ANALYTICAL MODEL FOR CAMERA DISTANCE RELATED 3D VIRTUAL VIEW DISTORTION ESTIMATION

Yijian Xiang[1], *Ngai-Man Cheung*[2], *Juyong Zhang*[1] *and Lu Fang*[1]

[1]University of Science and Technology of China
[2]Singapore University of Technology and Design

ABSTRACT

We propose an analytical model to estimate the depth-error-induced synthesis distortion in 3D video, taking into account the configuration of the cameras. In particular, the model mathematically relates the *Distance* between camera positions (reference view and virtual view) to the *Virtual View Distortion* (VVD), thus it is denoted as DVVD model. Specifically, the DVVD model accounts for two modules: *distribution of disparity errors* and *shift-induced distortion*. The former one is derived under a Laplacian distribution assumption of depth errors, and the latter one is estimated under a Quadratic model. We further propose a linear Steady-State model by performing Taylor series approximation of the DVVD model over a region of practical interest. Experiment results demonstrate that both the DVVD and Steady-State models are capable of estimating the relationship between VVD and the distance between virtual/reference view. Therefore, our model can effectively inform camera setup for capturing, in particular, the set-up of the cameras in situation where depth information will be compressed subsequently.

I. INTRODUCTION

I-A. Motivation

With new requirements for visual experience beyond traditional 2D video are proposed, 3D video technology has attracted great interest. In 3D video datasets, a texture-plus-depth representation [1] is often used to describe a 3D scene, where texture sequences (color images) are captured by multiple cameras at different locations, and associated depth sequences (gray scale images) are captured directly by depth cameras or estimated from texture data. To facilitate free selection of viewpoint, user-chosen virtual view sequences need to be synthesized from captured sequences, using depth image-based rendering (DIBR) [2].

Due to lossy compression [3], there always exist distortions between original scenes and acquired texture as well as depth data. While errors in texture data affect color information in synthesized virtual view, depth errors will lead to disparity errors in virtual view. In other words, there is a shift from the correct position for a pixel in virtual view due to associated inaccurate depth information. Referring to [4] [5], texture errors and depth errors can be modeled separately when considering their contribution to distortion in virtual view. Since influence of depth errors is more complicated, in this paper, we focus on how depth errors affects distortion in virtual view.

As noted in [6] [7] [8], synthesis distortion model is needed for R-D optimized bit-allocation in 3D video coding, especially under the condition of constrained bits resource. Also synthesis distortion model may help design the array of capturing cameras by suggesting a proper spacing [9]. Taking two cameras system for example, i.e. left and right reference views synthesizing an intermediate virtual view with associated weight, the synthesis distortion consists of three parts: distortions caused by errors in left and right reference views, and the correlation between the errors in the left reference view and errors in right reference view. Thus, the model of VVD caused by errors in one reference view is imperative, and this is the focus of this work.

I-B. Related Work

In previous work, Cheung [7] and Velisavljevic [10] derive models with distance between virtual view and reference view as a variable for VVD, based on stationary assumption of video signal. The parameters in their model are estimated from empirical results of synthesis distortion. However, the stationary assumption may not hold for sequences containing strong texture edges [5].

Fang and Cheung [5] propose a model based on Power Spectral Density (PSD) and spatial analysis. However, distance between reference view and virtual view is not explicitly expressed in the final model. Thus, it is not clear to see how distance may affect VVD. In our work, we explicitly discuss the relationship between camera distance and VVD.

Kim *et. al.* [11] focus on rate-distortion optimization for depth map coding and propose a local estimation sum of squared error (SSE), which is VVD at block level. The SSE is expressed in terms of variance of a video block and an autoregressive model for correlation coefficient. In our work, VVD is estimated on the whole frame, which is a global estimation.

Yuan *et. al.* [4] derive a VVD model covering texture and depth errors. The effects of errors in texture and depth data are decoupled, using an analysis based on Taylor expansion theory. In particular, VVD caused by depth errors is characterized by a linear model of mean squared error of disparity. A linear relationship between VVD and quantization step is applied to their optimal bit-allocation algorithm.

