Fast Edge-Preserving Depth Image Upsampler

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

**Abstract — ***In this paper, a new image upsampler is proposed to increase depth image resolution fast while preserving edge information. The proposed upsampler is based on a common edge region of color and depth images. In particular, if a vacant pixel in a higher resolution image grid belongs to the common edge region, it is assigned by a pixel selected from five candidates in a local window. A candidate is chosen by minimum cost evaluation; cost is computed by spatial, color, and range weighting functions. Otherwise, the vacant pixel is replaced with a pixel estimated using bilinear interpolation to speed up the process. In terms of a trade-off between depth image quality and computational complexity, experimental results show that the proposed upsampler outperforms other conventional methods, such as the bilinear interpolator and joint bilateral upsampler[1].*

Index Terms — Depth upsampling, common edge region, joint bilateral upsampling, bilinear interpolation, 3D video.

I. INTRODUCTION

A depth image is often available at a resolution lower than its corresponding color image in three-dimensional (3D) video applications. For instance, an advanced 3D TV system [1] reduces depth image resolution to make the best use of a transmission bandwidth, whereas color image resolution is maintained. In addition, depth images captured by active range cameras [2], [3] usually have 200×200 or 640×480 resolution due to many challenges in real-time distance measurement. In contrast, color images obtained from conventional video cameras have higher resolutions, such as 1024×768 or 1920×1080.

For practical purposes, depth image resolution should be the same as color image one [4]. Therefore, an efficient depth upsampler is necessary to convert depth image resolution from low to high. Conventional image upsamplers, such as the bilinear interpolator (BI) [5] and bilateral upsampler (BU) [6], can be directly used for depth image upsampling. However, these image upsamplers often make edges in upsampled depth images look like the shape of a staircase [7].

A joint bilateral upsampler (JBU) [8], [9] has been introduced to remove the staircase distortion. JBU refers to color data under the assumption that depth edges usually correspond to color edges. However, there are two main problems in JBU: *heavy computational complexity* and *visual artifacts*. Since JBU adopts color data additionally, it is much slower than the previous upsamplers. In our implementation, JBU takes about 25 seconds to convert depth image resolution from 480×270 to 1920×1080, whereas BI only needs about 0.02 seconds.

In addition, visual artifacts [10] are observed when the assumption of JBU fails; if depth edges correspond to homogeneous color areas, depth edges become blurred as the region marked by a rectangle in Fig. 1(c). Inversely, if homogenous depth areas are associated with color edges, depth data become distinguishable as the region marked by a circle in Fig. 1(c).

In this paper, a new depth image upsampler is proposed to resolve aforementioned problems of JBU. In order to reduce computational time while suppressing visual artifacts, the proposed upsampler deals with depth information in common edge regions separately; the common edge region is defined by the intersected edge areas of two dilated edge maps generated from a depth image and its color image.

In particular, if a vacant pixel in a higher resolution image grid is in the common edge region, it is replaced with a pixel selected from five candidates in a local window instead of considering all neighboring pixels; each candidate has its own cost computed by spatial, color, and range weighting functions. A candidate having minimum cost is selected to be assigned to the vacant pixel. If a vacant pixel is out of the common edge region, it is replaced with a pixel upsampled via BI. In this case, in order to avoid visual artifacts, color information is not considered.

The main contribution of our work is to provide a practical solution to upsample depth images using color information. As shown in Fig. 1(d), the proposed upsampler generates high-resolution depth images fast while preserving edge data.

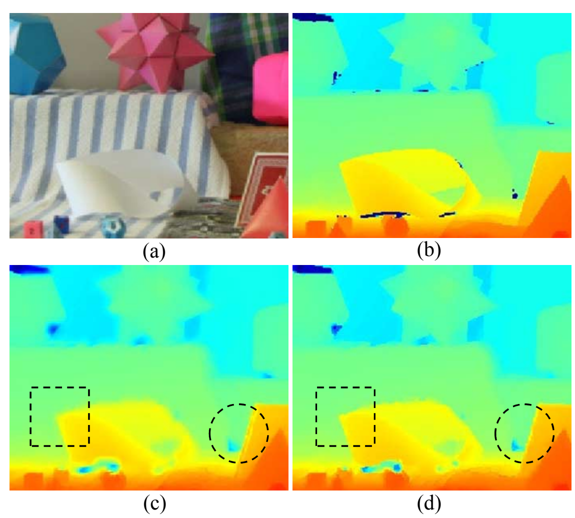


Fig. 1. Visual evaluation on *Moebius* data set [16]; (a) color image, (b) ground truth depth image, (c) result of JBU, and (d) result of the proposed upsampler. Input depth image resolution is 100×90. Output depth image resolution is 400×360. The runtime of JBU is 2.01 seconds, whereas the runtime of the proposed upsampler is 0.19 seconds.

[1] This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0030822).

S.-Y. Kim and Y.S. Ho are with the School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea (e-mail: {sykim75, hoyo}@gist.ac.kr).

Contributed Paper
Manuscript received 07/13/12
Current version published 09/25/12
Electronic version published 09/25/12.

II. RELATED WORK

Petschnigg *et al.* [11] have introduced the concept of a joint bilateral filter (JBF) to improve the quality of a non-flash image using its associated flash image. Then, Kopf *et al.* [8] have developed a joint bilateral upsampler (JBU) by extending the idea of JBF for an efficient image interpolation. JBU increases the resolution of a target image considering the photometric property of a high-resolution reference image. In depth image upsampling via JBU, the target image is a low-resolution depth image and the reference image is a high-resolution color image.

