ANALYSIS OF PIXEL-MAPPING ROUNDING ON GEOMETRIC DISTORTION AS A PREDICTION FOR VIEW SYNTHESIS DISTORTION

P. Carballeira, J. Cabrera, E. Ekmekcioglu, F. Jaureguizar, and N. García

ABSTRACT

We analyze the performance of the geometric distortion, incurred when coding depth maps in 3D Video, as an estimator of the distortion of synthesized views. Our analysis is motivated by the need of reducing the computational complexity required for the computation of synthesis distortion in 3D video encoders. We propose several geometric distortion models that capture (i) the geometric distortion caused by the depth coding error, and (ii) the pixel-mapping precision in view synthesis. Our analysis starts with the evaluation of the correlation of geometric distortion values obtained with these models and the actual distortion on synthesized views. Then, the different geometric distortion models are employed in the rate-distortion optimization cycle of depth map coding, in order to assess the results obtained by the correlation analysis. Results show that one of the geometric distortion models is performing consistently better than the other models in all tests. Therefore, it can be used as a reasonable estimator of the synthesis distortion in low complexity depth encoders.

Index Terms — 3D video, depth-image based rendering, depth coding, rate-distortion optimization.

1. INTRODUCTION

3D Video (3DV) and Free Viewpoint Video (FVV) expand the user's experience beyond what is offered by 2D video [1], offering a 3D depth impression of the observed scene and an interactive selection of the viewpoint. Generally, both systems use a data format that includes one or more video signals corresponding to the same scene (color signals from now on) and scene geometry information [2]. This data format allows the possibility of generating additional views on virtual camera positions. A widely adopted approach for the scene geometry is to use signals that describe the distance of the objects in the scene to the cameras that capture them. These signals are known as depth sequences. A common 3D Video format is the Multiview Video plus Depth (MVD), composed by *N* color sequences and *N* associated depth sequences [3].

The quality of 3DV systems depends, in addition to the quality of the transmitted views, on the quality of virtual views synthesized from decoded MVD data. Previous works have used distortion models on depth encoders that use view synthesis [4] or approximations of that view synthesis [5], taking into account the synthesis distortion to perform an efficient selection of depth encoding parameters. In order to avoid this computationally intensive synthesis on the depth encoder, it is desirable to predict view synthesis distortion directly from depth data characteristics.

Here, we focus on the distortion model for a low complexity depth encoder for real-time applications, and we base it on: (i) the geometric distortion on view synthesis derived from the depth coding error and (ii) the pixel-mapping precision used in the view synthesis process. On the one hand, lossy coding modifies the original values of depth map pixels, and such errors introduce a deviation in the pixel-mapping process of view synthesis algorithms. We refer to this deviation as geometric distortion. On the other hand, the fact that pixels are rounded to integer positions, (sub-pixel precisions may be used but this is equivalent to frame upsampling with a later integer pixel-mapping) may affect the final geometric distortion value, thus it has to be taken into consideration too.

We analyze the effect of including this rounding on the computation of the geometric distortion through an statistical analysis of the correlation between (i) different models to capture the rounding effect on the geometric distortion and (ii) the actual distortion of synthesized views. Then, we use the different geometric distortion models in a depth rate-distortion optimization (RDO) algorithm to assess the results obtained by the correlation analysis. Due to the lack of an unique objective quality metric that predicts the subjective quality of synthesized views, we evaluate the RD performance using several objective quality metrics. The correlation and depth RDO results are consistent and show that one of the geometric distortion models achieves an average correlation gain of 27% and 16% with respect to the other ones.

This paper is organized as follows: in Section 2 we present our geometric distortion model, in Section 3 we show the experimental results, and in Section 4 we present the conclusions.

