Neurocomputing 151 (2015) 663–673

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Depth map reconstruction and rectification through coding parameters for mobile 3D video system

You Yang [a,b], Huiping Deng [c,*], Jin Wu [c], Li Yu [a]

[a] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[b] Department of Automation, Tsinghua University, Beijing 100084, China
[c] School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China

ARTICLE INFO

Article history:
Received 15 November 2013
Received in revised form
21 April 2014
Accepted 26 April 2014
Available online 1 November 2014

Keywords:
Histogram rectification
Depth up-sampling
Edge preserving
Asymmetric coding
3-D video

ABSTRACT

Depth maps in mobile asymmetric 3D video system usually have the same video-rate as corresponding color video but lower resolution. In this system, depth maps will be re-scaled to the same resolution as the color video at the receiver side after decoding. In this case, blurring effect will appear after direct up-sampling process, causing artifacts in 3D scene reconstruction. In this paper, we propose a method for depth map reconstruction and rectification for the asymmetric coding scheme to avoid the blurring effect by using coding parameters in bit-streams. In the first step, we utilize an edge-oriented interpolation algorithm to improve the accuracy around edge-related regions for the up-sampled depth map. A weight model is built by edge and structural similarities, where edge similarity is a measurement between depth maps and their corresponding texture images, and structural similarity is for different depth maps. Based on the weight model, the up-sampling coefficients can be selected adaptively according to the edge orientation. After that, a block-based histogram rectification method is proposed to remove blurring artifacts caused by loop filter in decoder and up-sampling. A mapping function is built for the cumulative histogram of the up-sampled depth map and its original depth map to correct the error pixels in the up-sampled depth map. The block information of the histogram rectification method is extracted from the bit-stream at the decoder. The experimental results show that the proposed depth map reconstruction method has gains on both subjective and objective performances compared with existing methods. For subjective performance, our method can suppresses image blurring and preserves sharp edges in depth maps. As for objective performance, the proposed method achieves a maximum 2.93 dB and an average 1.60 dB PSNR gain compared the state-of-the-art methods.
© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Recent years have witnessed the rapid development of three-dimensional (3D) researches and applications in both academia and industry, including 3D movie and television in the field of video processing [1], video conference in visual communication [2–5], 3D natural scene modeling in computer vision and graphics [6–10], and many other fields. In all of these researches, depth map plays an important role because it records the coordinates in 3D space for all visible pixels. With the help of video-rate depth map, it is convenient to estimate the inter-view redundancies in 3D video compression, or identify the relationship of voxels in natural scene 3D modeling. If video-rate high quality and high resolution depth maps can be obtained, more details in 3D scenes can be recorded. However, there is a trade-off between quality and resolution in obtaining the video-rate depth maps, especially in mobile 3D applications [11–16].

In mobile 3D video communications, several approaches were proposed to handle the trade-off for video-rate depth maps, and they can be generally categorized into symmetric and asymmetric methods [17–21]. As for the symmetric methods, depth maps are treated as ordinary color video with only Y-component, and they are with video-rate high quality and high resolution. Traditional video coding techniques can be used to process the depth maps in this case. For example, the representation of multi-view video plus depth can be processed on H.264 and JMVC coding platform [22,17]. Further researches such as bit-allocation between depth map and color video can improve the quality of viewing experiences on the decoder side [23,24]. However, the pre-condition of frame-rate, quality and resolution for depth maps cannot be satisfied in some cases. For example, depth sensors (RGB-D or TOF cameras) can capture the 3D space coordinates in video-rate but lower resolution (LR) and quality [25–27]. In order to solve this problem, asymmetric methods were proposed when one of the

* Corresponding author.
E-mail address: denghuiping@wust.edu.cn (H. Deng).

http://dx.doi.org/10.1016/j.neucom.2014.04.088
0925-2312/© 2014 Elsevier B.V. All rights reserved.

conditions (i.e., video-rate, quality or resolution) cannot be satisfied. There are two main approaches for asymmetric methods, where one is asymmetric coding, and the other one is asymmetric processing [28–30]. As for asymmetric coding, depth maps are encoded with video-rate but LR, and these depth maps are obtained by down-sampling before decoding. At the decoder side, the LR depth maps are up-sampled to restore their original resolutions [28,29,31]. Asymmetric coding methods can be applied to the bandwidth-constrained environments due to their better coding performances compared to the symmetric methods. On the other hand, the main idea of asymmetric processing methods is that depth maps can be reconstructed at the decoder side rather than transmitted from the encoder side, and this can cause significant bandwidth cut with appreciate quality in mobile applications [32–35]. In these methods, it is assumed that the variation for a given dynamic scene is identical for both the depth and the color information of one viewpoint [11]. Specifically, the static objects in consequent color frames will not arouse variation of depth value. At the same time, the moving objects in consequent color frames cause depth value variation on them. Therefore, the status (i.e., static or motive) of object in consequent color frames can be used to describe the depth value variation in depth map.

Our motivation of this paper comes from the asymmetric method. As discussed above, depth maps in mobile 3D video system are usually with video-rate but LR, and these depth maps will be resized to the same resolution with the color video at the receiver side after decoding. Actually, blurring effect is resultant from direct up-sampling, leading artifacts in 3D dynamic scene reconstruction [11,30]. If asymmetric processing scheme is used, significant performance gain can only be found in very low bit-rate cases [32,36]. In many mobile applications, video-rate high quality and high resolution depth maps are needed at the receiver side, but only LR can be provided at the encoder side. Therefore, it stays challenge to obtain video-rate high quality and high resolution depth maps from low resolution ones in environment with limited bandwidth. In order to solve this challenge, many methods for high-quality depth map processing are proposed, where edge preservation is the most important goal. Traditional filters, including linear and bi-cubic interpolation, can be utilized for depth maps but better performance can be found on color images. Bilateral filter is commonly used for both depth and color images, and the weights of the filter are selected as a function of a photometric similarity of the neighboring pixels. Besides that, joint bilateral filters were proposed by using auxiliary information from high-resolution images, and these schemes are benefit for edge preservation. The scheme was also adopted by joint bilateral up-sampling (JBU), making this method has been widely utilized for depth map reconstruction in many applications [37,17,38]. Seung proposed an adaptive joint trilateral filter for both depth map and color image. In this method, the weighted model was determined by the local characteristics of depth map and color image [39]. However, depth maps processed by joint bilateral (or trilateral) filters are still suffered by unnecessary blur due to weight summation procedure. In order to preclude the summation, Yang et al. built a 3D cost of depth probability, and then iteratively applied a bilateral filter to the cost volume [40]. The output high resolution range depth map is obtained by taking the winner-takes-all approach on the weight cost volume. Wildeboer et al. proposed a JBU algorithm by utilizing the high-resolution texture video in the process of depth up-sampling [41]. They calculated a weight-cost based on pixel positions and intensity similarities. Min et al. proposed a weighted mode filtering (WMF) to enhance the depth spatial resolution by using the concept of JBU [42]. The weights are obtained by similarity measurement between reference and neighboring pixels, and they are used to construct a joint histogram. The final interpolated depth value is then determined by a global optimization among the depth candidates on this histogram.

In this paper, we propose a scheme for depth map reconstruction and rectification through coding parameters in 3D video system. This scheme is a joint method utilizing both asymmetric coding and processing algorithms, and has potentials in future mobile 3D video systems. In this scheme, we first utilize an edge-oriented interpolation algorithm to improve the accuracy around edge-related regions for the up-sampled depth map. A weight model is built by edge and structural similarities, where edge similarity is a measurement between depth maps and their corresponding texture images, and structural similarity is for different depth maps. Based on the weight model, the up-sampling coefficients can be selected adaptively according to the edge orientation. After that, a block-based histogram rectification method is proposed to remove blurring artifacts caused by two cases, including loop filter in decoder and up-sampling. A mapping function is built for the cumulative histogram of the up-sampled depth map and its original depth map to correct the error pixels in the up-sampled depth map. The block information of the

Table 1
Notations and definitions in the proposed method.

