JOINT TEXTURE AND DEPTH MAP CODING FOR ERROR-RESILIENT 3-D VIDEO TRANSMISSION

Pan Gao[⋆] *Wei Xiang*[†] *D. M. Motiur Rahaman*[‡] *Manoranjan Paul*[‡]

[⋆] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China
[†] College of Science and Engineering, James Cook University, Australia
[‡] School of Computing and Mathematics, Charles Sturt University, Australia

ABSTRACT

This paper addresses the problem of error-resilient source coding for 3-D video transmission over packet-loss networks. The proposed approach jointly optimizes the texture coding mode and the depth coding mode for each macroblock in the reference views. Firstly, a distortion model is developed to capture the effect of the texture distortion and depth distortion on the synthesized view. Then, joint optimization of texture and depth coding modes is derived based upon an operational rate-distortion framework using Lagrange multiplier method. In particular, a dual trellis-based algorithm is introduced in order to overcome the macroblock interdependencies of texture and depth map in the optimization procedure. Simulation results demonstrate that significant and consistent gains can be achieved over currently used techniques.

Index Terms— Texture coding mode, depth coding mode, error-resilient coding, 3-D video transmission

1. INTRODUCTION

3-D video has attracted considerable attention in recent years [1]. In order to obtain high coding efficiency, 3-D video compression strategies employ a multimode methodology [2]. As is well-known, in mono-video coding, there are mainly two coding modes, i.e., inter coding mode and intra mode. In 3-D video coding, in addition to these two traditional modes, several other coding modes have been developed to specifically adapt to the properties of the depth-based 3-D video coding. To exploit the redundancy between different camera views of the same scene, a straightforward extension of the inter mode, namely, the inter-view coding mode, is introduced for both texture and depth map coding. The inter-view mode differs from the inter mode in that, the picture in a different view is utilized as the reference picture of the current coding block for motion compensation. To further exploit the statistical dependencies between the texture and depth, the view synthesis prediction (VSP) mode has been proposed in enhanced texture coding [3], [4]. In VSP, the texture picture at the reference view is warped to the target view position using the available depth information, and then the synthetic reference view is used to predict the texture block at the target view.

By allowing multiple modes of operation to cater to different types of scene statistics, the rate-distortion performance of 3-D video can be significantly improved. However, the compressed 3-D video bit stream is highly susceptible to transmission error. When packet loss errors occur to texture and depth images, transmission errors can propagate temporally and spatially to subsequent frames due to the coding dependencies established in the inter and inter-view modes, respectively. Furthermore, the inter-component error propagation may take place if the VSP mode is employed. Finally, since the intermediate virtual views are usually synthesized from the decoded texture and depth images of the reference views, the existing errors in the texture and depth can still spread to the rendered views, which leads to catastrophic deterioration of the received 3-D video signal. As a consequence, in order to maintain high quality of synthesized views for auto-stereoscopic displays, it is crucial to develop error-resilient mechanisms at the encoder that can reduce the quality degradation caused by the unreliability of the networks.

There have been previous attempts to enhance the robustness of the 3-D video communication system to packet loss. Machiavello *et al.* introduced a block-level reference frame optimization scheme for loss-resilient depth map coding [5]. In this algorithm, the synthesized view distortion sensitivity to the reconstructed depth map errors is firstly approximated by a per-pixel quadratic weighting function, and then the depth reference block is selected to pro-actively minimize the expected synthesized view distortion subject to the bit rate constraint. Thereafter, this idea was extended to the encoding of both texture and depth map [6], and was augmented with an adaptive blending error concealment approach. In these two algorithms, inter-view error propagation on transmission distortion is not considered, and the impact of the rounding error incurred by sub-pixel mapping on view synthesis distortion is ignored. Also, these two methods optimize the texture and depth map in an approximately independent way. That is, the quadratic distortion model is employed for depth coding, and the conventional distortion metric is used for texture coding. In [7], in order to further improve the overall quality of reconstructed 3-D video in packet loss scenario, a loss-aware rate-distortion optimized coding mode switching scheme was presented for joint texture and depth map coding, in which the summative end-to-end distortions of both the rendered view and coded texture video are characterized and used as the distortion measure. The optimization of texture coding mode with respect to depth mode is achieved through a local exhaustive search where the intra, inter, and inter-view modes are considered. In this scheme, inter-component error propagation is not taken into account.

