Depth Up-Sampling via Pixel-Classifying and Joint Bilateral Filtering

Yannan Ren[1], Ju Liu[1,2,*], Hui Yuan[1,2], Yifan Xiao[1]
[1] School of Information Science and Engineering, Shandong University
Jinan 250100 - China
[e-mail: juliu@sdu.edu.cn]
[2] Suzhou Research Institute of Shandong University
Suzhou 215021 - China
*Corresponding author: Ju Liu

Received November 21, 2016; revised May 16, 2017; revised September 21, 2017; accepted February 5, 2018; published July 31, 2018

Abstract

In this paper, a depth image up-sampling method is put forward by using pixel classifying and jointed bilateral filtering. By analyzing the edge maps originated from the high-resolution color image and low-resolution depth map respectively, pixels in up-sampled depth maps can be classified into four categories: edge points, edge-neighbor points, texture points and smooth points. First, joint bilateral up-sampling (JBU) method is used to generate an initial up-sampling depth image. Then, for each pixel category, different refinement methods are employed to modify the initial up-sampling depth image. Experimental results show that the proposed algorithm can reduce the blurring artifact with lower bad pixel rate (BPR).

Keywords: Depth upsampling; pixel classifying; joint bilateral up-sampling; edge map; texture; smooth;

This work was supported in part by the National Natural Science Foundation of China (61571274, 61601268, 61771292), the Shandong Natural Science Fund for Distinguished Young Scholar (JQ201614), the Young Scholars Program of Shandong University (2015WLJH39), the Natural Science Foundation of Shandong Province (ZR2014FM012), the Shandong Provincial Key Research and Development Plan (2017CXGC1504,2015ZDXX0801A01).

http://doi.org/10.3837/tiis.2018.07.013 **ISSN : 1976-7277**

1. Introduction

Depth images can be cooperated with color images to obtain virtual image via depth image based rendering (DIBR) [1] technology. However, compared to the high resolution color image, many depth images usually retain in low resolution due to depth capturing techniques or limited transmission bandwidth. Therefore, depth up-sampling research has attracted more and more attention recently.

The up-sampled depth map can be obtained by conventional image super resolution, which directly convolves a low-resolution (LR) image with an interpolation kernel to obtain a target high-resolution (HR) image. However, there is an additional HR color input image in the 3DTV system, and based on the relationship between them there may be a better strategy for depth up-sampling research. In this system, there exists an assumption that depth edges or depth discontinuities in the HR depth map are consistent with color edges in HR color images, and an up-sampled depth map can be obtained by inputting an LR depth map and HR color image using joint image filtering techniques or image modeling algorithms.

In general, there are two main algorithms in contemporary depth up-sampling research: global up-sampling such as markov random field (MRF)-based image modeling [9–16] and local processing methods like filter-based depth map up-sampling [2–8,17,18].

Initially, [2] applied the high-resolution color image as a guide to improve the quality of up-sampled depth images by using joint bilateral filtering (JBF). Following that, some local processing approaches like the filter-based depth up-sampling method have been proposed [3–8]. In [3], a multistep joint bilateral depth up-sampler is presented that applies anti-alias pre-filtering. In [4], the authors applied geodesic curves to obtain the up-sampled depth map. In [5], a graph-based joint bilateral filter is proposed to generate an up-sampled depth map that enhances edges and reduces noise. In [6], an adaptive color-guided autoregressive model is proposed, and then the depth up-sampling task is formulated as a prediction error minimization problem. A weighted mode filtering method [7] is proposed to up-sample depth images based on a joint histogram using color similarity between reference pixels and neighboring pixels on the color image and the depth image. Furthermore, Jung et al. [8] applies the structure component of a high-resolution color image to reconstruct an HR depth image. Generally, the filter-based methods cost less computation and produce real-time results. However, these filter-based approaches may cause some problems such as blurring depth discontinuities and appearing blocking effect in depth homogeneous regions as shown in **Fig. 1**.

In addition, the global-based depth image up-sampling method takes the up-sampling problem as an optimization problem. *Diebel et al.* [9] proposed an MRF method that integrates both LR depth images and HR camera images. In this method, depth up-sampling problem is considered as a conjugate gradient optimization problem. *Hannemann et al.* [10] introduced amplitude values into an MRF model, the amplitude values can be obtained by time-of-Flight (TOF), and then used as a confidence measurement for depth values. In [11], a third layer is added to the MRF by encoding the image gradients as nodes in the graph. For dynamic scenes, *Zhu et al.* [12] employed a dynamic MRF to enhance the accuracy and robustness of depth recovery by propagating both the spatial and the temporal relationship in local neighbors. *Liu et al.* [13] presents a novel depth up-sampled work by defining an exponential function as the error norm in the smoothness term of MRF-based optimization framework. *Zuo et al.* [14] proposed an edge-based depth map super-resolution method via solving a labeling optimization problem in MRF to avoid the texture-copy artifacts and preserve the edges of

depth. In [15], a high-resolution edge map is constructed from those edges of the LR depth image through MRF optimization, and the up-sampled depth image is obtained through a modified joint bilateral filter. In [16], the HR depth map is generated by employing the dictionary of prior edge learned from an external database of high- and low-resolution examples. In general, these algorithms can produce better up-sampled depth map, however, they usually result in an over-smooth problem during the global optimization process. Moreover, it is also time consuming.

Recently, there is a new tendency for depth up-sampling research methods to focus on the local features of images and especially stress the importance of edge pixels. In [17], a common edge region is firstly defined based on color and depth images, and the pixels in the region are reassigned based on spatial, color and range united weighting functions. To refine the results of [2], an edge map is generated according to the variance of the depth map in [18], and then a local minimum filter is employed to refine the pixels in the edge map. *Kang et al.* [19] proposed a MRF-based depth up-sampler in which the energy function is incorporated with depth local features extracted from LR depth map and a color image. These new proposed depth up-sampled methods effectively improved the up-sampled depth map by mainly modifying the depth values of pixels in the edge regions, but they neglect other parts of the up-sampled depth image.

In order to get an up-sampled depth map with little blurring artifact, we present an effective depth up-sampling method via pixel classifying and jointed bilateral filtering (PCJBF). The contributions of this paper are provided as follows.

1). The detailed pixel-classifying scheme is presented by analyzing the edge points in both high-resolution color image and low-resolution depth image, in which the edge pixels close to the truth are extracted; meanwhile, other pixels of high-resolution image are further divided into three categories.
2). An effective optimization scheme is proposed for the edge regions. Experimental results show that it can reduce the blurring artifact and generate more smooth edge parts in up-sampled depth image than mostly JBF-based depth up-sampling algorithms.
3). Because the vibration of color values in the texture region will bring errors, the pixels in the defined texture set are refined by using an optimization scheme.

The rest of the paper is organized as follows. In section 2, the blurring artifact problem is described during JBF-based depth up-sampling algorithms. The proposed up-sampler is presented in section 3, Experimental results and conclusions are given in Section 4 and 5 respectively.

2. Blurring analysis of JBF-based depth up-sampling algorithms

Edge region plays a significant role in depth map. Consequently, depth up sampling should enhance the resolution of a depth map while not changing the original edge regions. The joint bilateral up-sampling (JBU) algorithm introduces an associated HR image into the LR resolution, which can enlarge the depth map and preserve the edges at the same time.