Suppose that there are a low-resolution depth image D^l and a high-resolution color image I^h. Let p and q denote coordinates of pixels in I^h, and $p_\downarrow$ and $q_\downarrow$ denote the associated coordinates in D^l. p is the center pixel in a local window $W\times W$. q is the neighboring pixel of p in the window where $q\in W\times W$. Formally, the new depth value D_p^h at p in an upsampled depth image D^h using JBU is computed by

$$D_p^h = \frac{\sum_{q\in W\times W} \kappa_{p,q}\ D_{q_\downarrow}^l}{\sum_{q\in W\times W} \kappa_{p,q}} \tag{1}$$

where $\kappa_{p,q}$ is a kernel weighting function [12].

$\kappa_{p,q}$ is defined by

$$\kappa_{p,q} = \varphi(\| p_\downarrow - q_\downarrow \|)\bullet\psi\ (\| I_p^h - I_q^h \|) \tag{2}$$

where φ and ψ are spatial and color weighting functions, respectively, and $\|\cdot\|$ is an Euclidean distance operator.

If an exponential function is used to model φ and ψ, those weighting functions are represented by

$$\varphi(n) = \exp(\frac{-n^2}{\sigma_\varphi}),\qquad \psi(n) = \exp(\frac{-n^2}{\sigma_\psi}) \tag{3}$$

where σ_φ and σ_ψ are the smoothing parameters of φ and ψ.

Recently, Riemens *et al.* [9] have presented a multi-step joint bilateral upsampler (M-JBU). M-JBU exploits intensity difference between a low-resolution color image and its high-resolution color image; image regions represented by great intensity difference are regarded as high frequency areas. The high frequency information is used to preserve depth edges. Suppose that there is a low-resolution color image I^l, which is created by downsampling I^h. Then, $\kappa_{p,q}$ in (2) for M-JBU is represented by

$$K_{p,q} = f(\| p_\downarrow - q_\downarrow \|)\ \ g(\| I_p^h - \breve{I}_q^h \|) \tag{4}$$

where $\breve{I}^h$ is the upsampled color image of I^l. The bilinear interpolator (BI) is employed to obtain $\breve{I}^h$ from I^l.

In M-JBU, the spatial function φ is represented by a box filter, which returns value 1 within a local window and value 0 outside it. The box filter reduces the effect of the spatial term of φ while increasing the effect of the color term of ψ.

Related to JBF, Yang *et al.* [13] have developed a fast post-processing based on JBF. Yang' work refines depth edges only considering color data associated with depth edge regions. However, in case that depth edges are corresponding to homogenous color areas, the previous work suffers from visual artifacts that depth data become blurred.

In addition, Lai *et al.* [14] and Cho *et al.* [15] have presented a joint multilateral filter (JMF) by adding a range term into the kernel weighting function $\kappa_{p,q}$. In JMF, $\kappa_{p,q}$ in (2) is represented by

$$\kappa_{p,q} = \varphi(\| p_\downarrow - q_\downarrow \|)\bullet\psi(\| I_p^h - I_q^h \|)\bullet\omega(\| D_p^h - D_q^h \|) \tag{5}$$

where ω is the range weighting function. Like (3), ω can be modeled by an exponential function as it follows:

$$\omega(n) = \exp(\frac{-n^2}{\sigma_\omega}) \tag{6}$$

where σ_ω is the smoothing parameters of ω.

III. PROPOSED DEPTH IMAGE UPSAMPLER

A. Depth Upsampler Structure

The proposed method is initially motivated by Yang's work [13] to reduce the computational time. However, there are differences between two methods in terms of the methodology. The main difference is the use of a region classification. The proposed upsampler is based on a common edge region to suppress visual artifacts fabricated from useless color data, whereas Yang's work uses all color data. In addition, Yang's work employs spatial and color weighting functions only to calculate the cost of each candidate, whereas the proposed method adds the range term for better cost evaluation. Finally, the aim of Yang's work is to refine depth data basically, whereas this work targets to upsample depth images.

Fig. 2 presents the overall flow of the proposed upsampler. First, a low-resolution depth image D^l is upsampled to the target resolution via BI. Then, a vacant pixel p in a bilinearly-interpolated image B^h is replaced with a pixel generated by two processes at p.

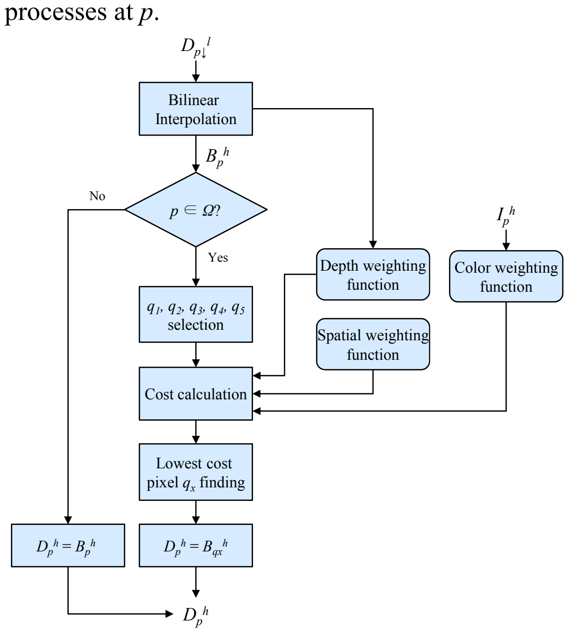


Fig. 2. Overall flow of the proposed upsampler.

