Resolution-independent Up-sampling for Depth Map Using Fractal Transforms

Meiqin Liu[12], Yao Zhao[12], Chunyu Lin[12], Huihui Bai[12] and Chao Yao[12]
[1] Institute of Information Science, Beijing Jiaotong University
[2] Beijing Key Laboratory of Advanced Information Science and Network Technology
Beijing, 100044 - China
[e-mail: {mqliu, yzhao, cylin, hhbai, 10112062}@bjtu.edu.cn]
*Corresponding author: Yao Zhao

Abstract

Due to the limitation of the bandwidth resource and capture resolution of depth cameras, low resolution depth maps should be up-sampled to high resolution so that they can correspond to their texture images. In this paper, a novel depth map up-sampling algorithm is proposed by exploiting the fractal internal self-referential feature. Fractal parameters which are extracted from a depth map, describe the internal self-referential feature of the depth map, do not introduce inherent scale and just retain the relational information of the depth map, i.e., fractal transforms provide a resolution-independent description for depth maps and could up-sample depth maps to an arbitrary high resolution. Then, an enhancement method is also proposed to further improve the performance of the up-sampled depth map. The experimental results demonstrate that better quality of synthesized views is achieved both on objective and subjective performance. Most important of all, arbitrary resolution depth maps can be obtained with the aid of the proposed scheme.

---

1. Introduction

Recently, three dimensional (3D) video coding has been a hot research topic which aims to enable a variety of display types and depth perception. To support the multi-view auto stereoscopic displays, Multi-View plus Depth (MVD) format is applied in the 3D video coding. MVD format includes multi-view texture videos and associated-per-pixel depth maps [1]. Different from the texture videos, depth maps which describe the distance between the camera and objects in the scene, are usually represented by 8-bit gray-scale values. Accordingly, depth maps generally have more homogeneous regions and sharp edges, compared with the corresponding texture images. In 3D video coding system, depth maps are not directly provided to viewers but are used to synthesize the virtual views by DIBR (Depth Image Based Rendering) algorithm [2]. Since depth maps have plenty of spatial redundancies, they can be encoded at a reduced resolution. In the 3D video standards such as 3D-AVC (3D-Advanced Video Coding), depth maps with half size of texture images [3] are adopted as default. At the end of that system, decoded depth maps need to be up-sampled to the full resolution. Moreover, the resolution of a current depth camera is relative low. For example, Mesa Imaging SR4000 has only QCIF (i.e., 176×144) resolution, while Kinect from Microsoft has relatively large resolution with 640×480. Compared with the texture images with the HD (i.e., 1920×1080) resolution, the up-sampling schemes are necessary so that each pixel in depth maps can correspond to that of texture images. Nevertheless, depth information is sensitive to the conventional up-sampling filters, especially for the sharp edges, while the edge information is very important in the synthesis process.

Depth up-sampling methods can be divided into two categories. One category is based on conventional interpolation methods such as nearest, bilinear and bicubic interpolation [3-4]. These interpolation methods are simple and fast, but they may suffer from artifacts around the edges which can cause a considerable impact on the synthesized view. Another category utilizes some auxiliary information to guide the up-sampling of depth maps, which targets at preserving the depth edges information [5-11]. In [5], a noise-aware upsampling filter is designed via the inherent noisy nature of real-time depth data and can help researchers to up-sample the low resolution 3D data to the image resolution of the video sensor. In [6], a depth reconstruction filter is proposed to recover object boundaries. In [7], an edge-preserving interpolation method is proposed for up-sampling-based depth coding, which uses edge similarity between the depth map and its corresponding texture image to suppress artifacts. Furthermore, the spatial and temporal correlation information from a depth and its associated texture video have been used to guide depth up-sampling in [8]. In [9], cross-view information is ulitized to assist the up-sampling at the decoder. In [10], up-sampling problem of depth map is formulated to a convex optimization problem by higher order regularization and guided by an isotropic diffusion tensor. In [11], using geodesic distances for up-sampling depth images, a joint filtering algorithm is proposed and approximated to achieve real-time performance. In [12], a joint adaptive bilateral filter is proposed to recover object boundaries and up-sample the low resolution depth map by checking whether the missing pixels belong to a common edge regions. In [13], a bundled-optimization scheme is proposed to process the complete chain from depth sensing to multiview dense depth map. The scheme first removes noises of low-resolution depth map and optimizes up-sampled depth map via Blocklet and clustering joint method. Despite the great advances in up-sampling for depth maps, some over-filters would cause misalignments between the depth maps and the texture images, which result in

some artifacts (such as geometry distortion) in the synthesized view and decrease the quality of virtual views. Meanwhile, these methods have seldom given full consideration to the geometrical features of depth maps and could only up-sample image with integer factor, which limits the up-sampling performance to some extent. Therefore, a high-quality and resolution-independent up-sampling algorithm for depth maps is needed.

As depth maps represent the distance instead of the intensity of the object, they are composed of smooth areas bounded by sharp edges [14-15]. These sharp edges are very significant to render virtual views with good quality and provide strong geometrical features. Fractal theory has been proposed as a geometrical tool to better capture geometrical properties and recover arbitrary resolution images [16]. It has played an important role in the seabed mapping [17], image compression [18], image synthesis and texture classification [19]. As an particular kind of image, depth map is self-referential and can be well represented by fractal theory. Self-referential feature means that a region in a depth map can find a similar match rigion within the same map [20]. In order to extract the fractal parameters which describe self-referential features, the proposed scheme first divides the depth map into small copies based on geometrical features of edges and then use fractal transform parameters to represent the relationship of these copies. In other words, the self-referential features of a depth map can be well described by fractal parameters. Due to the fractal *Contractive Mapping Theorem* and *Collage Theory*, the depth map can be approximately recovered by iteratively applying the fractal transforms to an arbitrary depth map [16, 21]. Meanwhile, these parameters represent only the relationship within a depth map and do not introduce inherent scale information. As a result, the fractal parameters are resolution-independent and can recover a depth map to arbitrary resolution. Therefore, we propose a depth up-sampling scheme which is based on fractal self-referential features and their resolution-independent characteristic. Experimental results show the proposed scheme can obtain both better objective and subjective quality of up-sampled depth maps and synthesized virtual views. Furthermore, due to the fractal resolution-independent features, the depth map can also be up-sampled fast to an arbitrary resolution image.