The LR depth image, the up-sampled depth image based on JBF and the corresponding HR color image are denoted as D_L, D_{JBU} and I respectively, and the up-sampling depth image can be described as followed:

$$D_{JBU}(p)=\frac{1}{k_p}\sum_{q_L\in\Omega}\exp\left(\frac{-\|p_L-q_L\|}{2\sigma_f^2}\right)\cdot\exp\left(\frac{-\|I(p)-I(q)\|}{2\sigma_g^2}\right)\cdot D_L(q_L) \tag{1}$$

$$k_p = \sum_{q_L \in \Omega} \exp\left(\frac{-\|p_L - q_L\|}{2\sigma_f^2}\right) \cdot \exp\left(\frac{-\|I(p) - I(q)\|}{2\sigma_g^2}\right)$$

where p and q are pixel point coordinates in up-sampled depth image D_{JBU} and HR color image I, p_L and q_L are the corresponding coordinates in the LR depth map D_L, $I(p)$ and $I(q)$ are the pixel values in the HR color image, Ω is the neighbor of p_L, represents the filter kernel and the σ_f and σ_g are the filter parameters.

In (1), if p and p_L are denoted as (x, y) and (x_L, y_L), respectively, and s is the up-sampled factor, then the pixel coordinate relation between the LR pixel and the LR depth map can be described as followed：

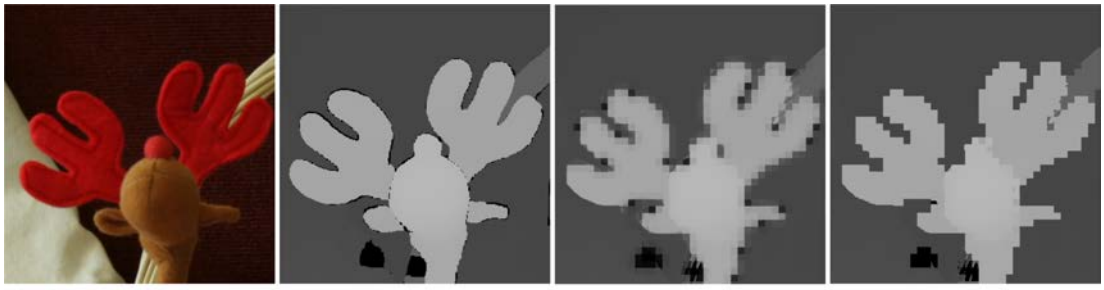
$$\begin{cases} x = s \times x_L + u, & u \in [0, s-1] \\ y = s \times y_L + v, & v \in [0, s-1] \end{cases} \tag{2}$$

Equation (2) indicates that a $s \times s$ block in the high-resolution image is projected to a same point in the LR depth map, and each pixel in this block has the same neighbor window in D_L. When there are edge points in the neighbor window, the blurring of sharp edge will be generated in the up-sampled depth map, as shown in **Fig. 1(c)**.

To solve the problem, *S. B. Lee et al.* [18] defined the edge map of the HR image according to the variance of LR depth map, and a local minimum filter to reduce the blurring of edge regions, which is described in the followed formula (3), refined the depth values in the edge map.

$$D_{DADU}(p) = \begin{cases} \arg\min_{d \in \Omega} |d - D_{JBU}(p)| & if\ \sigma_{D_L}^2(p_L) \geq th_D \\ D_{JBU}(p) & otherwise \end{cases} \tag{3}$$

where $\sigma_{D_L}^2$ represents the normalized variance map using the filter kernel Ω, and th_D is the threshold. Four candidates are directly chosen from the low resolution depth map, so new depth values are not created when the filter kernel straddles the depth discontinuity. This algorithm not only improves the blurring problem, but effectively reduces the bad pixel rates (BPR).



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

Fig. 1. Visual evaluation on part of reindeer [20]. Up-sampling factor is 8. (a) part of input image;(b) part of high resolution depth image;(c) part of up-sampled depth image based on JBU; (d) part of up-sampled depth image based on DADU.

However, the pixel coordinate relation between the HR map and the LR depth map still adheres to (2). In other words, a block of pixel points in the HR map is projected to the same pixel point in the LR map. Their final depth values may be refined and set as the same candidate, which bring staircase problem during rectifying blurring problem, as shown in **Fig. 1 (d)**.

In order to reduce the blurring artifact and staircase problems in the above two algorithms, we bring another edge map based on a high-resolution color image. Based on above two HR edge map, a detailed pixel category scheme is introduced. In the scheme, the new HR depth edge regions close to the truth are obtained which can effectively decrease the blocking effect. In the meantime, other pixels of HR depth image are classified into three different categories for subsequent up-sampled processing.

3. Depth up-sampling via pixel-classifying and joint bilateral filtering

The system diagram of the proposed method is shown in Fig. 2. In the system, from left to right, firstly, one HR edge map is generated through the edge map from the color image method by the input HR color image, meanwhile, the other HR edge map is obtained through the edge map from the depth image method by the input LR depth map. Secondly, the two HR edge maps are processed in the pixel classifier block, and then the pixels in the HR space are classified into four categories (i.e., edge points, edge-neighbor points, texture points and smoother points). At the same time, an initial up-sampling depth image is produced in the joint bilateral up-sampler block. Finally, the up-sampled depth map is obtained by using the depth up-sampler-based pixel category.

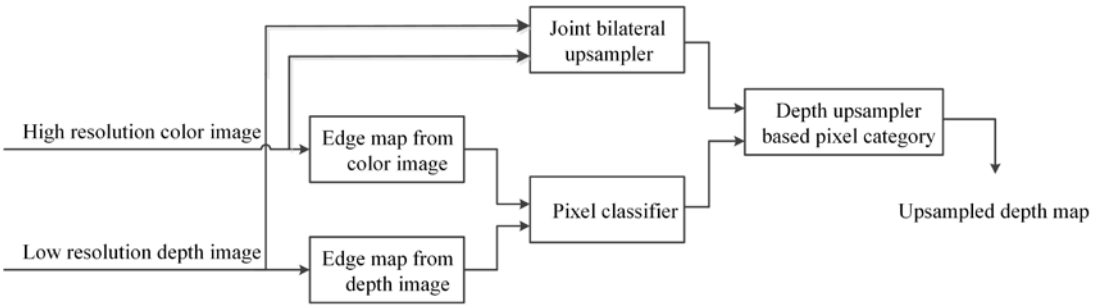


Fig. 2. Overall flow of PCJBF

3.1 The pixel-classifying scheme

Generally, the edge points in the depth map mean that huge depth abruption exists here, and the neighbor pixels of the edge point stand on the different depth levels. Considering the real application of the depth map, the edge pixels play a more important role than other pixels in the depth map. It is meaningful to obtain HR edge map close to the truth during the up-sampled depth map generation.

In [18], an effective depth edge map is obtained based on LR depth map as shown in (4).

$$E_d(p)=\begin{cases}1 & if\ \sigma^2_{D_L}(p_L)\geq th_d\\ 0 & otherwise,\end{cases} \tag{4}$$

where E_d is one HR edge map derived from LR depth map, and $\sigma_{D_L}(p_L)$ is the normalized variance of pixel point p_L in the LR depth map, th_d is a threshold. Obviously, the projection between the two kinds of coordinates will enlarge the true depth edge regions and bring some

error pixels in the true depth edge regions, leading to bad performance in further processing.

In addition, since the texture gradient in the color image is an effective depth cue, another HR edge map can be obtained from high resolution image as in (5) by using the maximum pixel color difference in neighborhood,

$$E_{\mathrm{c}}(p)=\begin{cases}1 & if \quad \left|\max\limits_{q\in N(p)} C(q)-\min\limits_{q\in N(p)} C(q)\right|\geq th_c \\ 0 & otherwise\end{cases} \tag{5}$$

where E_c is the other HR edge map derived from HR gray image, $C(q)$ is the gray value of the pixel q in the HR color image, *N(p)* is a window based on the pixel point *p* and th_c is a threshold. Though the texture region has no direct relevance with the true HR depth edge region, it can be auxiliary helpful to generate the HR depth edge map. More or less, to further decrease the error and take the other pixels into account, a detailed pixel category principle is proposed as follow.