First, if p belongs to common edge regions Ω, five candidates q_1, q_2, q_3, q_4, q_5 are selected and their costs are calculated based on spatial, color, and range weighting functions. When a candidate q_x has minimum cost among five candidates, the depth value D_p^h at p is assigned by the depth value B_{qx}^h at q_x. Second, if p is out of common edge regions Ω, D_p^h is assigned by B_p^h directly.

B. Common Edge Region

Common edge regions are defined by the intersected edge areas of two dilated edge maps generated from a depth image and its color image. Fig. 3 illustrates the extraction of the common edge region Ω.

Let E^D and E^I denote edge maps of D^l and I^l, respectively. In order to extract Ω, depth edge pixels in E^D and color edge pixels in E^I are initially set to zero. Non-edge pixels are set to a maximum value, e.g., 255 for an 8-bit grayscale image. Then, edge-expanded images T^D and T^I are created by applying a dilation operator [5] onto E^D and E^I.

Thereafter, a low-resolution region classification map J^l is generated by intersecting T^D and T^I as it follows:

$$J^l_{p_\downarrow} = T^D_{p_\downarrow} \cap T^I_{p_\downarrow}, where\ T^D_{p_\downarrow} == 0 \tag{7}$$

In order to refer J^l in depth image upsampling, it is needed to be spatially-interpolated to a higher resolution one J^h. For this, a near-pixel interpolator [5] is applied onto J^l. Formally, J^h is computed by

$$J^h_p = J^l_{[p_\downarrow \times r]} \tag{8}$$

where r is a scale factor and $[\cdot]$is a round down operator.

Finally, the common edge region Ω is defined as the zero pixels of J^h as it follows:

$$\Omega_p : p, \quad if\ J^h_p == 0 \tag{9}$$

In Fig. 3, the area colored in black on J^h is Ω, and the other area colored in white is disjoint edge regions.

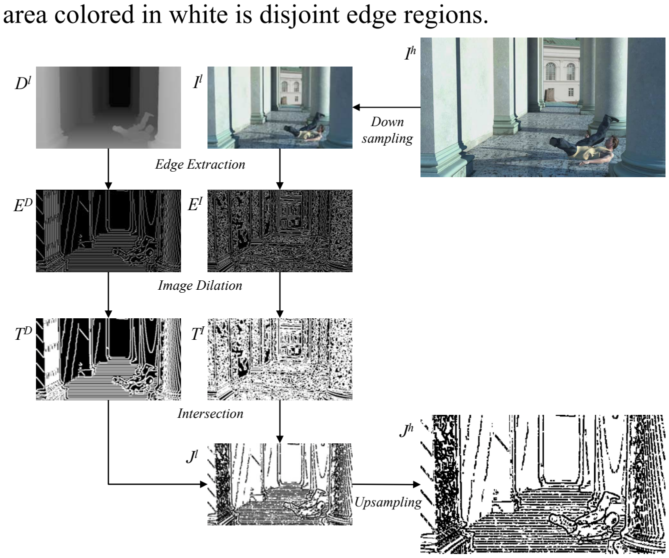


Fig. 3. Generation of a region classification map J^h to extract common edge region; Block colored areas on J^h is common edge regions Ω.

C. Depth Image Upsampling

After Ω is determined, D^l is upsampled to color image resolution using BI. Let B^h denote the bilinearly-interpolated depth image of D^l. If p is on Ω, then the depth value D_p^h at p is assigned by B_q^h at a neighboring pixel q in a local window. Otherwise, D_p^h at p is assigned by B_p^h at p.

In particular, if p belongs to Ω, five candidate pixels q_1, q_2, q_3, q_4, q_5 are selected in the local window; q_1, q_2, q_3, q_4, q_5 are the left, right, center, top, bottom pixels in the local window. When p is (x, y) coordinate, the candidates are defined by

$$\begin{aligned} &q_1 : (x-1, y) \\ &q_2 : (x+1, y) \\ &q_3 : (x, y) \\ &q_4 : (x, y-1) \\ &q_5 : (x, y+1) \end{aligned} \tag{10}$$

Each candidate has its own cost based on spatial, color, and range weighting functions φ, ψ, and ω. Formally, the cost $C_{p,q}$ at q with respect to p is calculated by

$$C_{p,q} = \varphi(\| p - q \|) \bullet \psi(\| I^h_p - I^h_q \|) \bullet \omega(\| B^h_p - B^h_q \|) \tag{11}$$

where φ, ψ, and ω are modeled by the exponential functions in (3) and (6), respectively, and $q \in W \times W$.

