance transform (CDT) values of color and depth pixels, and merged them to lost pixel regions. The merged regions are filled with neighboring depth information based on an edge-stopping convergence function; distance transform values of color edge pixels are used to form this function. In addition, a CDT-based joint multilateral filter (CDT-JMF) is used to remove noisy pixels. Experimental results show that the proposed method gives better performance than conventional hole filling methods and image filters*[1].

Index Terms —RGB-D camera, depth image filter, distance transform, mixed pixel.

I. Introduction

Following the significant advances in depth information acquisition technologies over the last few years, high performance RGB-D cameras [1], [2], [3] have been developed recently. The RGB-D camera produces a sequence of a color image and its depth map pair of a natural scene in real time. In general, the active TOF camera provides more accurate depth information in textureless and texture-patterned scenes than passive depth estimation methods [4].

However, the quality of the depth map captured by the RGB-D camera is degraded by following three problems [5], [6]; 1) *mixed pixels*, 2) *lost pixels*, and 3) *noisy pixels*. Fig. 1(d) presents mixed pixels in the 3D scene reconstructed by the color image and its depth map of *apple* RGB-D data [7]. When an IR ray hits the boundary of an object, the portion of the ray is reflected by the front object and the other part by the background objects. Both reflections are received by an RGB-D camera and result in a mixed measurement [8]. Mixed pixels usually lie on near the border of different objects.

Fig. 1(e) exhibits lost pixels in the 3D scene. The RGB-D camera has difficulty in obtaining depth data of shiny and dark surfaces, because IR rays reflected from these surfaces are weak or scattered. This phenomenon results in lost pixels in a depth map. Furthermore, since depth distortion by mixed pixels is too serious to employ in real applications, some RGB-D cameras provide a function to handle a part of mixed pixels as lost pixels. Thus, lost pixel recovery is an important task to be resolved.

Lastly, optical noise is often added to a depth map. The noise mostly occurs due to non-linear response of IR sensors and different reflectivity of IR rays on the variation of object surface materials. As shown in Fig. 1(f), noisy pixels are usually observed inside of objects.

Those inherent problems make it difficult to employ the active camera in various applications. Presently, the practical use of the RGB-D camera is limited in applications mainly involving foreground extraction [9] and motion tracking [1].

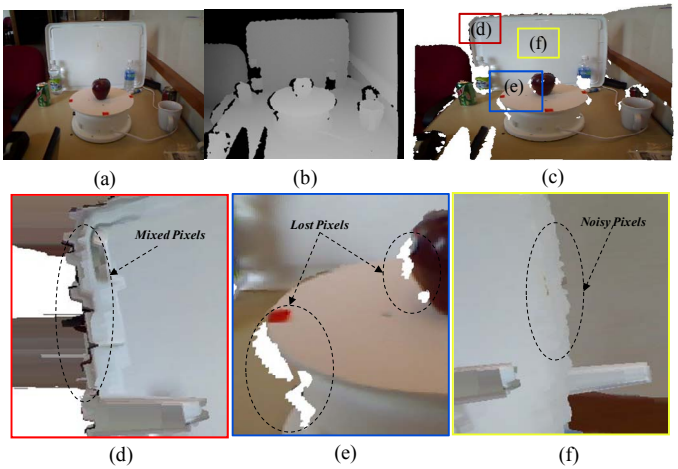


Fig. 1. Inherent problems of RGB-D cameras; (a) a color image, (b) a depth map of (a), (c) 3D scene reconstructed by (a) and (b), (d) mixed pixels, (e) lost pixels, and (f) noisy pixels.

In this paper, a new method is proposed to resolve such mixed, lost, and noisy pixel problems. For dealing with mixed and lost pixels, mixed pixel areas are merged with lost pixel

[1] This work was supported by the MOTIE (Ministry of Trade, Industry and Energy)/KEIT, Korea under System and Semiconductor Application Promotion Project and also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0009228).

S. -Y Kim is with Digital Media and Communication R&D Center, Samsung Electronics, Suwon, Republic of Korea (e-mail:sy1975.kim@samsung.com).

M. Kim is with the Department of Computer and Communications Engineering, Kangwon National University, Chuncheon, Kangwon, Republic of Korea (e-mail: manbae@kangwon.ac.kr).

Y.-S Ho is with the School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea (e-mail: hoyo@gist.ac.kr)

Contributed Paper
Manuscript received 07/01/13
Current version published 09/25/13
Electronic version published 09/25/13.

regions based on common distance transform (CDT) [10] values; CDT is a new transform that measures pixel-modal similarity of color and depth pixels. It is assumed that mixed pixels are located on textured areas having high pixel-modal similarity. Then, the merged region is filled by neighboring depth information based on an edge-stopping convergence function; distance transform [19] values of color edge pixels are used to form this function.

For noisy pixel removal, CDT-based joint multilateral filter (CDT-JMF) is proposed. CDT-JMF selects valid color data based on pixel-modal similarity; only the selected color data is used for depth map denoising. Due to the color data selection, noisy pixels can be minimized effectively while suppressing visual artifacts fabricated from useless color information.

The main contributions of this paper are three-folds: (a) *mixed pixel region detection based on CDT values*, (b) *lost pixel recovery using the DT-based edge-stopping convergence function*, and (c) *noisy pixel removal via CDT-JMF*. This paper is organized as follows. In Section II, related works are introduced briefly. Section III explains the proposed method in detail. After providing experimental results in Section IV, it is concluded in Section V.

II. Related Work

A. Hole Filling

Hole filling has been a challenging task over the last two decades. Many hole filling algorithms have been developed. For instance, Telea's image inpainting [11] is widely used to fill empty pixels in the field of computer vision and image processing. However, there are few works related to depth hole filling. A median filter [12] was introduced to fill the lost pixels. The method selects the exploitable depth pixels and carries out a median filter recursively. Recently, a joint bilateral filter (JBF)-based depth hole filling method [13] has been presented. The method recovers lost pixels using iterative JBF.

B. Depth Map Denoising

JBF [14] is often used to remove noisy pixels. Unlike a bilateral filter (BF) [15], JBF is based on color data instead of depth data. Formally, by considering color differences between the color value c_x at a pixel position x and its neighbors $\{c_n\}$ at x_n, the new depth value d_x^{new} at x is computed by

$$d_x^{new} = \frac{\sum_{n \in W \times W} K_S(\| x_n - x \|) K_C(\| c_n - c_x \|) d_n}{\sum_{n \in W \times W} K_S(\| x_n - x \|) K_C(\| c_n - c_x \|)} \quad (1)$$

where K_S and K_C are spatial and color weighting functions, S and C are smoothing parameters of K_S and K_C, respectively. If a Gaussian function is used to model K_S, S becomes its standard deviation. $W \times W$ is the size of local analysis window and $\|\cdot\|$ is an operator to calculate Euclidian distance.

