

Depth Enhancement Methods for Centralized Texture-Depth Packing Formats

YANG Jar-Ferr, WANG Hung-Ming,
and LIAO Wei-Chen

(Department of Electrical Engineering, Institute of Computer and Communication Engineering, National Cheng Kung University, 1 University Road, Taiwan 701, China)

Abstract

To deliver three-dimension (3D) videos through the current two-dimension (2D) broadcasting systems, the frame-compatible packing formats properly including one texture frame and one depth map in various down-sampling ratios have been proposed to achieve the simplest and most effective solution. To enhance the compatible centralized texture-depth packing (CTDP) formats, in this paper, we further introduce two depth enhancement algorithms to further improve the quality of CTDP formats for delivering 3D video services. To compensate the loss of color YCbCr 444 to 420 conversion of colored-depth, two efficient depth reconstruction processes based on texture and depth consistency are proposed. Experimental results show that the proposed enhanced CTDP depacking process outperforms the 2DDP format and the original CTDP depacking procedure in synthesizing virtual views. With the help of the proposed efficient depth reconstruction processes, more correct reconstructed depth maps and better synthesized quality can be achieved. Before the available 3D broadcasting systems, which adopt truly depth and texture dependent coding procedure, we believe that the proposed CTDP formats with depth enhancement could help to deliver 3D videos in the current 2D broadcasting systems simply and efficiently.

Keywords

3D videos; frame-compatible; 2D-plus-depth; CTDP

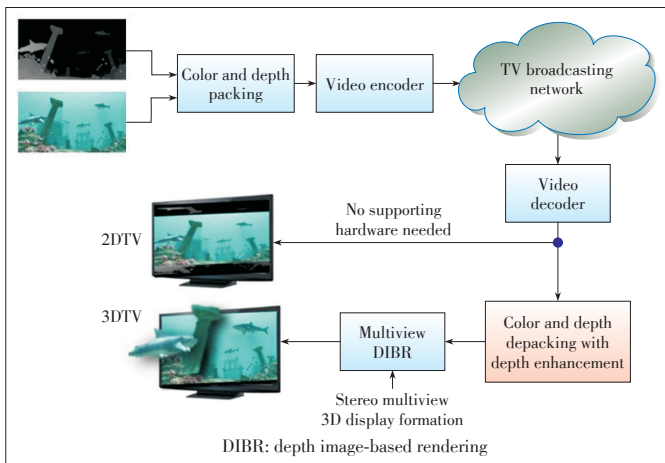
1 Introduction

Over past decades, more and more three-dimensional (3D) videos have been produced in the formats of stereo or multiple views with their corresponding depth maps. People desire to have more truthful and exciting experience through the true 3D visualizations.

In order to fit the traditional two-dimensional (2D) television (TV) programs, we need to modify the 3D videos to accommodate the certain constraints. Frame-packing is one of possible solutions to introduce 3D services in the current cable and terrestrial 2D TV systems. There are several well-known formats for packing the stereo views into 2D frame such as side-by-side (SbS), top-and-bottom (TaB), and checkerboard frame-compatible formats [1]–[4]. However, there exist two major problems, which slow down the development of the 3D TV services, in the existing frame-packing methods. The frame-compatible packing 3D videos of the stereo views mean that two texture images are gathered in one frame, which may make serious annoying effects on traditional 2D displays. Besides, stereo packing formats cannot support multi-view naked-eye 3D displays unless the stereo videos are further processed by real-time stereo matching methods [5], [6] and depth image-based rendering (DIBR) algorithms [7], [8]. To support multiview 3D displays, the 2D-plus-depth packing (2DDP) frame-compatible format, which arranges the texture in the left and the depth in the right, is suggested [9]. Once the color texture and depth arranged in the SbS fashion, the 2DDP format will bring even worse annoying visualization in 2D displays than the stereo packing formats. Recently, MPEG JCT-3V team proposed the latest coding standard for 3D video with depth [9]. However, it still needs some time to be deployed in current digital video broadcasting systems, which are with 2D and 3D capabilities.

To deal with the above problems, a novel frame-compatible centralized texture-depth packing (CTDP) formats for delivering 3D video services is proposed [10]. With AVS2 and HEVC video coders, the proposed CTDP formats [10] show better objective and subjective visual quality in 2D and 3D displays than the 2DDP format. In the CTDP format, the sub-pixel is utilized to store the depth information, while the texture information is arranged in the center of the frame to raise the 2D-compatible visual quality. However, the rearrangement will degrade the quality of the reconstructed depth map, especially when the video format with YCbCr space is 420 format with 4 Y components, one Cb component and one Cr component for each 4 color pixels. To further increase the visual quality, an efficient depth reconstruction process is also proposed in this paper. The frame structure of the CTDP method in cooperation with the current broadcasting system is shown in **Fig. 1**. Without any extra hardware, the 2D TV displays can also exhibit an acceptable 2D visual quality. For glasses or naked-eye 3D displays, we only need a simple CTDP depacking circuit followed by DIBR kernel to synthesize stereo or multiple views if the view-related sub-pixel formation of a naked-eye 3D display is given.

The rest of the paper is organized as follows. The CTDP formats are overviewed in Section 2. The proposed depth reconstruction process is described in Section 3. Experimental results to demonstrate the effectiveness of the proposed system are shown in Section 4. Finally, we conclude this paper in Sec-



▲ Figure 1. The broadcasting architecture by using the proposed enhanced CTDP format.

tion 5.

2 Centralized Texture-Depth Packing Formats

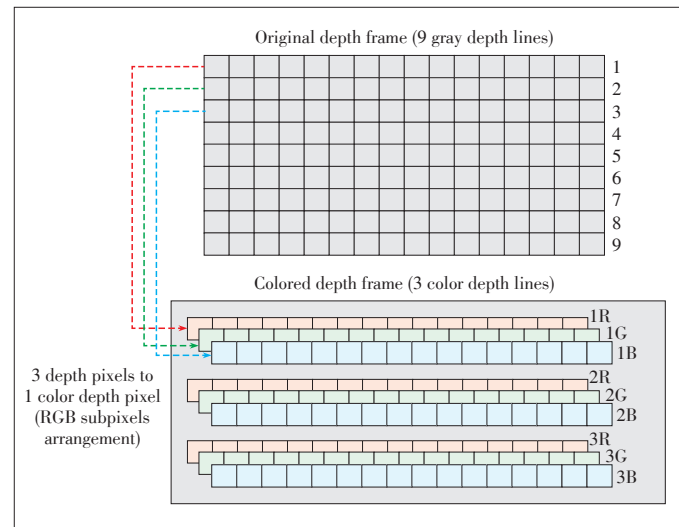
To achieve system compatibility, the basic concept of the CTDP method [10] is similar to frame compatible concept to pack texture and depth information together while keeping the same resolution as 2D videos. To solve the 2D visualization issue, we can arrange the texture in the center and the depth in two sides of the packed frame.