Then, we seek q_x that has the minimum cost among five candidate set Q = $\{q_1, q_2, q_3, q_4, q_5\}$. q_x is represented by

$$q_x : \underset{q_x \in Q}{\arg\min} \{C_{p,q_1}, C_{p,q_2}, C_{p,q_3}, C_{p,q_4}, C_{p,q_5}\}, \tag{12}$$

Finally, D_p^h is assigned by the depth value at q_x as it follows:

$$D^h_p = B^h_{q_x} \tag{13}$$

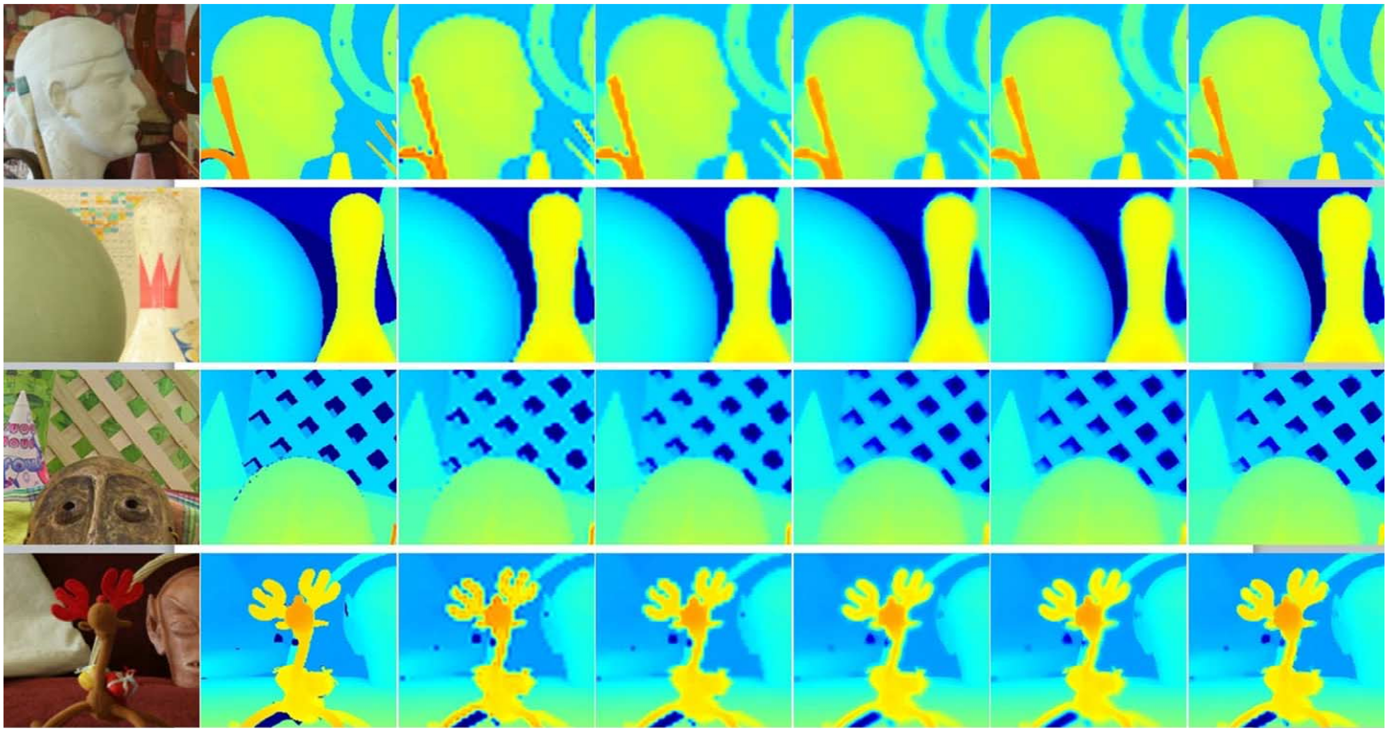
In case that p does not belong to Ω, D_p^h is assigned by B_p^h. In this situation, color information is not considered.

IV. Experimental Result

To evaluate the performance of the proposed upsampler, we tested with thirteen synthetic image sets having ground truth depth data [16]; these test data were *art*, *baby*, *barn*, *books, bowling*, *cone*, *dolls*, *flowerpots, laundry*, *moebius*, *reindeer*, *rocks*, and *sawtooth*.

Prior to the experiment, each ground truth depth image is downsampled by a factor of 4 and 16 to generate input low-resolution depth images; when the original depth image resolution is 440×360, resolutions of two input depth images become 220×180 and 110×90, respectively. For objective evaluation, after upsampling input depth images via BI [5], JBU [8], M-JBU [9], and the proposed upsampler, the quality of output depth images are measured by the peak signal-to-noise ratio (PSNR) based on ground truth depth data.

In the experiment, the size of a local window $W \times W$ for cost evaluation was set to 5×5. For JBU, σ_φ and σ_ψ in (3) were set to 2 and 0.1, respectively. For M-JBU, the box filter is used for φ, and σ_ψ was set to 0.1. For the proposed upsampler, σ_φ, σ_ψ, and σ_ω in (3) and (6) were set to 2, 0.1, and 0.1 to calculate the cost of each candidate.



(a) (b) (c) (d) (e) (f) (g)

Fig. 4. Results of *art*, *bowling*, *cone*, and *reindeer*. Downsampling factor is 16. (a) a part of the color image of each test image, (b) ground truth depth image of (a), (c) input low-resolution depth image, (d) results of BI, (e) results of JBU, (f) results of M-JBU, and (g) results of the proposed method.