I-C. Our Contributions

We propose to estimate VVD caused by depth errors in one reference view, taking into account the camera configuration. In particular, our proposed DVVD model is a function of distance between reference view (camera location) and virtual view, from which we can effectively analyze how distance affects VVD. More specifically, our model contains two modules: distribution of disparity errors and shift-induced distortion in reference view. Depth errors due to lossy compression can be modeled as Laplacian distributed random variables. Using the linear relationship between disparity errors and depth errors, we model the distribution of disparity errors. Shift-distortion, which refers to mean square error

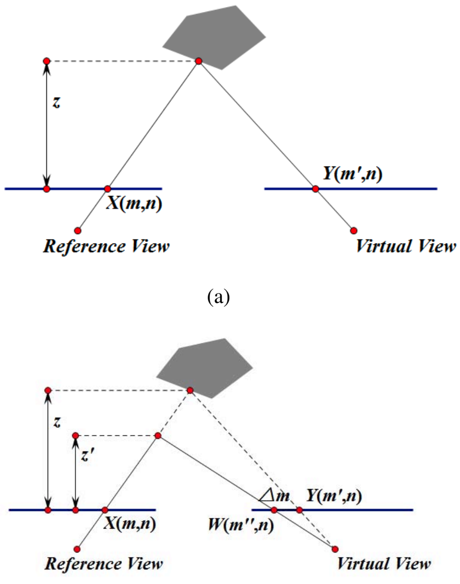
978-1-4799-5751-4/14/$31.00 ©2014 IEEE

(MSE) caused by the pixel position shifting, is estimated using a Quadratic model. For a region of practical interest, we further approximate DVVD model with Taylor series approximation and obtain a Steady-State model. Both the DVVD and Steady-State models are verified using sequences and tools from MPEG 3DV activities [12]. In summary, our contributions are:

- The proposed DVVD model takes camera configuration into account. It characterizes the relationship between VVD and the distance between reference / virtual views.
- A more precise quadratic model is specifically proposed to estimate shift-distortion in reference view, which is usually approximated by a linear model in previous work.
- A Steady-State model is proposed based on Taylor series approximation of the DVVD model within a particular range. The Steady-State model could estimate VVD well with simple closed-form linear formulation.

The rest of the paper is organized as follows: Section II discusses the proposed DVVD model and Steady-State model. The performance of proposed models is testified in Section III. Finally Section IV concludes our work.

II. PROPOSED DVVD MODEL



(a)

(b)

Fig. 1. (a) Warping model with uncoded depth map; (b) Warping model with encoded depth map.

Fig.1(a) suggests a basic model of warping pixels from reference view to virtual view. We have

$$m - m' = \frac{fb}{z}, \tag{1}$$

where f is focal length, b is distance between virtual view and reference view, z is physical depth value, m and m' are horizontal positions of a pixel before and after warping respectively.

Due to the compression of depth images, there are depth errors (denoted as Δd) in depth images, which lead to disparity errors (denoted as Δm) in virtual view accordingly. Fig.1(b) shows how depth error affects the position of a pixel in warping. For ease of analysis, we assume cameras are placed in a 1-D horizontal array, such that there is only horizontal disparity error $\Delta m = m' - m''$. Then we have

$$\Delta m = fb\left(\frac{1}{z'} - \frac{1}{z}\right), \tag{2}$$

where z' is reconstructed depth value after compression. In addition,

$$\frac{1}{z} = \frac{D(m,n)}{255}\left(\frac{1}{z_{near}} - \frac{1}{z_{far}}\right) + \frac{1}{z_{far}}, \tag{3}$$

where $D(m,n)$ denotes depth image, z_{near} and z_{far} are nearest and farthest depth values.

Substitute (3) into (2), we have

$$\Delta m = \frac{f\Delta d}{255}\left(\frac{1}{z_{near}} - \frac{1}{z_{far}}\right)b. \tag{4}$$

Examining (4), f, Z_{near} and Z_{far} are camera physical parameters that remain unchanged under different compression setups. b identifies the distance between virtual view and reference view, which reflects the camera configuration and is the prime element we intend to investigate. In other words, we expect to explore the relationship between virtual view distortion and distance b. The disparity error Δm and depth error Δd are highly related pair, i.e., Δm has a linear relation with Δd under the constant camera parameters.