Besides the above mentioned shortcomings, all the previous schemes share one common drawback, i.e., the encoder operation of each block is independently optimized. However, in a 3-D video coding system, since the displacement vector field exhibits a large spatial correlation, a differential pulse code modulation-based scheme is usually used for encoding the motion/disparity vectors. In this case, a bit rate coupling may occur among blocks. In addition, in the error concealment process, when a block is lost, the concealment motion vector for the block is always estimated from the displacement vectors of the neighboring blocks. This will lead to distortion dependency between adjacent blocks. Therefore, the

978-1-5090-2175-8/17/$31.00 ©2017 IEEE

resulting rate and distortion for a given block are often dependent on the selection of coding parameters in adjacent blocks.

In this paper, we propose a new rate-distortion based error resilient algorithm that optimize the texture and depth coding modes in the lossy environment. Different from the existing work, we explicitly consider the inherent correlation between texture and depth, and also the rate and distortion dependencies introduced between adjacent blocks in both texture and depth. Firstly, the optimization problem is formulated using the Lagrange multiplier method, where the overall expected view synthesis distortion induced by the texture error and depth error is employed as the optimization criteria. Then, a dual trellis is generated whose paths correspond to the full set of the combination of texture and depth modes of the corresponding reference view. Finally, the optimal path in the dual trellis, which has the minimum view synthesis cost, can be efficiently located using a deterministic dynamic programming solution.

The rest of this paper is organized as follows. In Section 2, we provide an overall distortion model that relates the texture distortion and depth distortion induced by packet losses to the distortion contributions in the synthesized view. Section 3 develops a Lagrangian-based method for jointly texture and depth map coding in the lossy environment. Experimental results are presented and discussed in Section 4. Finally, concluding remarks are drawn in Section 5.

2. OVERALL VIEW SYNTHESIS DISTORTION MODEL

Following the 3-D video coding standard specification and for ease of exposition, we assume that the sender usually transmits two pairs of textures and depth maps from two neighboring captured viewpoints, and at the receiver, an intermediate virtual view can be generated from the transmitted texture videos and depth maps via Depth-Image-Based Rendering (DIBR) technique [8]. In the view synthesis procedure, it is assumed that the virtual view and the original view are placed on a horizontal line and they are all pointing to the same direction, which is known as the 1-D parallel setting [9].

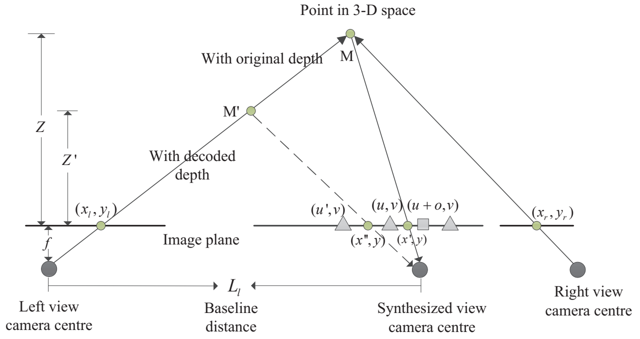

Fig. 1. Sub-pixel 3-D warping illustration with the distorted depth. In the image plane of the synthesized reference view, circles represent initially mapped sub-pixel positions, triangles represent the sub-pixel positions, and squares represent the rounding bound.

In a typical 3-D warping procedure, as shown in Fig. 1, when using the original depth map, the texture pixel (x_l, y_l) in the left coded reference view could be projected to "Point M" in the 3-D world coordinate by the depth value Z; then "Point M" could be projected to (x', y) in the virtual view; finally the pixel (x', y) is rounded to the neighboring position (u, v). Likewise, the texture pixel (x_r, y_r) in the right coded view can also be warped to (x', y). Given the two warped pixels, the pixel value at coordinate (x', y) (or (u, v)) is blended by the weighted average of the pixel values of (x_l, y_l) and (x_r, y_r). When using the decoded depth map, due to the reconstructed depth map errors caused by packet losses, the pixel (x_l, y_l) in the reference view may be projected to "$\mathrm{Point\ M'}$" in the 3-D world coordinate by the distorted depth value $\mathrm{Z'}$; then "$\mathrm{Point\ M'}$" could be projected to (x'', y) in the virtual view; and finally the (x'', y) could be rounded to the (u', v). The horizontal position difference between pixels (u, v) and (u', v) in the synthesized view is attributed clearly to depth map artifacts in the left view. In a similar manner, the effect of the depth distortion of the right view on the synthesized view can also be analyzed in terms of the horizontal shift. In contrast to the depth errors affecting the rendering position in the synthesized view, the texture errors caused by packet losses directly change the pixel intensity of the synthesized view.