If depth data are additionally considered by the depth value d_x and its neighbors $\{d_n\}$ in (1), a joint multilateral filter (JMF) [16] is formulated as follows:

$$d_x^{new} = \frac{\sum_{n \in W \times W} K_S(\| x_n - x \|) K_C(\| c_n - c_x \|) K_P(\| d_n - d_x \|) d_n}{\sum_{n \in W \times W} K_S(\| x_n - x \|) K_C(\| c_n - c_x \|) K_P(\| d_n - d_x \|)} \quad (2)$$

where K_P is the depth weighting function and P is the smoothing parameter of K_P.

In literature, Petschnigg *et al.* [14] developed JBF to compute the edge-stopping function using the flash image and non-flash image. Kopf *et al.* [17] employed the concept of JBF to upsample image resolution from low to high. Yang *et al.* [18] also presented iterative JBF to increase spatial image resolution of a depth map. For depth map denoising, Cho *et al.* [16] and Kim *et al.* [6] presented JMF for generating dynamic 3D human actors based on color and depth data.

III. CDT-Based Depth Image Filter

A. Overall Framework

Fig. 2 illustrates the overall block diagram of the proposed method. Blocks colored in yellow represent the flow to handle mixed and lost pixels. A process for minimizing noisy pixels is carried out through a group of blocks colored in blue.

For mixed and lost pixel recovery, the proposed method is carried out as follows: 1) Two edge maps are extracted from a depth map and its color image, 2) Distance transform (DT) is performed on both edge maps, 3) CDT values are derived by comparing DT values of color and depth pixels, 4) Mixed pixel areas are detected based on the CDT values and merged with lost pixel regions; a depth hole area to be filled is defined, and 5) Hole filling is carried out using an edge-stopping convergence function that is formulated by DT values of color edge pixels.

For noisy pixel removal, CDT-JMF is implemented as follows: 1) Scale factor w_n^c of K_C is calculated based on the CDT map, 2) K_C is modified to K_C' by w_n^c, 3) A weighting function is derived by combining K_S, K_C', and K_P, and 4) JMF is carried out based on the weighting function.

B. Mixed and Lost Pixel Recovery

Since mixed pixels are mismeasured depth data, mixed pixel areas and merged with lost pixel regions are removed. Lost pixel regions are easily detected by searching for zero depth areas in a depth map whereas mixed pixel regions are not.

A CDT map is utilized to detect mixed pixel areas. The CDT map represents pixel-modal similarity between a depth pixel and its corresponding color pixel. Pixel modality is measured by the DT values of pixels. It is assumed that mixed pixels are mostly located on textured areas having high pixel-modal similarity.

Prior to DT, as shown in Fig. 3(a) and Fig. 3(b), the color edge map E_C and the depth edge map E_D are extracted from input color image and its depth map using an edge detection operator [20]. Note that isolated edges are ignored by applying a median filter to input images before edge extraction.

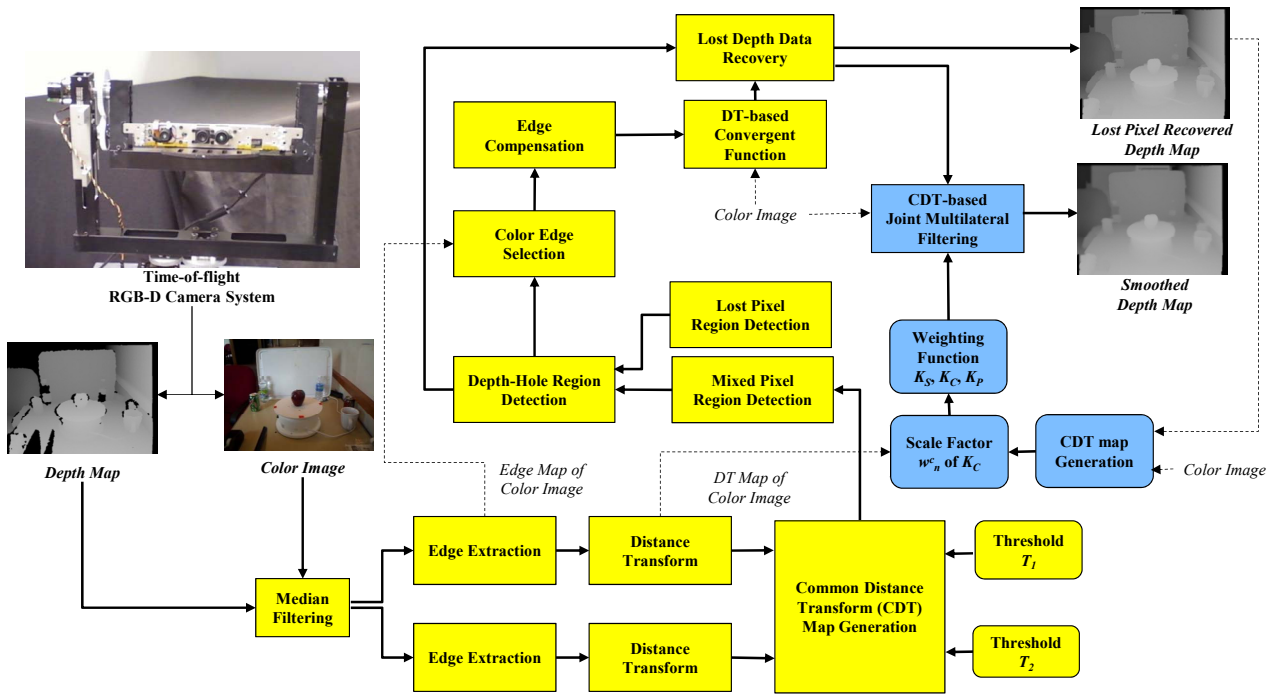


Fig. 2. Overall framework of 3D scene reconstruction using a RGB-D camera

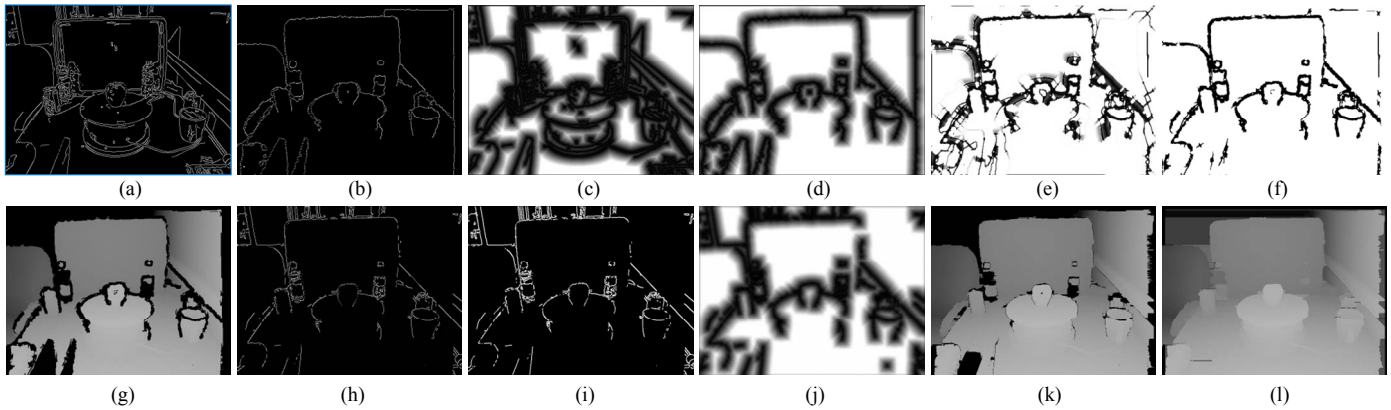


Fig. 3. The procedure of mixed and lost pixel recovery; (a) color edge map of Fig. 1(a), (b) depth edge map of Fig. 1(b), (c) color DT map, (d) depth DT map, (e) CDT map, and (f) mixed pixel region, (g) modified depth map *H*, (h) selected color edges, (i) compensated color edges, (j) DT-based convergence function *Z*, (k) depth hole filling using Z, and (l) depth hole filling for unknown region *U*.