2.1 Colored-Depth Frame

The depth frame is only a gray image with Y components. To pack the depth frame, the colored-depth frame is suggested to represent it [10]. Thus, the colored-depth frame can be treated as the normal color texture frame, which can be directly encoded by any 2D video encoders with three times efficiency. As shown in Fig. 2, three depth horizontal lines are treated as horizontal R, G, and B subpixel lines in the RGB colored-depth frame. Since the nearby depth values are very close, the RGB colored-depth frame will exhibit nearly gray visual sensation. After color subpixels packing in the vertical direction, the vertical resolution of RGB colored-depth frame becomes one third of the original resolution. In Fig. 2, for example, the nine depth lines have been packed into three RGB colored- depth lines. For the most video coders, the coding and decoding processes are conducted in YCbCr color space. Therefore, we apply the RGB to YCbCr color space conversion as

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2568 & 0.5041 & 0.0979 \\ -0.1482 & -0.2910 & 0.4392 \\ 0.4392 & -0.3678 & -0.0714 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

to transfer it to the YCbCr colored-depth frame [11]. It is noted that the sub-pixels in RGB space are with full resolution of (4, 4, 4). If the YCbCr space is with (4, 4, 4) format, the color



▲ Figure 2. Rearrangement of the depth frame into RGB colored depth frame in vertical direction.

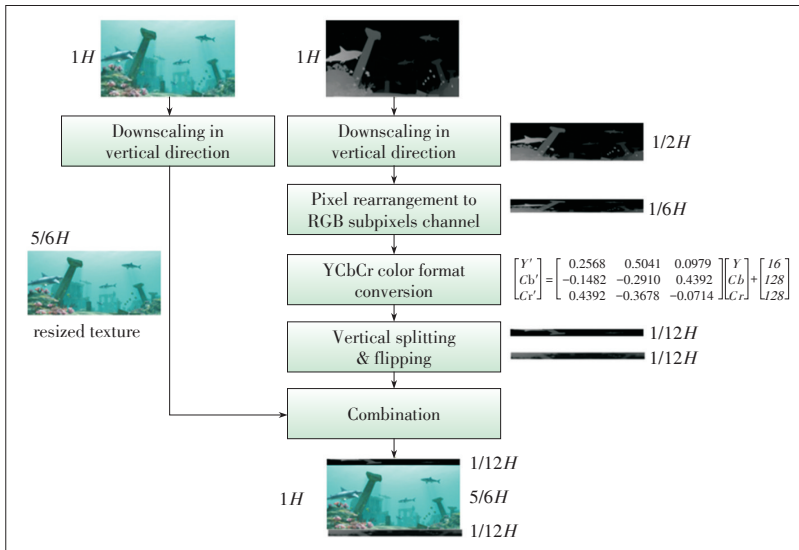
space transformation will not change the depth results with about ± 0.5 error due to the round-off errors in color space conversions. However, for the most video coders, the sub-pixels in YCbCr space could be in (4, 2, 0) or (4, 2, 2) format, where Cb and Cr components will be further downsampled. Even without coding errors, the YCbCr colored -depth frame might have slightly translation errors.

2.2 Centralized Texture-Depth Packing

Without loss of generality for frame-compatible packing, we assume that the vertical CTDP packing formats are desired. Then, we need to reduce the vertical resolutions of texture and depth separately such that the total packed resolution will remind the same, where the original horizontal resolution is H . If the reduction factors for texture and depth resolutions are a and b , we should choose reduction factors to satisfy $\alpha + (1/3)\beta = 1$ to achieve the frame compatible requirement [10]. For example, the reduction factors ($a = 3/4, b = 3/4$), ($a = 5/6, b = 1/2$), ($a = 7/8, b = 3/8$), ($a = 11/12, b = 1/4$), and ($a = 15/16, b = 3/16$) will satisfy the above frame compatible requirement. Fig. 3 shows the flowchart of the computation of generating the texture -5/6 CTDP format. First, we downscale the vertical resolution of texture and depth frames into five-sixths and one-second of the original resolution, respectively. By using the colored-depth concept, the resized depth frame with $1/2H$ can be further represented into RGB subpixels as suggested in Section 2.1 to reduce the vertical size to $1/6H$. Then, we can split the depth frame evenly into two separated parts with the size of $1/12H$. To make better coding efficiency and better 2D visualization, these two split colored - depth frames should be flipped vertically. The flipped depth frames will have better alignments to the texture frame and better visualization for 2D displays with visual shadow sensation. Finally, we obtain the texture -5/6 CTDP frame by combining the first

Depth Enhancement Methods for Centralized Texture-Depth Packing Formats

YANG Jar-Ferr, WANG Hung-Ming, and LIAO Wei-Chen



▲ Figure 3. The computation of the proposed frame compatible texture-5/6 CTD format.

flipped depth part (1/12H), the resized texture frame (5/6H), and the other flipped depth part (1/12H) from top to bottom sequentially.

The ratio of downscaling can also be changed to generate the other CTD formats [12]–[15]. For example, the reduction ratio of the texture frame could be 7/8 or 15/16. For texture-7/8 and texture-15/16 reduction ratios, the vertical resolutions of depth frames will be respectively downscaled to 3/8 and 3/16 to satisfy (2). Except the resizing factor, the packing procedures for texture-7/8 and texture-15/16 are similar to that of texture-5/6. If we want to attain horizontal CTD formats, all the resizing of texture and depth frame, the color-packed depth frame, slipping, and flipping procedures should be performed in the horizontal direction. The packed frame can be obtained by combining the first flipped depth part, the resized texture frame, and the other flipped depth part from left to right sequentially. The outlooks of the original texture, depth, and the CTD frames with different ratios and different orientations are shown in Fig. 4. It is noted that in the proposed CTD format, the width/height of the flipped depth part will be always in the horizontal/vertical CTD format, which helps avoid the compression artifact in texture and depth boundary. Please refer to [13] for more details of the arrangement.

2.3 Depacking CTD Formats

With respect to the packing procedure in Fig. 3, the flow diagram for de-

packing the texture - 5/6 CTD format is shown in Fig. 5. Once we receive the CTD format, we should first split the packed frame into three parts: the top flipped depth part, the central texture, and the bottom flipped depth part. For two flipped depth parts, we perform another vertical flipping and combined them into the whole texture-packed depth frame. The YCrCb colored-depth frame might need to upsample Cr and Cb components back to (4, 4, 4) format first. Then, we can convert it to (4, 4, 4) RGB colored-depth frame by