Considering the characters of the pixels in the color image and the depth map, we should take a classification of the pixels in the HR depth map to be generated. All the pixels in the up-sampled depth map to be generated are denoted as set S. For a pixel *p* in S, when $E_d(p)=1$ or $E_c(p)=1$, it can be considered in the HR edge map E_d or E_{c}. When some pixels in the HR depth image belong to above two edge maps set simultaneously, then they can be deemed as effective edge pixels close to the truth (denoted as set P_1); when they are only in an HR edge map E_d, then they can be regarded as the edge-neighbor pixels of edge points in the HR depth image (denoted as P_2); when other pixels are parts of E_{c}, they are texture regions in the HR color image (denoted as P_3) ; finally, the rest pixels are included in none of above two HR edge maps, then they can be looked upon as smooth regions (denoted as P_4) . The above classification process can be obtained by different combination of two edge maps E_d and E_{c} defined by (6). The whole pixel classifying process can be described in **Fig. 3**.

$$\begin{cases}P_1=\left\{p\middle|E_d(p)=1,\ E_c(p)=1,\ p\in S\right\} \\ P_2=\left\{p\middle|E_d(p)=1,\ E_c(p)=0,\ p\in S\right\} \\ P_3=\left\{p\middle|E_d(p)=0,\ E_c(p)=1,\ p\in S\right\} \\ P_4=\left\{p\middle|E_d(p)=0,\ E_c(p)=0,\ p\in S\right\}\end{cases} \tag{6}$$

3.2 The optimization scheme

After taking a classification of the pixels in depth map to be generated, different refining methods can be used in different regions whose initial depth are obtained by JBU. Taking the pixels characteristic into account, appropriate optimized schemes are proposed for coressponding pixels sets to obtain their up-sampled depth values, the whole optimization scheme is described as follows,

$$D_U(p)=\begin{cases}\arg\max\limits_{q_L\in N(p_L)}\left(\exp\left(\dfrac{-\|p_L-q_L\|}{2\sigma_s}\right)\cdot\exp\left(\dfrac{-\|I(p)-I(q)\|}{2\sigma_r}\right)\cdot D_L(q_L)\right) & \text{case I}\\ \arg\min\limits_{q_L\in\Omega}\left|D_L(q_L)-D_{JBU}(p)\right| & \text{case II}\\ \arg\min\limits_{q_L\in N(p_L)}\left|D_L(q_L)-D_{JBU}(p)\right| & \text{case III}\end{cases} \tag{7}$$

where D_U is the HR up-sampled depth map, D_L is input LR depth map, and D_{JBU} is up-sampled depth map by JBU algorithm. The definition of Ω is the same as formula (3), and $N(p)$ is the window of the pixel point p_L in the LR depth map. The case I, case II and case III are three different optimization functions.

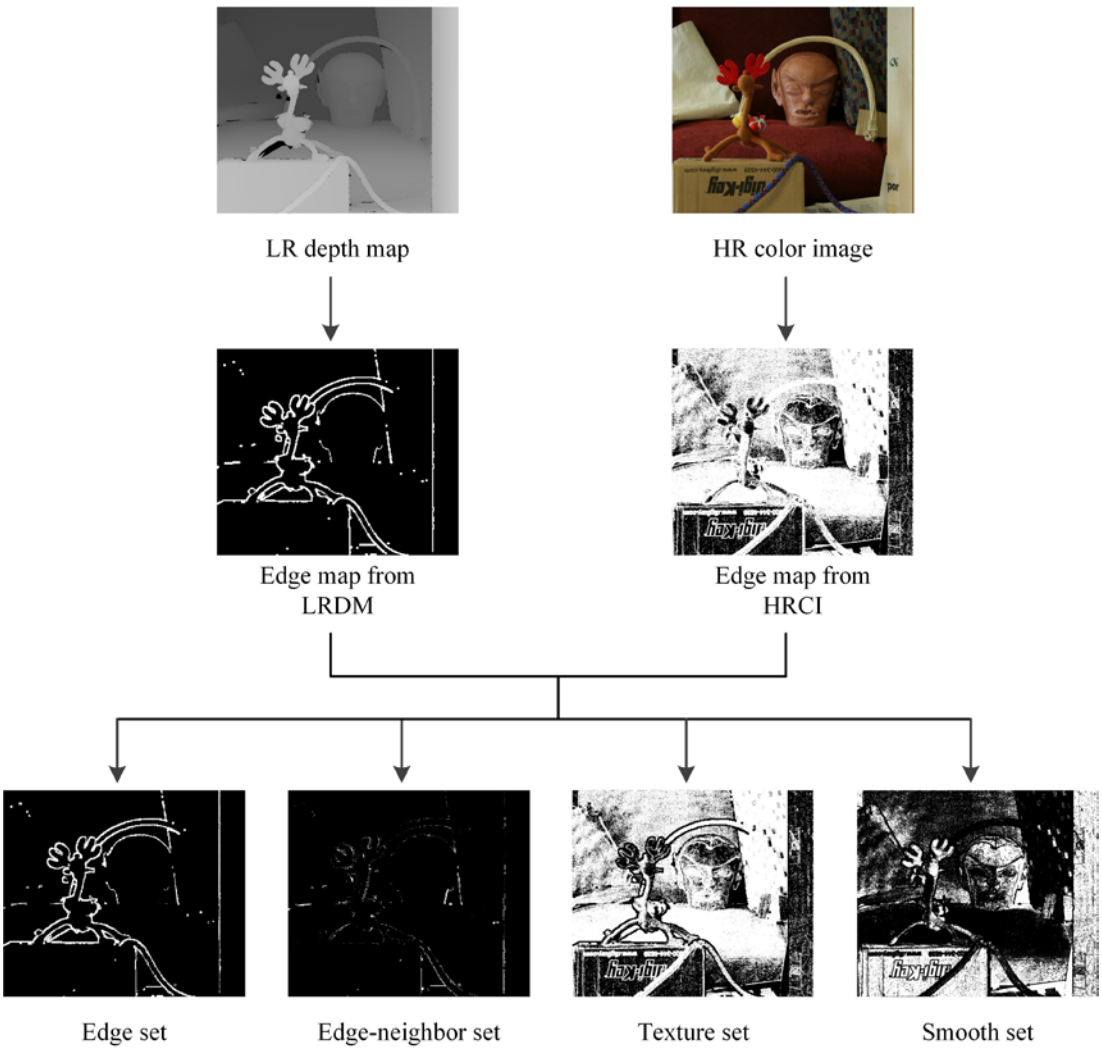


Fig. 3. Overview of the pixel classifying process

From the above analysis, the pixel in edge set P_1, where depth dicontiuety is consistent with HR color edge, is more possible to be the real edge point where depth abruption happened. Its initial up-sampled depth value from JBU is the weighted average value from the neighbor pixels of p_L. Generally, there exists a real edge point in the neighbor range, so the weighted average causes the blurring artifact. The optimization function is described as the case I in the formula (7), and the neighbor $N(p_L)$ of p_L is set as the candidate range in the LR coordinate system. Then, by synthesizing the space distance, color similarity and the depth value of each candidate pixel, the depth value of pixel p is set based on the candidate pixel with the max computing result as the final up-sampled depth value, which effectively improves the edge regions.

For edge-neighbor set P_2, where depth dicontiuety is inconsistent with the HR color edge, it is not easy to judge whether the pixel is a real edge point or a neighbor point of an edge pixel. It may be a better decision to keep the smoothness of those regions in the up-sampled depth map, therefore, we employ the minimum filter [18] to process the pixels for its good performance in these regions, just like the case II in (7) where a smaller value is chosen between the initial JBU and the real LR depth value as the final result of set P_2.