2. GEOMETRIC DISTORTION AND PIXEL-MAPPING PRECISION

The core of view synthesis, for a MVD data format is the well known Depth-Image Based Rendering (DIBR) algorithm [6], that uses 3D warping to project pixels from an original camera C_i^{orig} to a virtual camera C_j^{virt}. Without loss of generality, let us consider a parallel configuration of cameras, which happens to be a commonly used camera setting [3]. Then, 3D warping is restricted to horizontal pixel shifting [7]. Under those conditions, the position of a given pixel of C_i^{orig} is projected to C_j^{virt} by a horizontal disparity d:

$$d = \frac{f \times l}{Z} + du, \tag{1}$$

where f is the focal length, l is the distance between C_i^{orig} and C_j^{virt}, Z is the scene depth value of that pixel and du is the principal point horizontal difference. Note that in (1) a pixel may be

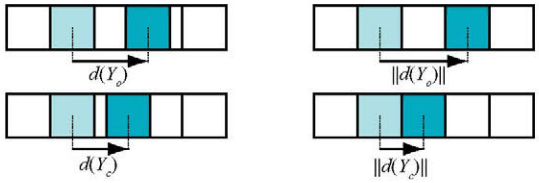


Figure 1. Example of rounding to integer-pixel positions in view synthesis algorithms for uncompressed and decoded depth pixel values. $||\cdot||$ indicates rounding to the nearest integer value.

mapped to a non-integer pixel position. For a fast implementation, rounding to an integer pixel position is performed whereas to favor synthesis quality, sub-pixel mapping may be preferred.

The scene depth value Z in (1) is represented on the gray-scale depth map by the following relationship:

$$Z = \frac{1}{\frac{Y}{255}\left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right) + \frac{1}{Z_{far}}}, \quad (2)$$

where Y is the pixel value of the depth image (between 0 and 255), and Z_{near} and Z_{far} define the depth range of the scene. From (1) and (2), the following relationship between a depth-pixel value Y and the disparity d can be derived:

$$d(Y) = \frac{f \times l}{255}\left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right) Y + \frac{f \times l}{Z_{far}} + du. \quad (3)$$

Even though view synthesis algorithms may use integer or sub-integer pixel precisions, the pixel-shifting in view synthesis that corresponds to a given value of $d(Y)$ has to be rounded according to the synthesis precision. Figure 1 shows an example of pixel-shifting with integer-pixel precision for an original pixel-depth value (Y_o) and for its decoded value (Y_c). If no rounding to integer-pixel positions is considered, the disparity error $\varepsilon_d(Y)$ is the following:

$$\varepsilon_d(Y) = d(Y_c) - d(Y_o), \quad (4)$$

and using (3):

$$\varepsilon_d(Y) = \frac{f \times l}{255}\left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right)(Y_c - Y_o). \quad (5)$$

However, rounding may be considered as part of $d(Y_c)$ and $d(Y_o)$ in (4), to capture geometric error between a view synthesized from decoded or uncompressed depth in practical DIBR implementations (Figure 1). Thus, $\varepsilon_d(Y)$ may not be strictly proportional to $(Y_c - Y_o)$. Rounding is effectively performed in view synthesis at the receiver using decoded data but also a virtual synthesis process with uncompressed data may be considered at the encoder. Our goal is to analyze whether incorporating this rounding to the geometric distortion error benefits its correlation with the synthesis distortion. Thus, we propose three different characterizations of ε_d depending on whether Y_o and/or Y_c are rounded to the nearest integer:

$$\begin{aligned} \varepsilon_d^{\mathbb{R}-\mathbb{R}}(Y) &= d(Y_c) - d(Y_o), \\ \varepsilon_d^{\mathbb{Z}-\mathbb{R}}(Y) &= \|d(Y_c)\| - d(Y_o), \\ \varepsilon_d^{\mathbb{Z}-\mathbb{Z}}(Y) &= \|d(Y_c)\| - \|d(Y_o)\|. \end{aligned} \quad (6)$$

Also, we compute the geometric distortion of a given depth region using the Sum of Absolute Error (SAE). Thus, according to (6) the geometric distortion of a depth map or depth map region ($D_{SAE}^{(\cdot)}$) is computed as follows:

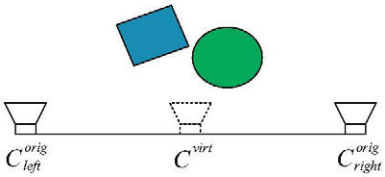


Figure 2. Parallel camera arrangement for the experimental tests, the virtual camera C^{virt} is located in the middle point between the two original cameras (C_{left}^{orig} and C_{right}^{orig}).