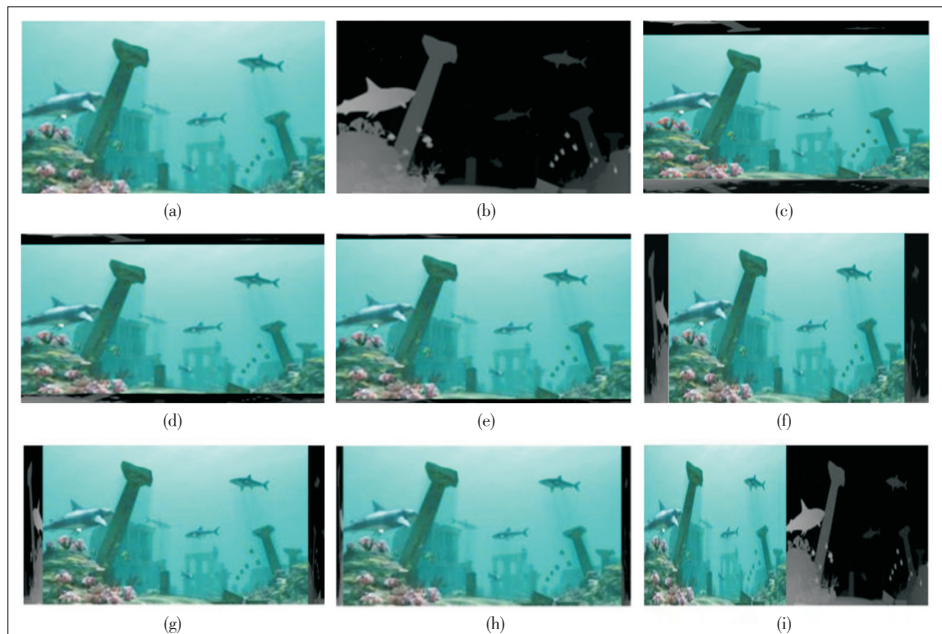
$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.1644 & -0.0001 & 1.5960 \\ 1.1644 & -0.3917 & -0.8130 \\ 1.1644 & 2.0173 & -0.0001 \end{bmatrix} \begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}. \quad (2)$$

After the color space conversion, The RGB colored-depth frame (1/6H) can be finally recovered to the resized depth frame (1/2H).

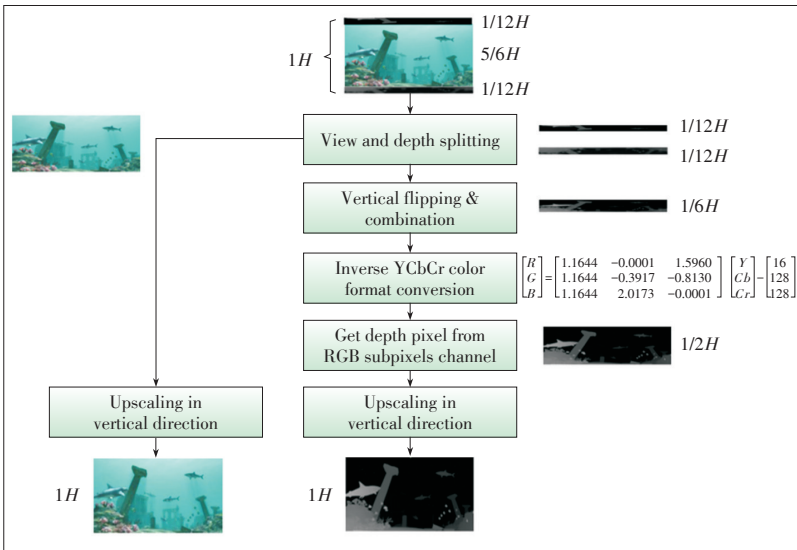
After 6/5 upscaling texture and 2/1 depth frames in the vertical direction, we finally depack the original texture and depth frames. Of course, a possible DIBR method should be used to generate all the necessary views. As for the other texture reduction ratios such as 7/8 and 15/16, all the procedures will be the same except the resizing factors of depth will be 3/8 and 3/16, respectively.

3 Depth Enhancement Algorithms

From the previous section, it is known that when the YCbCr space is (4, 2, 0) or (4, 2, 2) format, the YCbCr colored-depth frame will induce translation errors along the depth edges. To



▲ Figure 4. Schematics of original (a) texture, (b) depth; (c) vertical texture-5/6 CTD; (d) vertical texture-7/8 CTD; (e) vertical texture-15/16 CTD; (f) horizontal texture-5/6; (g) CTD, horizontal texture-7/8; (h) horizontal texture-15/16 CTD; and (i) 2DDP frame compatible formats.

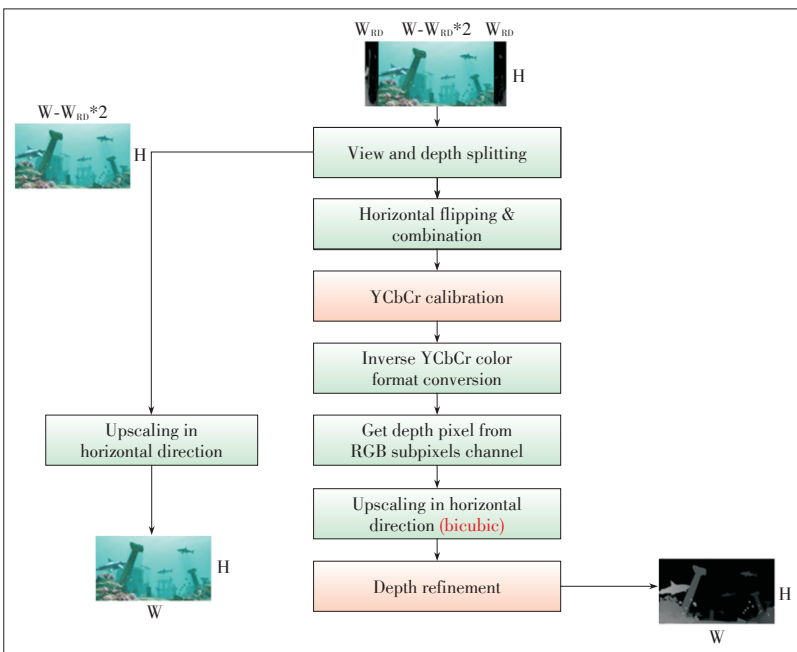


▲ Figure 5. The computation of the proposed texture-5/6 CTD de-packing procedure.

further reduce the depth edge errors, in this paper, we propose two efficient depth enhancement processes. The enhancement processes can be incorporated with the original de-packing process as shown in Fig. 6. The enhancement processes include YCbCr calibration, texture-similarity-based depth up-sampling and pattern-based down-sampling. Details of the enhancement algorithms are addressed in the following subsections.

3.1 YCbCr Calibration

When the YCbCr color space is (4, 4, 4), the color space transformation between RGB color space and YCbCr color space will only contain round-off errors in color space conver-



▲ Figure 6. Flowchart of the depth-enhanced CTD de-packing system.

sions. However, for the most video coders, the subpixels in the YCbCr color space might be (4, 2, 0) or (4, 2, 2) formats, where Cb and Cr components will be further down-sampled in order to save the bandwidth in broadcasting systems. At the de-packing side, we need to calibrate the translation errors between YCbCr (4, 4, 4) and YCbCr (4, 2, 0) and (4, 2, 2). For simplicity, we will illustrate our proposed system in YCbCr (4, 2, 0), however, the similar manner can still be applied for YCbCr (4, 2, 2).

Before we start to calibrate the YCbCr data, we first define some anchor pixels, which are shown in Fig. 7. The anchor pixels denote the pixels which have the correct Cb and Cr subpixel values.