For the texture points set P_3, where depth dicontiuety is still inconsistent with HR color edge,their low depth normalized vairances mean that the up-sampled depth values are almost equal. However, the vibration of color value will cause error calculation, especially in the JBU-based method. Generally, their up-sampled depth value should be the same as one neighbor of its projective pixel in the LR depth map. In order to keep the smoothness of up-sampled depth map, we employ the minimum filter [18] to process the pixels, and the range of its candidate pixel is extended to refine the final results. Similarly, we employ the same principle in the smooth set P_4 for the similar analysis shown as case III in (7).

4. Experimental results

In order to verify the effectiveness of the proposed method, 9 images (i.e., Baby, Books, Bowling, Cloth, Dolls, Laundry, Moebius, Reindeer and Wood) [20] were used for comparison. The ground truth depth images were down-sampled by factors of 2, 4 and 8 with the nearest-neighbor method to generate the LR depth images; when the original depth image resolution was 1104×1376, resolutions of three input depth images become 552×688, 276×344, and 138×172, respectively. We use bilinear interpolation (BI), JBU [2], fast edge-preserving depth image up-sampler (FEDIU) [17] and discontinuity adaptive depth upsampling (DADU) [18] for comparison. During the experiment, for JBU, the spatial Gaussian parameter is set to 0.5, and the range Gaussian parameter is set at 0.1, as described in [2]. The local window is set to 5×5. For FEDIU, the filter parameters are given on account of [18]. For the proposed method, parameters σ_f and σ_r are set at 0.2 and 0.1, respectively. The system configuration for experiments is 1.7Ghz CPU, 8GB RAM. The PCJBF algorithm is implemented using a common mathematical software like Matlab, and it is descirbed as the following Tabel 6.

Generally, PSNR（Peak Signal to Noise Ratio）and BPR (bad pixel rate) are global image quality objective evaluation indexes. Considering humane understanding to the distance is relative and the real application of depth maps, the quality of output depth images is measured based on the ground-truth depth maps. A smaller BPR value means better performance for a

depth up-sampling method.

From **Table 1**, it can be observed that the proposed method has lower BPRs than the other methods in most of the input images. However, the DADU has lower BPRs than the PCJBF in image *Baby* for scaling factor 2 , and in image *Wood* for scaling facor 2 and 4, respectively. Since there exist many depth holes(i.e. the pixels with zeroes depth values) in the input LR depth map and the HR truth depth map due to the object occlusion. These holes can cause many errors in the depth-based algorithms. The new BPR calculations without considering holes pixels will verify that the proposed method outperforms the DADU for input images *Baby* and *Wood* in scaling factor 2. We provide the new BPR without considering holes pixels in **Table 3** as follows. From **Table 3**, the proposed PCJBF method has generate better BPR results than the DADU.

Table 1. BPR (%) comparison

Scaling factors	Images	Up-sampling methods				
		BI	JBU	FEDIU	DADU	PCJBF
2	Baby	1.67	2.56	1.67	**0.98**	1.10
	Books	2.74	3.78	3.11	1.80	**1.32**
	Bowling	1.77	2.61	1.78	1.31	**1.07**
	Cloth	0.43	0.68	0.44	0.31	**0.25**
	Dolls	4.08	5.11	4.49	2.73	**1.70**
	Laundry	3.97	3.63	4.66	1.95	**1.39**
	Moebius	3.56	4.69	4.05	2.36	**1.66**
	Reindeer	2.54	3.30	2.86	1.46	**1.20**
	Wood	0.88	1.34	0.88	**0.32**	0.61
4	Baby	2.80	4.44	2.80	2.16	**1.99**
	Books	4.59	6.51	4.93	4.22	**2.46**
	Bowling	3.17	5.16	3.17	3.11	**2.03**
	Cloth	0.76	1.61	0.77	1.14	**0.42**
	Dolls	6.99	9.15	7.42	6.12	**3.11**
	Laundry	7.15	6.51	7.81	4.07	**2.54**
	Moebius	6.09	8.23	6.52	5.11	**3.04**
	Reindeer	4.47	6.60	4.81	3.77	**2.36**
	Wood	1.74	2.69	1.74	**0.69**	1.08
8	Baby	4.67	8.81	4.66	5.56	**4.64**
	Books	8.15	15.19	8.43	11.71	**7.39**
	Bowling	**5.67**	13.69	**5.67**	10.46	5.92
	Cloth	**1.49**	6.51	**1.49**	5.79	2.11
	Dolls	12.36	21.43	12.71	17.76	**10.12**
	Laundry	12.59	14.88	13.05	11.60	**7.41**
	Moebius	10.67	16.38	10.99	11.77	**7.82**
	Reindeer	8.07	17.21	8.34	13.12	**7.20**
	Wood	3.26	7.91	3.26	4.65	**2.18**

Table 2. BPR (%) comparison based on each pixel set

Images	Pixel category	Up-sampling methods			
		JBU	FEDIU	DADU	PCJBF
Books	P1	3.67	3.28	2.02	**1.77**
	P2	3.11	1.82	**1.28**	**1.28**
	P3	4.18	2.43	4.18	**2.36**
	P4	4.24	**0.90**	4.24	1.98
Dolls	P1	6.01	4.97	3.48	**2.49**
	P2	2.32	1.42	1.18	**1.18**
	P3	8.90	5.13	8.90	**4.55**
	P4	4.20	**1.18**	4.20	1.89
Laundry	P1	4.34	3.39	2.41	**2.0**
	P2	2.03	1.25	0.67	**0.67**
	P3	6.57	6.78	6.57	**3.84**
	P4	1.94	1.64	1.94	**0.90**
Moebius	P1	5.85	4.82	3.27	**3.01**
	P2	3.07	1.65	1.04	**1.04**
	P3	5.57	3.59	5.57	**3.05**
	P4	1.88	0.93	1.88	**0.71**
Reindeer	P1	6.23	5.36	**3.15**	3.26
	P2	1.61	1.11	0.62	**0.62**
	P3	6.04	**1.44**	6.04	2.23
	P4	3.32	**0.45**	3.32	1.09
Avg.BPRs for P1		5.22	4.46	2.87	**2.51**
Avg.BPRs for P2		2.428	1.45	**0.958**	**0.958**
Avg.BPRs for P3		6.252	3.874	6.252	**3.206**
Avg.BPRs for P4		3.116	**1.02**	3.116	1.314

From **Table 3**, it can be observed that the proposed method has lower BPRs than the other methods except the images *Bowling* and *Cloth* for Scaling factor 8. The corresponding pixel-category BPR results for the image *Cloth* and *Bowling* are shown in **Table 5**. There are more color texture regions in *Cloth*, which cause smaller BPRs in the FEDIU algorithm than the proposed method. The image *Bowling* has more smooth regions, which cause fewer BPRs in the FEDIU algorithm than the proposed method. The BI-based depth up-sampling algorithms may have better performance in regions where depth discontinuities are inconsistent with color edges, especailly with a bigger sacaling fator.