$$\begin{aligned} D_{SAE}^{\mathbb{R}-\mathbb{R}} &= SAE(\varepsilon_d^{\mathbb{R}-\mathbb{R}}\{Y_{i,j}\}), \\ D_{SAE}^{\mathbb{Z}-\mathbb{R}} &= SAE(\varepsilon_d^{\mathbb{Z}-\mathbb{R}}\{Y_{i,j}\}), \\ D_{SAE}^{\mathbb{Z}-\mathbb{Z}} &= SAE(\varepsilon_d^{\mathbb{Z}-\mathbb{Z}}\{Y_{i,j}\}), \end{aligned} \quad (7)$$

where $\{Y_{i,j}\}$ is the set of pixel values of that depth map region.

3. EXPERIMENTAL RESULTS

On our set of experiments, we consider the scenario depicted in Figure 2 with two original cameras in a parallel arrangement and one virtual camera in the middle position. The distance between C_{left}^{orig} and C_{right}^{orig} is 10 cm. An AVC encoder was used to encode depth (the mode setting has been fixed to intra-only mode). Both depth sequences are encoded independently. The multiview sequences used for the tests are the following: Newspaper, Kendo and Balloons (1024 × 768, 30 fps). These multiview sequences have been used by the MPEG group for the evaluation of 3D Video Coding (3DVC) technology [3]. The VSRS software [7] with integer-pixel accuracy was used for the view synthesis.

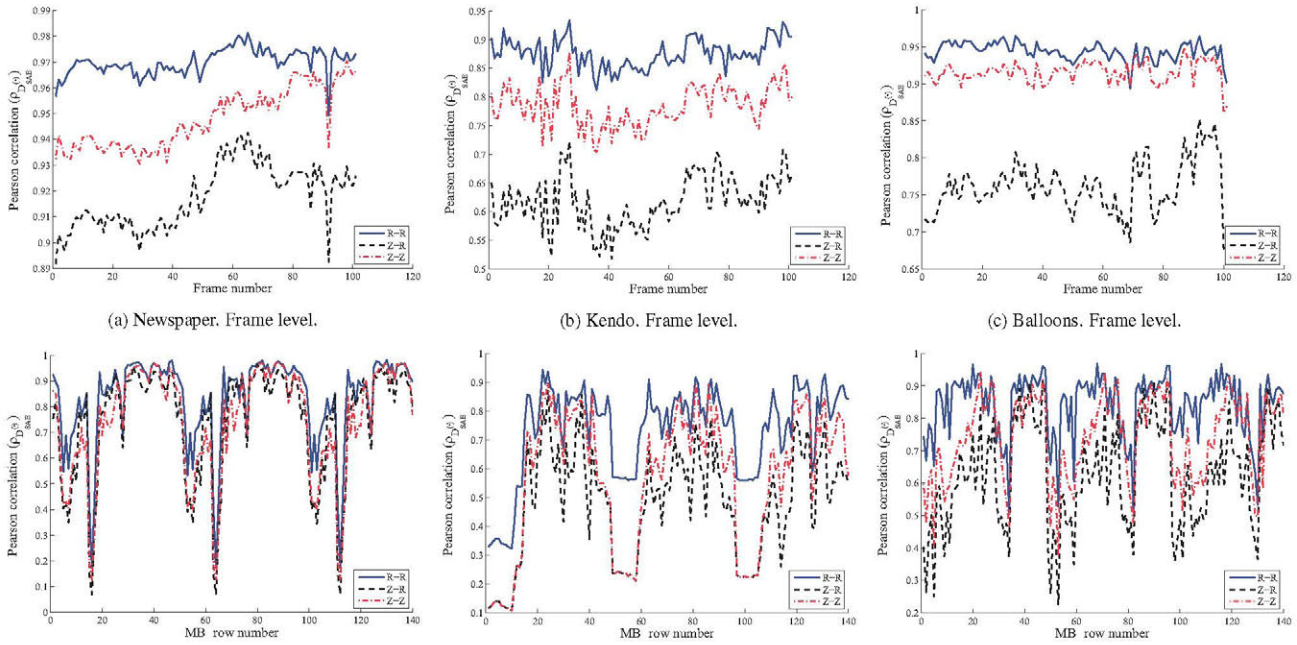
3.1. Correlation of geometric and synthesis distortions