Denote by $D_{V,t}^{u,v}$ the overall view synthesis distortion of pixel (u, v) at frame t in the synthesized view. Let $D_{T,t}^{x_i,y_i}$ denote the end-to-end distortion of texture pixel (x_i, y_i), and let $D_{D,t}^{x_i,y_i}$ denote the corresponding end-to-end depth distortion of pixel (x_i, y_i). We can write the combined resulting induced synthesized distortion from the corresponding two pixels from the left and right views as [7], [10]

$$D_{V,t}^{u,v} = \sum_{i \in \{l,r\}} (w_i^2 D_{T,t}^{x_i,y_i} + w_i^2 \psi_i f^2 L_i^2 C^2 \cdot D_{D,t}^{x_i,y_i} + w_i^2 \psi_i E(\varepsilon^2)) \tag{1}$$

where w_i denotes the weighting factor of the rendered image from the reference view i, ψ_i is the motion sensitivity associated with the image of the reference view i, f is the focal length, L_i is the baseline distance between the virtual view and the reference view i, ε is the rounding error, and $C = 1/255(1/Z_{\mathrm{near}} - 1/Z_{\mathrm{far}})$. As suggested by (1), the synthesized view distortion can be characterized as a linear combination of the distortions of the transmitted texture video and depth map of both views. When the compressed 3-D video is transmitted over the lossy network, different coding modes will exhibit different error propagation behaviours due to the varying coding dependencies. If the texture and depth pixels are coded using the conventional prediction modes, such as intra, inter, and inter-view modes, the expected texture and depth distortions can be estimated by the distortion modeling algorithms introduced in [7]. In the event that the texture pixel is coded with the VSP mode, the expected texture distortion can be calculated using the analytical model in [11]. It should be noted that the packet loss rate is implicitly included in the calculation of the expected texture and depth distortions.

3. PROBLEM FORMULATION AND DUAL-TRELLIS-BASED SOLUTION

After obtaining the expected overall view synthesis distortion, we now show how to address jointly rate-distortion optimal selection of texture and depth modes in 3-D video coding. Let $\chi = (X_1, \cdots, X_N)$ be a set of macroblocks (MBs) in a texture/depth frame of a reference view video sequence. Each MB in χ can be coded using only one of four possible texture modes given by the set $I_T = \{\mathrm{Intra}, \mathrm{Inter}, \mathrm{Inter-view}, \mathrm{VSP}\}$. Let $M_i^T \in I_T$ be the mode selected for MB X_i. Each MB is also associated with a depth coding mode, which is chosen from the set $I_D = \{\mathrm{Intra}, \mathrm{Inter}, \mathrm{Inter-view}\}$. Denote $M_i^D \in I_D$ as the depth mode selected for MB X_i. Then, for a given frame, the texture and depth modes assigned to the elements in χ are given by two correlated $N-$tuples, i.e., $\mathbf{M^T} = (M_1^T, \cdots, M_N^T)$ and $\mathbf{M^D} = (M_1^D, \cdots, M_N^D)$. The problem of finding the best combination sequence of texture and depth modes from one reference view for a given rate constraint R_c can be formulated as

$$\begin{aligned}&\min_{\mathbf{M^T},\mathbf{M^D}} E\left[D(\chi,\mathbf{M^T},\mathbf{M^D})\right]\\&\text{subject to } R(\chi,\mathbf{M^T},\mathbf{M^D}) \le R_c\end{aligned} \tag{2}$$

where $E\left[D(\chi,\mathbf{M^T},\mathbf{M^D})\right]$ and $R(\chi,\mathbf{M^T},\mathbf{M^D})$ represent the expected view synthesis distortion and total bit rate of texture and depth, respectively, resulting from the coding of χ with a vector choice of the combined texture and depth modes for each MB in a frame. $E\left[D(\chi,\mathbf{M^T},\mathbf{M^D})\right]$ can be calculated using (1) by summing up the expected distortion of all pixels within the frame, while $R(\chi,\mathbf{M^T},\mathbf{M^D})$ is attainable after real coding. Note that the coding mode pair $(\mathbf{M^T},\mathbf{M^D})$ is independently chosen for each reference view. When operating on the left view, the Inter-view and VSP modes are set to *NULL* in the texture mode set, and the Inter-view mode is set to *NULL* in the depth mode set.