The diagram of missing components in YCbCr (4, 2, 0) for all surrounding pixels is shown in Fig. 8. Each color means a set which Cb and Cr subpixel components are down-sampled. The black area means the missing Cb and Cr subpixels and they

can be given by:

$$Cb_{cal}(a,b) = \arg_{Cb_c} \min |Y_c - Y(a,b)|, \quad (3)$$

and

$$Cr_{cal}(a,b) = \arg_{Cr_c} \min |Y_c - Y(a,b)|, \quad (4)$$

where Y_c is a vector of the neighbor anchor pixels of the pixels $Y(a,b)$.

3.2 Texture-Similarity-Based Depth Up-Sampling

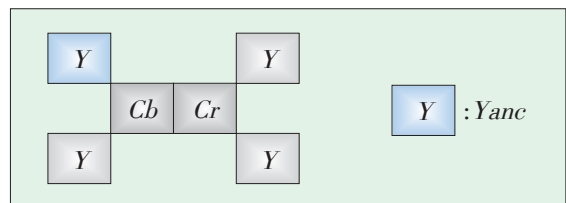
In order to preserve the continuity of the edge, the directional vectors are utilized to calculate the edge direction in the low-resolution (LR) depth and the corresponding high resolution (HR) texture image. The directional vectors of LR depth image and HR texture image can be formed as:

$$\overline{vd}_L = \sum_{\Omega} \exp\left(-\frac{D_E(x_L, y_L) - D_{\Omega}}{\sigma_V}\right) \times \vec{u}_{\Omega}, \quad (5)$$

and

$$\overline{vc} = \sum_{\Omega} \exp\left(-\frac{Y(x,y) - Y_{\Omega}}{\sigma_V}\right) \times \vec{u}_{\Omega}, \quad (6)$$

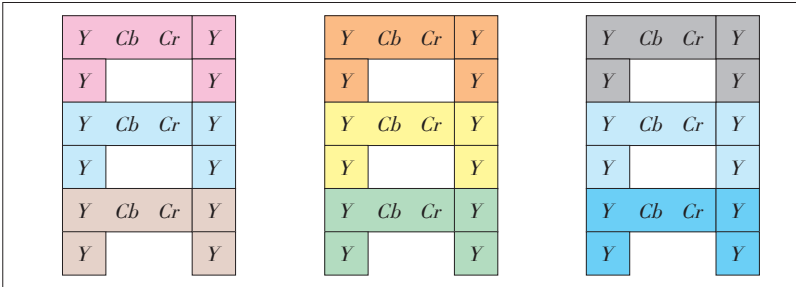
where \overline{vd}_L and \overline{vc} denote the directional vectors of the pixels in LR depth image and HR texture image, respectively, σ_V represents the standard deviation



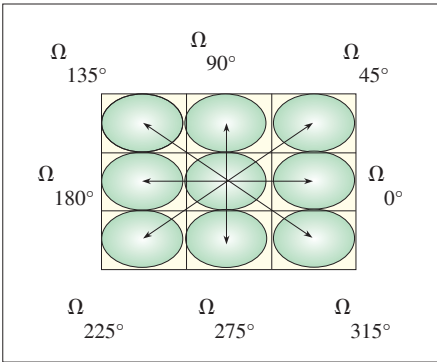
▲ Figure 7. Anchor pixels in YCbCr (4, 2, 0).

Depth Enhancement Methods for Centralized Texture-Depth Packing Formats

YANG Jar-Ferr, WANG Hung-Ming, and LIAO Wei-Chen



▲ Figure 8. Missing Cb and Cr subpixels in YCbCr (4, 2, 0).



◀ Figure 9. The diagram of the neighbor pixels, Ω .

of the directional vector function, Ω denotes the 8 neighbor pixels of the target pixel (Fig. 9), D_E represents the combined depth, which is obtained from previous step, Y is the brightness of the texture image, and \vec{u}_Ω is the unit vector corresponding to the neighbor pixels Ω in 8 directions.

Before up-sampling the depth image, the directional vectors are first transformed from Cartesian coordinate system to Spherical coordinate system. The transform function is given by:

$$r = \sqrt{x_\partial^2 + y_\partial^2}, \quad (7)$$

and

$$\theta = \arctan\left(\frac{y_\partial}{x_\partial}\right), \quad (8)$$

where x_∂ and y_∂ denote the coordinate of reconstructed depth at high resolution. For example, at vertical texture-11/12 CTDP, $x_\partial = 4x$ and $y_\partial = y$. However, the resolution of directional vectors in depth image is smaller than the resolution of directional vectors in texture image. The bilinear interpolation [16] is utilized to scale up the depth directional vector to the resolution of the texture image. After that, The interpolated depth image is formed as:

$$D_{up} = \begin{cases} \frac{1}{T_{up}} \sum D_E(q) \times \psi(D_E(p) - D_E(q)), & \text{if } \|Vd(\theta) - Vc(\theta)\| < \frac{\pi}{8} \text{ or } Vc(r) < 1 \\ \text{hole}, & \text{else} \end{cases}, \quad (9)$$

where T_{up} denotes the normalized factor, p is the target pixel which needs to be scale up, q is the neighbor pixels of the target pixel, and $Vd(\theta)$ is the value of θ in the scaled $Vd_L(\theta)$. ψ

denotes the Gaussian weight function and can be given as:

$$\psi(n) = \exp\left(-\frac{n^2}{\sigma_\psi}\right). \quad (10)$$

The basic concept of the depth interpolation is to compare the directional vectors of the depth image and the texture image. The weighted summation of the LR depth is utilized to interpolate the HR depth if the directional vectors of the depth image and the texture image are similar. Otherwise, the pixels in HR depth are regarded as holes, which are filled in the step of hole-filling. The function of hole-filling is given as:

$$D_{hole-filling}(x,y) = \begin{cases} \arg_{D_{up}} \xi(\min(\Delta Pc(\theta))), & \text{if } (x,y) \ni \text{holes} \\ D_{up}(x,y), & \text{else} \end{cases}, \quad (11)$$

where $\Delta Pc(\theta)$ denotes the difference of the degree between Pc_θ and 8 neighbor pixels. ξ represents the selection function of the hole-filling and it can be formed as:

$$\xi(m) = \begin{cases} Y(m), & \text{if } \|Y - Y(m)\| < TH_Y \\ \xi(m) + 1, & \text{else} \end{cases}, \quad (12)$$

where Y denotes the brightness of the target pixel, $Y(m)$ denotes the brightness of the neighbor pixels in m direction, TH_Y is the threshold to control the selection range, and $\xi(m) + 1$ represents the next pixel in m direction.

3.3 Pattern-Based Down-Sampling

In order to contain texture image and depth image in one single frame, both depth image and texture image need to be down-sampled. For the depth image, the bilinear and bi-cubic convolution methods are utilized to down-sample the depth image. However, the weighted summation strategy in bilinear and bi-cubic convolution leads to the blur of the down-sampled data. Hence, we propose two sampling patterns to down-sample the depth image without fusing the data. There are the direct line pattern and slant line pattern.