Notation	Description
D	Interpolated HR depth map, size: $2M \times 2N$
$w \times w$	Window size
d	Down-sampled LR depth map, size: $M \times N$
h_d	Histogram of d
(x, y)	Pixel coordination
c_d	Cumulative histogram of d
$D_{2x,2y}; D_{2x+2,2y};$	Neighbor pixels
h_D	Histogram of D
$D_{2x,2y+2}; D_{2x+2,2y+2}$	Diagonal pixels
$\mathbf{K} = [k_0, k_1, k_2, k_3]$	Interpolation coefficients
c_D	Cumulative histogram of D
p_n	Weight model
D_c	Rectified depth map

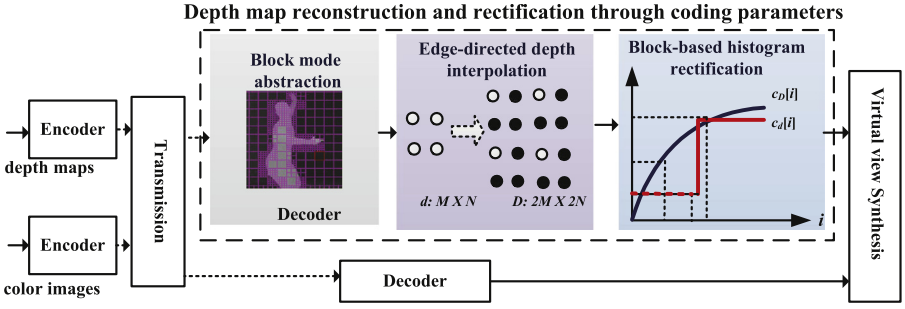

Fig. 1. Framework of the proposed method.

histogram rectification method is extracted from the bit-stream at the decoder. The contribution of our work in this paper is in three-fold.

- We propose a depth map reconstruction and rectification method through coding parameters for mobile 3D video system. In this method, low-resolution depth maps are up-sampled and rectified by using the encoding information in bit-streams.
- We utilize an edge-oriented interpolation algorithm to solve the low spatial resolution problem for compressed depth maps. Edge and structure similarities are taken into account to build a weight model, where edge similarity is for depth map and the corresponding texture image, and structure similarity is for depth maps. Based on the weight model, the up-sampling coefficients can be estimated through the local covariance of the down-sampled depth map, and then be adapted to the edge orientation for the goal of edge preservation.
- A block-based histogram rectification method is utilized to remove the fake edges, which are the side effects of asymmetric coding and processing. Actually, the local cumulative histogram of original depth map is piecewise linear, whereas will be transitional after being up-sampled because of compression losses. We build a mapping function between the cumulative histograms of the up-sampled depth map and its original depth map to correct the error pixels in the up-sampled depth map.

The remainder of this paper is organized as follows. Section 2 describes our proposed method in details. After that, experimental results and conclusions are given in Sections 3 and 4, respectively.

2. Proposed method

In this paper, we present a scheme for depth map reconstruction and rectification for asymmetric coding scheme by using coding parameters in bit-streams. The proposed method utilizes coding parameters as side information to reconstruct the depth maps while preserving the original depth edges.

The overview of the framework is provided in Fig. 1. For each input low resolution (LR) depth maps and higher resolution (HR) color images, we encode them independently. At the decoder, the decoded LR depth map is up-sampled to restore the original resolution by using an edge-preserving depth interpolation. After that, a block-based histogram rectification is used to eliminate the blurring artifacts caused by previous interpolation. Because the parameter of block partition size in bit-stream is helpful in distinguishing objection edges, this parameter is extracted for our histogram rectification.

2.1. Edge-preserving depth up-sampling

Let d denote the input LR depth map of size $M \times N$, and D is the HR depth map after up-sampling to size $2M \times 2N$. We up-sample the depth map by factor of 2 for simplicity. For clarity, we list the notations to be used throughout the paper in Table 1. We utilize the fourth-order linear interpolation to include the nearest four neighbors along the diagonal directions. We directly copy d to D by $D_{2x,2y} = d_{x,y}$. After that, we complete D from d by two steps. The first step is to interpolate $D_{2x+1,2y+1}$ from its nearest four neighbors $D_{2x,2y}$, $D_{2x+2,2y}$, $D_{2x,2y+2}$ and $D_{2x+2,2y+2}$ along the diagonal directions of a square lattice. The second step is to interpolate other missing pixels $D_{2x+1,2y}$ and $D_{2x,2y+1}$ from a rhombus lattice in the same way after a 45° rotation of the square grid. For example, $D_{2x+1,2y+1}$ is calculated as

$$y_{2i+1,2j+1} = k_0 y_{2i,2j} + k_1 y_{2i,2j+2} + k_2 y_{2i+2,2j} + k_3 y_{2i+2,2j+2}, \quad (1)$$

where k_0, k_1, k_2, and k_3 are interpolation coefficients.

There is an assumption in natural statistical image representation that the mean and variance for a region centered by one pixel equals to the local mean and variance of all pixels for a certain range surrounding [43,44]. Comparing to natural images, depth map has more homogenous regions. Therefore, it is a natural selection to treat depth maps as locally stationary. For example, the optimal minimum mean squared error (MMSE) linear interpolation is based on this assumption, and it can effectively remove noises while preserving important image features (e.g., edges). Under this assumption, depth image can be modeled by a locally stationary Gaussian process. Therefore, according to classical Wiener filtering theory, the optimal MMSE linear interpolation coefficients $\mathbf{K} = [k_0, k_1, k_2, k_3]$ can be given by

$$\mathbf{K} = \mathbf{R}^{-1}\mathbf{r} \quad (2)$$

where $\mathbf{R} = E[\mathbf{D}\mathbf{D}^{\mathbf{T}}]$, $\mathbf{D} = [D_{2x,2y}, D_{2x+2,2y}, D_{2x,2y+2}, y_{2x+2,2y+2}]^{\mathbf{T}}$, and $\mathbf{r} = E[D_{2x+1,2y+1}\mathbf{D}]$ is the local covariance at the high-resolution level.

By exploiting the similarity between the HR covariance and the LR covariance, $\mathbf{R}$ and $\mathbf{r}$ can be estimated from a local window of its LR depth map. As shown in Fig. 2, we estimate $\mathbf{R}$ and $\mathbf{r}$ based on the local statistics from a window with size $w \times w$ and centered by

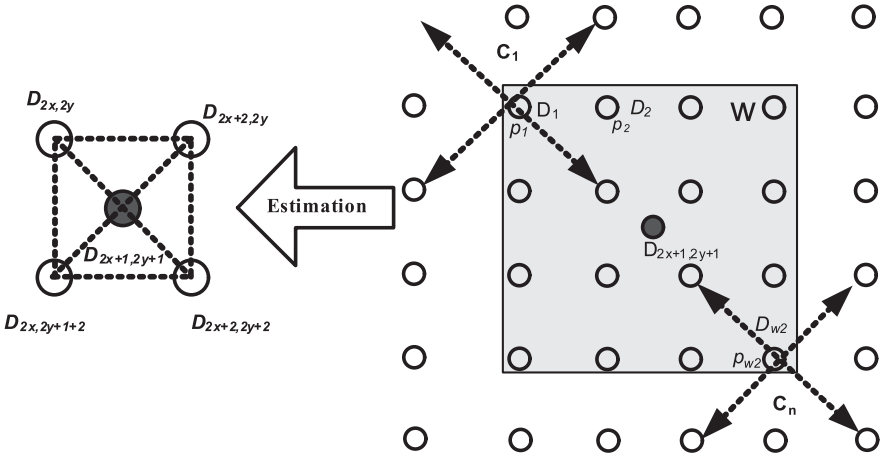

Fig. 2. Covariance estimation based on local statistics from a local window.

the interpolated pixel location, and the model is

$$\hat{R} = \sum_{n=1}^{w^2} p_n \mathbf{c}_n^T \mathbf{c}_n, \quad \hat{r} = \sum_{n=1}^{w^2} p_n \mathbf{c}_n^T \mathbf{c}_n \tag{3}$$

where p_n is the weight of sample D_n, $\mathbf{c}_n$ is a 4×1 matrix whose samples are the nearest four neighbors of D_n along the diagonal directions.