To measure the Virtual View Distortion (VVD) caused by depth errors (denoted as $E[Z^2]$), we can compute the Mean Square Error (MSE) between virtual images that are generated without depth error (Y in Fig. 1(a)) and with depth error (W in Fig. 1(b)), i.e.,

$$\begin{aligned} E[Z^2(m,n)] &= E[Y(m,n) - W(m,n)]^2 \\ &= E[Y(m,n) - Y(m - \Delta m, n)]^2. \end{aligned} \tag{5}$$

After the identical transformation, (5) can be represented as

$$E[Z^2(m,n)] = E_{\Delta m}\{E\{[Y(m,n) - Y(m - \Delta m, n)]^2 | \Delta m\}\}, \tag{6}$$

where $E_{\Delta m}(\cdot)$ is expectation operator applied to variable Δm and the conditional expectation $E\{[Y(m,n) - Y(m - \Delta m, n)]^2 | \Delta m\}$ measures the distortion between virtual view pixel $Y(m,n)$ and shifted virtual view pixel $Y(m - \Delta m, n)$ when Δm remains constant. Let us denote $E\{[Y(m,n) - Y(m - \Delta m, n)]^2 | \Delta m = i\}$ as $MSE_{\Delta m = i}$, then (6) is rewritten as

$$E[Z^2(m,n)] = \sum_{i=-\infty}^{\infty} p(\Delta m = i) MSE_{\Delta m = i}. \tag{7}$$

Examining (7), the estimation of depth-error-induced VVD can be divided into two modules: distribution of disparity errors ($p(\Delta m = i)$) and shift-distortion ($MSE_{\Delta m = i}$). In the following subsections, we will elaborate them respectively.

II-A. Distribution of Disparity Errors

Without loss of generality, we show the distribution of representative empirical depth errors in Fig 2, where depth errors are caused by lossy compression. It can be approximated that the depth error is modeled as Laplacian distributed random variables, i.e.,

$$f(\Delta d | \mu, \beta) = \frac{1}{2\beta} \exp\left(-\frac{|\Delta d - \mu|}{\beta}\right), \tag{8}$$

where μ is the mean value of depth errors, $\beta = \sqrt{\sigma^2/2}$, and σ^2 is variance.

To simplify the derivation, we set $\mu = 0$, i.e., the depth errors are zero-mean variables. Such assumption is based on the

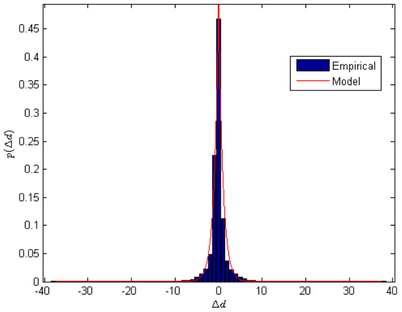


Fig. 2. Example for the fitting of representative empirical depth error distribution (blue) and Laplacian distribution model (red).

experimental facts that maximum likelihood estimation of mean value μ is approximately to be zero for most of frames.

Due to the linear relationship between depth error Δd and disparity error Δm in (4), Δm follows Laplacian Distribution as well,

$$f(\Delta m|kb\mu, kb\frac{\sigma}{\sqrt{2}}) = \frac{1}{2kb\frac{\sigma}{\sqrt{2}}} \exp\left(-\frac{|\Delta m - kb\mu|}{kb\frac{\sigma}{\sqrt{2}}}\right), \tag{9}$$

where

$$k = \frac{f}{255}(\frac{1}{z_{near}} - \frac{1}{z_{far}}). \tag{10}$$

Given that the disparity error Δm is discrete, we integrate continuous Laplacian Distribution to acquire discrete distribution of Δm. For example, we integrate probability density function in an interval such as $(-0.5, 0.5)$ to get the probability mass at $\Delta m = 0$. Consequently, the probability mass function of Δm is

$$p(\Delta m = i) = \begin{cases} \left[1 - \exp\left(-\frac{1}{\sqrt{2}kb\sigma}\right)\right], & \text{for } i = 0; \\ \frac{1}{2}\left[\exp\left(-\frac{2|i|-1}{\sqrt{2}kb\sigma}\right) - \exp\left(-\frac{2|i|+1}{\sqrt{2}kb\sigma}\right)\right], & \text{for } i \neq 0, \end{cases} \tag{11}$$

where $p(\Delta m = i)$ reaches top at $i = 0$, and shows exponential descending in both sides.