In general, the constrained discrete optimization problem is very difficult to solve. However, since the distortion measure employed in 3-D video coding is additive, the cost function and the rate constraint can be decomposed into a sum of terms over the elements, and rewritten using an unconstrained Lagrangian formulation. The objective cost function then becomes

$$\min_{\mathbf{M^T},\mathbf{M^D}} \sum_{i=1}^{N} J(X_i,\mathbf{M^T},\mathbf{M^D}) \tag{3}$$

where $J(X_i,\mathbf{M^T},\mathbf{M^D})$ is the Lagrangian cost function for MB X_i and is given by

$$J(X_i,\mathbf{M^T},\mathbf{M^D}) = E\left[D(X_i,\mathbf{M^T},\mathbf{M^D})\right] + \lambda R(X_i,\mathbf{M^T},\mathbf{M^D}) \tag{4}$$

For an appropriate value of the Lagrangian multiplier λ, the convex hull solution to (3) corresponds to an optimal solution to (2) for a particular value of R_c [12]. Because of the monotonic relationship between λ and rate, the appropriate λ can be found by using the bisection iterative search or other fast search algorithms [13].

However, because of the total rate and view synthesis distortion dependencies manifested in the $E\left[D(X_i,\mathbf{M^T},\mathbf{M^D})\right]$ and $R(X_i,\mathbf{M^T},\mathbf{M^D})$ terms, the solution to (3) is not trivial. Without further assumption, the resulting distortion and rate associated with a particular MB is inextricably coupled to the chosen modes for every other MB in χ. For example, as done in [7], both the rate and distortion for a given MB are assumed to be impacted only by the content of the current MB and its respective operational texture and depth modes. As a result, the optimization problem of (3) can be easily minimized by independently selecting the best texture and depth modes for each MB. However, this assumption neglecting the block-to-block dependency is rather restrictive and not the case for 3-D video coding.

In order to make the optimization problem more tractable while guaranteeing source coding performance, we assume the 1-D dependency in this paper, i.e., the total influence on rate and distortion for any particular MB is limited to that from the immediately preceding MB. It should be noted that this simplified dependency exactly accords with the way MBs are coded in 3-D video coding standards. For instance, when the motion vector is differentially coded using depth-based motion vector prediction (DMVP) [14], the rate term for a given MB is dependent not only on the current modes but on the modes of the adjacent MBs. Likewise, the error concealment strategy, typically estimating the lost motion vector by resorting the motion information of the neighboring blocks, results in distortion dependency between adjacent MBs. Further, the geometry errors caused by depth errors usually prevent a bijective mapping in 3-D warping, which may lead to the view synthesis distortion of a particular rendered block simultaneously relating to the modes from the current MB and neighboring MB of the reference views [15]. Under this condition, the rate and distortion for MB X_i is dependent on the texture and depth modes selected for both MB X_i and X_{i-1}, in which each Lagrangian term can be written as

$$\begin{aligned}J(X_i,M_i^T,M_i^D) = \min_{M_{i-1}^T,M_{i-1}^D} \Big\{ & J(X_{i-1},M_{i-1}^T,M_{i-1}^D)\\ & + E\left[D(X_i,M_{i-1}^T,M_{i-1}^D,M_i^T,M_i^D)\right]\\ & + \lambda R(X_i,M_{i-1}^T,M_{i-1}^D,M_i^T,M_i^D)\Big\}\end{aligned} \tag{5}$$