The rest of the paper is organized as follows. In Section 2, the framework of the proposed scheme is first given, and then the last component of the framework is detailed in the following two subsections. The performance of the proposed scheme is reported and compared with other up-sampling methods in Section 3.We conclude the paper in Section 4.

2. Proposed Scheme

Inspired by the recent work on fractal geometry theory, we apply the fractal self-referential features on the depth maps, and propose a novel resolution-independent up-sampling scheme for depth map. Due to the better edge preserving advantage of fractal, synthesized views with the up-sampled depth have better quality. In addition, the low resolution depth map could be up-sampled to any resolution.

2.1 Framework of the Proposed Scheme

Different from natural images, depth maps generally have more homogeneous regions and sharper edges than natural texture images. Thereby, there are more fractal self-referential features in depth maps. Motivated by this, some extracted fractal parameters from depth maps are used to describe self-referential features in our proposed approach. Here, such fractal parameters can also be used to represent the relationship among different scales of a depth map itself. With these parameters, an arbitrary resolution depth map can be reconstructed.

The details about the proposed framework are described in **Fig. 1**. In the conventional resolution-reduction video coding scheme [3], for the sake of reducing the bit cost on depth maps, an original depth map with full resolution is firstly down-sampled to low resolution. Then, standard video codec (such as H.264/AVC, JMVC) are separately used to compress and decompress the low-resolution depth maps. In the decoder side, the decoded low-resolution depth map is up-sampled to a high resolution. To reconstruct the up-sampled depth map with good performance, we apply the fractal theory to implement the up-sampling operator. Firstly, for the decoded depth map y, we extract its fractal parameters to describe the self-referential features, and then apply the fractal reconstruction method to re-generate the depth map. In this framework, a depth map y' which has the same size as y is reconstructed based on the extracted fractal parameters, and this process should be implemented prior to up-sampling. Then, the residual information between y and y' by calculating $y - y'$ can be obtained, which denoted as e. With resolution-independent advantage, the fractal parameters can be used to reconstruct the depth map with arbitrary size, and the up-sampled version is denoted as Y. Considering the possible distortion caused in the feature extracting process, the residual information will be used to enhance the up-sampled depth map. Notice that, the residual information e is up-sampled to the same size as Y by traditional interpolation methods.

Some details about the fractal feature extraction algorithm and resolution-independent based up-sampling algorithm will be provided as follows.
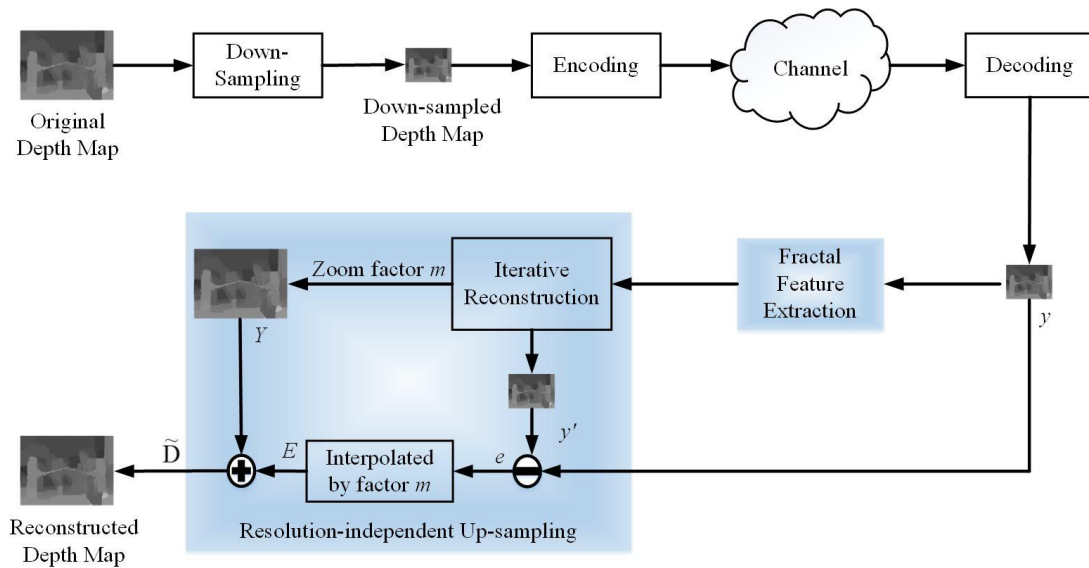

Fig. 1. Framework of the proposed scheme

2.2 Fractal Feature Extraction Algorithm

The fractal theory was proposed in 1988 by M.F. Barnsley, and it can effectively approximate a real world image via fractal parameters [16], which are also called Iterated Function System. Depth map is one kind of images with smooth regions and sharp edges, which contains fractal self-referential and resolution-independent features as well. Thereby a depth map can be represented by fractal parameters.

To take a concrete example, let a depth map x be one element of a complete metric space $(\mathbb{R}, d)$ (where d is given metric). The extracting of fractal self-referential feature is to

construct a contractive mapping $T:\mathbb{R} \to \mathbb{R}$ with contractive factor s, which satisfies the following Equations (1), (2) [22].

Given two arbitrary depth maps $x_1, x_2 \in \mathbb{R}$, if there exists $s \in [0,1)$, such that

$$d\big(T(x_1), T(x_2)\big) \leq s \cdot d(x_1, x_2) \tag{1}$$

Then T admits a unique fixed-point $x^* \in \mathbb{R}$,

$$x^* = T(x^*) \tag{2}$$

The fixed point theorem of contractive mapping can guarantee the existence and uniqueness of fixed point x^* for fractal transforms T, and x^* can be reached by iteration of T on any initial $x_o \in \mathbb{R}$.

Practical extraction process is ensured by the following equation,

$$\mathrm{d}(x, x^*) \leq \frac{1}{1-s} d(x, T(x)) \tag{3}$$

If the left side of the Equation (3) is minimized, the approximate fractal transforms T is extracted to represent the self-referential feature of the depth map x. In practical applications, the bound on right side $d(x, T(x))$ is minimized instead of $d(x, x^*)$ and therefore extracted fractal transforms T is an approximate representation of the depth map x_{ori}. i.e. $T(x) \approx x$.