Table 3. BPR (%) comparison without considering hole pixels

Scaling factors	Images	Up-sampling methods				
		BI	JBU	FEDIU	DADU	Proposed
2	Baby	1.21	2.0	1.22	0.72	**0.63**
	Books	2.11	3.0	2.11	1.38	**0.64**
	Bowling	1.34	2.07	1.35	1.0	**0.57**
	Cloth	0.37	0.61	0.37	0.26	**0.18**

	Dolls	3.41	4.32	3.42	2.32	**0.93**
	Laundry	3.6	3.18	3.6	1.71	**0.98**
	Moebius	2.97	3.95	2.97	1.99	**0.98**
	Reindeer	2.14	2.82	2.14	1.18	**0.71**
	Wood	0.64	0.93	0.64	0.23	**0.18**
4	Baby	2.23	3.83	2.22	1.72	**1.42**
	Books	3.80	5.69	3.80	3.58	**1.69**
	Bowling	2.59	4.47	2.60	2.72	**1.41**
	Cloth	0.69	1.54	0.70	1.08	**0.35**
	Dolls	6.18	8.28	6.17	5.58	**2.28**
	Laundry	6.71	6.02	6.71	3.75	**2.09**
	Moebius	5.35	7.37	5.34	4.63	**2.26**
	Reindeer	2.14	2.82	2.14	1.18	**0.71**
	Wood	1.33	1.94	1.33	0.56	**0.3**
8	Baby	4.05	8.21	4.05	5.03	**4.03**
	Books	7.33	14.41	7.33	11.03	**6.60**
	Bowling	**4.86**	12.85	**4.86**	10.02	5.15
	Cloth	**1.42**	6.45	**1.42**	5.72	2.04
	Dolls	11.52	20.64	11.51	17.14	**9.27**
	Laundry	12.11	14.4	12.10	11.23	**6.95**
	Moebius	9.82	15.53	9.82	11.1	**7.0**
	Reindeer	7.5	16.6	7.5	12.72	**6.51**
	Wood	2.50	6.51	2.50	4.45	**0.67**

Table 4. BPR (%) comparison without holes based on each pixel set for scaling factor 4

Images	Pixel category	Up-sampling methods			
		JBU	FEDIU	DADU	Proposed
Books	P1	2.13	1.79	0.86	**0.54**
	P2	1.18	0.52	0.36	**0.36**
	P3	1.40	1.22	1.41	**0.56**
	P4	0.93	0.24	0.93	**0.21**
Dolls	P1	3.51	2.81	1.5	**0.66**
	P2	1.0	0.49	0.35	**0.35**
	P3	2.85	2.50	2.87	**1.1**
	P4	0.84	0.32	0.84	**0.15**
Laundry	P1	2.48	1.86	0.9	**0.69**
	P2	0.82	0.41	0.15	**0.15**
	P3	2.44	4.14	2.44	**1.19**
	P4	0.25	0.27	0.25	**0.05**
Moebius	P1	3.32	2.52	1.41	**0.98**
	P2	1.03	0.46	0.24	**0.24**
	P3	2.44	2.05	2.46	**0.93**
	P4	0.5	0.27	0.5	**0.1**
Reindeer	P1	1.82	1.47	0.44	**0.43**
	P2	0.28	0.13	0.04	**0.04**
	P3	0.54	0.46	0.54	**0.19**
	P4	0.17	0.08	0.17	**0.04**

Avg.BPRs for P1	2.65	2.09	1.02	**0.66**
Avg.BPRs for P2	0.862	0.4	0.23	**0.23**
Avg.BPRs for P3	1.93	2.07	1.94	**0.79**
Avg.BPRs for P4	0.54	0.24	0.54	**0.11**

Table 5. Pixel-category BPR (%) comparison without holes for scaling factor 8

Images	Pixel category	Up-sampling methods			
		JBU	FEDIU	DADU	Proposed
Bowling	P1	2.64	1.97	1.32	**1.23**
	P2	3.37	**1.70**	1.94	1.94
	P3	0.98	0.32	0.99	**0.17**
	P4	5.61	**0.81**	5.65	1.76
Cloth	P1	1.56	0.91	0.91	**0.64**
	P2	0.13	**0.05**	0.06	0.06
	P3	4.33	**0.44**	4.33	1.26
	P4	0.42	**0.02**	0.42	0.08

To further demonstrate the effectiveness of pixel category, we recalculate its BPRs including holes and obtain the average BPRs for each pixel subset shown in **Table 2**. The down-sampling factor is 4. From **Table 2**, the PCJBF has bigger BPR performance than the other algorithms in some cases. Then, the BPRs based on each pixel set without holes pixels are recalculate in **Table 4** as follows.

From **Table 4**, The average BPRs of output depth map in P_1 are 2.65, 2.09, 1.02and 0.66 for JBU, FEDIU, DADU and the proposed method, respectively. This outcome indicates that the proposed method has better BPRs performance by as much as about 1.99, 1.34 and 0.36 over JBU, FEDIU and DADU, respectively, on average in edge regions. Besides, we obtain the average BPRs for the texture subset. The average BPRs of up-sampled depth map in P_3 are 1.93, 2.07, 1.94 and 0.79 for JBU, FEDIU, DADU and the proposed method, respectively. The proposed method has lower BPRs by as much as about 1.14, 1.28 and 1.15 over JBU, FEDIU and DADU, respectively, on average in texture regions, which still proves that the vibration of color values in the texture region will add errors. The average BPRs of up-sampled depth map in P_4 are 0.54, 0.24, 0.54 and 0.11 for JBU, FEDIU, DADU and the proposed method, respectively. We determine the PCJBF has fewer average BPRs for each pixel set than the other methods.

Table 6. The description of the PCJBF algorithm

The PCJBF algorithm
Input: 2D LR depth map D_L , its size is h × w;
Input: 2D HR image I; the image size is H × W;
Output: 2D HR depth map D;
Parameters and notations
D_0: an HR depth map based on JBU method
E_c : an HR edge map derived from LR depth map;
E_d : an HR edge map derived from HR color image;
p: the pixeles in HR space;
p_L: the p's corresponding pixels in LR space;
s: up-sampling factor;

```
/*PCJBF Algorithm*/
/*Intialize D*/
Compute D0 using Eq. (1);
/* compute the two HR edge map, respectively.*/
for i ← 1 to H
  for j ← 1 to W
    Compute the normalized variance σ_{D_L}(p_L), then obtain E_d(i,j) using Eq. (4);
end
end
for i ← 1 to H
  for j← 1 to W
    Compute E_c(i,j) using Eq. (5);
end
end
/* obtain the up-sampled depth map*/
for i ← 1 to H
  for j← 1 to W

    if (p∈E_d and p∈E_c )
            then Compute D(i,j) using the case Ⅰ of Eq. (7);
    else if (p∈E_d and p∉E_c )
          then Compute D(i,j) using case Ⅱ of Eq. (7);
    else
                then Compute D(i,j) using case Ⅲof Eq. (7);
      end
end
end
```

In addition, The staircase problem is mainly generated by the resolution gap. By using the other edge map based on a high-resolution color image, the HR pixels corresponding to a same LR pixel are classified into edge points and edge-neighbor pixels. The BPR results of P_1 set in Table 4 show that the proposed PCJBF has more accurate HR edge results than DADU and other methods, which effectively decrease the blurring artifact. In **Fig. 10**, there are two rows of figures, the up-row figure in the first column is part of input HR color image *Dolls*, and the down-row in the first column is part of the HR true depth map. Form column (b) to column (f) are the results of BI, JBU, FEDIU, DADU and PCJBF. The depth edges are getting smoother in PCJBF than the other methods, which means that PCJBF can reduce the staircase blurring artifact.

As a result, the proposed depth up-sampler has the best performance among the comparative methods in terms of BPRs evaluation. We can observe that the proposed method can provide better results than the other methods in many cases.

In the following, the subjective comparisons of *Baby*, *Books*, *Bowling*, *Cloth*, *Dolls*, *Laundry*, *Moebius*, *Reindeer* and *Wood* are shown as **Fig. 4**, for the case of the down-sampling factor 4. **Fig. 4(a)** and **Fig. 4(b)** show the input HR color image and the ground truth depth maps. **Fig. 4(b)**, **(c)**, **(d)**, **(e)** and **(f)** exhibit the results of BI, JBU, FEDIU, DADU and the proposed method. **Fig. 5** and **Fig. 8** show the subjective comparison of the above images for the case of the down-sampling factors 8 and 2, respectively.