For a given original camera position, and a depth-frame region $\{Y_{i,j}\}$, this depth data is encoded with all possible Quantization Parameter (QP) values. Then, $D_{SAE}^{(\cdot)}(QP)$ is computed for the decoded data. Also, the associated synthesis distortion $D_{SSE}^{Synth}(QP)$ is computed using the decoded depth data, uncompressed original color images and uncompressed depth of the other original view. Then, we compute the Pearson correlation ($corr(X,Y)$) between the geometric and synthesis distortion as follows:

$$\rho_{D_{SAE}^{(\cdot)}} = corr(\{D_{SAE}^{(\cdot)}(QP)\}, \{D_{SSE}^{Synth}(QP)\}), \quad (8)$$

where $\{D_{SAE}^{(\cdot)}(QP)\}$ and $\{D_{SSE}^{Synth}(QP)\}$ are arrays of distortion values for all QP values for that depth region.

We have evaluated this correlation for different sizes of $\{Y_{i,j}\}$: (i) a whole frame and (ii) a full row of MBs within a frame (MB-row). Some example of the obtained correlation results are shown in Figure 3. Each graph shows the value of $\rho_{D_{SAE}^{(\cdot)}}$ for different frames or MB-row positions. The results show that $D_{SAE}^{\mathbb{R}-\mathbb{R}}$ has the highest correlation with D_{SSE}^{Synth} among the three geometric distortion metrics both at frame and MB-row levels. For the analyzed sequences and at the frame level, $\rho_{D_{SAE}^{\mathbb{R}-\mathbb{R}}}$ is within a range of high values: $[0.81, 0.98]$. Looking at the MB-row level, the variation in the correlation value range is significantly higher, where the correlation might severely deteriorate at several MB-rows. Those are MB-rows with homogeneous color for which depth coding distortion does not entail synthesis distortion. The results of the average



(a) Newspaper. Frame level. (b) Kendo. Frame level. (c) Balloons. Frame level.

(d) Newspaper. MB-row level. (e) Kendo. MB-row level. (f) Balloons. MB-row level.

Figure 3. Pearson correlation of the three different distortion metrics ($D^{\mathbb{R}-\mathbb{R}}_{SAE}$, $D^{\mathbb{Z}-\mathbb{R}}_{SAE}$, $D^{\mathbb{Z}-\mathbb{Z}}_{SAE}$) and the synthesis distortion (D^{Synth}_{SSE}) at the frame and MB-row levels. Results for the left original camera in all the cases and 100 frames.

Table 1. Values of $\Delta\rho(\mathbb{R}-\mathbb{R},\mathbb{Z}-\mathbb{R})$ and $\Delta\rho(\mathbb{R}-\mathbb{R},\mathbb{Z}-\mathbb{Z})$ on the MB-row level for all tested sequences, and both original camera positions.

Sequence	View	$\Delta\rho(\mathbb{R}-\mathbb{R},\mathbb{Z}-\mathbb{R})$	$\Delta\rho(\mathbb{R}-\mathbb{R},\mathbb{Z}-\mathbb{Z})$
Newspaper	Left	14.82	13.57
	Right	8.16	3.94
Kendo	Left	54.94	36.05
	Right	30.31	20.78
Balloons	Left	34.45	13.07
	Right	19.55	9.28
Average		27.04	16.11

correlation gain for the MB-row case are shown in Table 1, and computed using:

$$\Delta\rho(\mathbb{R}-\mathbb{R},(\cdot))(\%) = 100 \times \frac{\rho_{D^{\mathbb{R}-\mathbb{R}}_{SAE}} - \rho_{D^{(\cdot)}_{SAE}}}{\rho_{D^{(\cdot)}_{SAE}}}. \quad (9)$$

The results in Table 1 show that $D^{\mathbb{R}-\mathbb{R}}_{SAE}$ is the best option to be used in depth RDO algorithms, as it shows the best correlation with the synthesis distortion. Furthermore, it has the lowest complexity as it is linearly proportional with the depth coding error.