(a) Depth map   (b) Square partition

(c) Triangle partition   (d) Delaunay partition
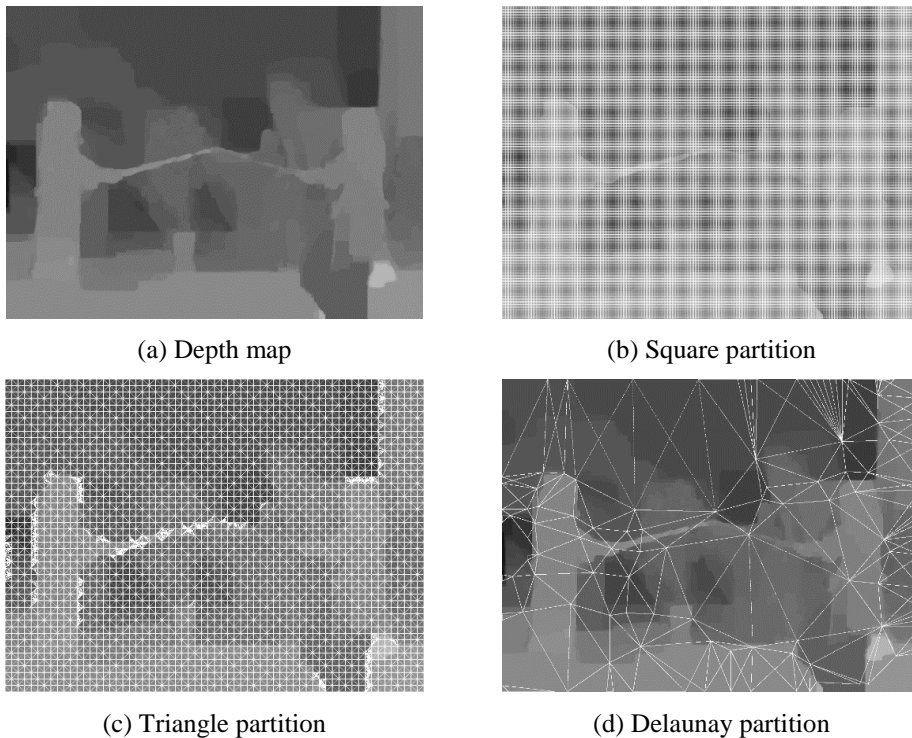
Fig. 2. Different partition methods of a depth map (e.g., "*Kendo*")

In practical application, constructing fractal transforms T is equivalent to the extracting of the union of a series of contractive transforms. i.e. $\bigcup_{i=1}^{N} T_i$, where T_i is the ith contractive

transform. Therefore, in fractal theory, the first step is the partition of the depth map. Different partition methods of a depth map (e.g., "*Kendo*") are shown in **Fig. 2**. Square partition is the simplest method, but it is not the most efficient one. Based on the complexity of depth maps, irregular partition methods, such as triangle partition, Delaunay partition and freely-shaped partition [23], can well represent the sharp edges and geometry features of a depth map and are very useful to improve the synthesized quality.

In order to extract the transforms parameters T, depth map x is divided into two classes of blocks, named range block R_i ($i = 1: N$, N is the number of range blocks) and domain block D_j($j = 1: M$, M is the number of domain blocks). The size of domain block is usually larger than that of range block to meet the contraction need of fractal fixed point theorem. All the range blocks constitute the whole depth map x and cannot be overlapped($i.e., R_i \cap R_k = \emptyset, \cup_{i=1}^{N} R_i = x$), while domain blocks can be overlapped ($i.e., D_j \cap D_l \neq \emptyset$).

After partition of depth map, extraction procedure of each T_i is separately shown in **Fig. 3**. The figure suggests that two blocks (labeled by red lines) in depth map x are similar to each other and self-referential features really exist in depth map.
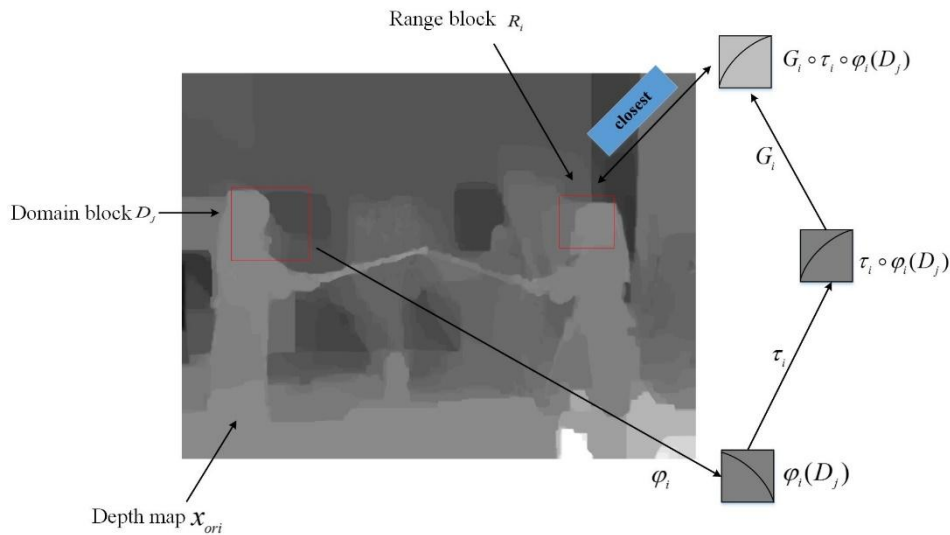

Fig. 3. Extraction process of fractal parameters

For each R_i, a contractive affine mapping T_i is searched in all domain blocks to find the closest D_j. Specifically, D_j is equivalent to sequential geometric transform G_i, isometric transform τ_i and luminance transform φ_i, i.e., $T_i = G_i \circ \tau_i \circ \varphi_i$. The next step is to find the minimization between the range block R_i and the transformed domain block $G_i \circ \tau_i \circ \varphi_i(D_j)$ in all possible G_i, τ_i, φ_i and all domain blocks to make D_j be similar enough to R_i, i.e. $\min_{G_i,\tau_i,\varphi_i,D_j} d(R_i - G_i \circ \tau_i \circ \varphi_i(D_j))$ [18]. Similar matching process is done for all the range blocks to get other T_is. Thus all the fractal feature parameters $T = \cup_{i=1}^{N} T_i$, such as mappings (G_i, τ_i, φ_i) and positions of the matching domain blocks, are extracted. The details of the proposed fractal feature extraction can be found in **Algorithm 1**.