For the DT, edge pixels in E_C and E_D are initially set to zero, whereas non-edge pixels are set to infinity. Formally, based on *a*-*b* distance transform (*a*-*b* DT), the DT value $dt_{i,j}^{k}$ at iteration k is computed by

$$dt_{i,j}^{k} = \min\ [dt_{i-1,j-1}^{k-1}+b, dt_{i-1,j}^{k-1}+a, dt_{i-1,j+1}^{k-1}+b, \\ dt_{i,j-1}^{k-1}+a,\quad dt_{i,j}^{k-1},\quad dt_{i,j+1}^{k-1}+a, \\ dt_{i+1,j-1}^{k-1}+b, dt_{i+1,j}^{k-1}+a, dt_{i+1,j+1}^{k-1}+b] \tag{3}$$

where a and b controls the strength of distance transform.

Fig. 3(c) and Fig. 3(d) demonstrate the color DT map DT^C and the depth DT map DT^D, respectively. It is observed that DT values of pixels close to edges are assigned small numbers. In contrast, pixels far from edges have great DT values. These DT values indicate that the latter may belong to a homogenous area whereas the former to a textured area.

Suppose that there are the depth DT value DT_x^D at x and its corresponding color DT value DT_x^C. If DT_x^D is equal to or similar to DT_x^C, they belong to either homogenous or textured areas. Otherwise, one pixel may be in a homogenous area while the other pixel is in a textured area. In this manner, pixel-modal similarity between a depth pixel and its color pixel is measured.

Formally, a CDT value DT_x^J at x in a CDT map DT^J is calculated by

$$DT_x^J = \begin{cases} 0 & if \quad DT_x^D > T_2 \,\&\, DT_x^C > T_1 \\ DT_x^D & if \quad \left|DT_x^D - DT_x^C\right| \le T_2 \\ 255 & otherwise \end{cases} \tag{4}$$

where T_1 and T_2 are common homogenous and textured region detection threshold, respectively. Fig. 3(e) presents an example of the CDT map.

If $DT_x^J = 0$, then x may be on common homogenous region of the depth map and the color image. In contrast, if $0 < DT_x^J < T_2$, x may be on common textured region and a mixed pixel. Therefore, mixed pixel area M is defined by

$$M_x = \begin{cases} 1 & if \quad 0 < DT_x^J < T_2 \\ 0 & otherwise \end{cases} \tag{5}$$

If M_x is equal to one, then x belongs to mixed pixel areas. Otherwise, x is present in non-mixed pixel regions. Fig. 3(f) shows an example of M.

Then, M is merged with lost pixel regions. As shown in Fig. 3(g), the merged region is defined as a depth hole area, which will be filled by neighboring depth information. The depth map H including the depth hole area is represented by

$$H_x = \begin{cases} 0 & if \quad M_x = 1 \\ d_x & otherwise \end{cases} \tag{6}$$

For efficient depth hole filling with H, predicting lost depth edges is required. However, since there exists no depth edge information in the depth hole area, it is difficult to predict the potential edges. Color edges corresponding with the depth hole area are regarded as lost depth edges. These color edges are stored to an edge map E_H, as shown in Fig. 3(h).

However, some isolated edge pixels in E_H might cause undesirable depth hole filling results. In order to connect the isolated pixels, E_H is expended by a dilation operation [23] and then the dilated edges are thinned [24]. The changed edge map E'_H is shown in Fig. 3(i).

An edge-stopping convergence function Z is considered. Z is estimated by the DT map DT^H of E'_H and color data. Z plays a role in terminating the pixel traversal during depth hole filling. Formally, Z is expressed by

$$Z_x = \sum_{n \in W \times W} K_S(\| x_n - x \|) K_H(\| c_n - c_x \|) K_H(\| dt_n^h - dt_x^h \|) dt_n^h \tag{7}$$

where K_S and K_H are Gaussian functions with standard deviation S and H, respectively. dt_x^h is the DT value in DT^H at x and c_x is the color value at x. Fig. 3(j) depicts the convergence function as an image.

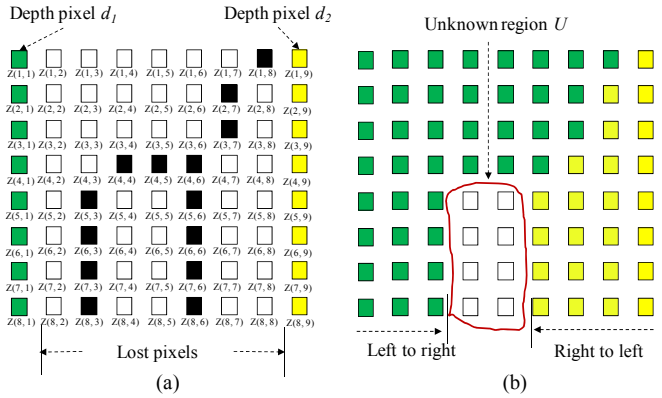


Fig. 4. Depth hole filling; (a) an example of lost pixel situation, (b) the procedure of depth hole filling.

Depth hole filling is carried out horizontally and vertically. Fig. 4 illustrates the procedure of horizontal depth hole filling based on Z. In Fig. 4(a), lost pixels are the black and white color pixels. Depth pixel d_1 (green color) and d_2 (yellow color) are the exploitable depth information to be used. Black color pixels have the lowest convergence value in each row.

As shown in Fig. 4(b), the direction of the traverse is left to right. For instance, in the first row, from a starting point (1, 1), the traverse continues to (1, 8) with the lowest convergence value. Then, the traversed pixel position is filled by d_1. Thereafter, the direction is changed to right-to-left and the traverse continues until meeting the black color pixel. Then depth holes are filled by d_2. In this manner, the depth holes are traversed and filled by d_1 and d_2. Fig. 3(k) shows the result of horizontal and vertical depth hole filling.

Note that unknown region U in Fig. 4(b) occurs during the depth hole filling. Unfortunately, there is no exploitable depth information to recover the region. Simply, an average value of d_1 and d_2 is used to fill these lost pixels. Fig. 3(l) shows the final result by the proposed depth hole filling.

C. Noisy Pixel Removal

CDT-JMF is presented to remove noisy pixels with the aid of their color data. CDT-JMF selects valid color data based on the pixel-modal similarity; only the selected color data is used for noisy pixel removal.