Fig. 4. Results of *Baby*, *Books*, *Bowling*, *Cloth*, *Dolls*, *Laundry*, *Moebius*, *Reindeer* and *Wood*. Down-sampling factor is 4. (a) high resolution color image, (b) ground truth high resolution depth image, (c) results of BI, (d) results of JBU, (e) results of FEDIU, (f) results of DADU and (g) results of the proposed method.

Fig. 5. Results of *Baby*, *Books*, *Bowling*, *Cloth*, *Dolls*, *Laundry*, *Moebius*, *Reindeer* and *Wood*. Down-sampling factor is 8. (a) high resolution color image, (b) ground truth depth image of (a), (c) results of BI, (d) results of JBU, (e) results of FEDIU, (f) results of DADU and (g) results of the proposed method.

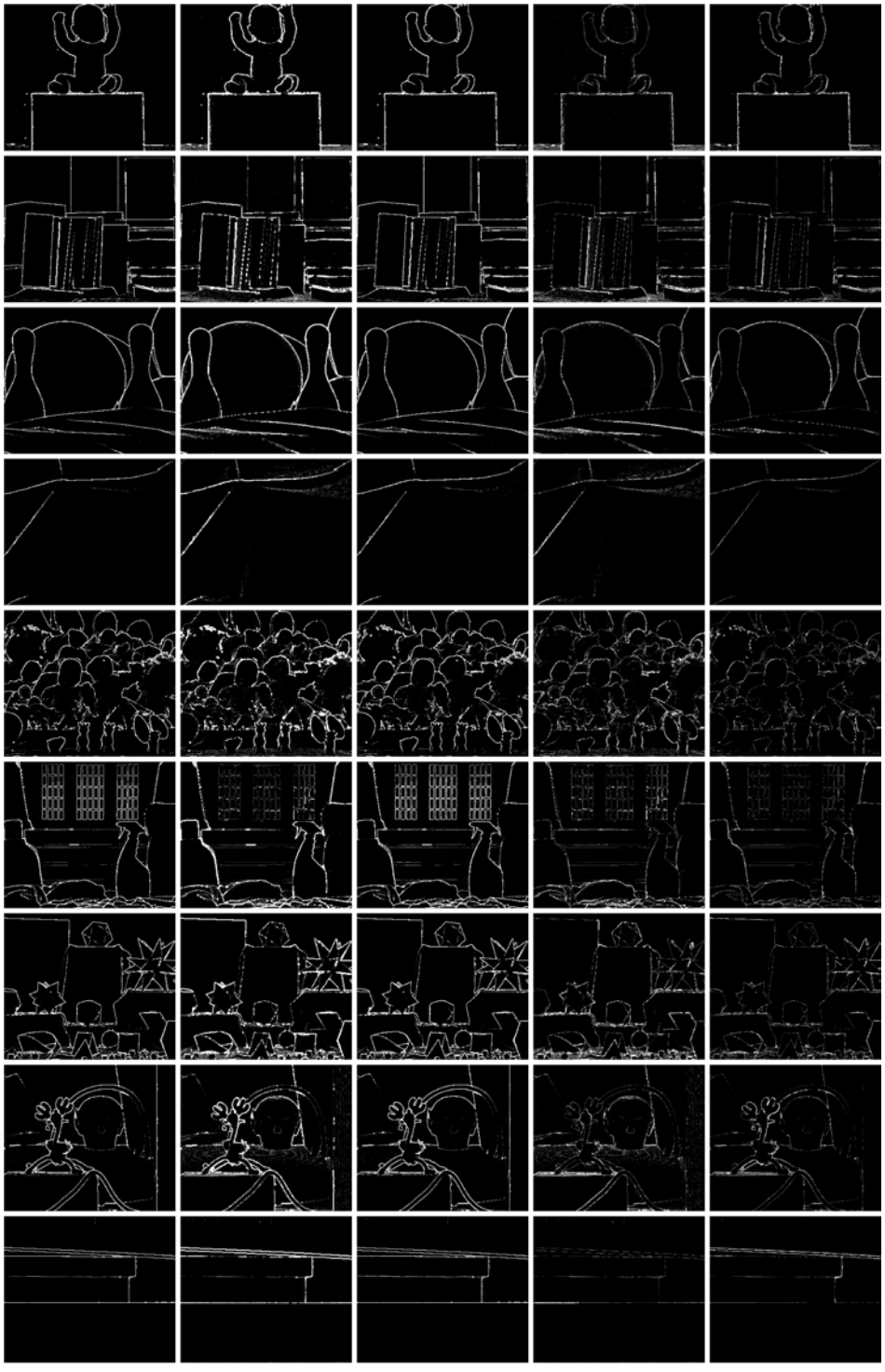


Fig. 6. Error map results of *Baby*, *Books*, *Bowling*, *Cloth*, *Dolls*, *Laundry*, *Moebius*, *Reindeer* and *Wood*. Down-sampling factor is 4. (a) results of BI, (b) results of JBU, (c) results of FEDIU, (d) results of DADU and (e) results of the proposed method.