Algorithm 1: Extraction algorithm of fractal self-referential parameter T_i

Input: depth map x.
1. Partition x_{ori} into range blocks $R_i\left(R_i \cap R_k = \emptyset, \cup_{i=1}^{N} R_i = x\right)$and domain blocks $D_j\left(D_j \cap D_l \neq \emptyset\right)$.
2. For $i = 1: N$ (N is the number of range blocks)
     For $j = 1: M$(M is the number of domain blocks)
   $D_j \xrightarrow{Geometric\ transform} \varphi_i(D_j) \xrightarrow{Isometry} \tau_i \circ \varphi_i(D_j) \xrightarrow{Luminance\ transform} G_i \circ \tau_i \circ \varphi_i(D_j)$
3. Compute and search the minimum difference between the range block and the transformed domain block $\min_{G_i, \tau_i, \varphi_i, D_j} d(R_i - G_i \circ \tau_i \circ \varphi_i(D_j))$.

Output: extracted fractal parameter T_i (G_i, τ_i, φ_i and the position of $D_j, i = 1: N$).

2.3 Resolution-independent Up-sampling Algorithm

2.3.1 Iterative Reconstruction Algorithm

This section is devoted to the reconstruction of a depth map from T. From the above section, extracted fractal parameters T_i consist of (G_i, τ_i, φ_i) $(i = 1: N)$ and the positions of the matching domain block. T_i s clearly represent the mapping relationship between two types of blocks in the depth map itself and are irrelevant to the resolution of the depth map, i.e., the fractal parameters T do not introduce inherent scale. **Algorithm 2** represents the iterative reconstruction method.

Algorithm 2: Iterative reconstrution algorithm

Input: An arbitary depth map x_0, extracted fractal parameters (G_i, τ_i, φ_i and the position of $D_j, i = 1: N$), zoom factor m.
1. According to zoom factor n, initialize the size of x_0 and positions of range block and domain block (detailed in **Fig. 4**).
2. For the number of iterations $k = 1: n$ (n is the maximum of iterative process), compute
$$x_k = \bigcup_{i=1}^{N} G_i \circ \tau_i \circ \varphi_i (x_{k-1})$$
3. Compute the difference d between x_k and x_{k+1}, If $d(x_k, x_{k+1}) < Tol$ (Tol is a predefined amount), end the loop. Else go to the step 2.
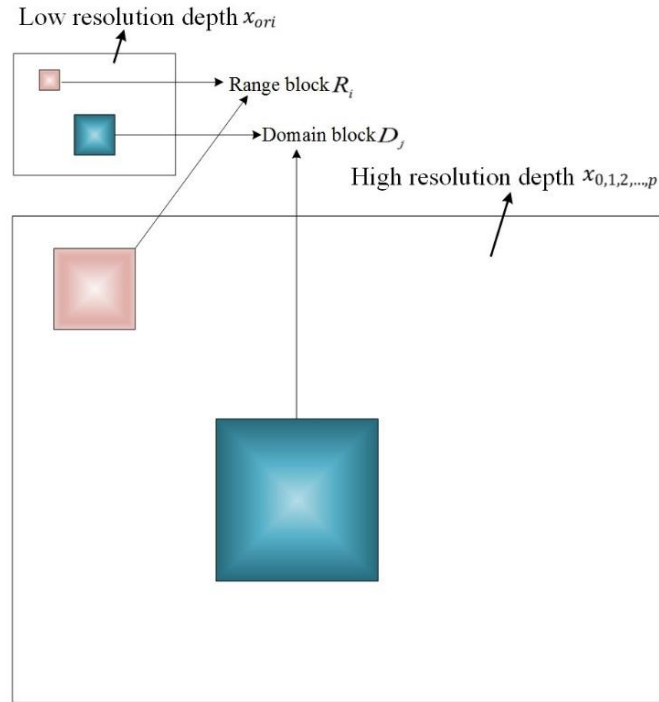4. End

Output: x_n(the needed high resolution depth maps).

Fig. 4. Zoom process of fractal parameters (*m*=4)

Based on Equation (3), reconstructed depth map x^* which is as close as the original depth map x, can be found by starting with an arbitrary depth map $x_0 \in \mathbb{R}$ and defining a sequence $\{x_n\}$

$$x_n = T(x_{n-1}) \tag{4}$$

Then

$$\lim_{n\to\infty} x_n = x^* \tag{5}$$

where n is iteration number. This means that the reconstructed depth map is the fixed point x^* of fractal parameters T and can be achieved by an iteration process. Specifically, as extracted parameters T represent the relationship between range block R_i and domain block D_j, the size of initial depth map x_0 and the positions of range block and its domain block zoom in m times in order to up-sample the low resolution depth x to high resolution depth x_n. The intensity of x_0 can be arbitrary value. **Fig. 4** shows the process of zooming in fractal parameters. The high resolution depth map x_n, which is the fixed point x^* of T, can be recovered by repeatedly applying T to x_0. For the first iteration, i.e. n=1, $x_1 = T(x_0)$ is obtained. Then x_1 is the input of the second iteration, i.e. n=2, $x_2 = T(x_1)$. The iteration process doesn't come to end until the difference between x_k and x_{k+1} differs by less than a predefined amount (i.e. $d(x_k, x_{k+1}) < Tol$, Tol is a predefined amount). Or else, the procedure continues to next iterative reconstruction.

2.3.2 Enhancement with residual information

It should be noted that the extraction process is a procedure that searches the minimum distortion between the reference block and the transformed domain block. Since it cannot

promise that R_i is definitely equal to $G_i \circ \tau_i \circ \varphi_i(D_j)$, some distortion will be introduced in the following reconstruction process. To avoid this extra distortion, residual information is recorded by calculating the difference between the reference image and the reconstructed image. In our approach, an enhancement processing is added as shown in **Fig. 1**. Using **Algorithm 2**, a reconstructed image y' which has the same size with y can be generated. The residual information can be calculated by

$$e = y - y' \tag{6}$$

Here, e can be used to enhance the reconstructed image at an arbitrary resolution. Since the residual information has the same size with the decoded depth map y, it should be up-sampled by the traditional interpolation methods to the corresponding resolution, when the fractal features are applied to up-sample depth maps.