For instance, if the CDT map value is zero, the degree of pixel-modal similarity of a depth pixel and its color pixel is very high. i.e., the depth pixel is on a homogenous region in a depth map and its corresponding color pixel may be also a homogenous region in a color image. Therefore, the color data can be directly used for denoising.

In contrast, if the CDT value is infinity, e.g., 255 for an 8-bit gray-scaled image, the pixel modalities of both pixels are not identical. In this case, CDT-JMF only uses depth information. Otherwise, the degree of pixel-modal similarity is moderate. Then, the amount of color information to be used is determined by DT_x^J. Basically, the greater DT_x^J is, the less the proposed filter uses color information.

Formally, CDT-JMF is represented by

$$d_x^{new} = \begin{cases} \dfrac{\sum_{n \in W \times W} K_S(\| x_n - x \|) K_C(w_n^c \cdot \| c_n - c_x \|) K_P(\| d_n - d_x \|) d_n}{\sum_{n \in W \times W} K_S(\| x_n - x \|) K_C(w_n^c \cdot \| c_n - c_x \|) K_P(\| d_n - d_x \|)} & if \quad DT_n^J < T_1 \\ \dfrac{\sum_{n \in W \times W} K_S(\| x_n - x \|) K_P(\| d_n - d_x \|) d_n}{\sum_{n \in W \times W} K_S(\| x_n - x \|) K_P(\| d_n - d_x \|)} & otherwise \end{cases} \tag{8}$$

where w_n^c is a scale factor for K_C and $w_n^c \geq 1$.

For K_S, a box filter is used, which returns value 1 within $W \times W$ and value 0 outside it. The box filter is used for reducing the effect of the spatial term while increasing the

effect of the color term. The box filter is expressed by

$$K_S(x)=\begin{cases}1 & if \quad x\in W\times W\\ 0 & otherwise\end{cases} \tag{9}$$

For K_C and K_P, exponential functions are used as follows:

$$K_C(x)=e^{-\frac{x^2}{C^2}},\quad K_P(x)=e^{-\frac{x^2}{P^2}} \tag{10}$$

where C and P are smoothing parameters of K_C and K_P.

w_n^c is derived directly from DT_n^J in the CDT map as follows:

$$w_n^c=\begin{cases}1 & if \quad DT_n^J\le a\\ e^{\log\beta\cdot\frac{DT_n^J-a}{T-a}} & if \quad DT_n^J<T_1\end{cases} \tag{11}$$

where a controls the strength of DT, β indicates the maximum scale factor with $\beta>1$, and T_1 is the threshold in (4).

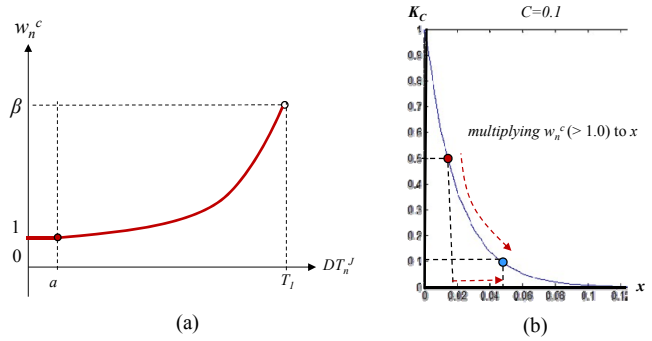


Fig. 5. Relationship between w_n^c and K_C; (a) Scale factor w_n^c based on DT_n^J and (b) K_C is changed by w_n^c.

Fig. 5(a) and Fig. 5(b) show the graphs of w_n^c and the variation of K_C with respect to w_n^c. If $DT_n^J \le a$, then $w_n^c = 1$. In this case, the degree of pixel-modal similarity of a depth pixel and its color pixel is high. In (8), since $\|c_n\text{-}c_x\|$ is multiplied by w_n^c (= 1), K_C is unchanged.

On the other hand, if $a < DT_n^J < T_1$, then $w_n^c > 1$. In this situation, the degree of pixel-modal similarity is not high but moderate. As shown in Fig. 5(b), w_n^c is multiplied by $\|c_n\text{-}c_x\|$ for moving K_C toward zero. Hence, the effect of color data is reduced during depth map denoising.

Fig. 6 illustrates the comparison of a Gaussian filter (GF), a BF, a JBF and the proposed CDT-based JMF. In a 5×5 depth map (Fig. 6(b)), the interest pixel whose depth value is 8 is located at (3, 3). It is assumed that the gray pixels belong to an object and that three noisy pixels are present in this depth map. Two noisy pixels at (2, 2) and (4, 1) have a value of 8. Although their pixel values (8) are similar to other depth values, the pixels are isolated and out of the object represented by depth value 8 or 9. In addition, the other pixel located at (4, 4) is a noisy pixel because its depth value 6 is different from object depth value 8 or 9.

Fig. 6(a) shows a weighting scale according to color. In GF, K_S (Fig. 6(e)) is only used as shown in Fig. 6(i). In BF, since the shape K_P is so precipitous, the depth pixel at (4, 4) will be ignored. Therefore, BF uses the combination weighting function of K_S and K_P (Fig. 6(f)) for depth map denoising as shown in Fig. 6(j). In JBF, K_C (Fig. 6(g)) is used instead of K_P as shown in Fig. 6(k). In proposed CDT-based JMF, K_C is changed to K'_C by multiplying it by w_n^c that is computed based on a CDT map (Fig. 6(d)). The result of the proposed filter is shown in Fig. 6(l). In this comparison, it is noticeable that the proposed filter only uses noiseless depth pixels to estimate a new depth value at (3, 3) whereas the other filters use the noisy pixels at (2, 2), (4, 1), and (4, 4). As a result, the new depth value at (3, 3) computed by the proposed method will be almost 9.

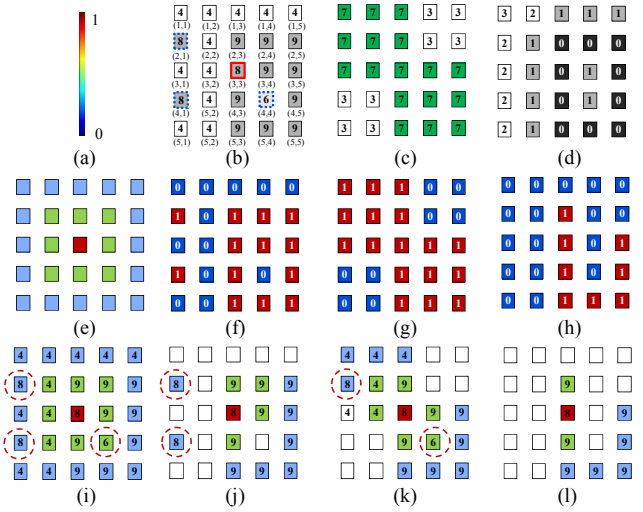


Fig. 6. Illustration of the comparison of GF, BF, JBF and proposed CDT-based JBF; (a) weighting scale, (b) depth map, (c) color image, (d) CDT map, (e) spatial weighting function K_S, (f) range weighting function K_P, (g) color weighting function K_C, (h) K'_C, (i) GF, (j) BF (K_S×K_P), (k) JBF (K_S×K_C), and (l) Proposed (K_S×K'_C×K_P).