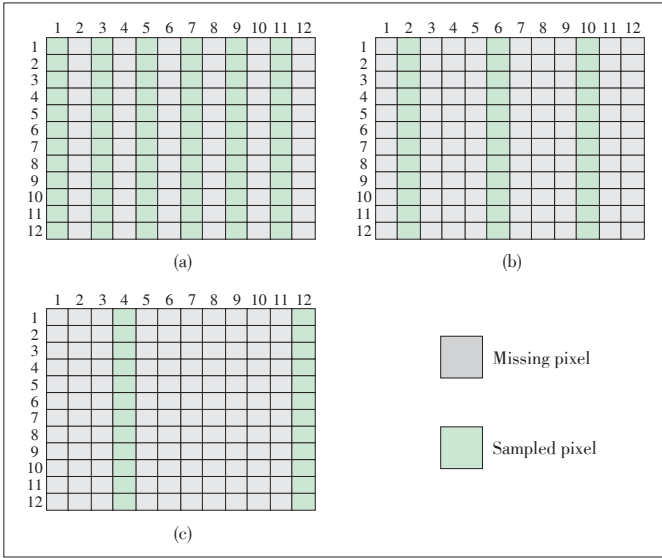
1) Direct line pattern

The sampling strategy of direct line pattern is to grab pixels in the straight line direction. According to the characteristic of the CTDP format, the reduction of the resolution is only in either horizontal or vertical direction. The function of direct line pattern is given as:

$$D_{down}(x,y) = D_{origin}(\partial_{hor} \times x - [\partial_{hor}/2], \partial_{ver} \times y - [\partial_{ver}/2]), \quad (13)$$

where ∂_{hor} and ∂_{ver} are the factors of down-sampling ratios in horizontal direction and vertical direction, respectively. For CTDP format usage, either ∂_{hor} or ∂_{ver} is equal to 1, while the other one denotes the down-sampling ratio in packing procedure. $[x]$ is the floor function, which means the largest integer not greater than x . The direct line pattern in horizontal direction with 2, 4, 8 down-sampling ratio is shown in Fig. 10.

2) Slant line pattern



▲ Figure 10. The direct line pattern in horizontal direction of (a) down-sampling factor 2; (b) down-sampling factor 4; and (c) down-sampling factor 8.

The sampling strategy of slant line is to grab pixels in 45 degree direction. The function of direct line pattern is given as:

$$D_{down}(x, y) = D_{origin}(\partial_{hor} \times x - (\partial_{hor} - y), y), \quad (14)$$

or

$$D_{down}(x, y) = D_{origin}(x, \partial_{ver} \times y - (\partial_{ver} - x)). \quad (15)$$

Equ. (14) is utilized to down-sample the depth image in horizontal direction, while the down-sampling of the vertical direction follows (15). The slant line sampling pattern is suitable for down-sampling the depth image both in vertical and horizontal direction, which is shown in Fig. 11 with 2, 4, 8 down-sampling ratios.

With the down-sampling by the direct line pattern, the up-sampling function in de-packing procedure needs to be modified as:

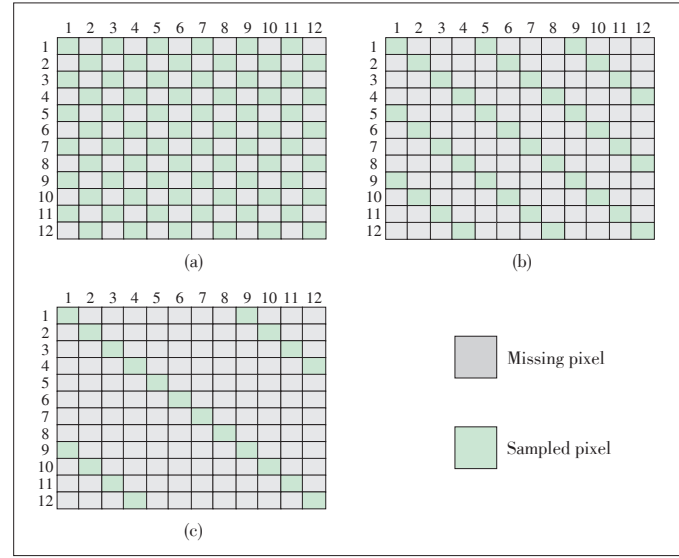
$$D_{up} = \begin{cases} D_E(p), & \text{if } p \in \text{sampled data} \\ \frac{1}{T} \sum D_E(q) \times \psi(D_E(p) - D_E(q)), & \text{else if } \|Vd(\theta) - Vc(\theta)\| < \frac{\pi}{8} \text{ or } Vc(r) < 1. \\ \text{hole}, & \text{else} \end{cases} \quad (16)$$

Because of the pattern-based sampling strategy, the pixels of the up-sampled depth are directly copied from the LR depth if there are located at position of the direct line pattern.

4 Experimental Results

4.1 Performance Evaluation of CTDFormat with Respect to 2DDP Format

In order to verify the coding performances of the proposed CTDFormat with respect to the 2DDP format, we conducted



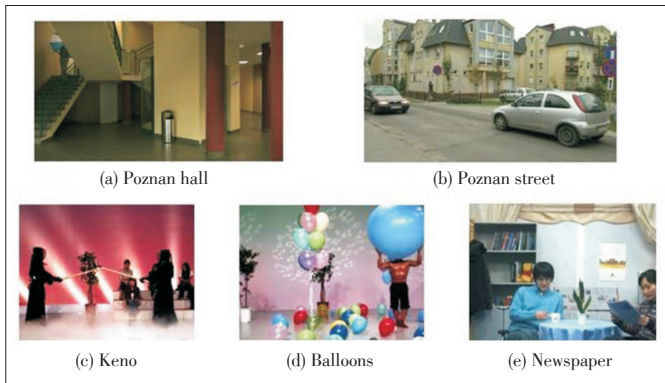
▲ Figure 11. The slant line pattern of (a) down-sampling factor 2; (b) down-sampling factor 4; (c) down-sampling factor 8.

a set of experiments to evaluate performances of packing methods in cooperation with a specific video coder (AVS2) in terms of the peak signal-to-noise ratio (PSNR), bitrate qualities of the depacked texture and depacked depth frames, and their synthesized virtual views. In the experimental simulations, we use five MPEG 3D video sequences, which are Poznan Hall, Poznan Street, Kendo, Balloons, and Newspaper sequences as shown in Figs. 12a–12e, respectively.