We note that there is a special case for us where the covariance estimation scheme in NEDI [45] with each sample inside the $w \times w$ window has the same weight $p_n = 1/w^2$. In edge-preserving depth map interpolation, the samples $\mathbf{d} = [D_1, D_2, \ldots, D_{w^2}]^{\mathbf{T}}$ selected for coefficients calculation should have similar geometric structure (i.e. edge direction) with the region centered by the interpolated pixel $D_{2x+1,2y+1}$. Otherwise, in the presence of a sharp edge, severe artifacts will appear if the sample is interpolated crossing rather than aligning the edge direction.

Based on previous analyses, we built a weighted model for each sample and make samples adaptively according to the local characteristics of depth map. In this model, pixel distance, intensity difference, and texture similarity are designed according to the geometric similarity within depth maps and the photometric similarity between the depth map and corresponding texture image. These three weights are mutually independent, and then can be pooled by a linear model as follows:

$$p_n = a p_n^c + b p_n^d + c p_n^t, \tag{4}$$

where a, b and c are model parameters. The variable p_n^c is related to the distance between the current and center pixel positions (i.e., (x_n, y_n) and (x_c, y_c)), and it is measured by the Euclidean distance as

$$dist(n) = \sqrt{(x_c - x_n)^2 + (y_c - y_n)^2} \tag{5}$$

and given by

$$p_n^c = \frac{\max\_dist - dist(n)}{\max\_dist - \min\_dist}, \tag{6}$$

where max_*dist* and min_*dist* are the maximum and minimum pixel distance within the window W, respectively. The variable p_n^d in Eq. (4) is related to the absolute difference $dif(n) = |D_n - D_c|$ between the current and center pixel values (i.e., D_n and D_c) in depth map, and given by

$$p_n^d = \frac{\max\_difD - difD(n)}{\max\_difD - \min\_difD}, \tag{7}$$

where max_*difD* and min_*difD* are the maximum and minimum pixel difference within the window W, respectively. Similar to p_n^d,

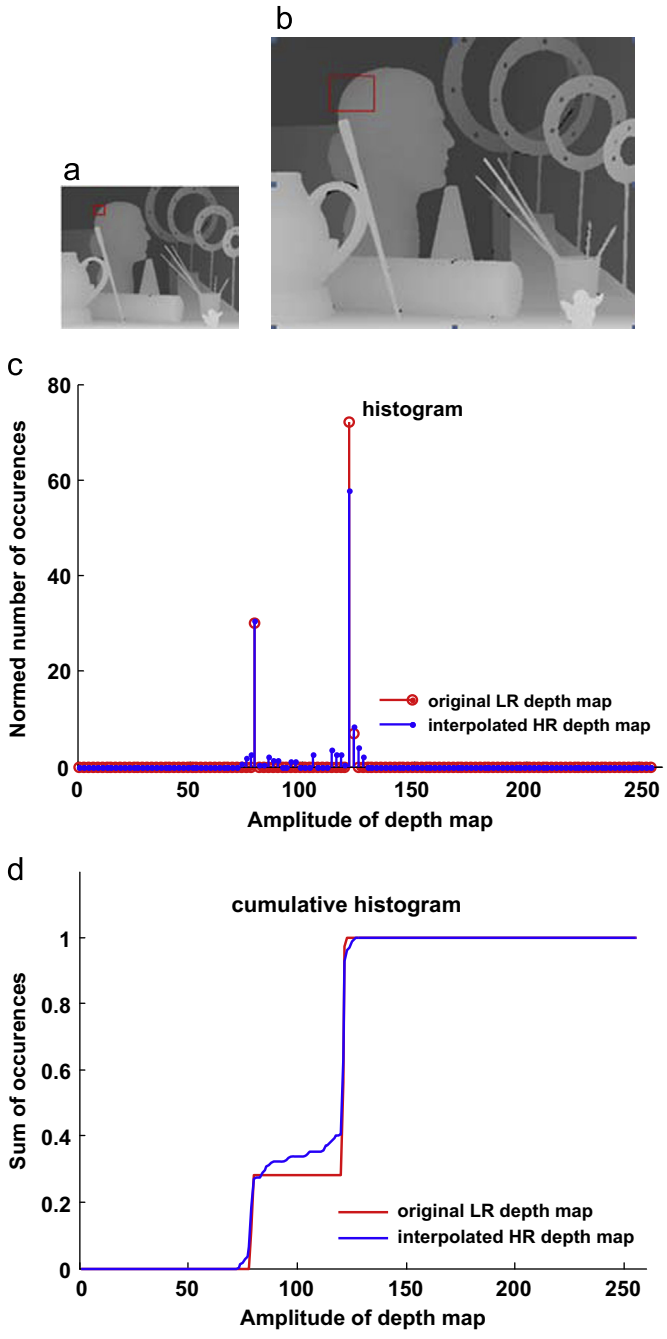

Fig. 3. Histogram and cumulative histograms of depth map. (a) The original LR depth map. (b) The up-sampled depth map. (c) Histogram of (a) and (b). (d) Cumulative histogram of (a) and (b).

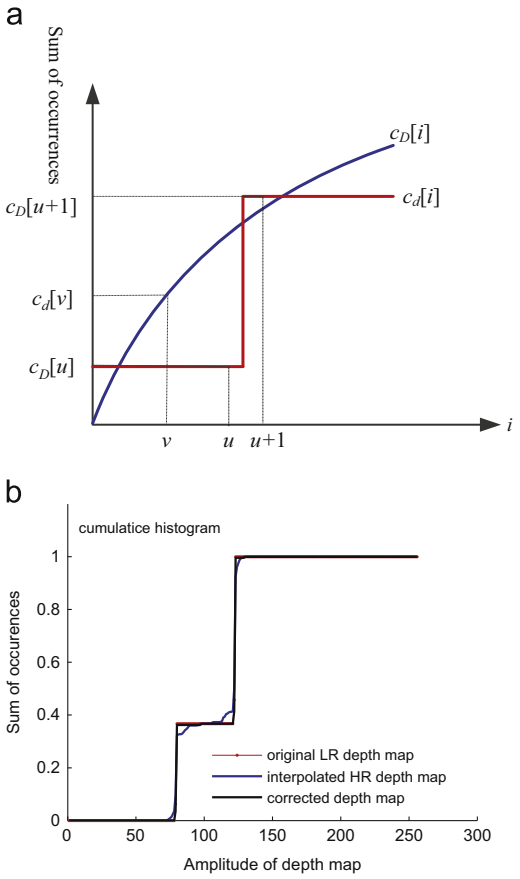

Fig. 4. Cumulative histogram analysis for depth map. (a) Cumulative histogram mapping and (b) Cumulative histogram of rectified depth map.

the variable p_n^t is the similarity between the current and center pixel values (i.e., I_n and I_c) in texture image. It is measured by the absolute difference $difT(n) = |I_n - I_c|$ and given by

$$p_n^t = \frac{\max\_difT - difT(n)}{\max\_difT - \min\_difT}, \tag{8}$$

where max_*difT* and min_*difT* indicate the maximum and minimum texture intensity differences within the window W, respectively. With the help of texture similarity, we can utilize the corresponding texture information to the edges in depth map if artifacts appear around edge regions.

With the weight model in Eq. (4), we can estimate $\mathbf{R}$ and $\mathbf{r}$ using (3). In our experiment, we set parameters $a = b = c = 1$. Consequently the interpolation coefficients $\mathbf{K} = [k_0, k_1, k_2, k_3]^{\mathbf{T}}$ needed in Eq. (1) can be obtained from (2) as

$$\mathbf{K} = \left(\sum_{n=1}^{w^2} p_n \mathbf{c}_n^T \mathbf{c}_n \right)^{-1} \sum_{n=1}^{w^2} p_n \mathbf{c}_n^T y_n. \tag{9}$$

2.2. Block-based histogram rectification

Up-sampled depth map suffers from fake edges crossing foreground and background. These undesired depth values will result in visual artifacts in DIBR [46]. In order to handle these fake edges, we propose a histogram matching-based method to rectify the processed depth map.