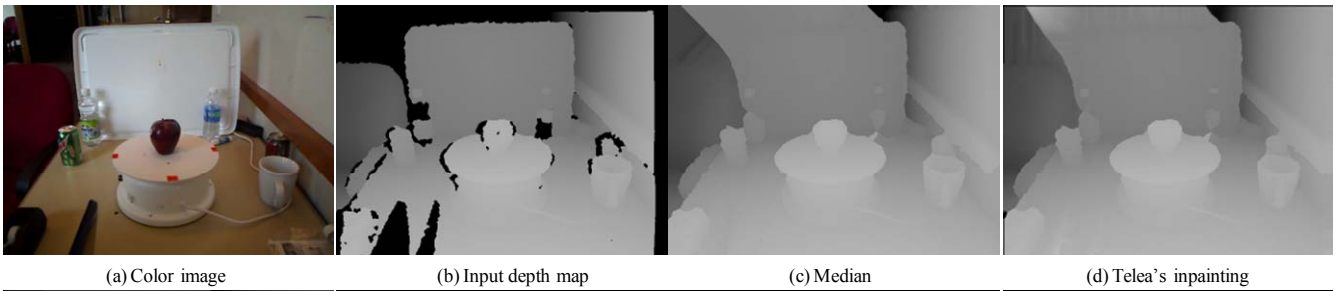
IV. Experimental Result

For evaluating the performance of the proposed depth hole filling method, the proposed method was tested with RGB-D datasets provided by [7]. Three datasets were selected; the proposed method was compared to four other methods: *apple*, *kitchen*, and *meeting*.

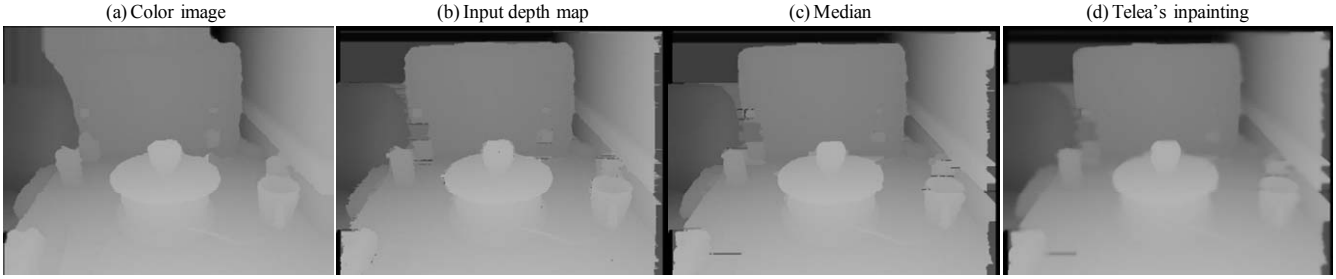
The proposed method was compared to four other methods: median filtering [12], Telea's inpainting [11], Navier-Stokes inpainting [21], and JBF-based methods [13]. For CDT map generation, T_1 is set to 18 and T_2 to 54 in (4). *9-10* DT is used for pixel modality estimation. For median filtering, local analysis window size 5×5 was used. For inpainting methods, the range parameter was set to 1. In JBF method, $S = 2$ and $C = 0.1$, respectively.

Fig. 7, Fig. 8, and Fig. 9 show the result of *apple*, *kitchen*, and *meeting* by the comparative methods. As observed from the result of *apple*, the proposed method (Fig. 7(g) and Fig. 7(h)) leads to better depth information near an apple and an plate than median filtering (Fig. 7(c)), image inpainting (Fig. 7(d) and Fig. 7(e)), and JBF method (Fig. 7(f)).

In addition, as shown in Fig. 8, the depth data of the hat in *kitchen* is better recovered by the proposed method than the other methods. Furthermore, the result of *meeting* shows an apparent difference between the proposed method and the others in recovering depth data of the part to a chair.

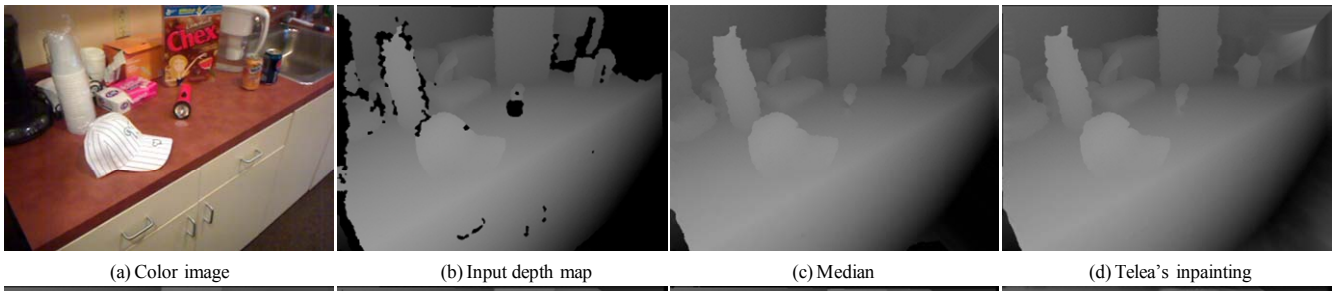


(a) Color image (b) Input depth map (c) Median (d) Telea's inpainting

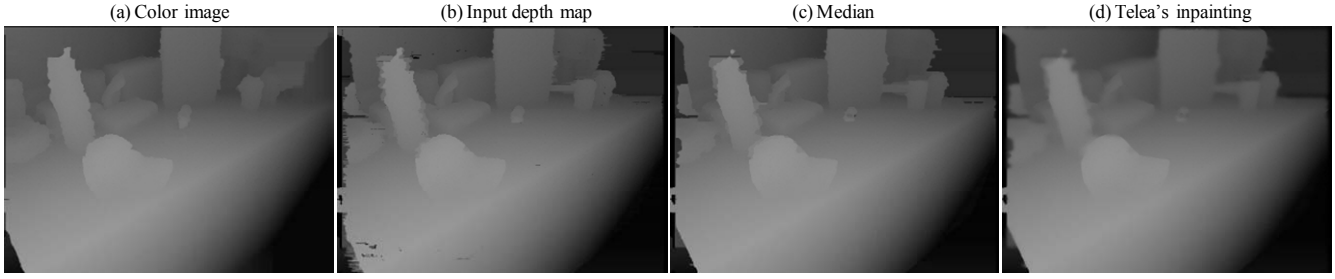


(e) Navier-Stokes image inpainting (f) JBF method (g) Proposed depth-hole filling (h) Proposed method + CDT-JBF

Fig. 7. Result of *apple* RGB-D image.