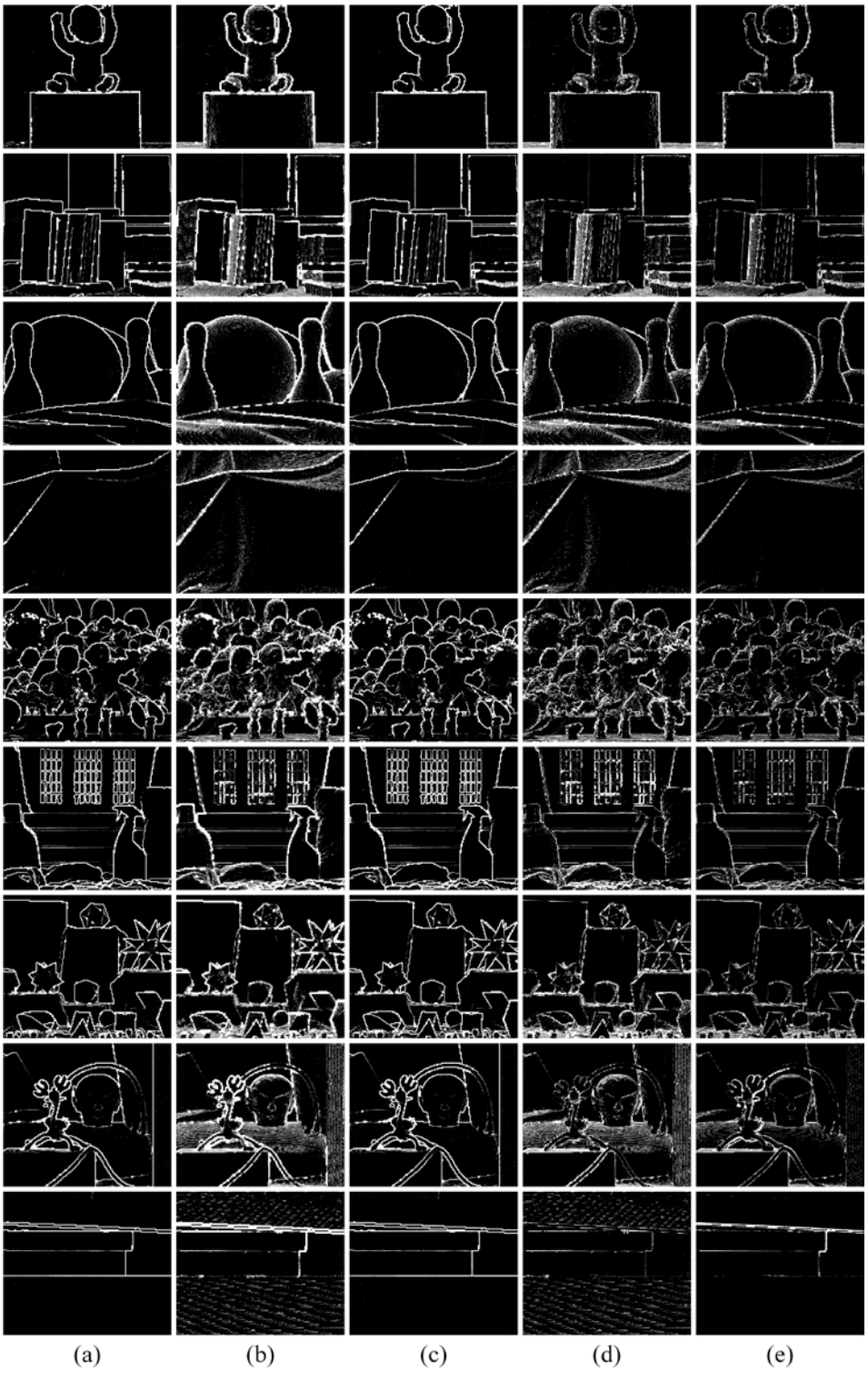


Fig. 7. Error map results of Baby, Books, Bowling, Cloth, Dolls, Laundry, Moebius, Reindeer and Wood. Down-sampling factor is 8. (a) results of BI, (b) results of JBU, (c) results of FEDIU, (d) results of DADU and (e) results of the proposed method.

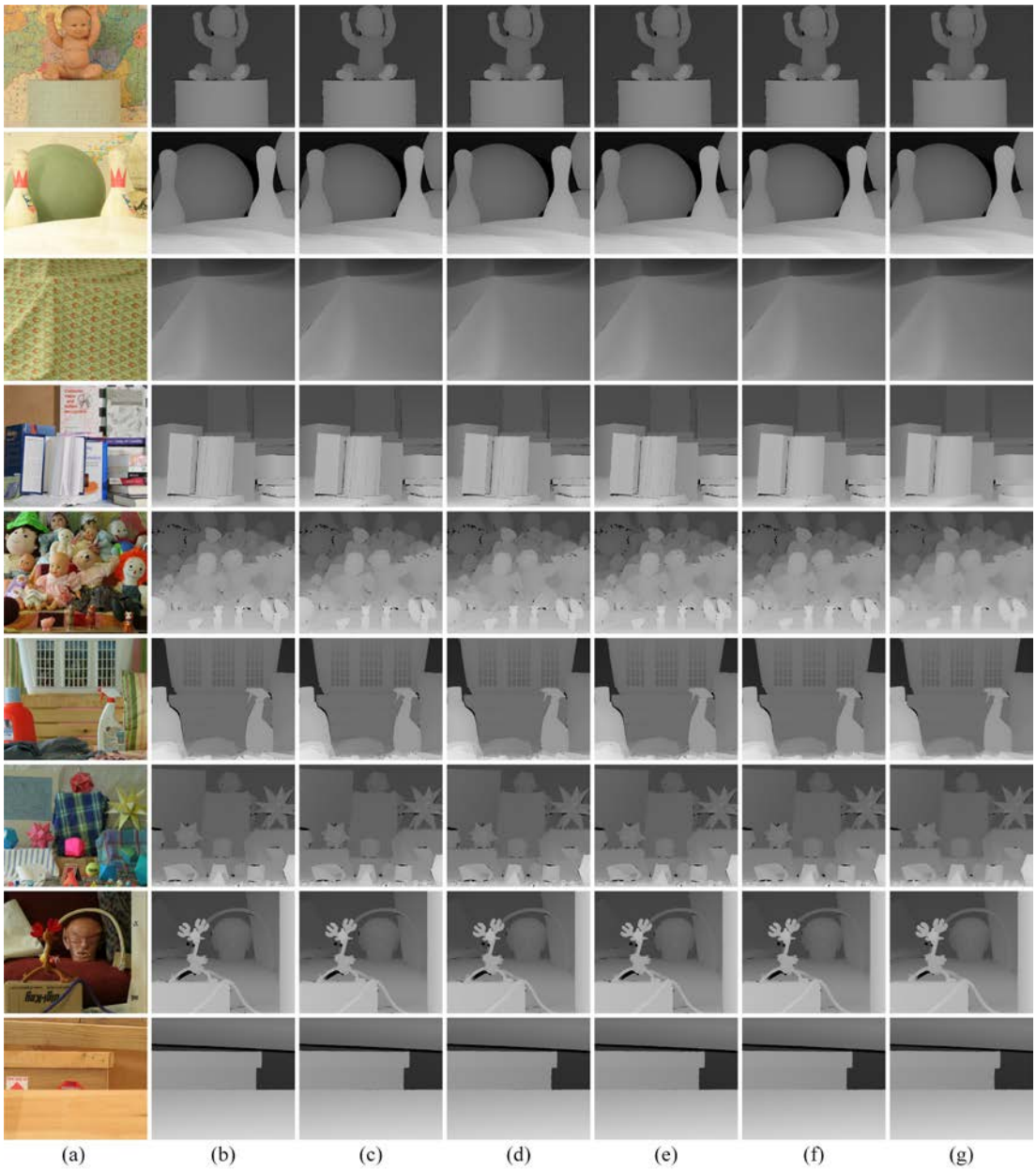


Fig. 8. Results of *Baby*, *Books*, *Bowling*, *Cloth*, *Dolls*, *Laundry*, *Moebius*, *Reindeer* and *Wood*. Down-sampling factor is 2. (a) high resolution color image, (b) ground truth depth image of (a), (c) results of BI, (d) results of JBU, (e) results of FEDIU, (f) results of DADU and (g) results of the proposed method.

Fig. 9. Error map results of *Baby*, *Books*, *Bowling*, *Cloth*, *Dolls*, *Laundry*, *Moebius*, *Reindeer* and *Wood*. Down-sampling factor is 2. (a) results of BI, (b) results of JBU, (c) results of FEDIU, (d) results of DADU and (e) results of the proposed method.

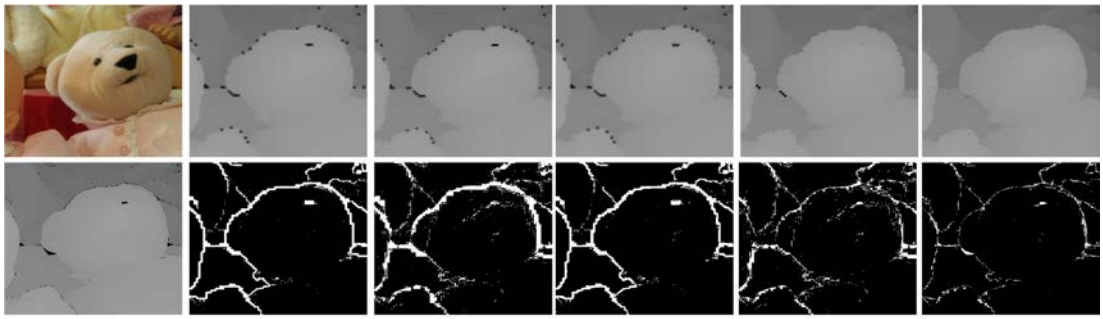


Fig. 10. Part results of Dolls. Down-sampling factor is 4. From up to down, (a) part of HR color image and part of ground true HR depth map, (b) part results of HR depth map and error map by BI , (c) part results of HR depth map and error map by JBU, (d) part results of HR depth map and error map by FEDIU, (e) part results of HR depth map and error map by DADU and (f) part results of HR depth map and error map by PCJBF.