Histogram of an image represents the number of pixels inside a given region and the cumulative histogram indicates a probability distribution of the pixels lies below a certain value. Different from natural images, depth map smoothly varies inside objects but has sharp discontinuities on boundaries. Therefore, the cumulative histogram of original depth map is piecewise linear and smooth. However, the cumulative histogram of the up-sampled depth map is transitional for the new coming pixels caused by interpolation procedure. Fortunately, the probability distribution of pixel value for both LR and HR depth maps is approximately equal after a normalization step. In this case, the cumulative histograms for both LR and HR depth maps are approximately equal. Based on these characteristics, we can rectify the error pixels in up-sampled depth map by matching on cumulative histograms.

Histogram rectification can be categorized into global and local approaches [47]. A global histogram approach usually derives a single mapping function and applies it to all pixels in an entire image. A local approach derives and applies the mapping function for each pixel adaptively according to the characteristic of a local region. Because the interpolation-induced distortion in depth map only appears in local region such as objection boundaries, there is no obvious difference between the global histograms of LR depth map and its interpolated HR depth map. Furthermore, the block partition parameter in bit-stream is useful to distinguish depth boundaries and smooth regions, and thus we extract the parameter for the block-based histogram rectification. For each block in depth map, we rectify the error pixels in the up-sampled depth map by cumulative histogram matching of the distorted HR block to its original LR depth map.

We first extract the histogram of a block from the original LR depth map, and then count the number of pixels in the block according to their values:

$$h_d[v] = \frac{1}{m \cdot n} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \delta(v, d(x,y)) \tag{10}$$

with

$$\delta(v, d(x,y)) = \begin{cases} 1 & \text{if } v = d(x,y) \\ 0 & \text{else} \end{cases} \tag{11}$$

where m denotes the width and n is the height of the block.

Table 2
Images used in experiments.

Low resolution			High resolution		
Images	Color image resolution	Depth image resolution	Images	Color image resolution	Depth image resolution
ConesQ	450 × 375	225 × 187	ConesH	1800 × 1500	900 × 750
teddyQ	450 × 375	225 × 187	teddyH	1800 × 1500	900 × 750
ArtQ	463 × 370	231 × 185	ArtH	1390 × 1110	695 × 555
Venus	434 × 383	217 × 191	books	1390 × 1110	695 × 555

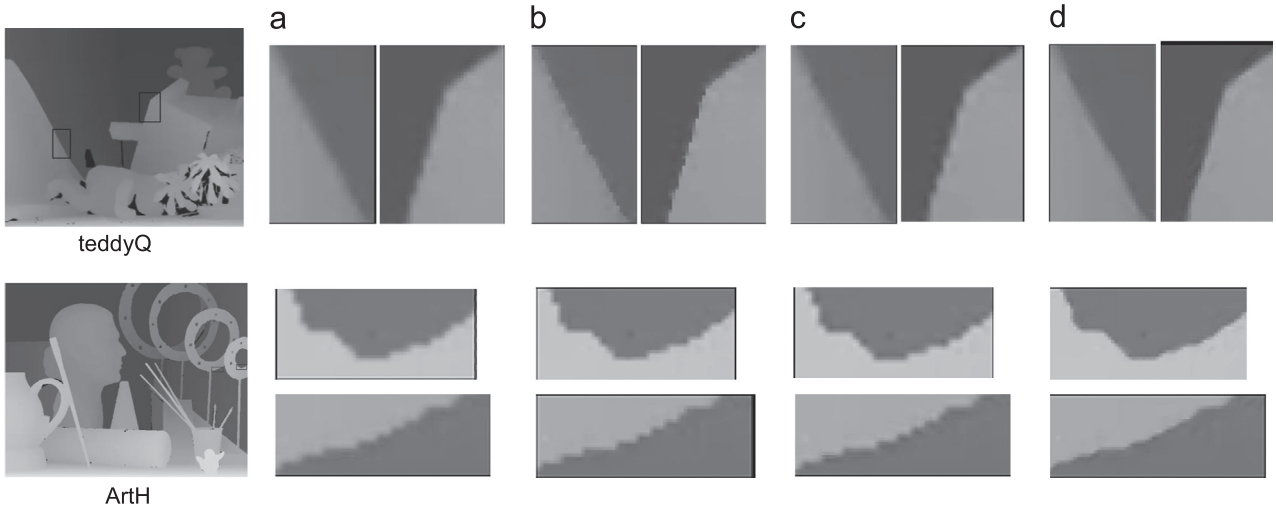

Fig. 5. Depth maps processed by different up-sampling methods. (a) Bilinear. (b) JBU. (c) NEDI and (d) EPDU.

The cumulative histogram of the original LR depth map is created by

$$c_d[v] = \sum_{i=0}^{v} h_d[v] \tag{12}$$

The histogram $h_D[v]$ and the cumulative histogram $c_D[v]$ of the up-sampled HR depth map are calculated in the same manner. Fig. 3 shows an example of histograms and cumulative histograms of an original LR image block and its up-sampled HR image block. As shown in Fig. 3(a), depth value smoothly varies inside objects and has sharp discontinuities on the object boundaries. In this case, the block crossing boundary contains both foreground and background depth values, and so the original LR histogram contains two peaks, as shown by Fig. 3(c). Correspondingly, the cumulative histogram of the original LR depth map is piecewise linear as shown in Fig. 3(d). When the depth value is interpolated, it results in fake edges crossing the foreground and background as shown in Fig. 3(b), and brings noise peaks as shown in Fig. 3(d). In this case, the cumulative histograms of the original LR is a desired shape. Therefore, we will utilize the cumulative histogram matching for depth map rectification.

Table 3
Comparative performance on PSNR for different methods.

Test images	Bilinear	JBU	NEDI	EPDU	PSNR gain (dB)
ConesQ	28.59	28.28	28.27	28.96	**0.69**
teddyQ	28.09	27.7	27.66	28.3	**0.64**
ArtQ	28.99	28.93	28.86	31	**2.14**
Venus	38.43	38.71	38.42	41.35	**2.93**
Average gain					**1.60**
ConesF	27.5	27.4	27.39	27.94	**0.55**
teddyF	27.77	27.65	27.66	28.07	**0.41**
ArtF	27.43	27.36	27.31	28.67	**1.36**
books	26.76	26.59	26.57	26.9	**0.33**
Average gain					**0.66**