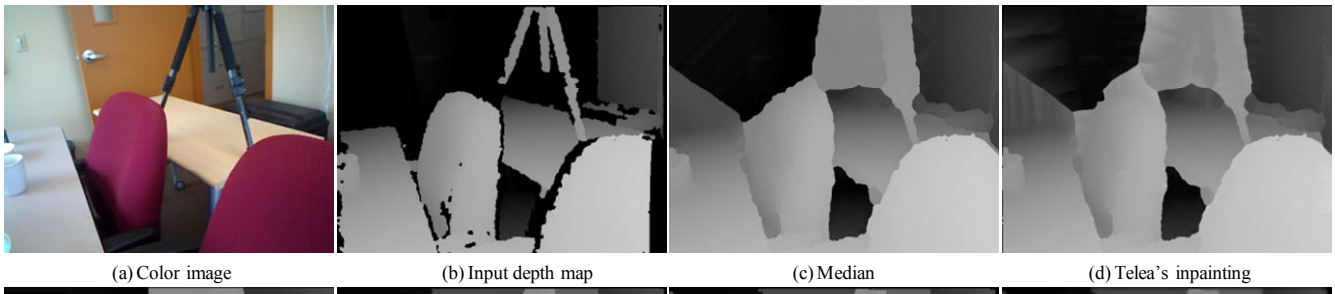


(a) Color image (b) Input depth map (c) Median (d) Telea's inpainting

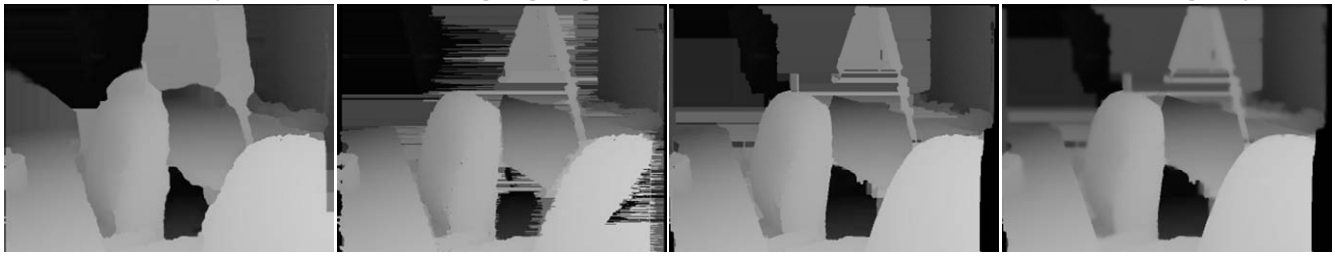


(e) Navier-Stokes image inpainting (f) JBF method (g) Proposed depth-hole filling (h) Proposed method + CDT-JBF

Fig. 8. Result of *kitchen* RGB-D image.



(a) Color image (b) Input depth map (c) Median (d) Telea's inpainting



(e) Navier-Stokes image inpainting (f) JBF method (g) Proposed depth-hole filling (h) Proposed method + CDT-JBF

Fig. 9. Result of *meeting* RGB-D image.

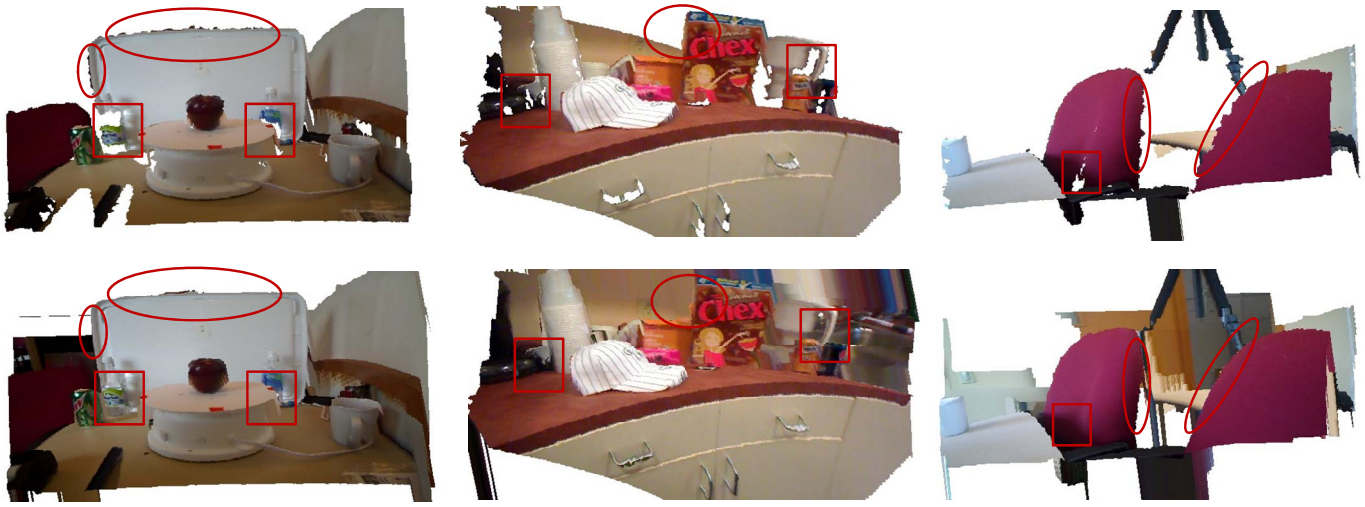


Fig. 10. Result of 3D scene reconstruction; (Row 1) the original 3D scenes and (Row 2) Recovered 3D scenes by the proposed method