The AVS2 coding conditions are followed by the instruction suggested by the AVS workgroup while the QPs are set to 27, 32, 38, and 45 for Intra frames [17]. Under All Intra (ai), Low Delay P (ldp), Random Access (ra) test conditions, Tables 1 and 2 show the average BDPSNR and BDBR [18] performance for different kinds of CTDFormat with respect to the 2DDP format achieved by AVS2. For calculating the PSNR of the 2DDP format, we first separate the texture and depth frames from the 2DDP frame and upsample them to the original image size $W \times H$. By using the recovered texture and depth frames from 2DDP frame and the original uncompressed texture and depth frames, the PSNR can therefore be calculated. Similarly, the PSNR of CTDFormat is calculated by using the texture and depth frames recovered from CTDFormat frame and the original uncompressed texture and depth frames. From Tables 1 and 2, we can see that the proposed texture-5/6, 7/8, and 15/16 CTDFormat formats have much better PSNR and bitrate saving in texture when comparing with the 2DDP format, which means our CTDFormat can achieve better visual quality in 2D displays when only texture frames are viewed. In addition, the depth quality for CTDFormat will become worse while the resizing factors getting bigger. Besides the comparisons of original texture and depth achieved by different packing formats, we also compare the quality of synthesized virtual view with respect to the 2DDP format. It is noted that the reference synthesized vir-

Depth Enhancement Methods for Centralized Texture-Depth Packing Formats

YANG Jar-Ferr, WANG Hung-Ming, and LIAO Wei-Chen



▲ Figure 12. Five texture and depth frame compatible packing formats.

▼ Table 1. Averaged BDPSNR performances

	BDPSNR					
	Recovered texture performance			Synthesized virtual view performance		
	5/6	7/8	15/16	5/6	7/8	15/16
ai	2.2693	2.2587	2.34168	0.6150	0.3664	-0.7732
ra	2.5868	2.71322	2.86354	0.7371	0.7489	-0.2754
ldp	2.2079	2.42228	2.54206	0.3920	0.6005	-0.3928
avg	2.3546	2.464733	2.582427	0.5813	0.5720	-0.4805

▼ Table 2. Averaged BDBR performances

	BDBR					
	Recovered texture performance			Synthesized virtual view performance		
	5/6	7/8	15/16	5/6	7/8	15/16
ai	-55.3282	-46.5149	-47.9069	-22.9490	-7.92186	56.0625
ra	-61.0706	-59.5852	-61.6926	-24.6686	-24.6019	30.0742
ldp	-48.7147	-54.3501	-56.2303	-9.9111	-19.6368	36.5666
avg	-55.0378	-53.4834	-55.2766	-19.1762	-17.3869	40.9011

tual view for calculating the PSNR is also obtained by the original uncompressed texture and depth frames. The DIBR setting for virtual view synthesis is shown in Table 3. As to the quality of the synthesized virtual view, the texture-5/6 and 7/8 CTDP formats after the DIBR process show better BDPSNR and BDBR performances than 2DDP format. It is noted that all synthesized views do not perform any depth enhancement and depth preprocessing, and the hole filling used in the DIBR pro-

▼ Table 3. DIBR settings for virtual view synthesis

Sequence	Resolution	Frames	Coded view	Synthesized view
Poznan hall	1920*1088	200	6	5
Poznan street	1920*1088	250	4	3
Kendo	1024*768	300	3	4
Balloons	1024*768	300	3	5
Newspaper	1024*768	300	4	6

cess is the simple background extension.

In summary, the texture qualities BDPSNR and BDBR in Tables 2 and 3 can be treated as the objective quality indices in 2D displays, while the virtual view qualities can be the objective quality indices in 3D displays. The results show that the proposed texture-5/6 and 7/8 CTDP format will be the better choices for the broadcasters. The texture-3/4 CTDP format has better 3D performance while texture-7/8 CTDP format achieves better 2D performance.

4.2 Performance Evaluation of Depth Enhancement for CTDP Format

To verify the proposed depth enhancement mechanism, we first show the reconstructed depth from original and depth-enhanced CTDP formats. The RD curves for different ratios of CTDP formats are shown in Fig. 13. It can be seen that the proposed refined CTDP format can always achieve better performance. The gains between the depth-enhanced CTDP and the original CTDP formats are increased while the ratio of texture is increased.

For the subjective evaluation, the partial portions of the reconstructed depth for Shark sequence are shown in Fig. 14. It can be seen that the depth can be reconstructed well especially for the edge region by using the depth enhancements.

In the following, we will compare the synthesis results. The partial portions of the generated views are shown in Fig. 15. From the results, the proposed CTDP format can successfully preserve the edges well of the synthesis views without the jaggly noise.

4.3 Comparison with Different Depth Interpolation Methods

The comparison results of different depth interpolation methods are shown in Table 4 for Shark sequence at all-intra (ai) coding condition with QP=32. The symbols of Bi and BC denote the bilinear and bi-cubic convolution interpolation methods, respectively. The methods of JBU [19] and FEU [20] are the texture-similarity based depth interpolation methods. The proposed depth up-sampling method has better PSNR and SSIM results for reconstructed depth images in vertical-11/12 CTDP and vertical-23/24 CTDP formats. For the vertical-5/6 CTDP format, the proposed depth up-sampling method can also provide better reconstructed depth images.

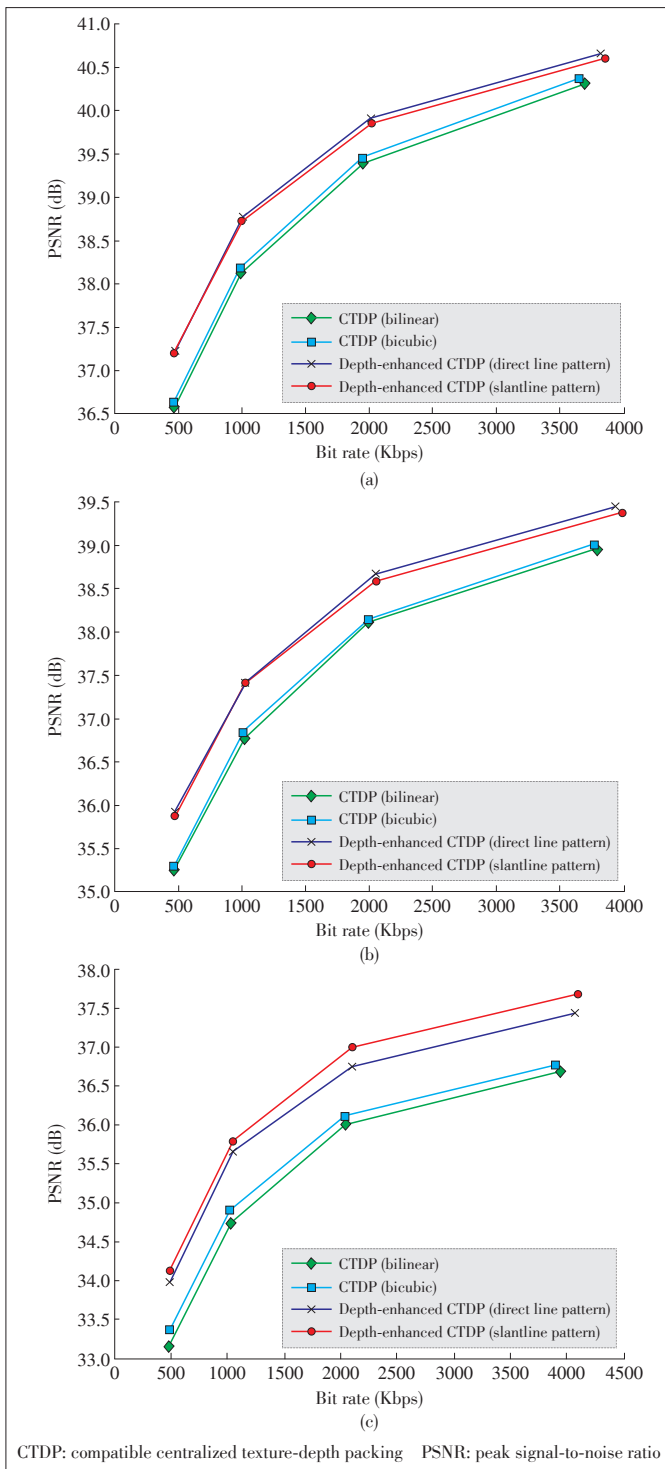
The comparison results of partial reconstructed depth with different depth interpolation methods are shown in Fig. 16. The reconstructed depth images of bilinear and bi-cubic convolution interpolation methods have serious jaggly noise among the edges. It can be seen that the proposed depth up-sampling method can outperform other methods with better edges.