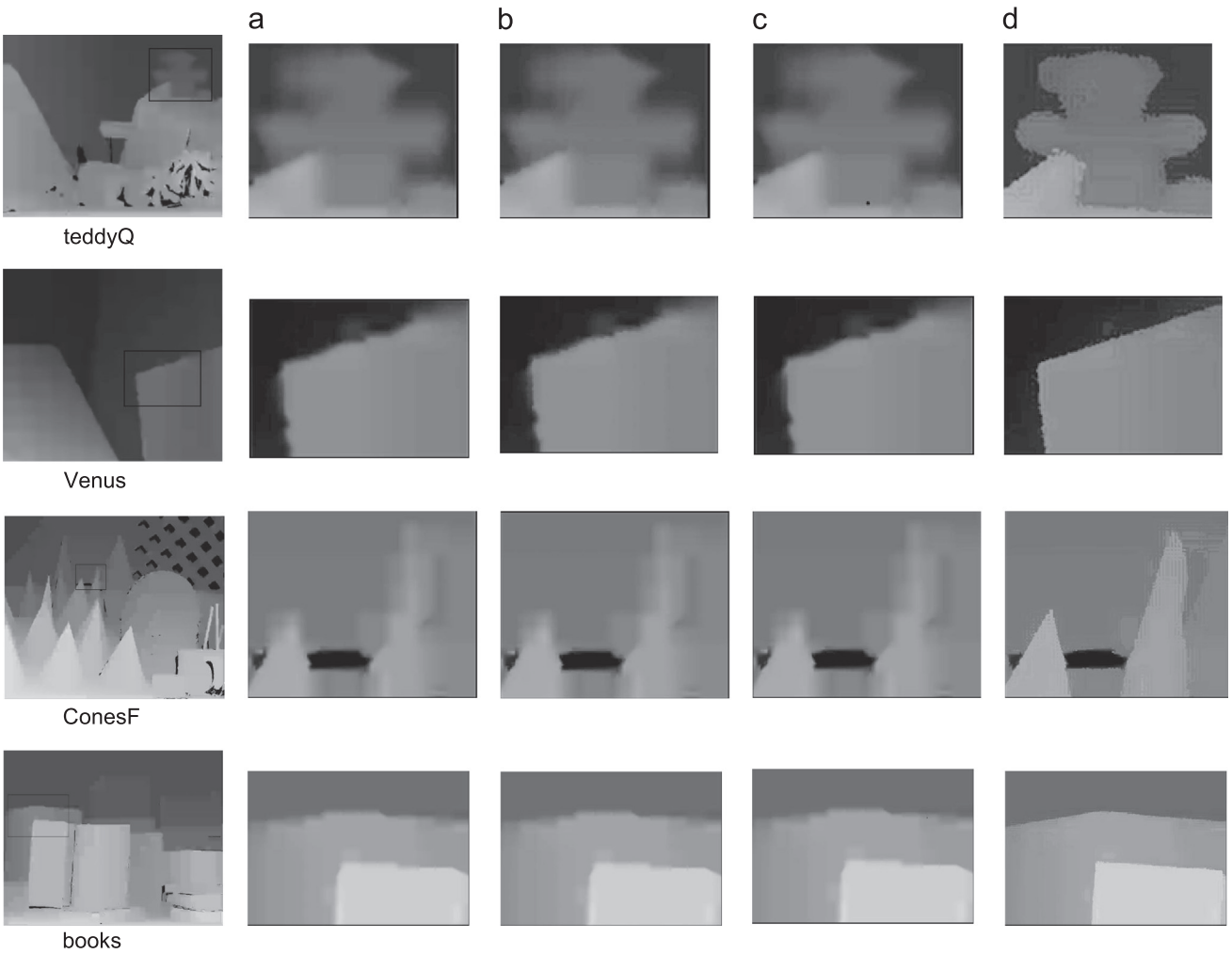

Fig. 6. Reconstructed depth maps with different methods. (a) Bilinear. (b) JBU. (c) NEDI and (d) EPDU.

Based on the cumulative histograms, we derive a mapping function for the cumulative histograms of the original and the up-sampled depth maps. Let $Map[v]$ denote the mapping function, as illustrated in Fig. 4(a), the mapping function can be formulated by (13)

$$Map[v]=u, \quad \text{with} \quad c_d[u]<c_D[v]<c_d[u+1] \tag{13}$$

The mapping is then applied to the up-sampled depth map for the rectified depth map $D_c(x,y)$:

$$D_c(x,y)=Map[D(x,y)] \tag{14}$$

The cumulative histogram of the rectified depth map is shown in Fig. 4(b). As can be found, the cumulative histogram of the rectified depth map is close to the one of the original LR depth map, and the pixel value in interval of [75, 125] in the interpolated depth map are now rectified to be 75.

3. Experimental results

We test the performance of the proposed method by the dataset of Middlebury [48]. For each depth map, we down-sampled it by factor of 2 and then treated the down-sampled depth map as the input LR depth map. Table 2 shows the detail information of those dataset. The input LR depth maps and HR color images are separately encoded by using JMVC8.5 as shown in Fig. 1. For each depth map, the quantization parameter (QP) in the encoder is set as 24, 32, 40 and 48.

3.1. Performance on edge preservation

In this part of experiments, we compare the performance of the proposed algorithm (i.e., 'EPDU' for short) with bi-linear, JBU [41], and NEDI [45] algorithms. Fig. 5 shows the comparisons. As can be found in Fig. 5(a), the results of bilinear are not distinguishable around object boundaries. After that, as Fig. 5(b) shows, the results of JBU suffer from noticeable dotted and jagged artifacts along edges. On the other hand, NEDI improves visual quality around the edges. As shown in Fig. 5(d), EPDU has better performance on object boundary preservation. The reason for EPDU has better performance comes from two-fold. First, EPDU employs NEDI as the basis of interpolation in the up-sampling procedure. It tunes the interpolation coefficients to match the arbitrary edge orientations. Therefore, EPDU has much better visual quality around object boundaries, especially when comparing to the benchmark algorithms bilinear and JBU. Second, EPDU takes advantage of the geometric similarity within depth maps as well as the photometric

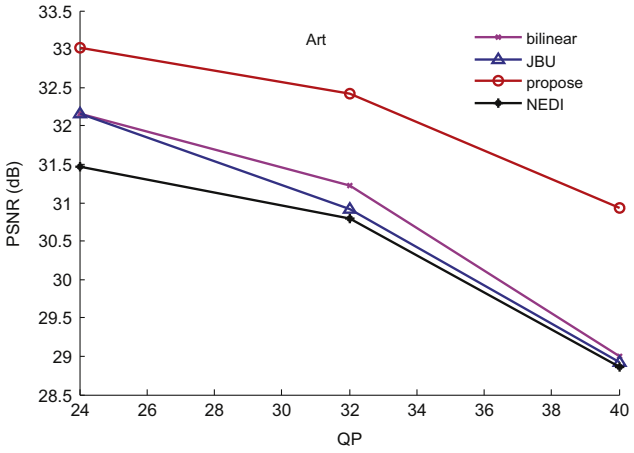

Fig. 8. RD curves of different up-sampling methods.

Table 4
Performance on PSNR of histogram rectification.

Test images	Bilinear	Bilinear+histogram	PSNR gain (dB)	JBU	JBU+histogram	PSNR gain (dB)
conesQ	28.59	29.43	**0.84**	28.27	28.89	**0.62**
Teddy	28.29	28.87	**0.58**	27.85	28.05	**0.20**
Art	28.99	31.27	**2.28**	28.93	30.94	**2.01**
Venus	38.43	41.26	**2.83**	38.71	40.98	**2.27**
Average gain			**1.63**	Average gain		**1.28**

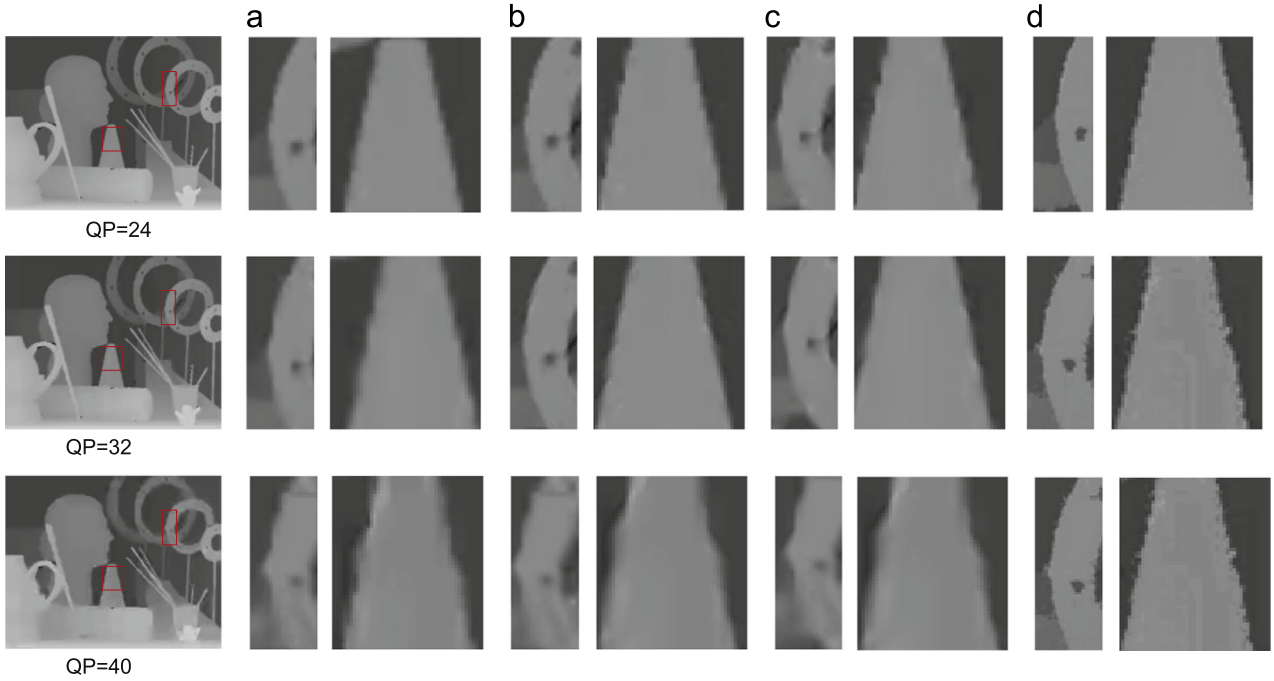

Fig. 7. Up-sampled depth maps with different QPs.

similarity between the depth map and the corresponding color image. The weighted model for each sample utilize the similarities, leading EPDU can be adaptive in selecting the weight according to the edge orientations. With this weight model, we can utilize the geometric and intensity similarities to improve the accuracy around depth boundaries.