In order to further demonstrate the effectiveness of the proposed method visually, the error maps of the above images for scaling factors 4 and 8 are shown in **Fig. 6** and **Fig. 7**, respectively. In **Fig. 6**, **Fig. 7** and **Fig. 9**, columns (a), (b), (c), (d) and (e) exhibit the results of BI, JBU, FEDIU, DADU and the proposed method from left to right. The error maps show the differences between the up-sampled depth maps and the truth HR depth map. Generally, the depth errors are mainly in the regions where depth discontinuity is inconsistent with the HR color edges (i.e., the regions where depth values are disrupted and the color texture regions). So, there are many similar shape structures in **Fig. 6**, **Fig. 7** and **Fig. 9**. When the color values of these lines are bigger, then the lines in the error maps are brighter, there are more pixels in the regions, and the error values are higher. From Table 3, the BPR values increase with the increasing scaling factor; similarly, the lines with similar shapes get brighter from **Fig. 9** to **Fig. 6** to **Fig. 7**. In the figures, the depth errors near the edge regions were effectively reduced in the results of the proposed method compared with the other methods, and the depth errors in the homogeneous regions were also reduced. As a result, we can observe that the proposed method can obtain more smoothing effects than other methods from those figures.

5. Conclusion

In this paper, a novel depth up-sampling method is proposed based on pixel classifying and joint bilateral filtering. By analyzing the edge points both in the high resolution color image and low resolution depth map, pixels in depth maps can be classified into four categories: edge points, edge-neighbor points, texture points and smooth points. First, joint bilateral up-sampling (JBU) method is used to generate an initial up-sampling depth image. Then, for each category, different refinement methods are employed to modify the initial up-sampling depth image. Extensive experiments demonstrates that the proposed method outperforms conventional algorithms, JBU, FEDIU and DADU, and it can effectively reduce the over blurred depth regions and improve the up-sampled depth quality.

Reference

[1] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3d-tv," in *Proc. of SPIE, Stereoscopic Displays and Virtual Reality Systems XI*, Vol.5291, pp.93-104, 2004. Article (CrossRef Link)

[2] J. Kopf, M.F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. On Graphics*, Vol.26 No.3, pp.1-6, 2007. Article (CrossRef Link)

[3] A.K. Riemens, O.P. Gangwal, B. Barenbrug, and R.M. Berretty, "Multi-step joint bilateral depth upsampling," in *Proc. of SPIE, Visual Communications and Image Processing*, Vol.7257, pp.72570M, 2009. Article (CrossRef Link)

[4] M. Liu, O. Tuzel, Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.169-176, October, 2013. Article (CrossRef Link)

[5] Y. Wang, A. Ortega, "A graph-based joint bilateral approach for depth enhancement," in *Proc. of IEEE Conf. on Acoustic, Speech and Signal Processing*, pp.885-889, July, 2014. Article (CrossRef Link)

[6] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autogressive model," *IEEE Trans. On Image Processing*, Vol.23 No.8, pp.3443-3458, 2014. Article (CrossRef Link)

[7] D. Min, J. Lu and M.N. Do, "Depth video enhancement based on weighted mode filtering," *IEEE Trans. On Image Processing*, Vol.21 No.3, pp.1176-1190, 2012. Article (CrossRef Link)

[8] C. Jung, S. Joo and S. M. Ji, "Depth map upsampling with image decomposition," *Electronics Letter*, Vol.51 No.22, pp.1782-1784, 2015. Article (CrossRef Link)

[9] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," *Advances in Neural Information Processing Systems*, Vol.18, pp.291-298, 2006. Article (CrossRef Link)

[10] W. Hannemann, A. Linarth, B. Liu and G. Kokai, "Increasing depth lateral resolution based on sensor fusion," *International Journal of Intelligent Systems Technologies and Applications*, Vol.5 No.3, pp.393-401, 2008. Article (CrossRef Link)

[11] B. Huhle, S. Fleck and A. Schilling, "Integrating 3D time-of-flight camera data and high resolution images for 3DTV applications," in *Proc. of 3DTV Conference*, pp.1-4, May, 2007. Article (CrossRef Link)

[12] J. Zhu, L. Wang, J. Gao and R. Yang, "Spatial-temporal fusion for high accuracy depth maps using dynamic MRFs," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol.32, pp.899-909, 2010. Article (CrossRef Link)

[13] W. Liu, S. Jia, P. Li, X. Chen, J. Yang and Q. Wu, "An MRF-based depth upsampling: upsample the depth map with Its own property," *IEEE Signal Processing Letters*, Vol.22, pp.1708-1712, 2015. Article (CrossRef Link)

[14] Y. Zuo, P. An, S. Zheng, Z. Zhang, "Depth upsampling method via markov random fields without edge-misaligned artifacts," in *Proc. of IEEE Conference on Image Processing*, pp.2324-2328, September, 2015. Article (CrossRef Link)

[15] J. Xie, R. S. Feris and M. Sun, "Edge-guided single depth image super resolution," *IEEE Trans. on Image Processing*, Vol.25 No.1, pp.428-438, 2016. Article (CrossRef Link)

[16] D. Ferstl, M. Rüther, H. Bischof, "Variational depth superresolution using example-based edge representations," in *Proc. of IEEE Conf. on Computer Vision*, pp.513-521, December, 2015. Article (CrossRef Link)

[17] S. Kim and Y. Ho, "Fast edge-preserving depth image upsampler," *IEEE Trans. on Consumer Letters*, Vol.58, pp.971-976, 2012. Article (CrossRef Link)

[18] S. B. Lee, S. Kwon and Y. S. Ho, "Discontinuity adaptive depth upsampling for 3D video acquisition," *Electronics Letters*, Vol.49, pp.612-1614, 2013. Article (CrossRef Link)

[19] Y. S. Kang, S. B. Lee and Y.S. Ho, "Depth map up-sampling using depth local features," *Electronics Letters*, Vol.50, pp.170-171, 2014. Article (CrossRef Link)

[20] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, Vol.47, pp.7-42, 2002. Article (CrossRef Link)

Yannan Ren received her B.E. degree in electrical engineering, and M.S. degree in communication engineering from Lanzhou University, China, in 2000 and 2003, respectively. She is currently a Ph.D. candidate in the School of Information Science and Engineering at Shandong University. Her research interests include 2D-to-3D conversion, depth map generation, depth up-sampling, image understanding.
E-mail: ren_yannan@hotmail.com

Ju Liu received his B.S. and M.S. degrees both in Electronic Engineering from Shandong University (SDU), Jinan, China, in 1986 and 1989, respectively and his Ph.D degree in Signal Processing from Southeast University (SEU), Nanjing, China, in 2000. Since July 1989, he has been with the Department of Electronic Engineering of SDU. At present he is a professor with the School of Information Science and Engineering at SDU. His current research interests include space-time processing in wireless communication, blind signal separation, and multimedia communications.
E-mail: juliu@sdu.edu.cn

Hui Yuan received the B.E. and Ph.D. degree in telecommunication engineering from Xidian University, Xi'an, China, in 2006 and 2011, respectively. He was a Post-Doctoral Fellow with the Department of Computer Science, City University of Hong Kong, Hong Kong, from 2013.01 to 2014.12. He is currently an Associate Professor with the School of Information Science and Engineering, Shandong University (SDU), Jinan, China. His current research interests include video coding, multimedia communication.
E-mail: Huiyuan@sdu.edu.cn

Yifan Xiao received the B.E. degree in electrical engineering from Shandong University (SDU), Jinan, China, in 2015. She is currently pursuing the M.E. degree in information science and engineering, at Shandong University. Her research interests include super-resolution, frame rate up-conversion and image processing.
E-mail: xiaoyifran@163.com