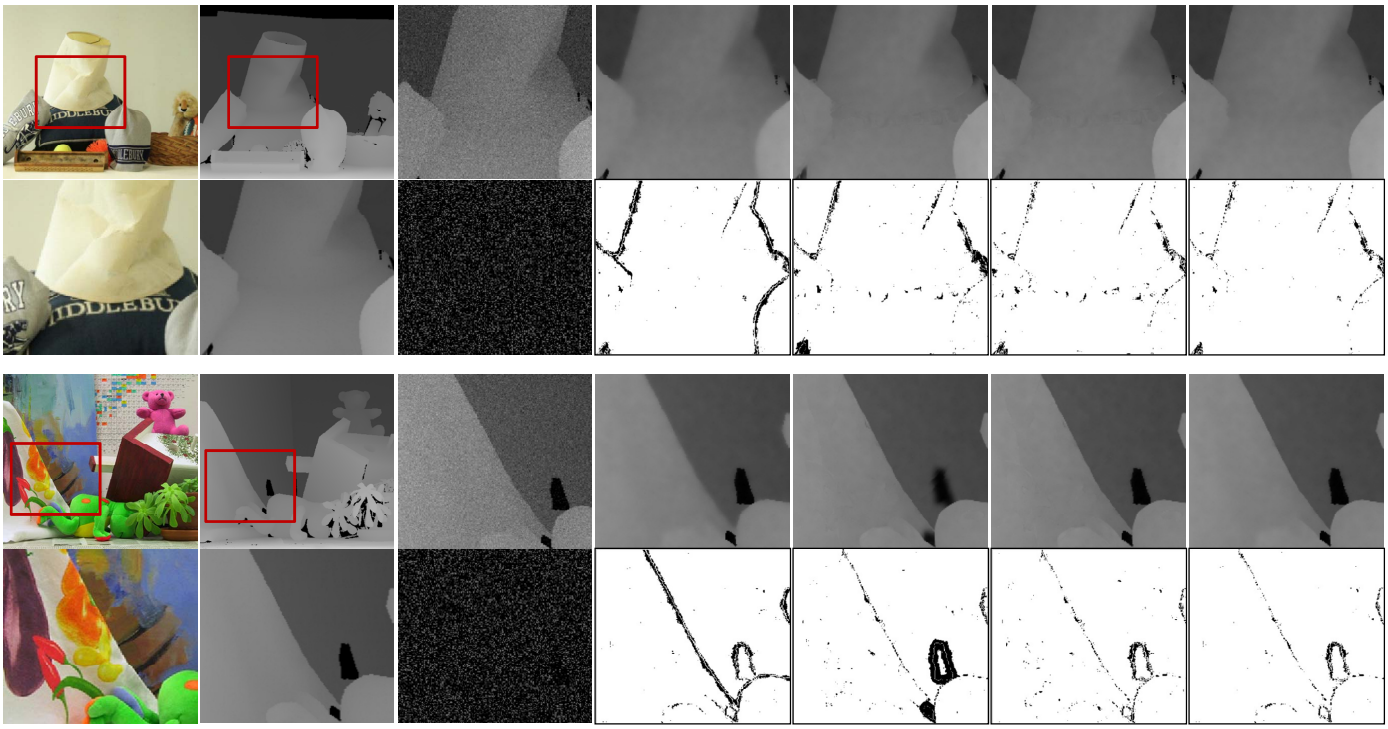


Fig. 11. Result of *midd1* and *teddy*; (a) color image and a part of it, (b) ground truth depth map and a part of it, (c) a part of artificially-generated noisy depth map and its difference map with (b), (d) result of BF and its difference map with (b), (e) result of JBF and its difference map with (b), (f) result of JMF and its difference map with (b), (g) result of the proposed method and its difference map with (b).

Fig. 10 shows the result of 3D scene reconstruction using a depth image-based modeling method [22]. The first row of Fig. 10 shows the original 3D scene generated by a raw depth map and its color image. The second row shows the recovered 3D scene generated by the enhanced depth map via the proposed depth hole filling. It is observed that the lost pixels marked by rectangles are restored and mixed pixels marked by circles are minimized.

For assessing the improvement of the depth accuracy of the proposed CDT-JMF, it was tested with ground truth image sets provided from [4]. Ten datasets were selected; *baby1*, *bowling*, *cloth*, *flowerpots*, *lampshade*, *midd1*, *monopoly*, *rocks1*, *teddy*, and *wood1*. For the experiment, Gaussian noise (standard deviation σ=20) is artificially added to the ground truth data to generate noisy depth maps.

The values of some parameters needed for CDT-JMF are set as follows: $\beta = 1.5$ in (11), $W = 11$ in (8), $P = 0.1$ and $C = 0.1$ in (10). Peak signal to noise ratio (PSNR) measurement based on ground truth data was employed for an objective evaluation of depth quality improvements. The proposed method was compared to bilateral filter (BF) [15], JBF [17], and JMF [16]. For BF, the parameter S is set to 2 and P is set to 0.1. For JBF and JMF, $S = 2$, $P = 0.1$, and $C = 0.1$.

Table 1 shows the average PSNR comparison of the comparative methods. The average PSNRs are 35.61 *dB*, 32.18 *dB*, 36.33 *dB*, and 36.77 *dB* for BF, JBF, JMF and the proposed CDT-JMF, respectively. This result indicates that the proposed method outperforms other comparative methods by 1.16 *dB*, 4.59 *dB*, and 0.44 *dB* on average.

TABLE I
PSNR COMPARISON (UNIT: *dB*)

Test data	BF	JBF	JMF	Proposed
baby1	31.83	31.83	36.84	37.05
bowling	36.11	34.33	36.67	36.89
cloth	39.40	38.37	38.04	39.21
flowerpots	33.88	27.98	34.73	34.59
lampshade	34.91	31.95	35.16	36.14
midd1	34.88	31.08	35.23	35.32
monopoly	35.97	33.17	36.60	37.02
rocks1	34.88	30.57	35.36	35.57
teddy	37.35	28.87	38.37	38.54
wood1	36.91	33.14	36.27	37.35
Average	**35.61**	**32.13**	**36.33**	**36.77**

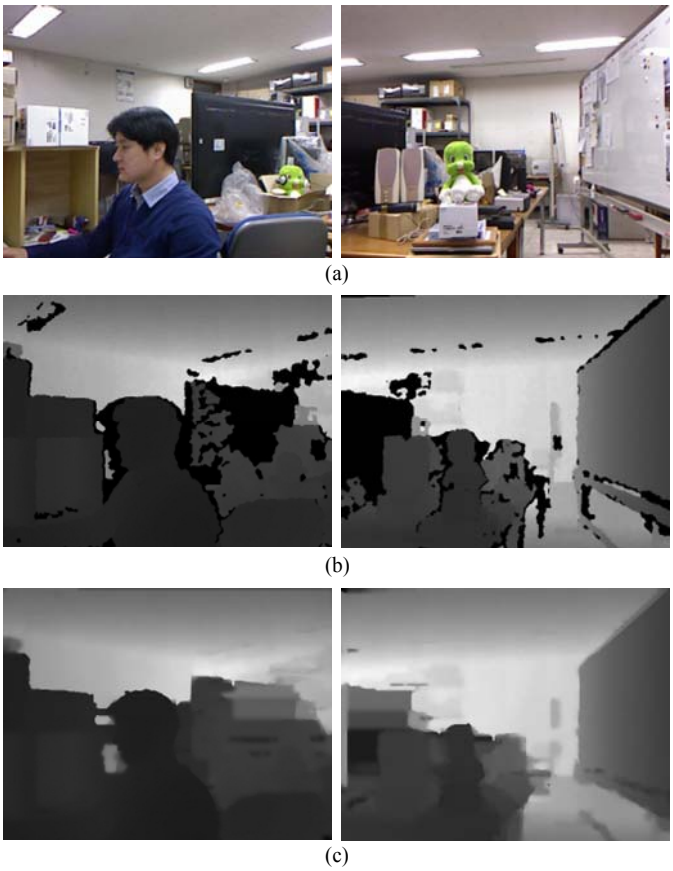


Fig. 12. Result of captured RGB-D images; (a) captured color images, (b) captured depth maps, and (c) output depth maps by the proposed method.

Fig. 11 presents the result of *midd1* and *teddy*. As shown in Fig. 11(d), BF blurs depth information on object boundaries more than other methods. In the case of JBF and JMF, some visual artifacts are observed in homogeneous areas. Those artifacts are occurred by transferring texture detail in the color image to homogeneous areas in the depth map. On the other hand, when the difference maps of BF, JBF, JMF are compared with the proposed method one, it is noticeable that the proposed method refines the depth data on object boundaries while suppressing visual artifacts of JBF and JMF.