At last, by iteratively reconstructing the high resolution depth map, we can get the depth map Y. The residual information e is up-sampled to the same resolution with Y, denoted as E The final reconstructed depth map $\widetilde{D}$ can be enhanced by adding the interpolating residual information E, as

$$\widetilde{D} = Y + E \tag{7}$$

3. Experimental Results and Analysis

In this section, the extensive experiments have been carried out to evaluate the proposed up-sampling scheme. Four standard 3D test sequences with the resolution 1024×768, namely, *Book-Arrival*, *Balloons*, *Newspaper* and *Kendo* and one *Dancer* sequence with 1920×1088 have been selected in our experiments. For *Book-Arrival*, view 8 and view 10 are selected as references and view 9 is set as the virtual view. For *Balloons* and *Kendo*, view 1 and view 3 are selected as references and view 2 as the virtual view. For *Newspaper*, view 2 and view 4 are selected as references and view 3 as the virtual view. And for *Dancer*, view 1 and view 5 are selected as references and view 3 as the virtual view. Virtual view is synthesized by View Synthesis Reference Software (VSRS 3.5) [24]**.** Here, 50 frames of each view are selected for experiments.

Each sequence is down-sampled by three zoom factors as 2, 4 and 8 in the horizontal and vertical direction, respectively. Here, the traditional up-sampling methods: Nearest (abbr. N), Bilinear (abbr. L) and Bicubic (abbr. C) are used as anchors to evaluate the proposed method (abbr. F). Furthermore, we adopt the same experimental configuration as that of the two benchmark algorithms, denoted as "JEDU" [8] and "CDU" [9]. The detailed results are shown as follows.

3.1 Up-sampling with zoom factor 2

When zoom factor $m = 2$, the proposed scheme is compared with "JEDU" and "CDU" separately using *Book-Arrival* and *Newspaper*. Similar as [8], for *newspaper* sequence, view 4 and view 6 are selected as references and view 5 as virtual view. H.264/AVC reference software is used as the codec. Depth maps are encoded with three different bit-rates 150kbps, 250kbps, 350kbps for *Book-Arrival*, while 400 kbps, 600 kbps, 800 kbps for *newspaper*. In the decoder side, the decoded depth maps are up-sampled by "JEDU" and the proposed scheme,

after which the up-sampled depth maps and corresponding textures are used to synthesize the virtual view by VSRS 3.5. **Fig. 5** shows the R-D performance for *Book-Arrival* and *Newspaper* sequences, respectively. The Peak-Signal-to-Noise-Ratio (PSNR) is used to evaluate the synthesized performance of both methods. Compared with "JEDU", there are about 0.6dB-0.82dB gain for *Book-Arrival* and 0.2-0.6dB gain for *Newspaper*, respectively.

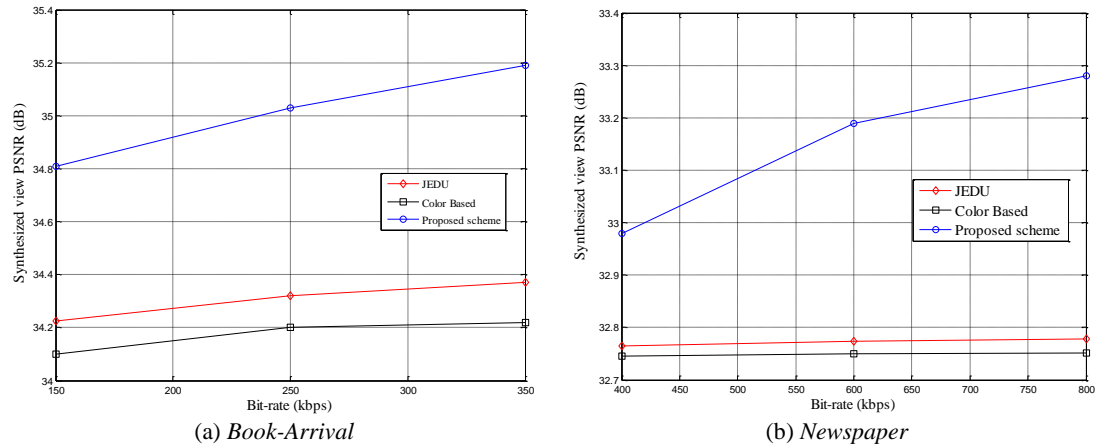(a) *Book-Arrival*   (b) *Newspaper*
Fig. 5. R-D curves for synthesised view: (a) *Book-Arrival*, (b) *Newspaper* ($m = 2$)

Furthermore, sequences *Dancer* and *Balloons* are selected to evaluate the performance of "CDU" and the proposed scheme. JMVC 6.0 [25] is adopted as the codec. Four quantization parameters (QP) 32, 36, 38 and 42 are set in the encoder. The delta QP is set as zero in all layers. **Fig. 6** shows the R-D performance of the synthesized view. For *Dancer*, the synthesized quality of the proposed scheme is 0.03dB-0.15dB larger than that of CDU. For *Balloons* the quality of synthesized virtual view is better than that of CDU, up to 0.2dB.

(a) Dancer   *(b) Balloons*
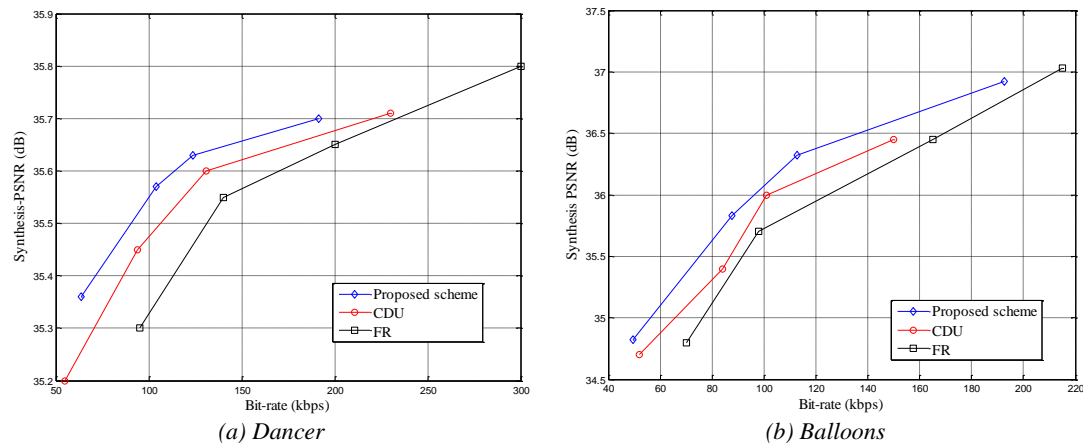**Fig. 6.** Performance of synthesized view: (a) *Dancer* (b) *Balloons* ($m = 2$)

3.2 Up-sampling with zoom factor 4

When $m = 4$, the proposed scheme is compared with classical up-sampling methods. Here, four test sequences, *Book-Arrival*, *Balloons*, *Newspaper* and *Kendo*, are firstly down-sampled by zoom factor 4, then encoded and decoded by H.264/AVC reference software. Four QPs (26,

31, 36 and 41) have been used. The objective results of the synthesized views have been shown in **Table 1**.