3.2. Performance on edge rectification

After the experiment on edge preservation, we further evaluate the performance of the proposed block-based histogram rectification method. The objective and subjective performances are given by Table 3 and Fig. 6. Compared with NEDI algorithm, our proposed

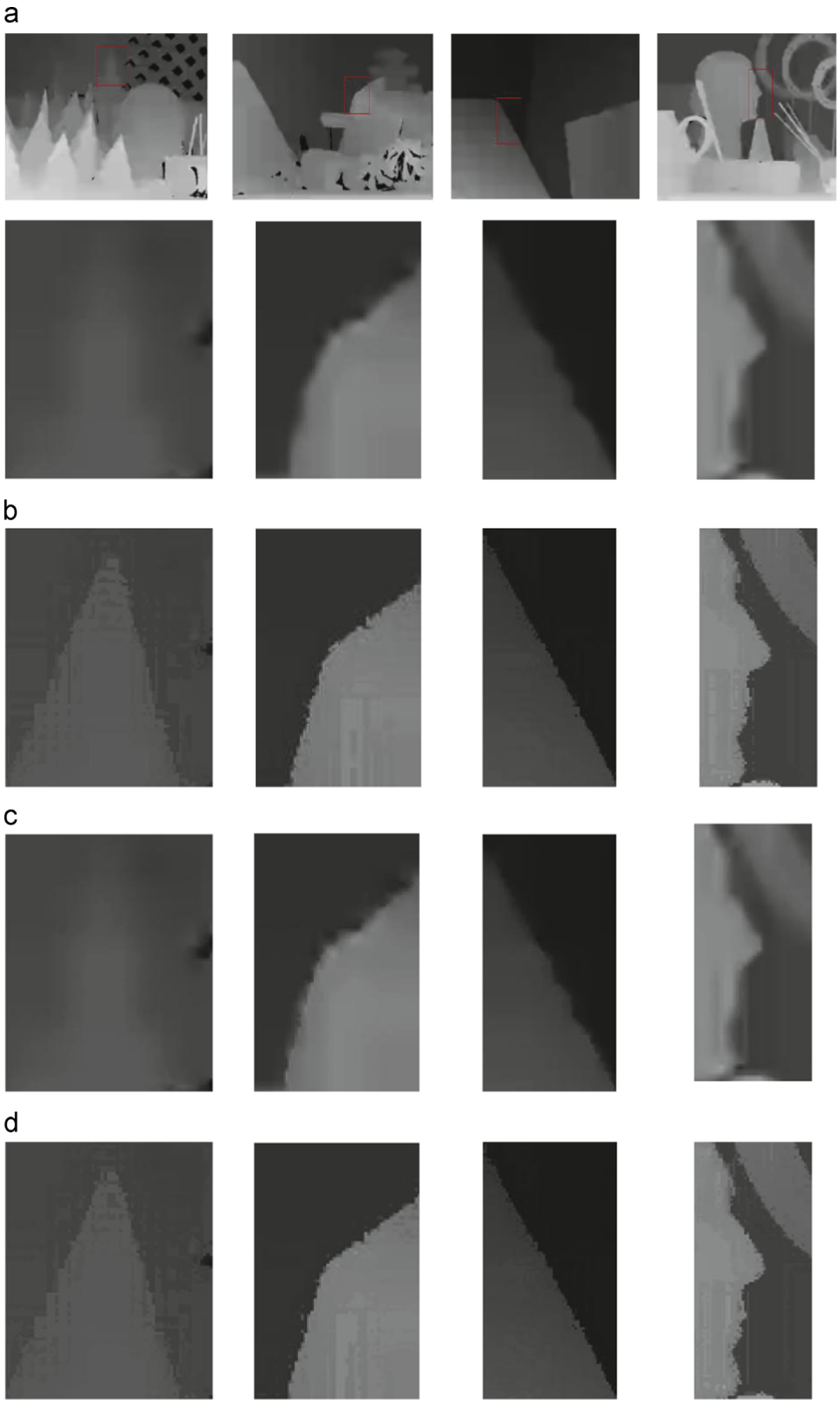

Fig. 9. Reconstructed depth maps with different methods. (a) Bilinear. (b) Bilinear+histogram rectification. (c) JBU and (d) JBU+histogram rectification.

method can achieve 1.60 dB average PSNR gain and up to 2.93 dB for LR, while 0.66 dB on average and up to 1.36 dB for HR. Besides that the perceptual quality is also improved by our method. Fig. 7 shows the visual performance of the reconstructed depth maps. The depth maps of bilinear, JBU and NEDI methods still suffer from blurring effect along object boundaries. However, our method effectively suppresses the blurring effects around object boundaries, and then it preserves the sharp edges well. The reason for this better performance is that our method rectifies the blurred regions around object boundaries by histogram matching. Generally, depth errors in an asymmetric coding scheme come from both depth map up-sampling and compression. Depth map up-sampling is unable to eliminate depth compression-induced distortions. In this case, our histogram matching method can rectify the compression-induced artifacts by adjusting the local histogram of the distortion block to an original reference one. If the local cumulative histogram of a corrected image is similar to the original one, the pixels in the local block is also rectified as close as the original pixels. Therefore, our proposed method can achieve better performance than the benchmarks.

3.3. Performance on QPs

Besides the previous experiments, we test the proposed method over three QPs (24, 32 and 40) for 'ArtQ' image to evaluate the effect of our method under different bit rates. As shown in Fig. 8, the objective quality decreases with increasing QP. At higher QP (i.e., lower bit rate), the objective quality of bilinear, JBU and NEDI methods is very closer and lower. However, we find that our method can achieve a relative higher objective quality even at lower bit rates. For subjective quality, as shown in Fig. 7, bilinear, JBU and NEDI methods suffer from serious edge blurring effect at QP 40. On the contrary, these artifacts are well handled by our method. On the other hand, our method can also preserve the depth boundaries when QP is 40, and the performance is also better than JBU and NEDI. The results suggest that our method can offer better performance even in the case of very low bit-rate. Actually, the level of noise and distortion in decoded image is proportional to the value of QP. In this case, the performance of bilinear, JBU and NEDI methods are affected by this fact, because the quality of up-sampled depth map mainly depends on the input. On the other hand, our method reconstructs the depth map by combining an edge preserving interpolation with a block-based histogram rectification method. The block-based histogram rectification method corrects the errors by matching cumulative histogram of the distorted block, and rectifies the blurred pixels to a correct one.

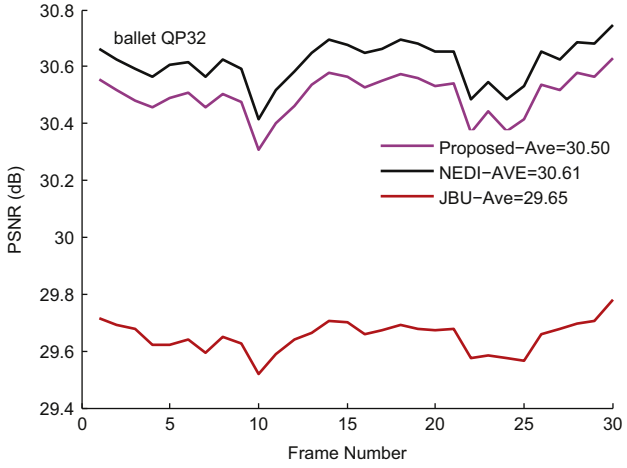

Fig. 10. PSNR performance on Ballet sequence.