TABLE I
PSNR COMPARISON (UNIT: *dB*, DOWNSAMPLING FACTOR: 4)

Test data	BI	JBU	M-JBU	Proposed
art	34.10	32.24	32.09	34.53
baby	38.76	36.15	36.10	38.12
barn	42.51	41.35	40.5	43.96
books	33.54	31.82	31.64	33.54
bowling	37.19	34.08	34.06	38.02
cone	32.75	29.82	29.76	32.95
dolls	35.54	33.14	33.08	35.07
flowerpots	30.94	29.01	28.96	31.56
laundry	36.61	34.43	34.39	37.02
moebius	34.27	31.91	31.81	34.07
reindeer	35.41	33.36	33.33	35.73
rocks	33.61	31.22	31.14	33.81
sawtooth	40.69	41.00	40.84	42.48
Avg. PSNR	35.84	33.81	33.67	36.22

TABLE II
PSNR COMPARISON (UNIT: *dB*, DOWNSAMPLING FACTOR: 16)

Test data	BI	JBU	M-JBU	Proposed
art	30.37	30.53	30.63	31.12
baby	35.54	35.06	35.16	35.47
barn	39.03	39.69	38.81	41.11
books	30.85	30.86	30.77	31.06
bowling	32.87	32.68	32.98	33.71
cone	29.33	28.76	28.79	29.44
dolls	32.76	32.50	32.57	32.48
flowerpots	27.24	27.56	27.84	28.21
laundry	33.25	33.21	33.34	33.69
moebius	31.05	30.88	30.92	31.07
reindeer	32.11	32.17	32.33	32.89
rocks	30.10	29.94	29.98	30.39
sawtooth	37.07	38.76	39.56	40.28
Avg. PSNR	32.43	32.51	32.59	33.15

For generating common edge regions Ω, we employed the *Canny* edge detector [17]; the low and high thresholds for edge detection were set to 50 and 150, respectively, and 7×7 window kernel was used.

Fig. 4 shows the results of *art*, *bowling*, *cone*, and *reindeer* for the case of the downsampling factor 16. Fig. 4(a), Fig. 4(b), and Fig. 4(c) exhibit a part of the color image of each test dataset, ground truth depth data of the part, and the input low-resolution depth image. From the result of BI in Fig. 4(d), a staircase distortion is observed; depth data on object boundaries looks like the shape of a staircase. In contrast, the results of JBU, M-JBU, and the proposed method minimize the staircase distortion.

From the results of JBU and M-JBU in Fig. 4(e) and Fig. 4(f), depth data blurring is noticeable on object boundaries. For instance, in *bowling*, the bowling pin has similar color information with the background, whereas their depth data are quite different each other. During JBU and M-JBU, similar color data affects the depth data to be blurred on the bowling pin boundary. In contrast, since the proposed upsampler excludes the situation by referring common edge regions Ω, it minimizes depth data blurring on object boundaries.

Table 1 and Table 2 show the result of PSNR comparison. When the downsampling factor is 4, the average PSNRs of output depth images are about 35.84 *dB*, 33.81 *dB*, 33.67 *dB*, and 36.22 *dB* for BI, JBU, M-JBU, and the proposed method, respectively. This outcome indicates that the proposed upsampler has higher PSNRs as much as about 0.38 *dB*, 2.41 *dB*, and 2.55 *dB* than BI, JBU, and M-JBU on average.

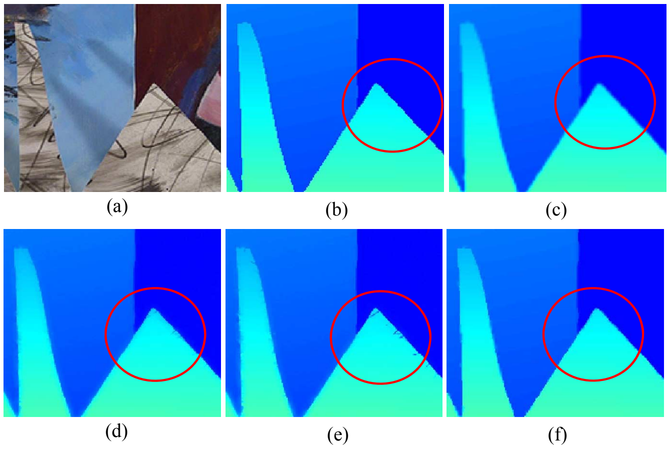


Fig. 5. Visual artifacts of JBU and M-JBU; (a) color image, (b) ground truth depth image, (c) result of BI, (d) result of JBU, (e) result of M-JBU, and (f) result of the proposed method.

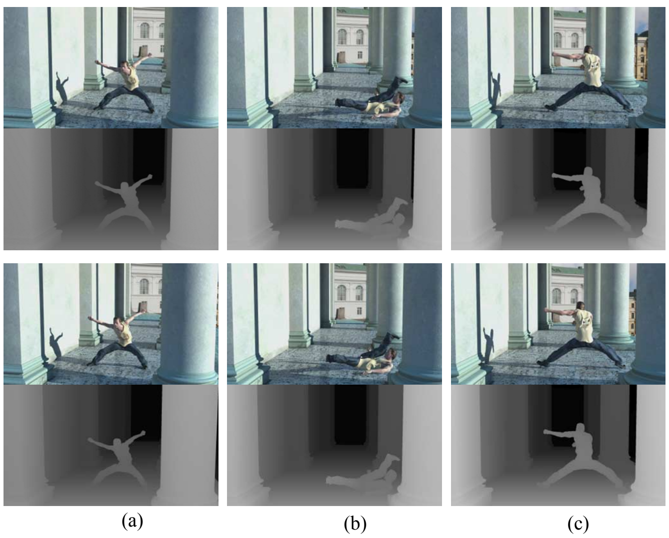


Fig. 6. *Undo_Dancer* test sequence, (Row 1) and (Row 2) are the color and depth images of the frame for the 1st view. (Row 3) and (Row 4) are the color and depth image of the frame for the 9th view; (a) the 100th frame (b) the 150th frame, and (c) the 250th frame.