   Table 1 has shown objective quality (measured by PSNR and percentage of average PSNR increment) of synthesized views. The results indicate the good performance of the proposed scheme, for example, the proposed scheme achieved over 4% gain than the classical nearest interpolation method for all QPs on the sequence *Balloons*. Meanwhile, all subjective quality of synthesized views using the proposed scheme can be better than that of the traditional interpolation methods, as shown in **Fig. 7** and **Fig. 8**. Here, **Fig. 7** is the subjective evaluation of up-sampled depth maps for all the methods. **Fig. 8** is the subjective measurement of corresponding virtual views. From these figures, it can be found that the proposed scheme could keep much sharper edges (labeled by red circle) than other compared methods.

Table 1. Quality of synthesized views ($m = 4$) (dB)

Sequences	QP	Virtual View								
		N_N	N_F	PAPI	L_L	L_F	PAPI	C_C	C_F	PAPI
Kendo	26	37.16	37.61	**1.21%**	38.17	38.76	**1.55%**	38.27	38.82	**1.44%**
	31	36.36	36.82	**1.27%**	36.72	37.36	**1.74%**	37.23	37.85	**1.67%**
	36	34.56	35.13	**1.65%**	34.86	35.53	**1.92%**	35.11	36.02	**2.59%**
	41	33.83	34.42	**1.74%**	33.86	34.66	**2.36%**	34.04	35.11	**3.14%**
Balloons	26	36.78	38.63	**5.03%**	37.28	39.06	**4.77%**	37.32	39.65	**6.24%**
	31	35.01	37.1	**5.97%**	35.41	37.23	**5.14%**	35.72	37.57	**5.18%**
	36	33.64	35.12	**4.40%**	33.79	35.36	**4.65%**	34.2	35.97	**5.18%**
	41	32.24	33.58	**4.16%**	32.89	33.64	**2.28%**	32.95	33.72	**2.34%**
Newspaper	26	35.22	35.58	**1.02%**	35.5	35.79	**0.82%**	35.65	35.93	**0.79%**
	31	34.01	34.93	**2.71%**	34.18	35.06	**2.57%**	34.38	35.35	**2.82%**
	36	33.35	33.65	**0.90%**	33.39	33.78	**1.17%**	33.42	33.86	**1.32%**
	41	32.65	32.73	**0.25%**	32.82	32.98	**0.49%**	32.89	33.14	**0.76%**
Book Arrival	26	34.02	35.12	**3.23%**	34.67	35.45	**2.25%**	34.76	35.77	**2.91%**
	31	33.1	34.05	**2.87%**	33.25	34.13	**2.65%**	33.41	34.73	**3.95%**
	36	32.35	33.16	**2.50%**	32.39	33.41	**3.15%**	32.43	33.72	**3.98%**
	41	31.02	32.03	**3.26%**	32.12	33.22	**3.42%**	32.2	33.52	**4.10%**

Note:
N_N suggests down-sampling and up-sampling methods be both nearest interpolation.
N_F suggests down-sampling be nearest interpolation and up-sampling method be proposed method.
L_L suggests down-sampling and up-sampling methods be both bilinear interpolation.
L_F suggests down-sampling bilinear interpolation and up-sampling method be proposed method.
C_C suggests down-sampling and up-sampling methods be both bicubic interpolation.
C_F suggests down-sampling bicubic interpolation and up-sampling method be proposed method.
PAPI is abbr. of Percentage of Average PSNR Increment

(a) Nearest up-sampled scheme   (b) Bilinear up-sampled scheme

(c) Bicubic up-sampled scheme   (d) The proposed up-sampled scheme

Fig. 7. Subjective quality of up-sampled decoded depth maps by four methods ($m = 4$)

(a) Nearest up-sampled scheme

(b) Bilinear up-sampled scheme

(c) Bicubic up-sampled scheme

(d) The proposed up-sampled scheme

Fig. 8. Subjective quality of corresponding virtual views by four methods ($m = 4$)

3.3 Up-sampling with zoom factor 8

When $m = 8$, the similar experiments as $m = 4$ are executed to evalute the performance of the proposed scheme . Experimental results are shown seperately in **Table 2**, **Fig. 9** and **Fig. 10**. **Table 2** is objective quality comparison of four up-sampling schemes for synthesized view. It shows that when the scale improved, our proposed approach can preserve the up-sampling performance to some extent. Therefore, more gains are obtained. **Fig. 9** and **Fig. 10** show subjective performance of four up-sampling methods.
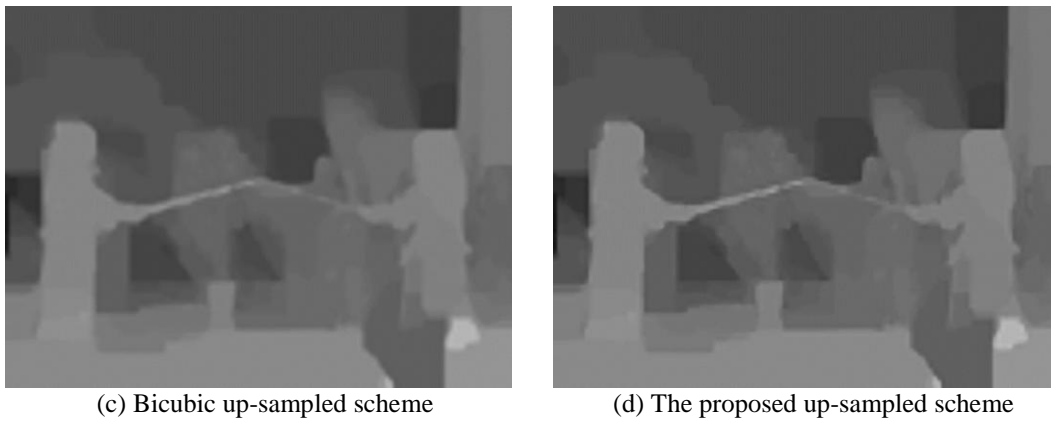
Table 2. Quality of Synthesized views ($m = 8$) (dB)