3.4. Performance on joint processing

In fact, the proposed histogram rectification method can be combined with other up-sampling algorithms to enhance quality of up-sampled depth map. In this part of experiment, we combine our method by bilinear and JBU (i.e., 'bilinear+histogram' and 'JBU+histogram' for short). The results are listed in Table 4, and it can be found that up to 2.83 dB PSNR gain is obtained for 'Venus'. In general, the performance for 'bilinear+histogram' is superior to bilinear, and 'JBU+histogram' is superior to JBU. Fig. 9 also offers the subjective comparisons on different methods. The results suggest that the methods of 'bilinear+histogram' and 'JBU+histogram' can offer sharp object boundaries. As we analyzed above, up-sampling algorithms just suppress edge blurring in the interpolation procedure. Moreover, histogram rectification not only removes fake edges from up-sampling, but also corrects errors from depth map compression. Therefore, the proposed method is effective to improve the depth map quality in up-sampling.

3.5. Failure cases

Our proposed method can have better performance when the accuracy of original depth maps is well. For example, the depth maps provided by Middlebury are obtained by structured light, and these test images can fit our method better than other comparative algorithms. However, the performance drops down for our method when the accuracy of original depth maps is low, and the comparative algorithms can perform better for this case.

In this part, we test the depth maps provided by MPEG 3DV group. The obtained subjective and objective performances are shown in Figs. 10 and 11. We can find that our method does not have the best performance in this case, both on the PSNR in Fig. 10 and the visual quality in Fig. 11. The depth maps were originally obtained by manual operations on color video. After that, post-processing methods, such as lower-band and other filters, were utilized for the purpose of visual comfort in consequent virtual view rendering. Therefore, the provided depth maps are better for rendering and visual quality, although blurring effect on object boundaries are common in these depth maps. In this case, the

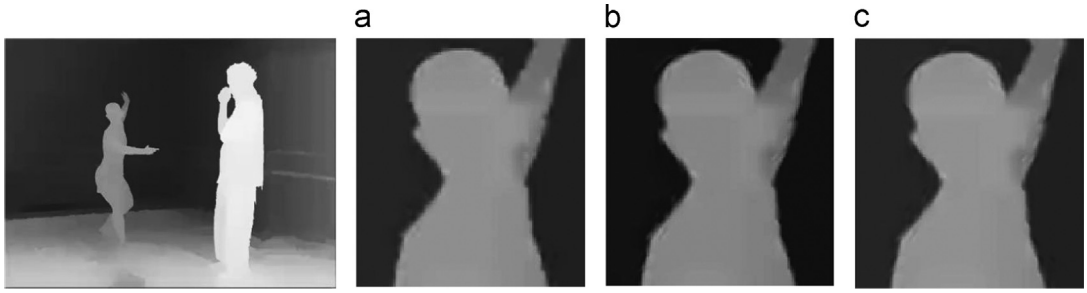

Fig. 11. Subjective performance on Ballet sequence.

performance of our method is slightly affected by the distortions in depth maps, and the PSNR lost is 0.11 dB when comparing to NEDI method. Although the performance is affected, our method can also have 0.85 dB gain than JBU method.

4. Conclusions

In this paper, we have presented a depth map reconstruction and rectification through coding parameters for mobile 3D video application. Different from the existing asymmetric coding and processing for down-sampled depth maps, our method up-sample and rectify the down-sampled depth map by using the encoding information in bit-stream. The proposed scheme is composed by depth edge preserving interpolation and block-based histogram rectification. For the edge preserving interpolation, we build a weighted model for each sample that incorporates geometric similarity as well as the intensity similarity in both the depth map and its corresponding color image, thus allowing adaptively select interpolation coefficients according to the edge orientation. Because the block information in depth map encoder is useful in distinguishing object boundaries and smooth regions, we extract block size parameters from bit-stream for the block-based histogram rectification. For each depth block, we rectify the blurred pixels in depth map by matching cumulative histogram. The experimental results suggest that the proposed method has better performance on edge preservation and noise removing when comparing to benchmark methods.

Acknowledgments

This work is supported in part by the National Science Foundation of China under Grant nos. 61170194, 61231010 and 61202301.