II-B. Shift-Distortion Estimation

Note that we propose to estimate the rendering error caused by the depth errors without actually performing any warping / synthesis of the virtual view. In other words, Y is in fact unknown. Fortunately, the content of X and Y would be similar when view distance b is constrained in a reasonable distance, as they represent the same scene at slightly different view angles [5] [11]. Therefore, we approximate $MSE_{\Delta m=i}$ using the distortion between $X(m,n)$ and $X(m - \Delta m, n)$ instead of $Y(m,n)$ and $Y(m - \Delta m, n)$,

$$MSE_{\Delta m=i} \simeq \frac{1}{MN} \sum_{(m,n)} [X(m,n) - X(m - \Delta m, n)]^2, \tag{12}$$

where $M \times N$ is the spatial dimension of a sequence.

Inspired from Kim *et. al.* [11] and Velisavljevic *et. al.* [10], we have a model of covariance of texture image pixels,

$$E\left[X(m,n)X(m + \Delta m, n)\right] = s_t^2\left(1 - (1-\rho)\cdot|\Delta m|\right), \tag{13}$$

where $s_t^2 = E[X^2(m,n)] = E[X^2(m + \Delta m, n)]$ and ρ is a covariance of the texture image pixels for a unit shift $k = 1$. From Eq. (13), we can derive a linear model for shift-distortion,

$$E[X(m,n) - X(m - \Delta m, n)]^2 = 2s_t^2(1-\rho)|\Delta m|. \tag{14}$$

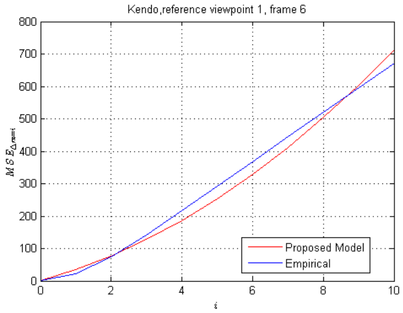


Fig. 3. Empirical $MSE_{\Delta m=i}$(Blue) and proposed model $a_1 i + a_2 i^2$ (Red).

However, (13) works based on an assumption that video signal is stationary and zero mean, which does not hold for sequence containing strong texture edges [5], as pixel intensity changes quickly across strong texture edge than that in non-edge region and correlation statistics around edge region is much different from non-edge region. To make the shift-distortion model more flexible, we specifically add a Quadratic term to the linear model, such that the shift-distortion is modeled as

$$MSE_{\Delta m=i} = a_1 i + a_2 i^2, \tag{15}$$

where the parameters a_1 and a_2 are image characteristics dependent, which can be well estimated from representative empirical values of $MSE_{\Delta m=i}$ based on Least Square Estimation (LSM). As illustrated in Fig.3, the model of (15) describes the tendency of $MSE_{\Delta m=i}$ well.

II-C. Conclusion of DVVD Model

We combine models of $p(\Delta m = i)$ in Section II-A and $MSE_{\Delta m=i}$ in Section II-B to derive our DVVD model for $E[Z^2(m,n)]$ in (7). After solving summation of infinite series, we have

$$E[Z^2] = \frac{a_1 \exp\left(-\frac{1}{\sqrt{2}kb\sigma}\right)}{1 - \exp\left(-\frac{\sqrt{2}}{kb\sigma}\right)} + a_2 \frac{\exp\left(-\frac{1}{\sqrt{2}kb\sigma}\right) + \exp\left(-\frac{3}{\sqrt{2}kb\sigma}\right)}{\left[1 - \exp\left(-\frac{\sqrt{2}}{kb\sigma}\right)\right]^2}, \tag{16}$$

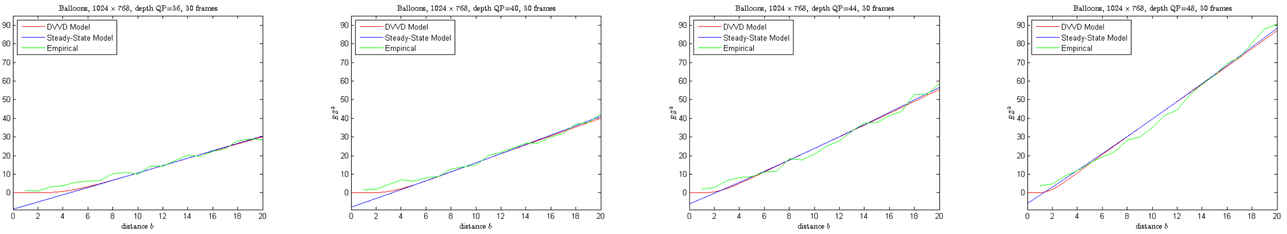
where both of the two terms after a_1 and a_2 increase monotonously with distance b between virtual view and reference view. Due to the complex formulation in (16), we further propose a Steady-State model by performing Taylor series expansion over (16) within interested range, and adopting low order expansion terms as follows