This unconstrained optimization over finite discrete sets can be solved with the Dynamic Programming (DP) technique using the Viterbi Algorithm (VA) [16]. Prior to establishing a DP recursion formula, a trellis has to be constructed for a given synthesized frame. Considering the optimization problem formulated on a MB-by-MB basis for each texture and depth frame, the nodes in the trellis are given by the elements in sets I_T and I_D of all the MBs in the given frame. The ith MB represents the ith stage in the trellis. At each stage, there are 12 trellis states, each representing a combination of texture and depth modes $\left\{M_i^T,M_i^D\right\}$ for the corresponding ith MB. Since the control parameter set which influences the rate and distortion for a node is a dual combination of prediction modes of two different coding signals, the admissible states of the nodes are henceforth called the dual states in this paper to distinguish from the traditional trellis state with only one control parameter, and the resulting trellis from the dual states is termed the *dual trellis*. The transitional costs from node $\left\{M_{i-1}^T,M_{i-1}^D\right\}$ to node $\left\{M_i^T,M_i^D\right\}$ are given by the Lagrangian cost terms specified in (5). Once the trellis is formed, the VA is applied to trace the shortest path. Using the VA, only one incoming path (i.e., the minimal rate-distortion cost path) is kept for each dual state at each stage. The accumulated cost and the path are recorded. The shortest path can be found at the end of the stage of the trellis, and the optimal trellis states are backtracked. The corresponding dual trellis is constructed in Fig. 2, where each transition within the ith MB stands for a $\left\{M_i^T,M_i^D\right\}$ pair, representing the intrinsic dependency between the texture and depth coding modes. The connections between the adjacent MBs represent the block-wise rate and distortion dependencies introduced by the differential encoding and error concealment techniques, respectively. Note that each MB in the reference view is indicated by a dashed box. These dashed boxes are not part of the trellis used for the DP algorithm but indicate the set of admissible state values (i.e., $I_T \times I_D$) for the individual MB. This graph naturally facilitates the calculation of the $J(X_i,M_i^T,M_i^D)$ for every path, which enables us to efficiently solve (3) using the VA.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The 3D-AVC reference software 3D-ATM v6.0 [17] is employed to encode both the multi-view video sequences and depth maps, while the View Synthesis Reference Software (VSRS) 3.5 [18] is used to render the synthetic reference view at the encoder and virtual intermediate views at the decoder. The standard multi-view video sequences “Lovebird1”, “Newspaper”, and “GT_Fly” are chosen for our simulations. Among these sequences, for “Lovebird1” sequence, the views 6 and 8 are adopted as the left and right views, respectively. For “Newspaper” sequence, views 4 and 6 are served as the left and right views, respectively. For “GT_Fly” sequence, views 5 and 9 are used as the left and right views to render the virtual view. The

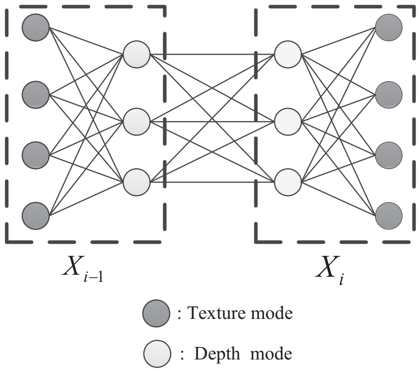

Fig. 2. Dual trellis representation used for jointly selecting texture and depth coding modes for the entire frame in the reference view.

resolution of the first two sequences is 1024×768, while the remaining one is of full HD (1920×1080). For each sequence, each view is encoded with a group of pictures (GOP) size of 30 frames, where the first frame in the left view is coded as an I-frame, and the remaining frames are coded as P and B frames. In 3-D video coding, the coding order $T_0D_0T_1D_1$ is used, where T_i and D_i are the texture and depth components of the i^{th} view, respectively, corresponding to the left or right views. The depth maps are selected to be coded at full resolution. The encoder uses variable block-size motion and disparity estimation, with a search range of 64 pels. The default QP set of (40, 35, 30, 25) is used. Backward VSP is enabled. The virtual views are generated with half-pel precision and symmetric rounding.