3.2. Rate-Distortion Optimization

To validate the conclusions extracted in Section 3.1, we have implemented a basic depth RDO algorithm at the MB level. For each depth MB, the QP that meets the following condition is selected:

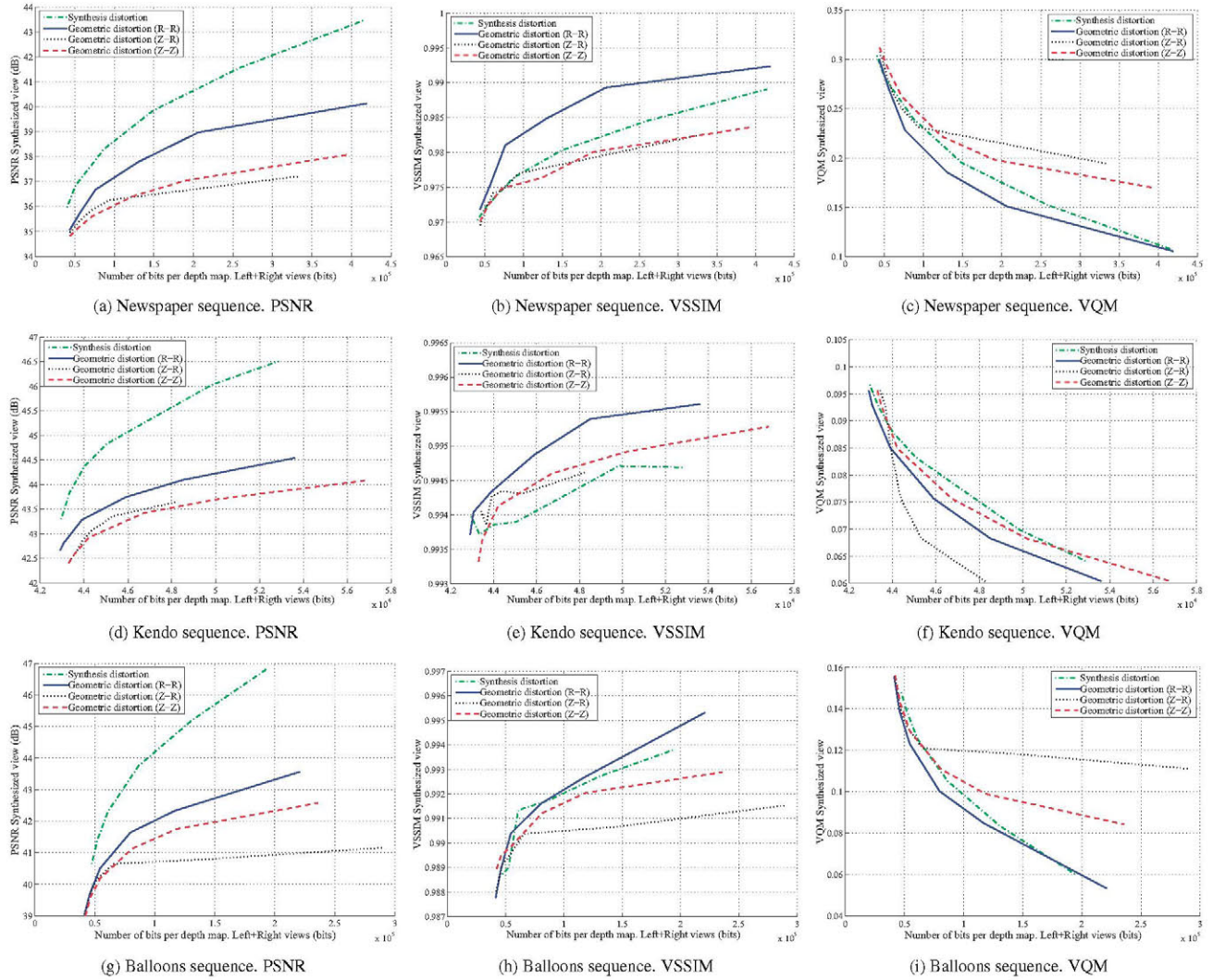
$$\min(bits(QP)) | D^{(\cdot)}_{SAE}(QP) < T_D, \quad (10)$$

where T_D is a predefined distortion threshold value (to adjust rate point). The resulting decoded depth maps at each rate point were then used to synthesize a virtual view at C^{virt} using the uncompressed original color images. The obtained synthesized views are evaluated using different video quality metrics: PSNR, VSSIM [8], and VQM [9]