Sequences	QP	Virtual View								
		N_N	N_F	PAPI	L_L	L_F	PAPI	C_C	C_F	PAPI
Kendo	26	34.43	36.32	**5.49%**	37.37	38.77	**3.75%**	37.71	38.86	**3.05%**
	31	33.85	35.39	**4.55%**	35.67	37.32	**4.63%**	36.33	37.92	**4.38%**
	36	32.23	34.43	**6.83%**	34.2	35.58	**4.04%**	34.71	36.02	**3.77%**
	41	31.09	33.74	**8.52%**	33.66	35.01	**4.01%**	33.84	35.09	**3.69%**
Balloons	26	33.98	37.46	**10.24%**	35.15	37.65	**7.11%**	35.52	38.02	**7.04%**
	31	32.41	35.84	**10.58%**	33.56	36.16	**7.75%**	33.72	36.46	**8.13%**
	36	31.34	34.53	**10.18%**	31.9	34.52	**8.21%**	32.22	34.96	**8.50%**
	41	30.3	33.25	**9.74%**	30.98	33.64	**8.59%**	31.15	33.73	**8.28%**
Newspaper	26	33.82	34.78	**2.84%**	33.25	35.18	**5.80%**	34.02	35.58	**4.59%**
	31	33.27	34.38	**3.34%**	32.87	34.59	**5.23%**	33.38	34.76	**4.13%**
	36	32.76	33.82	**3.24%**	32.1	33.98	**5.86%**	32.62	34.06	**4.41%**
	41	31.95	33.19	**3.88%**	31.99	33.36	**4.28%**	32.13	33.52	**4.33%**
Book Arrival	26	33.28	34.31	**3.09%**	33.78	34.89	**3.29%**	33.92	34.96	**3.07%**
	31	32.62	33.69	**3.28%**	32.25	33.79	**4.78%**	32.4	33.85	**4.48%**
	36	31.87	33.01	**3.58%**	32.09	33.23	**3.55%**	32.11	33.35	**3.86%**
	41	30.74	31.85	**3.61%**	31.21	32.54	**4.26%**	31.92	32.84	**2.88%**

(a) Nearest up-sampled scheme   (b) Bilinear up-sampled scheme

(c) Bicubic up-sampled scheme   (d) The proposed up-sampled scheme

Fig. 9. Subjective quality of up-sampled by four methods for decoded depth maps ($m = 8$)

(a) Nearest up-sampled scheme   (b) Bilinear up-sampled scheme

(c) Bicubic up-sampled scheme   (d) The proposed up-sampled scheme

Fig. 10. Subjective quality of synthesized views from four up-sampled methods ($m = 8$)

During the extraction of fractal parameters, it will cost some time on finding the best matching domain block for each range block and increase the up-sampling time. **Table 3** shows the comparison time of four up-sampling methods and the proposed scheme for a depth map is 0.11-0.21s on average slower than other methods. As a matter of fact, the extraction process of fractal parameters costs much time on extraction of the fractal parameters. However, as depth maps generally have much homogeneous regions and sharp edges, it can help shorten the extraction process. Meanwhile, different high resolution depth maps can be reconstructed

with the once extraction of fractal parameters and fast iteration with the aid of advantages of fractal theory, such as fast iterative reconstruction. Therefore, up-sampling time of the proposed scheme can be further shorten.

Table 3. Comparison time of up-sampling schemes

| Sequences	Up-sampling time for depth map (s)			
	Nearest	Bilinear	Bicubic	proposed
Kendo	0.14	0.18	0.23	0.35
Balloons	0.16	0.21	0.25	0.36
Newspaper	0.18	0.22	0.29	0.41
Book-Arrival	0.19	0.24	0.31	0.39

4. Conclusion

In this paper, a novel up-sampling scheme for depth maps has been proposed based on the fractal self-referential and resolution-independent features. We first give the framework of the proposed scheme, and then detail the processes of the parameters extraction and resolution-independent up-sampling. Experimental results show that the proposed up-sampling scheme provides good performance on the synthesized virtual views with different zoom factors. Moreover, experiments of influence on encoder and up-sampled time of depth map are further added to verify the efficiency of our scheme.

References

[1] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proc. of IEEE International Conference on Image Processing(ICIP)*, pp. 201-204, September 16-19, 2007. Article (CrossRef Link)

[2] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Proc. of SPIE 5291, Stereoscopic Displays and Virtual Reality Systems XI*, pp. 93-104, May 21, 2004. Article (CrossRef Link)

[3] 3D-AVC Test Model 5, *Document JCT3V-C1003, ITU-TSG16WP3 and ISO/IEC JTC1/SC29/WG11*, January, 2013. Article (CrossRef Link)

[4] E. Ekmekcioglu, S. T. Worrall, and A. M. Kondoz, "Bit-rate adaptive downsampling for the coding of multi-view video with depth information," *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D video*, pp.137-140, May 28-30, 2008. Article (CrossRef Link)

[5] D.Chan, H.Buisman, C.Theobalt, S.Thrun, "A noise-aware filter for real-time depth upsampling," in *Proc. of Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SFA2 2008*, 2008. Article (CrossRef Link)

[6] KJ. Oh, S. Yea, A. Vetro, and YS. Ho, "Depth reconstruction filter and down/up sampling for depth coding in 3-D video," *IEEE Signal Processing Letters*, vol.16, no.9, pp.747-750, September, 2009. Article (CrossRef Link)

[7] H. Deng, L. Yu, and Z. Xiong, "Edge-preserving interpolation for down/up sampling-based depth compression," in *Proc. of IEEE Conference on Image Processing (ICIP)*, pp.1301-1304, September 30-October 3, 2012. Article (CrossRef Link)

[8] H. Deng, L. Yu, J. Qiu and J. Zhang, "A joint texture/depth edge-directed up-sampling algorithm for depth map coding," in *Proc. of IEEE International Conference on Multimedia and Expro(ICME)*, pp.646-650, July 9-13, 2012. Article (CrossRef Link)

[9] Q.Liu, Y.Yang, R.Ji, Y.Gao and L.Yu, "Cross-View Down/Up-Sampling Method for Multiview Depth Video Coding," *IEEE Signal Processing Letters*, vol.19, no.5, pp.295-298, May, 2012. Article (CrossRef Link)