5 Conclusions

In this paper, we proposed depth enhancement processes for

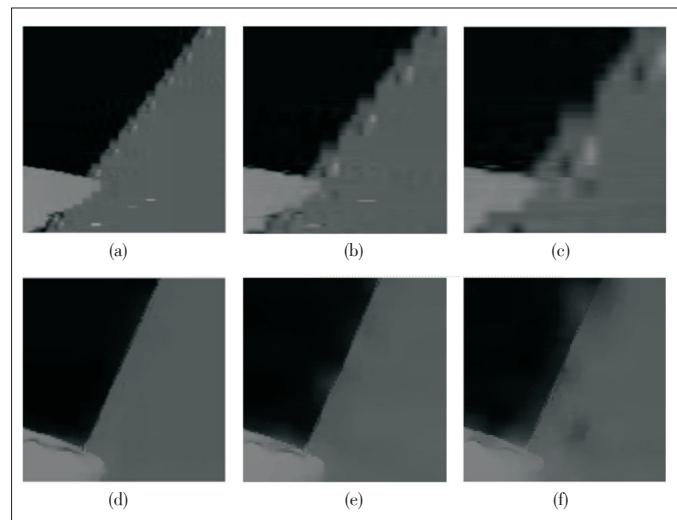
Depth Enhancement Methods for Centralized Texture-Depth Packing Formats

YANG Jar-Ferr, WANG Hung-Ming, and LIAO Wei-Chen

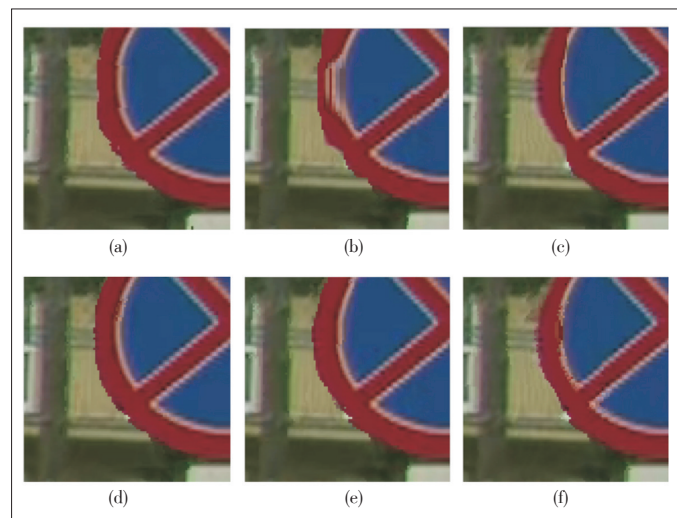


▲ Figure 13. RD curves of reconstructed depth from the original CTDTP depacking process and the proposed depth-enhanced CTDTP depacking process for (a) texture-5/6; (b) texture-11/12; (c) texture-23/24 formats.

CTDP formats [10]. The CTDTP formats can be comfortably and directly viewed in 2DTV displays without the need of any extra computation. However, the CTDTP formats slightly suffer from the depth discontinuities for high texture ratios. Comparing to



▲ Figure 14. Partial portions of reconstructed depth in S10 Shark with the original CTDTP depacked: (a) texture-5/6; (b) texture-11/12; (c) texture-23/24 formats and the proposed depth enhanced CTDTP depacked; (d) texture-5/6; (e) texture-11/12; (f) texture-23/24 formats.



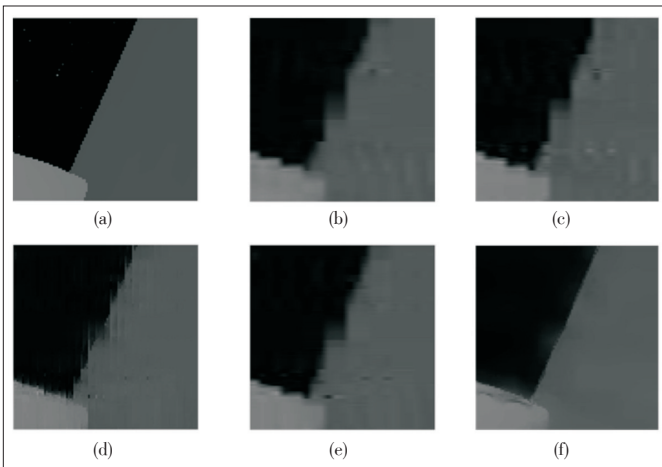
▲ Figure 15. Partial synthesis views of S02 Poznan Street with original CTDTP depacking; (a) texture-5/6; (b) texture-11/12; (c) texture-23/24 and the proposed depth-enhanced CTDTP depacking; (d) texture-5/6; (e) texture-11/12; (f) texture-23/24.

▼ Table 4. The PSNR and SSIM comparison of different depth interpolation methods in S10 Shark at All Intra (ai) QP=32

PSNR (dB)	Bi	BC	JBU	FEU	Proposed
Vertical-5/6 CTDTP	33.7164	33.5138	33.6252	33.7135	33.5239
Vertical-11/12 CTDTP	31.7651	31.6581	32.0857	31.7850	32.4422
Vertical-23/24 CTDTP	29.8525	29.7836	30.3490	29.8953	30.5411
SSIM	Bi	BC	JBU	FEU	Proposed
Vertical-5/6 CTDTP	0.9361	0.9334	0.9361	0.9368	0.9397
Vertical-11/12 CTDTP	0.9147	0.9122	0.9192	0.9158	0.9270
Vertical-23/24 CTDTP	0.8914	0.8879	0.8959	0.8925	0.9068

Depth Enhancement Methods for Centralized Texture-Depth Packing Formats

YANG Jar-Ferr, WANG Hung-Ming, and LIAO Wei-Chen



▲ Figure 16. The comparison of partial reconstructed depth with different depth interpolation methods for vertical-11/12 CTDTP format: (a) ground truth; (b) bilinear; (c) bi-cubic convolution; (d) JBU [19]; (e) FEU [20]; (f) proposed.

the 2DDP format, the CTDTP formats with the same video coding systems, such as AVS2 (RD 6.0) and HEVC [10], show better coding performances in texture and depth frames and synthesized virtual views. To further increase the visual quality, in this paper, the depth enhancement methods, including YCbCr calibration and texture-similarity-based depth up-sampling, are proposed. Experimental results reveal that the proposed depth enhancement can efficiently help to increase the depacking performances of the CTDTP formats to achieve better reconstructed depth images and better synthesis views as well. With the aforementioned simulation results, we believe that the proposed depth enhanced CTDTP depacking methods will be a greatly-advanced system for current 2D video coding systems, which can provide 3D video services effectively and simply.