As shown in Table 2, when the downsampling factor is 16, the averages of PSNRs are 32.43 *dB*, 32.51 *dB*, 32.59 *dB*, and 33.15 *dB* for BI, JBU, M-JBU, and the proposed method. The PSNR gains of the proposed method are approximately 0.72 *dB*, 0.64 *dB*, and 0.56 *dB* higher than BI, JBU and M-JBU. As a result, the proposed upsampler has the best performance among the comparative methods in terms of PSNR evaluation.

Note that JBU and M-JBU have even lower PSNRs than BI in Table 1 because of the presence of visual artifacts. Fig. 5 displays visual artifacts created by useless color data in *sawtooth*. In Fig. 5(d) and Fig. 5(e), depth data around the sawtooth edges marked by circles is fabricated by the texture data in the color image. In contrast, the proposed upsampler suppresses the visual distortion, as shown in Fig. 5(f).

For the sake of computational time comparison of BI, JBU, M-JBU, and the proposed method, the average processing time for those test data is calculated. The test was done with a personal computer equipped with CPU 2.67 GHz and Ram 32 GB. Note that source codes for implementing such methods are not optimized and well-organized.

Table 3 shows the average computational time. Runtimes are approximately 0.01 *s.*, 1.95 *s.*, 1.96 *s.*, and 0.21 *s.* for BI, JBU, M-JBU, and the proposed method, respectively. BI is much faster than the other methods, since BI considers only depth data. The proposed upsampler is the second fastest. Especially, the proposed method is faster than JBU and M-JBU as much as about 9 times. Consequently, in terms of a trade-off between upsampled depth data quality and computational complexity, the proposed upsampler has better performance than the other methods.

TABLE III
RUNTIME COMPARISON

Factor	BI	JBU	M-JBU	Proposed
4	**0.01 *s.***	**1.94 *s.***	**1.95 *s.***	**0.19 *s.***
16	**0.01 *s.***	**1.96 *s.***	**1.97 *s.***	**0.22 *s.***
Average	**0.01 *s.***	**1.95 s.**	**1.96 s.**	**0.21 s.**

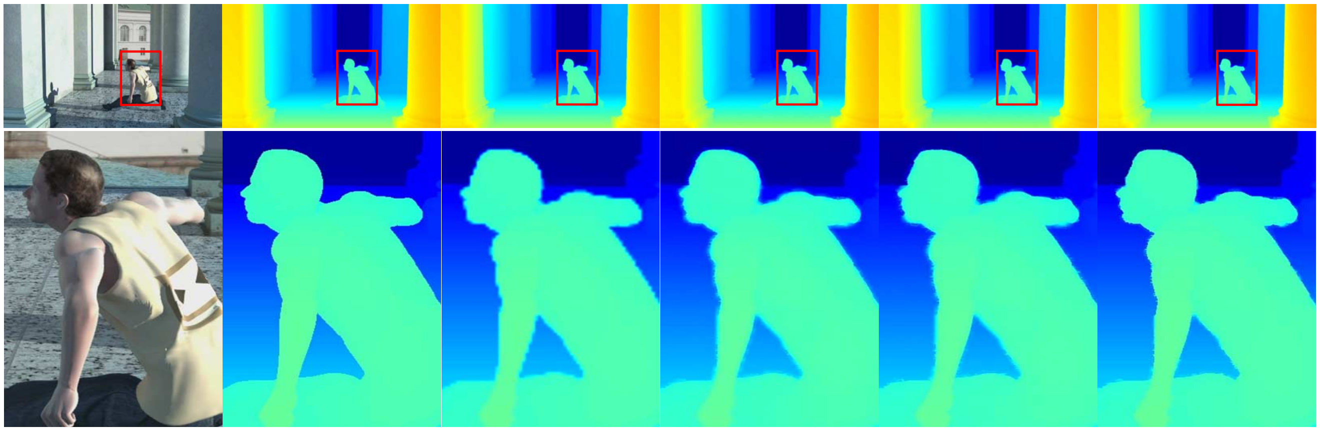
Another experiment has been performed on a multiview video-plus-depth *Undo_Dancer* [18], which is a computer-generated imagery with ground truth depth data. *Undo_Dancer* is composed of 9-view video-plus-depths; a video-plus-depth is a sequence of color and depth image pair. Each-view video-plus-depth consists of 250 frames with 1920×1088 resolution. Fig. 6 displays the 100th, 200th and 250th frames of the 1st and 9th view of *Undo_Dancer*.

Among nine views, depth images of the 1st and 9th view are first downsampled by a factor of 16; the resolution of input depth images is 480×272. Then, input depth images are upsampled to the original resolution 1920×1088 by BI, JBU, M-JBU, and the proposed upsampler. Finally, the output depth images and their corresponding color images of the 1st and 9th view are used to generate the 5th view color images using depth image-based rendering [19], [20].

For objective evaluation, we have measured PSNRs of upsampled depth images based on the ground truth depth data at the 1st and 9th view. In addition, PSNRs of virtually-synthesized color images at the 5th view were calculated based on the original color images at the same view.