[10] D. Ferstl, C.Reinbacher, R.Ranftl and M.Ruether, "Image guided depth upsampling using anisotropic total generalized variation," in *Proc. of IEEE Conference on Computer Vision (ICCV)*, pp. 993-1000, December 1-8, 2013. Article (CrossRef Link)

[11] MY. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.169-176, June 23-28, 2013. Article (CrossRef Link)

[12] J. Kim, G. Jeon, and J. Jeong, "Joint-adaptive bilateral depth map upsampling," *Signal Processing: Image Communication*, vol. 29, no. 4, pp. 506-513, April, 2014. Article (CrossRef Link)

[13] Y.Yang, X.Wang, Q.Liu, M.Xu, L.Yu, "A bundled-optimization model of multiview dense depth map synthesis for dynamic scene reconstruction," *Information Sciences*, vol.320, pp.306-319, November, 2014. Article (CrossRef Link)

[14] C.Zhu, Y.Zhao, L.Yu and M.Tanimoto, "3D-TV system with depth-image-based rendering," *Springer-Verlag, New York*, 2012. Article (CrossRef Link)

[15] H.Bai, M.Zhang, A.Wang, Y.Zhao, "Multiple Description Video Coding Using Correlation Optimized Temporal Sampling," *Science China Information Sciences*, vol.57, no.5, pp.1-10, May, 2014. Article (CrossRef Link)

[16] M.Barnsley, A.Vince, "Developments in fractal geometry," *Bulletin of Mathematical Sciences*, vol.3, no.2, pp.299-348, August 2013. Article (CrossRef Link)

[17] M.Berry, "Benefiting from fractals," in *Proc. of Symposia in Pure Math*, vol.72, pp.31-33, 2004. Article (CrossRef Link)

[18] M.F.Barnsley, "Fractal image compression," *Notice of the AMS*, vol.43, no.6, pp.657-662 June, 1996. Article (CrossRef Link)

[19] Y.Fisher, *Fractal image compression*, *Springer*, New York, 1995. Article (CrossRef Link)

[20] Y. Xu, H. Ji, C. Fermüller, "Viewpoint invariant texture description using fractal analysis," *International Journal of Computer Vision*, vol.83, pp.85-100, June, 2009. Article (CrossRef Link)

[21] J. Lei, S. Li, C. Zhu, M. Sun and C. Hou, "Depth coding based on depth-texture motion and structure similarities," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.25, no.2, pp.275-286, February, 2015. Article (CrossRef Link)

[22] Y.Zhao, B.Yuan, "A new affine transformation: its theory and application to image coding," *IEEE Transaction on Circuits and Systems for Video Technology*, vol.8, no.3, pp. 269-274, June, 1998. Article (CrossRef Link)

[23] Y.Sun, Y.Zhao, B.Yuan, "Region-based fractal image coding with freely-shaped partition," *Chinese Journal of Electronics*, vol.13, no.3, pp.506-511, July, 2004. Article (CrossRef Link)

[24] M. Tanimoto, T. Fujii, and K. Suzuki, "View synthesis algorithm in view synthesis reference software 3.5 (VSRS3.5) Document M16090, ISO/IEC JTC1/SC29/WG11 (MPEG)," May, 2009.

[25] Y.Chen, P.Pandit, S.Yea and C.Lim, "Draft reference software for MVC," in *Proc. of Joint Video Team (JVT) of ISO/IEC/MPEG and ITU-T/VCEG, Doc. JVT-AE 207*, London, U.K., 2009. Article (CrossRef Link)
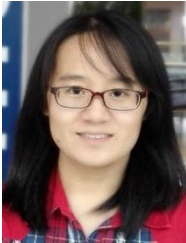
Meiqin Liu was born in Shanxi Province, China. She received her BS degree from Changsha University of Science and Technology, China, in 2004, and ME degree from Beijing Jiaotong University (BJTU), China, in 2007. She is currently a lecturer at institute of information and science in BJTU where she pursures the Ph.D. degree in signal and information processing since 2011. From 2014 to 2015, she was a visiting scholar in Simon Fraser University, Canada. Her research interests include fractal image coding, image/video compression and 3D video coding and processing.

Yao Zhao received the BS degree from Fuzhou University, China, in 1989, and the ME degree from Southeast University, Nanjing, China, in 1992, both from the Radio Engineering Department, and the PhD degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), China, in 1996. He became an associate professor at BJTU in 1998 and became a professor in 2001. From 2001 to 2002, he was a senior research fellow with the Information and Communication Theory Group, Faculty of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands. He is currently the director of the Institute of Information Science, BJTU. His current research interests include image/video coding, digital watermarking and forensics, and video analysis and understanding. He serves on the editorial boards of several international journals, including as associate editors of IEEE Transactions on Cybernetics, IEEE Signal Processing Letters, and an area editor of Signal Processing: Image Communication (Elsevier), etc. He was named a distinguished young scholar by the National Science Foundation of China in 2010, and was elected as a Chang Jiang Scholar of Ministry of Education of China in 2013. He is an IEEE senior member and an IET Fellow.

Chunyu Lin was born in Liaoning Province, China.He received the Ph.D. degree from Beijing Jiaotong University, Beijing, China, in 2011. From 2009 to 2010, he was a Visiting Researcher with the ICT Group, Delft University of Technology,
Delft, The Netherlands. From 2011 to 2012, he was a Postdoctoral Researcher with the Multimedia Laboratory, Ghent University, Ghent, Belgium. He is currently an Associate Professor with Beijing Jiaotong University. His current research interests include image/video compression and robust transmission, 2D-to-3D conversion, and 3D video processing.

Huihui Bai received her B.S. degree from Beijing Jiaotong University, China, in 2001, and her Ph.D. degree from Beijing Jiaotong University, China, in 2008. She is currently an associate professor in Beijing Jiaotong University. She has been engaged in R&D work in video coding technologies and standards, such as HEVC, 3D video compression, multiple description video coding (MDC), and distributed video coding (DVC).

Chao Yao received the B.S. degree in computer science from Beijing Jiaotong University, Beijing, China, in 2009, where he is currently pursuing the Ph.D. degree in signal and information processing at the Institute of Information Science. His current research interests include video compression and processing, 3D video coding, and computer vision.