Fig. 12 shows the result of color images and depth maps captured by an RGB-D camera [1]. As shown in Fig. 12(a) and Fig. 12(b), the captured scene is more complicated than *apple*, *kitchen*, and *meeting*. The proposed method provides restored depth data for objects in the scene, but the recovery of large depth hole areas, such as the black color TV, still remains problematic.

V. CONCLUSION

In this paper, a new method was presented to resolve inherent mixed, lost, and noisy pixel problems of time-of-flight RGB-D cameras. A common distance transform (CDT) map was used to detect the mixed pixel region and CDT-based joint multilateral filter (CDT-JMF) was developed to minimize noisy pixels. For mixed and lost pixel recovery, a convergence function based on distance transform was presented. Experimental results show that the proposed depth hole filling produces better refined depth information than median filtering, image inpainting, and JBF methods. In addition, in terms of noisy pixel removal, based on ten test depth maps, PSNR gains of the proposed method are approximately 1.16 *dB*, 4.57 *dB*, and 0.44 *dB* greater than bilateral filter, JBF, and JMF.

REFERENCES

[1] J. Shotton, A. Fitzgibbon, M. Cook, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Proc. of IEEE Computer Vision and Pattern Recognition*, 2011, pp.1297 - 1304.

[2] E.K. Lee, Y.S. Ho, "Generation of multi-view video using a fusion camera system for 3D displays," *IEEE Trans. Consum. Electron.,* vol. 56, no. 4, pp. 2797-2805, 2010.

[3] G.J. Iddan and G. Yahav, "3D imaging in the studio and elsewhere" in *Proc. of SPUE Videometrics and Optical Methods for 3D Shape Measurements*, 2001, pp. 48-55.

[4] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Jour. of Computer Vision*, vol. 47, no. 1-3, pp. 7-42, 2002.

[5] A. A. Dorrington, A. D. Payne, and M. J. Cree, "An evaluation of time-of-flight cameras for close range methodology applications," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Part 5 Commission V Symposium, 2010.

[6] S.Y. Kim, J. Cho, A. Koschan, and M. A. Abidi, "3D video generation and service based on a TOF depth sensor in MPEG-4 multimedia framework," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1730-1738, 2010.

[7] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view RGB-D object dataset," in *Proc. of IEEE International Conference on Robotics and Automation*, 2011, pp. 1817 - 1824.

[8] C. Ye, "Mixed pixels removal of a laser rangefinder for mobile robot 3-D terrain mapping," in *Proc. of IEEE International Conference on Information and Automation*, 2008, pp. 1153-1158.
[9] R. Grabb, C. Tracey, A. Puranik, and J. Davis, "Real-time foreground segmentation via range and color imaging," in *Proc. of IEEE Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1-5.
[10] S.Y. Kim, W. Cho, A. Koschan, and M. Abidi, "Depth map enhancement using adaptive steering kernel regression based on distance transform," *Lecture Notes in Computer Science*, vol. 6938, pp. 291-300, 2011.
[11] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of Graphics Tools*, vol. 9, no. 1, pp. 25-36. 2003.
[12] S.Matyunin, D. Vatolin, Y. Berdnikov, and M. Smirnov, "Temporal filtering for depth maps generated by kinect depth camera," in *Proc. of 3DTV Conference*, 2011, pp.1-4.
[13] M. Camplani and L. Salgado, "Efficient spatio-temporal hole filling strategy for kinect depth maps," in *Proc. of SPIE*, Three-Dimensional Image Processing and Applications II, 82900E, 2012.
[14] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 664-672, 2004.
[15] M. Elad, "On the origin of the bilateral filter and ways to improve it," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1141–1150, 2002.
[16] J. Cho, S.Y. Kim, Y.S. Ho, and K. H. Lee, "Dynamic 3D human actor generation method using a time-of-flight depth camera," *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1514-1521, 2008.
[17] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, pp.1-6, 2007.
[18] Q. Yang, R. Yang, J. Davis, and David Nistér, "Spatial-depth super resolution for range images," in *Proc. of IEEE Computer Vision and Pattern Recognition*, 2007, pp.1-8.
[19] G. Borgefors, "Hierarchical chamfer matching: a parametric edge matching algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 6, pp.849-865, 1988.
[20] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, 1986.
[21] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-strokes, fluid dynamic, and image video inpainting," *Proc. of IEEE International Conference on Pattern Recognition*, 2001, pp. 335-362.
[22] S.Y. Kim, S.B. Lee, and Y.S. Ho, "Three-dimensional natural video system based on layered representation of depth maps," *IEEE Trans. Consum. Electron.*, vol. 52, no. 3, pp. 1035-1042, 2006.
[23] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Prentice Hall, 2002.
[24] L. Lam, S.W. Lee, and C. Y. Suen, "Thinning methodologies – a comprehensive survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 9, pp.869-885, 1992.

BIOGRAPHIES

Sung-Yeol Kim received his B.S. degree in Information and Telecommunication engineering from Kangwon National University, South Korea, in 2001, and M.S. and Ph.D degree in Information and Communication Engineering at the Gwangju Institute of Science and Technology (GIST), South Korea, in 2003 and 2008, respectively. From 2009 to 2011, he was with imaging, robotics, and intelligent system lab at The University of Tennessee at Knoxville (UTK), USA, as a research associate. He is currently working for Digital Media and Communication R&D Center at Samsung Electronics. His research interests include digital image processing, depth image-based modeling and rendering, computer graphic data processing, 3DTV and realistic broadcasting.

Manbae Kim received the B.S. degree in Electronic Engineering from Hanyang University, South Korea, in 1983, and M.S. and Ph.D degrees in Electrical Engineering Department from the University of Washington, Seattle, USA, in 1986 and 2001, respectively. In 2001, he joined Samsung Electronics as a senior research staff. During 2001-2009, he was involved in the development of 3D image processing technologies. In 2009, he joined the Kangwon National University, Chuncheon, South Korea as a faculty member. He is now a professor and has been involved in various research projects covering stereoscopic image/video processing, 3D human factors, and stereoscopic conversion.

Yo-Sung Ho (M'81-SM'06) received both B.S. and M.S. degrees in electronic engineering from Seoul National University (SNU), Korea, in 1981 and 1983, respectively, and Ph.D. degree in Electrical and Computer Engineering from the University of California, Santa Barbara, in 1990. He joined the Electronics and Telecommunications Research Institute (ETRI), Korea, in 1983. From 1990 to 1993, he was with Philips Laboratories, Briarcliff Manor, New York, where he was involved in development of the advanced digital high-definition television (AD-HDTV) system. In 1993, he rejoined the technical staff of ETRI and was involved in development of the Korea direct broadcast satellite (DBS) digital television and high-definition television systems. Since 1995, he has been with the Gwangju Institute of Science and Technology (GIST), where he is currently a professor in the Information and Communications Department. His research interests include digital image and video coding, image analysis and image restoration, advanced coding techniques, digital video and audio broadcasting, 3D television, and realistic broadcasting.