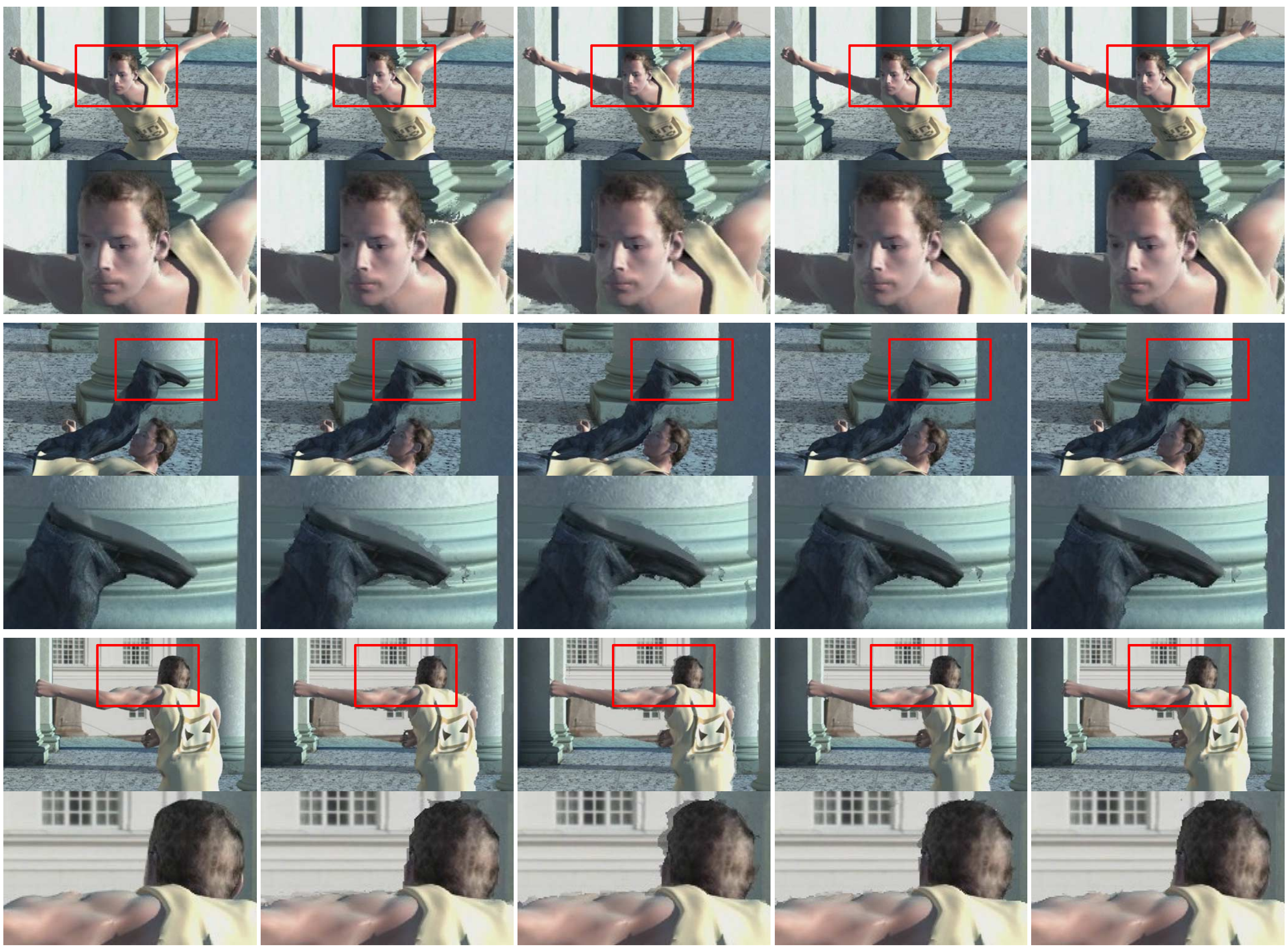
Fig. 7 demonstrates the result of depth image upsampling for the 180th frame of *Undo_Dancer*. When the result of BI in Fig. 7(c) is compared to the ground truth depth image in Fig. 7(b), the staircase distortion is observed. In addition, from the results of JBU and M-JBU in Fig. 7(d) and Fig. 7(e), depth edges become discrete like thorns on a rose stem. In contrast, as shown in Fig. 7(f), the proposed method restores depth edges and reduces visual artifacts.

Fig. 8 displays view synthesized results of the 100th, 200th and 250th frame at the 5th view generated by color and upsampled depth images at the 1st and 9th view. As shown in the 2nd, 4th, and 6th rows, which are magnified by the region marked by rectangles in the 1st, 3rd, and 5th rows, our method generates higher quality virtual views than the other methods.



(a) (b) (c) (d) (e) (f)

Fig. 7. Results of *Undo_Dancer* for the 180th frame of the 1st view; (a) color image, (b) ground truth depth image, (c) results of BI, (d) results of JBU, (e) results of M-JBU, and (f) results of the proposed method.



(a) (b) (c) (d) (e)

Fig. 8. Results of synthesized view generation at the 5th view. (Row 1) and (Row 2) are synthesized images of 100th frame and their magnified image at the 5th view, respectively. (Row 3) and (Row 4) are synthesized images of 200th frame. (Row 5) and (Row 6) are synthesized images of 250th frame; (a) the original color image at the 5th view, (b) results of BI, (c) results of JBU, (d) results of M-JBU, and (e) results of the proposed method.