References

[1] J.-F. Yang, H.-M. Wang, K.-I. Liao, L. Yu, and J.-R. Ohm, "Centralized texture-depth packing formats for effective 3D video transmission over current video broadcasting systems," *IEEE Transactions on Circuits and Systems for Video Technology*, submitted for publication.

[2] Dolby Laboratories, Inc. (2015). *Dolby Open Specification for Frame-Compatible 3D Systems* [Online]. Available: <http://www.dolby.com>

[3] ITU. (2015). *Advanced Video Coding for Generic Audio—Visual Services* [Online]. Available: <http://www.itu.int>

[4] G. Sullivan, T. Wiegand, D. Marpe, and A. Luthra, "Text of ISO/IEC 14496-10 advanced video coding (third edition)," ISO/IEC JTC 1/SC 29/WG11, Redmond, USA, Doc. N6540, Jul. 2004.

[5] G. J. Sullivan, A. M. Tourapis, T. Yamakage, and C. S. Lim, "ISO/IEC 14496-10: 200X/FPDAM 1," ISO/IEC JTC 1/SC 29/WG11, Apr. 2009.

[6] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment," *IEEE Transactions on Pattern Analysis and Matching Intelligence*, vol. 16, no. 9, pp.920–932, Sept. 1994. doi:10.1109/34.310690.

[7] K. Zhang, J. Lu, and G. Lafuit, "Cross-based local stereo matching using orthogonal integral images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 7, pp.1073– 1079, Jul. 2009. doi: 10.1109/TCSVT.2009.2020478.

[8] S.-C. Chan, H.-Y. Shum, and K.-T. Ng, "Image-based rendering and synthesis," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 22–33, Nov. 2007. doi: 10.1109/MSP.2007.905702.

[9] T.-C. Yang, P.-C. Kuo, B.-D. Liu, and J.-F. Yang, "Depth image-based rendering with edge-oriented hole filling for multiview synthesis," in *Proc. International*

Conference on Communications, Circuits and Systems, Chengdu, China, Nov. 2013, vol. 1, pp. 50–53. doi: 10.1109/ICCCAS.2013.6765184.

[10] Philips 3D Solutions, "3D interface specifications, white paper," Eindhoven, The Netherlands, Dec. 2006.

[11] *Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios*, ITU-R BT.601-5, 1995.

[12] J.-F. Yang, K.-Y. Liao, H.-M. Wang, and Y.-H. Hu, "Centralized texture-depth packing (CTDP) SEI message syntax," Joint Collaborative Team on 3D Video Coding Extensions of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Strasbourg, France, Doc. no. JCT3V-J0108, Oct. 2014.

[13] J.-F. Yang, K.-Y. Liao, H.-M. Wang, and C.-Y. Chen, "Centralized texture-depth packing (CTDP) SEI message," Joint Collaborative Team on 3D Video Coding Extensions of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Geneva, Switzerland, Doc. no. JCT3V-K0027, Feb. 2015.

[14] J.-F. Yang, H.-M. Wang, Y.-A. Chiang, and K. Y. Liao, "2D frame compatible centralized color depth packing format (translated from Chinese)," AVS 47th Meeting, Beijing, China, AVS-M3225, Dec. 2013.

[15] J.-F. Yang, H.-M. Wang, K.-Y. Liao, and Y.-A. Chiang, "AVS2 syntax message for 2D frame compatible centralized color depth packing formats (translated from Chinese)," AVS 50th Meeting, Nanjing, China, AVS-M3472, Oct. 2014.

[16] H. C. Andrews and C. L. Patterson, "Digital interpolation of discrete images," *IEEE Transaction on Computers*, vol. 25, no. 2, 1976.

[17] X.-Z. Zheng, "AVS2-P2 common test conditions (translated from Chinese)," AVS 46th Meeting, Shenyang, China, AVS-N2001, Sep. 2013.

[18] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," Austin, USA, Doc. VCEG-M33 ITU-T Q6/16, Apr. 2001.

[19] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Transaction on Graphics*, vol. 26, no. 3, Article 96, Jul. 2007. doi:10.1145/1275808.1276497.

[20] S.-Y. Kim and Y.-S. Ho, "Fast edge-preserving depth image upsampler," *Journal of Consumer Electronics*, vol. 58, no. 3, pp. 971–977, Aug. 2012. doi: 10.1109/TCE.2012.6311344.

Manuscript received: 2015-11-12

Biographies

YANG Jar-Ferr (jefyang@mail.ncku.edu.tw) received his PhD degree from the University of Minnesota, USA in 1988. He joined the National Cheng Kung University (NCKU) started from an associate professor in 1988 and became a full professor and distinguished professor in 1995 and 2007. He was the chairperson of Graduate Institute of Computer and Communication Engineering during 2004–2008 and the director of the Electrical and Information Technology Center 2006–2008 in NCKU. He was the associate vice president for Research and Development of the NCKU. Currently, he is a distinguished professor and the director of Technologies of Ubiquitous Computing and Humanity (TOUCH) Center supported by National Science Council (NSC), Taiwan, China. Furthermore, he is the director of Tomorrow Ubiquitous Cloud and Hypermedia (TOUCH) Service Center. During 2004–2005, he was selected as a speaker in the Distinguished Lecturer Program by the IEEE Circuits and Systems Society. He was the secretary, and the chair of IEEE Multimedia Systems and Applications Technical Committee and an associate editor of *IEEE Transaction on Circuits and Systems for Video Technology*. In 2008, he received the NSC Excellent Research Award. In 2010, he received the Outstanding Electrical Engineering Professor Award of the Chinese Institute of Electrical Engineering, Taiwan, China. He was the chairman of IEEE Tainan Section during 2009–2011. Currently, he is an associate editor of *EURASIP Journal of Advances in Signal Processing* and an editorial board member of *IET Signal Processing*. He has published 104 journal and 167 conference papers. He is a fellow of IEEE.

WANG Hung-Ming (ming@video5.ee.ncku.edu.tw) received the BS and PhD degrees in electrical engineering from National Cheng Kung University (NCKU), Taiwan, China in 2003 and 2009, respectively. He is currently a senior engineer of Novatek Microelectronics Corp., Taiwan, China. His major research interests include 2D/3D image processing, video coding and multimedia communication.

LIAO Wei-Chen (a800812momo@gmail.com) received the BS and MS degrees in electrical engineering from National Cheng Kung University (NCKU), Taiwan, China in 2013 and 2015, respectively. His major research interests include image processing, video coding and multimedia communication.