Each predictively coded frame is partitioned into slices, where each depth slice contains nine horizontal rows of MBs, and each texture slice includes three horizontal rows of MBs due to higher associated bit rates. Each coded slice is then carried in a single packet according to the real-time transfer protocol specifications [19]. The random packet loss pattern is employed to simulate packet losses, and the loss of any two packets is independent [20]. To simulate the channel, at each packet loss rate, 100 packet loss patterns are randomly generated. In our experiments, the error concealment method introduced in [21] is employed for texture and depth. When a block is lost, the motion vector of the missing block is estimated as the median of the motion vectors of the neighboring three blocks. If the neighboring blocks are also lost, the estimated motion vector is set to zero. For the objective video quality assessment, the average peak signal-to-noise ratio (PSNR) of the synthesized views are measured. The PSNR of synthesized view is computed between the virtual view images synthesized by the uncompressed texture and depth images and the decoded texture and depth images.

In order to evaluate the effectiveness of the proposed rate-distortion optimized joint texture and depth map coding algorithm, two state-of-the-art error-resilient algorithms for 3-D video transmission are compared: Rate-Distortion Optimized Reference Frame Selection (RDO-RPS) [6], Rate-Distortion Optimized Intra Update (RDO-IU) [7]. For fair comparison, all the test schemes use the same bit rate budget for each sequence during encoding. To meet the transmission bit budget, for RDO-RPS and RDO-IU, the advance adaptive rate control described in [22] is adopted to bring the total bit rate as close as possible to the target bit rate, while in the proposed algorithm, the fine-tuning rate is accomplished by adjusting λ. Table 1 shows a summary of the average PSNR gain of the proposed algorithm over RDO-RPS and RDO-IU for a variety of sequences and packet loss rates. The PSNR gain is obtained using the popular Bjontegaard Delta-PSNR method in [23]. From Table 1, it is evident that the proposed algorithm yields significant and consistent gains over the other competing methods. To summarize, the proposed algorithm achieves an average PSNR gain of 1.05 dB over RDO-RPS, and outperforms RDO-IU by 0.54 dB on average. The reason for these gains is as follows. The proposed algorithm considers the overall rate-distortion behavior of the whole synthesized frame, instead of the specific characteristics of each individual texture MB, each depth MB (as done in RDO-RPS), or each combination of them (as done in RDO-IU). Thanks to the global optimization of each pair of texture and depth MBs, the proposed algorithm can always achieve the minimum average view synthesis distortion at the frame level.

Table 1.
Average PSNR gain of the proposed algorithm over RDO-RPS and RDO-IU with a variety of packet loss rates.

Sequence	Loss rate	Y-PSNR (dB)	
		Proposed over RDO-RPS	Proposed over RDO-IU
Lovebird1	5%	0.89	0.47
	10%	0.87	0.51
Newspaper	5%	1.16	0.73
	10%	1.28	0.69
GT_Fly	5%	1.02	0.60
	10%	1.09	0.54

As a visual comparison, we also provide the subjective evaluations using the various test algorithms. Fig. 3 shows the subjective view rendering quality of the Newspaper sequence at the packet loss rate of 10%. In order to clearly show the visual difference among these test algorithms, we provide the image segment of the synthesized frame that is more vulnerable to transmission error for illustration. It can be observed that the proposed algorithm can indeed lead to significantly improved view rendering quality as compared to the other two reference methods.

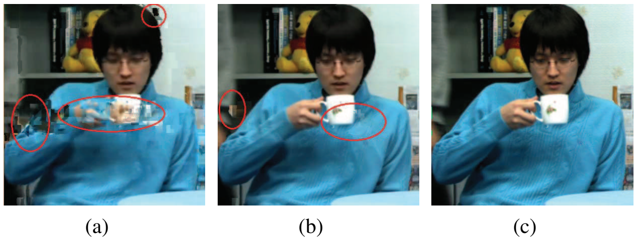
(a) (b) (c)

Fig. 3. Subjective quality comparison of the image segment of the synthesized frame 49 of the Newspaper sequence at the loss rate of 10%. (a) RDO-RPS, (b) RDO-IU, (c) The proposed algorithm.

5. CONCLUSIONS

In this paper, we proposed a new algorithm that jointly selects the texture and depth coding modes to optimize the error resilience performance of 3-D video. Firstly, the joint optimization of texture and depth is formulated, for a given bit budget, as a problem of how to select the texture and depth modes on a block-by-block basis within each frame of the reference views such that the overall view synthesis distortion is minimized. Then, the constrained optimization problem is converted to an associated unconstrained cost function using the Lagrangian multiplier method. Finally, we develop a dual-trellis-based algorithm to efficiently find the optimal solution. Simulation results confirm that the performance gains of the proposed algorithm over the existing work are considerable and consistent.