References

[1] A. Fernando, S.T. Worrall, et al., 3DTV: Processing and Transmission of 3D Video Signals, John Wiley & Sons, 2013.
[2] R. Ji, Y. Gao, R. Hong, Q. Liu, D. Tao, X. Li, Spectral-spatial constraint hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 52 (3) (2013) 1811–1824.
[3] Y. Sung, K. Cho, Development and evaluation of wireless 3d video conference system using decision tree and behavior network, EURASIP J. Wirel. Commun. Netw. 2012 (1) (2012) 1–14.
[4] M. Tekalp, et al., Peer-to-peer system design for adaptive 3d video streaming, IEEE Commun. Mag. 51 (5) (2013) 108–114.
[5] Y. Liu, S. Ci, H. Tang, Y. Ye, Application-adapted mobile 3d video coding and streaming—a survey, 3DR Rev. 3 (1) (2012) 1–6.
[6] S. Chaudhuri, V. Koltun, Data-driven suggestions for creativity support in 3d modeling, in: ACM Transactions on Graphics (TOG), vol. 29, ACM, 2010, p. 183.
[7] J. Liebelt, C. Schmid, Multi-view object class detection with a 3d geometric model, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010, pp. 1688–1695.
[8] Y. Gao, M. Wang, D. Tao, R. Ji, Q. Dai, 3-d object retrieval and recognition with hypergraph analysis, IEEE Trans. Image Process. 21 (9) (2012) 4290–4303.
[9] Y. Gao, J. Tang, R. Hong, S. Yan, Q. Dai, N. Zhang, T.-S. Chua, Camera constraint-free view-based 3-d object retrieval, IEEE Trans. Image Process. 21 (4) (2012) 2269–2281.
[10] Y. Gao, Q. Dai, N.-Y. Zhang, 3d model comparison using spatial structure circular descriptor, Pattern Recognit. 43 (3) (2010) 1142–1151.
[11] Y. Yang, Q. Liu, R. Ji, Y. Gao, Dynamic 3d scene depth reconstruction via optical flow field rectification, PloS One 7 (2012) e47041.
[12] C. Rougier, E. Auvinet, J. Rousseau, M. Mignotte, J. Meunier, Fall detection from depth map video sequences, in: Toward Useful Services for Elderly and People with Disabilities, Springer, 2011, pp. 121–128.
[13] Y. Gao, M. Wang, Z.-J. Zha, Q. Tian, Q. Dai, N. Zhang, Less is more: efficient 3-d object retrieval with query view selection, IEEE Trans. Multimed. 13 (5) (2011) 1007–1018.
[14] L. Zhang, Y. Gao, R. Zimmermann, Q. Tian, X. Li, Fusion of multichannel local and global structural cues for photo aesthetics evaluation, IEEE Trans. Image Process. 23 (3) (2014) 1419–1429.
[15] Y. Gao, M. Wang, R. Ji, X. Wu, Q. Dai, 3d object retrieval with Hausdorff distance learning, IEEE Trans. Ind. Electron. 61 (4) (2014) 2088–2098.
[16] Y. Gao, Q. Dai, View-based 3-d object retrieval: challenges and approaches, IEEE Multimed. Mag. (2014) 1.
[17] H. Deng, L. Yu, Z. Xiong, Edge-preserving interpolation for down/up sampling-based depth compression, in: IEEE International Conference on Image Processing, IEEE, 2012, pp. 1301–1304.
[18] C. Fehn, P. Kauff, S. Cho, H. Kwon, N. Hur, J. Kim, Asymmetric coding of stereoscopic video for transmission over T-DMB, in: 3DTV Conference, IEEE, 2007, pp. 1–4.
[19] G. Saygili, C.G. Gurler, A.M. Tekalp, Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3d video streaming, IEEE Trans. Broadcast. 57 (2) (2011) 593–601.
[20] H. Brust, A. Smolic, K. Mueller, G. Tech, T. Wiegand, Mixed resolution coding of stereoscopic video for mobile devices, in: 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, IEEE, 2009, pp. 1–4.
[21] Y. Yang, Q. Liu, Y. Gao, B. Xiong, L. Yu, H. Luan, R. Ji, Q. Tian, Stereotime: a wireless 2d and 3d switchable video communication system, in: Proceedings of the 21st ACM International Conference on Multimedia, ACM, 2013, pp. 473–474.
[22] P. Merkle, A. Smolic, K. Muller, T. Wiegand, Multi-view video plus depth representation and coding, in: IEEE International Conference on Image Processing. IEEE, 2007, pp. 201–204.
[23] H. Yuan, Y. Chang, J. Huo, F. Yang, Z. Lu, Model-based joint bit allocation between texture videos and depth maps for 3-d video coding, IEEE Trans. Circuits Syst. Video Technol. 21 (4) (2011) 485–497.
[24] Y. Liu, Q. Huang, S. Ma, D. Zhao, W. Gao, Joint video/depth rate allocation for 3d video coding based on view synthesis distortion model, Signal Process.: Image Commun. 24 (8) (2009) 666–681.
[25] P. Henry, M. Krainin, E. Herbst, X. Ren, D. Fox, RGB-D mapping: using kinect-style depth cameras for dense 3d modeling of indoor environments, Int. J. Robot. Res. 31 (5) (2012) 647–663.
[26] T. Leyvand, C. Meekhof, Y.-C. Wei, J. Sun, B. Guo, Kinect identity: technology and experience, Computer 44 (4) (2011) 94–96.
[27] J.-H. Cho, S.-Y. Kim, Y.-S. Ho, K. Lee, Dynamic 3d human actor generation method using a time-of-flight depth camera, IEEE Trans. Consum. Electron. 54 (4) (2008) 1514–1521.
[28] F. Shao, G. Jiang, M. Yu, K. Chen, Y.-S. Ho, Asymmetric coding of multi-view video plus depth based 3-d video for view rendering, IEEE Trans. Multimed. 14 (1) (2012) 157–167.
[29] L. Pinto, P. Assuncao, Asymmetric 3d video coding using regions of perceptual relevance, in: International Conference on 3D Imaging (IC3D), IEEE, 2012, pp. 1–6.
[30] Q. Liu, Y. Yang, R. Ji, Y. Gao, L. Yu, Cross-view down/up-sampling method for multiview depth video coding, IEEE Signal Process. Lett. 19 (5) (2012) 295–298.
[31] L. Zhang, M. Song, Q. Zhao, X. Liu, J. Bu, C. Chen, Probabilistic graphlet transfer for photo cropping, IEEE Trans. Image Process. 22 (2) (2013) 802–815.
[32] Y. Yang, Q. Liu, R. Ji, Y. Gao, Remote dynamic three-dimensional scene reconstruction, PloS One 8 (5) (2013) e55586.
[33] Y. Gao, M. Wang, Z.-J. Zha, J. Shen, X. Li, X. Wu, Visual-textual joint relevance learning for tag-based social image search, IEEE Trans. Image Process. 22 (1) (2013) 363–376.
[34] R. Ji, H. Yao, W. Liu, X. Sun, Q. Tian, Task-dependent visual-codebook compression, IEEE Trans. Image Process. 21 (4) (2012) 2282–2293.
[35] R. Ji, L.-Y. Duan, J. Chen, H. Yao, J. Yuan, Y. Rui, W. Gao, Location discriminative vocabulary coding for mobile landmark search, Int. J. Comput. Vis. 96 (3) (2012) 290–314.
[36] Q. Liu, Y. Yang, Y. Gao, R. Ji, L. Yu, A Bayesian framework for dense depth estimation based on spatial–temporal correlation, Neurocomputing 104 (2013) 1–9.
[37] D.-Y. Kim, S.-Y. Kim, Y.-S. Ho, Edge refinement of up-sampled depth images using joint bilateral filtering, in: International Conference on SMA, IEEE, 2012, pp. 1–2.
[38] X. Yan, Y. Yang, G. Er, Q. Dai, Depth map generation for 2d-to-3d conversion by limited user inputs and depth propagation, in: 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2011, IEEE, 2011, pp. 1–4.
[39] S. Jung, Enhancement of Image and Depth Map Using Adaptive Joint Trilateral Filter, 2013, pp. 258–269.
[40] Q. Yang, R. Yang, J. Davis, D. Nister, Spatial-depth super resolution for range images, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007, pp. 1–8.
[41] M.O. Wildeboer, T. Yendo, M.P. Tehrani, T. Fujii, M. Tanimoto, Color based depth up-sampling for depth compression, in: Picture Coding Symposium (PCS), IEEE, 2010, pp. 170–173.
[42] D. Min, J. Lu, M.N. Do, Depth video enhancement based on weighted mode filtering, IEEE Trans. Image Process. 21 (3) (2012) 1176–1190.
[43] D.T. Kuan, A.A. Sawchuk, T.C. Strand, P. Chavel, Adaptive noise smoothing filter for images with signal-dependent noise, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-7 (2) (1985) 165–177.
[44] J.-S. Lee, Digital image enhancement and noise filtering by use of local statistics, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-2 (2) (1980) 165–168.
[45] X. Li, M.T. Orchard, New edge-directed interpolation, IEEE Trans. Image Process. 10 (10) (2001) 1521–1527.
[46] C. Fehn, Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3d-tv, in: Electronic Imaging, International Society for Optics and Photonics, 2004, pp. 93–104.

[47] U. Fecker, M. Barkowsky, A. Kaup, Histogram-based prefiltering for luminance and chrominance compensation of multiview video, IEEE Trans. Circuits Syst. Video Technol. 18 (9) (2008) 1258–1267.
[48] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. J. Comput. Vis. 47 (1–3) (2002) 7–42.

You Yang received the B.S and M.S. degrees in Applied Mathematics from Chinese University of Mining and Technology in 2000 and 2003, respectively, and the Ph. D. degree in Computer Science and Technology from Institute of Computing Technology, Chinese of Academy of Sciences in 2009. He was a Research Fellow in the Department of Automation, Tshinghua University from 2009 to 2011, and now is a faculty in the Department of Electronics and Information Engineering, Huazhong University of Science and Technology. His research interests include digital multimedia computing, three-dimensional image modeling and representation, multi-view and free viewpoint video, 3-D television, and image quality assessment in above fields.

Huiping Deng received a B.S. degree in electronics and information engineering, an M.S. degree in communication and information system from Yangtze University, Jingzhou, China, in 2005 and 2008, respectively. She received the Ph.D. degree in the Electronics and Information Engineering Department, Huazhong University of Science and Technology in 2013, and now is a faculty in the School of Information Technology and Engineering of Wuhan University of Science and Technology. Her research interests are video coding and computer vision, currently focusing on three dimensional video (3DV).

Jin Wu received the B.S. degree in electronics and information engineering from Huazhong University of Science and Technology (HUST) in 1988. She received the M.S. degree from University of Science and Technology Beijing in 1997, and the Ph.D. degree from Huazhong University of Science and Technology in 2006. In 1997, she joined School of Information Science and Engineering of Wuhan University of Science and Technology, where she has a professor since 2005. Her current research interests include image processing and pattern recognition.

Li Yu received the B.S. degree in electronics and information engineering, the M.S. degree in communication and information system and the Ph.D. degree in electronics and information engineering, all from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1995, 1997, and 1999, respectively. In 2000, she joined the Electronics and Information Engineering Department, HUST, where she has a professor since 2005. She is a co-sponsor of China AVS standard special working group and working as the key member of China AVS standard special working group. Her team has applied more than 10 related patents and submitted 79 proposals to AVS standard organization. Her current research interests include multimedia communication and processing, computer network, wireless communication.