$$E[Z^2] = w_0 + w_1(b - b_0) \tag{17}$$

where b_0 is expanding point and

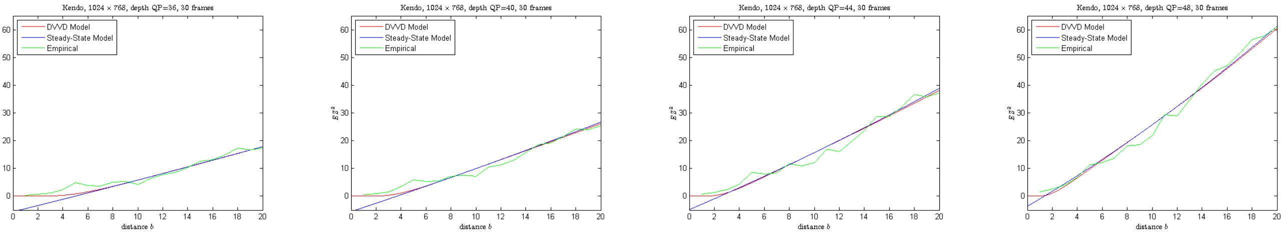
$$w_0 = a_1 \frac{\exp\left(-\frac{1}{\sqrt{2}kb_0\sigma}\right)}{1 - \exp\left(-\frac{\sqrt{2}}{kb_0\sigma}\right)} + a_2 \frac{\exp\left(-\frac{1}{\sqrt{2}kb_0\sigma}\right) + \exp\left(-\frac{3}{\sqrt{2}kb_0\sigma}\right)}{\left[1 - \exp\left(-\frac{\sqrt{2}}{kb_0\sigma}\right)\right]^2}. \tag{18}$$

Similarly, we have w_1. Note that we expand at the midpoint of the distance interval that we are interested in. For example, in our experiments, we explore the distance interval $(0, 20)$, thus we expand at $b_0 = 10$. Examining Eq. (18), the coefficients w_0 and w_1 are independent of distance b, but related to k (given in (10)) and σ. Note that k is constant when cameras array are well calibrated, and σ is related to the distribution of depth errors. With the Steady-State model, the depth-error-induced VVD can be easily estimated



(a) $QP = 36$ (b) $QP = 40$ (c) $QP = 44$ (d) $QP = 48$

Fig. 4. Comparisons of DVVD model (red curve), Steady-State model (blue curve), and the empirical case (green curve) for Balloons sequence under different depth comression (depth QP) settings.



(a) $QP = 36$ (b) $QP = 40$ (c) $QP = 44$ (d) $QP = 48$

Fig. 5. Comparisons of DVVD model (red curve), Steady-State model (blue curve), and the empirical case (green curve) for Kendo sequence under different depth comression (depth QP) settings.

using a closed-form linear model. Both the DVVD and Steady-State models take account of the camera configuration and reflect the visualized relation between the virtual view distortion ($E[Z^2]$) and view distance b.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, both the DVVD and Steady-State models are verified by comparing to the empirical case using representative 3D test sequences [12]. As we focus on depth-error-induced VVD estimation, the associated texture maps are uncoded, while depth maps are encoded with JMVC Encoder 8.3.1 [13]. We consider Quantization Parameter (QP) 36, 40, 44 and 48 for encoding depth maps. Similar to state-of-art work, VSRS 3.5 [14] is used to warp reference view to virtual view for generating the empirical depth-error-induced VVD.

Fig. 4 and 5 show the comparisons between our models and empirical results for Balloons and Kendo (1024×768) sequences. The results of DVVD model and Steady-State model are computed from (16) and (17) respectively. It can be shown that DVVD model (red curve) is capable in estimating the empirical (green curve) relationship between depth-error-induced VVD and the camera distance (b) under almost all the QP cases.