6. REFERENCES

[1] K. Muller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 643-656, Apr. 2011.

[2] G. Tech, Y. Chen, K. Muller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the multi-view and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, Jan. 2016.

[3] D. Tian, F. Zhou, and A. Vetro, "CE1.h: Backward View Synthesis Prediction using Neighboring Blocks," *ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V-C0152,* Geneva, CH, Jan. 2013.

[4] P. Gao and W. Xiang, "Disparity vector correction for view synthesis prediction-based 3-D video transmission," *IEEE Trans. Multimedia,* vol. 17, no. 8, pp. 1153–1165, Aug. 2015.

[5] B. Macchiavello, C. Dorea, E. M. Hung, G. Cheung, and W. Tan, "Reference frame selection for loss-resilient depth map coding in multiview video conferencing," in *Proc. SPIE Visual Inf. Process. Commun.*, 2012.

[6] B. Macchiavello, C. Dorea, E. M. Hung, G. Cheung, and W. Tan, "Loss-resilient coding of texture and depth for free-viewpoint video conferencing," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 711-725, Apr. 2014.

[7] P. Gao and W. Xiang, "Rate-distortion optimized mode switching for error-resilient multi-view video plus depth based 3-D video coding," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 1797-1808, Nov. 2014.

[8] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3-D-TV," in *Proc. 11th SPIE Stereoscopic Displays Virtual Reality Syst.*, Jan. 2004, pp. 93-104.

[9] "Call for Proposal on 3D Video Coding Technology," ISO/IEC JTC1/SC29/WG11, MPEG, Doc. N12036, Geneva, Switzerland, March, 2011.

[10] 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. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 485-497, Apr. 2011.

[11] P. Gao, W. Xiang, and L. Zhang, "Transmission distortion modeling for view synthesis prediction based 3-D video streaming," in *Proc. Int. Conf. Acoust., Speech, Signal Processing (ICASSP),* Apr. 2015, pp. 1448–1452.

[12] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 36, no. 9, pp. 1445–1453, Sep. 1988.

[13] A. De Abreu, G. Cheung, P. Frossard, and F. Pereira, "Optimal Lagrange multipliers for dependent rate allocation in video coding," in *arXiv: 1603.06123*, Mar. 2016.

[14] M. M. Hannuksela, D. Rusanovskyy, W. Su, L. Chen, R. Li, P. Aflaki, D. Lan, M. Joachimiak, H. Li, and M. Gabbouj, "Multiview-video-plus-depth coding based on the advanced video coding standard," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3449-3458, Sep. 2013.

[15] G. Tech, K. Muller, H. Schwarz, and T. Wiegand, "Partial depth image based re-rendering for synthesized view distortion computation," *IEEE Trans. Circuits Syst. Video Technol.*, published online, Jan. 2017, DOI: 10.1109/TCSVT.2016.2631568.

[16] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximations," *IEEE Trans. Image Process.,* vol. 3, no. 1, pp. 26–40, Jan. 1994.

[17] 3D-ATM reference software version 6.0, "http://mpeg3dv.nokiaresearch.com/svn/mpeg3dv/tags/3DV-ATMv6.0/."

[18] ISO/IEC JTC1/SC29/WG11, 3DV/FTV EE2: Report on VSRS Extrapolation, Doc. M18356, Guangzhou, China, 2010.

[19] S. Wenger, "H.264/AVC over IP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 645-656, Jul. 2003.

[20] S. Wenger, "Proposed Error Patterns for Internet Experiments," ITU-T VCEG document Q 15-I-16r1, Oct. 1999.

[21] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 966-976, Jun. 2000.

[22] Z. Li, W. Gao, F. Pan, S. Ma, K. P. Lim, G. Feng, X. Lin, S. Rahardja, H. Lu, and Y. Lu, "Adaptive rate control with HRD consideration," Joint Video Team document JVT-H014, May 2003.

[23] G. Bjontegaard, *Calculation of Average PSNR Differences Between RD Curves*, document VCEG-M33, ITU-T SG16/Q6 (VCEG), Austin, TX, Apr. 2001.