Table 4 shows performance evaluation for *Undo_Dancer* by the average PSNR of upsampled depth images at the 1st and 9th view, the average PSNR of synthesized color images at the 5th view, and the average computational time. First, the average PSNRs for BI, JBU, M-JBU, and the proposed upsampler are 41.3 *dB*, 41.8 *dB*, 42.3 *dB*, and 43.0 *dB*. Consequently, the average PSNR gains for the proposed method are approximately 1.7 *dB*, 1.2 *dB*, and 0.7 *dB* more than BI, JBU, and M-JBU.

Second, in quality comparison of virtually-generated color images at the 5th view, the average PSNR for the proposed upsampler are higher by 3.1 *dB*, 2.4 *dB*, and 0.9 *dB* more than BI, JBU, and M-JBU. As a result, our method outperforms the other methods in terms of virtual view synthesis.

Finally, in computational time comparison, average runtimes of BI, JBU, M-JBU, and the proposed upsampler for 250 frames are 0.02 *s*., 24.5 *s*., 24.6 *s*., and 0.93 *s*., respectively. Color data-based methods, such as JBU, M-JBU, and the proposed method, are much slower than BI. However, the proposed upsampler reduces the gap between color data-based methods and BI while improving the quality of upsampled depth images.

TABLE IV
PERFORMANCE EVALUATION OF *Undo_Dancer* (250 FRAMES)

Evaluation		BI	JBU	M-JBU	Proposed
Average PSNR	*1st View Output Depth Image*	41.5 *dB*	42.0 *dB*	42.4 *dB*	43.2 *dB*
	9th View Output Depth Images	41.1 *dB*	41.7 *dB*	42.2 *dB*	42.8 *dB*
	5th View Virtual Color Images	31.8 *dB*	32.5 *dB*	34.0 *dB*	34.9 *dB*
Average Runtime		0.02 *s*.	24.5 *s*.	24.6 *s*.	0.93 *s*.

V. CONCLUSION

In this paper, a new method has been proposed to upsample depth images with the aid of color information. The proposed upsampler was based on common edge regions of color and depth images. Based on thirteen test images having ground truth depth data, the average PSNRs of the proposed method were approximately 0.55 *dB*, 1.52 *dB*, and 1.55 *dB* higher than the bilinear interpolator (BI), joint bilateral upsampler (JBU), and multi-step joint bilateral upsampler (M-JBU). In addition, based on a multiview video-plus-depth, the average PSNR gains of the proposed method were about 1.2 *dB*, 1.7 *dB*, and 0.7 *dB* more than BI, JBU, and M-JBU. Furthermore, in terms of virtual view synthesis, the proposed upsampler is more effective than BI, JBU, and M-JBU.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11, "Call for proposals on 3D video coding technology," *MPEG document* N12036, 2011.

[2] C. Lee, H. Song, B. Choi, and Y. S. Ho, "3D scene capturing using joint stereo camera with a time-of-flight camera for 3D displays," *IEEE Trans. Consum. Electron.*, 57(3):1370-1379, 2011.

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

[4] C. Fehn, R. Barré, and S. Pastoor, "Interactive 3-D TV- concepts and key technologies," *Proceedings of the IEEE*, 94(3):524-538, 2006.

[5] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Prentice Hall, 2002.

[6] K.W. Hung and W.C. Siu, "Improved image interpolation using bilateral filter for weighted least square estimation," in *Proc. of IEEE International Conference on Image Processing*, pp. 3297-3300, 2010.

[7] S. Schuon, C. Theobalt, J. Davis, S. Thrun, "High-quality scanning using time-of-flight depth superresolution," in *Proc. of IEEE Computer Vision and Pattern Recognition Workshops*, pp. 1-7, 2008.

[8] J. Kopf, M.F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, 26(3):1-6, 2007.

[9] A.K. Riemens, O.P. Gangwal, B. Barenbrug, and R.M. Berretty, "Multi-step joint bilateral depth upsampling," in *Proc. of SPIE*, vol. 7257, pp. 72570M, 2009.

[10] O.P. Gangwal and B. Djapic, "Real-time implementation of depth map post-processing for 3D-TV in dedicated hardware," in *Proc. of IEEE International Conference on Consumer Electronics*, pp. 173-174, 2010.

[11] 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.*, 23(3): 664-672, 2004.

[12] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.*, 16(2): 349-366, 2007.

[13] Q. Yang, L. Wang, and N. Ahuja, "A constant-space belief propagation algorithm for stereo matching," *Proc. of IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1458-1465, 2010.

[14] P. Lai, D. Tian, and P. Lopez, "Depth map processing with iterative joint multilateral filtering," in *Proc. of Picture Coding Symposium*, pp. 9-12, 2010.

[15] 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.*, 54(4): 1514-1521, 2008.

[16] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Jour. of Computer Vision*, 47(1-3): 7-42, 2002.

[17] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6): 679-698, 1986.

[18] ISO/IEC JTC1/SC29/WG11, "Extension of existing 3DV test set toward synthetic 3D video content," *MPEG document* N19221, 2011.

[19] H. Y. Shum, S. B. Kang, and S. C. Chan, "Survey of image-based representations and compression techniques," *IEEE Trans. Circuits Syst. Video Technol.*, 13(11): 1020-1037, 2003.

[20] L. Zhang and W. Tam, "Stereoscopic image generation based on depth images for 3D TV," *IEEE Trans. Broadcast.*, 51(2): 191-199, 2005.

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. His research interests include digital image processing, depth image-based modeling and rendering, computer graphic data processing, 3DTV, and realistic broadcasting.

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.