In terms of Steady-State model (blue curve), we notice that:

- For relatively large b, the Steady-State model fits extremely well with the DVVD model as well as the empirical situation. With the decrease of b, a gap arises. This is because we merely consider low order (linear) terms in Steady-State model, which fails to revel the non-linear characteristic under small b.
- For small distance b, the result of Steady-State model is always an under-estimation of the empirical result. This can be explained that under small distance b, the result of (17) may be negative, and those negative values significantly lower down the value of estimated distortion. A possible solution is to set negative values as zero, such that the Steady-State model will be a piecewise linear model.
- Different depth QP tends to correspond to different 'slope' in empirical case. While examining the formulation of our Steady-State model, different QP leads to different σ, then leads to different w_1, and therefore different slope. In other words, such observation fully demonstrates the usefulness of our Steady-State model.

IV. CONCLUSION AND FUTURE WORK

By modeling the distribution of disparity errors based on Laplacian distribution approximation of depth errors, and modeling shift-distortion using a quadratic model, We propose a camera distance related DVVD model as well as a Steady-State model to estimate depth-error-induced virtual view distortion. The proposed model effectively illustrates a visualized relation between VVD and camera configuration (distance between virtual view and reference view), which can be utilized to inform camera setup for capturing, and how cameras should be set up in situation where depth information will be compressed subsequently.

V. ACKNOWLEDGMENT

The material reported in this document is supported by the SUTD-MIT International Design Centre (IDC). Any findings, conclusions, recommendations, or opinions expressed in this document are those of the author(s) and do not necessary reflect the views of the IDC.

This work has been supported in part by Natural Science Foundation of China (NSFC) under contract (No. 61303151, 61303148, 61331015, 61390514), Specialized Research Fund for the Doctoral Program of Higher Education under contract (No. 20133402120029, 20133402120002), the Fundamental Research Funds for the Central Universities under contract No. WK2100060012.

VI. REFERENCES

[1] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *IEEE International Conference on Image Processing (ICIP)*, 2007.

[2] C. Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," in *Electronic Imaging*, 2004.

[3] A. Sánchez, G. Shen, and A. Ortega, "Edge-preserving depth-map coding using graph-based wavelets," in *Asilomar Conference on Signals, Systems and Computers (ACSSC)*, 2009.

[4] H. Yuan, Y. Chang, J. Huo, F. Yang, and Z. Lu, "Model-based joint bit allocation between texture videos and depth maps for 3-d video coding," *IEEE Trans. on Circuits and Systems for Video Technology (ITCSVT)*, vol. 21, no. 4, pp. 485–497, 2011.

[5] L. Fang, N. M. Cheung, D. Tian, A. Vetro, H. Sun, and O. C. Au, "An analytical model for synthesis distortion estimation in 3d video," *IEEE Trans. on Image Processing (TIP)*, vol. 23, no. 1, pp. 185–199, 2013.

[6] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Mueller, T. Wiegand, et al., "The effects of multiview depth video compression on multiview rendering," *Signal Processing: Image Communication*, vol. 24, no. 1, pp. 73–88, 2009.

[7] G. Cheung, V. Velisavljevic, and A. Ortega, "On dependent bit allocation for multiview image coding with depth-image-based rendering," *IEEE Trans. on Image Processing (TIP)*, vol. 20, no. 11, pp. 3179–3194, 2011.

[8] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map distortion analysis for view rendering and depth coding," in *IEEE International Conference on Image Processing (ICIP)*, 2009.

[9] N.M. Cheung, D. Tian, A. Vetro, and H. Sun, "On modeling the rendering error in 3d video," in *IEEE International Conference on Image Processing (ICIP)*, 2012.

[10] V. Velisavljevic, G. Cheung, and J. Chakareski, "Bit allocation for multiview image compression using cubic synthesized view distortion model," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2011.

[11] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map coding with distortion estimation of rendered view," in *IS&T/SPIE Electronic Imaging*, 2010.

[12] MPEG Video and Requirement group, "Call for proposals on 3d video coding technology," *Tech. Rep., MPEG, 2011, MPEG N12036*.

[13] Alexis Michael Tourapis, Athanasios Leontaris, K Suhring, and Gary Sullivan, "H. 264/14496-10 avc reference software manual," *Doc. JVT-AE010*, 2009.

[14] M. Tanimoto, T. Fujii, and K. Suzuki, "View synthesis algorithm in view synthesis reference software 2.0 (vsrs2.0)," *Tech. Rep., MPEG, 2009, MPEG M16090.*