In this comparison, we have used as ground truth the RD results of an equivalent depth RDO algorithm that uses synthesis distortion ($D^{Synth}_{SSE}(QP)$), i.e., the SSE of the target synthesized view obtained from the original depth image modified by the coded MB pixels in the corresponding MB position. The video quality results for the synthesized views versus the number of bits of the encoded depth maps are shown on Figure 4. The comparative RD results show consistent results with the correlation results for all quality measures and bitrates. Among the graphs, the one corresponding to $D^{\mathbb{R}-\mathbb{R}}_{SAE}$ shows the best RD performance, except for part of the bitrate range and the VQM metric in the case of the Kendo sequence. Furthermore, while for PSNR $D^{Synth}_{SSE}(QP)$ outperforms all geometric distortion approaches, for the other measures $D^{\mathbb{R}-\mathbb{R}}_{SAE}$ outperforms these ground truth results.

4. CONCLUSIONS AND FUTURE WORK

The correlation results and the performance on depth RDO algorithms show that the proposed geometric distortion model can be used as a reasonable estimator of the distortion of synthesized views. In this sense, $\varepsilon^{\mathbb{R}-\mathbb{R}}_d(Y)$ is the best option among the alternatives that we have presented to combine the geometric distortion derived from depth coding error and pixel-precision on view synthesis. Furthermore, this metric requires the lowest complexity level among the three as it is linearly proportional to the depth coding error. Thus, the suggested metric can be employed for a low-complexity but high quality depth map compression engine, which does not have to carry out DIBR process using color images. Nevertheless, the analysis of the correlation at a MB-level has shown that there are cases in which the correlation between the geometric distortion and synthesis distortion is low. Thus, further research to address these cases is encouraged.



(a) Newspaper sequence. PSNR (b) Newspaper sequence. VSSIM (c) Newspaper sequence. VQM

(d) Kendo sequence. PSNR (e) Kendo sequence. VSSIM (f) Kendo sequence. VQM

(g) Balloons sequence. PSNR (h) Balloons sequence. VSSIM (i) Balloons sequence. VQM

Figure 4. RD performance of depth encoders using the analyzed geometric distortion metrics and synthesis distortion in the RDO cycle. The objective quality of the synthesized view is plotted against the bitrate of the coded depth bitstreams.

5. ACKNOWLEDGEMENT

This work has been partially supported by the Ministerio de Economía y Competitividad of the Spanish Government under project TEC2010-20412 (Enhanced 3DTV). Also, P. Carballeira wishes to thank the Comunidad de Madrid for a personal research grant.

6. REFERENCES

[1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards," in *Proc. of IEEE ICME*, Jul. 2006, pp. 2161–2164.

[2] M. Tanimoto, M.P. Tehrani, T. Fujii, and T. Yendo, "Free-Viewpoint TV," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 67–76, Jan. 2011.

[3] ISO/IEC JTC1/SC29/WG11, "Report on Experimental Framework for 3D Video Coding," *output doc. N11631*, Guangzhou, China, Oct. 2010.

[4] P. Carballeira, G. Tech, J. Cabrera, K. Müller, F. Jaureguizar, T. Wiegand, and N. García, "Block based Rate-Distortion analysis for quality improvement of synthesized views," in *Proc. of IEEE 3DTV-CON*, Jun. 2010, pp. 1–4.

[5] D.V.S.X. De Silva, W.A.C. Fernando, S.T. Worrall, and A.M. Kondoz, "A novel depth map quality metric and its usage in depth map coding," in *Proc. of IEEE 3DTV-CON*, May 2011, pp. 1–4.

[6] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems*, Jan. 2004, vol. 5291, pp. 93–104.

[7] D. Tian, P. Lai, P. Lopez, and C. Gomila, "View synthesis techniques for 3D video," *Applications of Digital Image Processing XXXII*, vol. 7443, no. 1, pp. 74430T, 2009.

[8] K. Seshadrinathan and A.C. Bovik, "A Structural Similarity Metric for Video Based on Motion Models," in *Proc. of IEEE ICASSP*, Apr. 2007, vol. 1, pp. I–869–I–872.

[9] M.H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, Sep. 2004.