MODEL-BASED INTRA CODING FOR DEPTH MAPS IN 3D VIDEO USING A DEPTH LOOKUP TABLE

Fabian Jäger, Mathias Wien, Philipp Kosse

Institut für Nachrichtentechnik, RWTH Aachen University

ABSTRACT

3D Video is a new technology, which requires the transmission of depth data alongside conventional 2D video. The additional depth information allows to synthesize arbitrary viewpoints at the receiver and enables adaptation of the perceived depth impression and driving of multi-view auto-stereoscopic displays. In contrast to natural video signals, depth maps are characterized by piecewise smooth regions bounded by sharp edges along depth discontinuities. Conventional video coding methods tend to introduce ringing artifacts along these depth discontinuities, which lead to visually disturbing geometric distortions in the view synthesis process. Preserving the described signal characteristics of depth maps is therefore a crucial requirement for new depth coding algorithms.

In this paper a novel model-based intra-coding mode is presented, which works as an addition to conventional transform-based intra coding tools. The proposed intra coding mode yields up to 42% BD-rate savings in terms of depth rate and up to 2.5% in terms of total rate. The average bitrate savings are approximately 24% for depth rate and 1.5% for the total rate including texture and depth.

Index Terms **—** Video coding, Depth map coding, Model-based coding, 3D Video

1. INTRODUCTION

Recent developments in the field of 3D display technologies need to synthesize additional arbitrary viewpoints based on the limited number of available decoded views of a video sequence. To allow for this extent of flexibility, depth information needs to be available at the receiver side and consequently needs to be coded in addition to conventional 2D video data.

These additional depth maps show different signal characteristics compared to natural video data. Moreover, distortions in depth maps have an indirect impact on the visual quality of the displayed video as they are used to synthesize new views of the same scene and are never shown to the user themselves. Compressing depth maps with algorithms optimized for natural 2D videos may result in ringing artifacts along depth discontinuities, which then introduce geometric distortions in the synthesized views. Thus, new compression algorithms have to be developed that are adapted to signal characteristics of depth maps.

Previous work on compression of depth data regarded depth data as gray-colored video and used conventional transform-based video coding algorithms as found in H.264/AVC [1]. It was shown that these conventional coding tools yield relatively high compression efficiency in terms of PSNR, but at the same time introduce ringing artifacts along sharp edges in the original depth maps. These artifacts result in geometric distortions in the view synthesis stage, as will be shown exemplarily in Section 5.

More recent depth compression algorithms approximate depth map signal characteristics by a partitioning of the frame into triangular meshes [2] or platelets [3] and modeling each segment by an appropriate 2D function. These pure model-based approaches can also be combined with conventional transform-based tools by introducing an additional coding mode, like the sparse-dyadic (SD) mode [4]. Here, an SD-coded block is partitioned into two segments, which are described by two constant depth values. This concept can be further extended by incorporating structural information from the corresponding texture component as proposed with Depth Modeling Modes [5].

As the preservation of depth discontinuities is most important when compressing depth maps, another approach is to losslessly compress the location of these discontinuities and approximate the piecewise smooth regions, as previously proposed [6]. The disadvantage of this approach is the inability of reaching low bitrates due to the lossless encoding of depth contours.

In this paper a novel Depth Model-based Coding (DMC) method is presented as an addition to the 3D extension of High Efficiency Video Coding (HEVC) [7] and adds another intra-coding mode to the available HEVC intra coding tools. For DMC-coded blocks residual values are coded in the pixel-domain. Therefore, a new Depth Lookup Table (DLT) is proposed, which allows to further reduce the required bitrate for coding the residual.

The remainder of this paper is structured as follows. Section 2 introduces the main concepts of the proposed depth map coder, including explanations of specific parts of the prediction scheme. In Section 4 the concept of the Depth Lookup Table (DLT) is introduced and explained how it influences the coding of depth value residuals. Experimental results comparing the proposed approach with a reference implementation are shown in Section 5 before Section 6 concludes the paper with a summary and an outlook on future investigations.

2. GENERAL CONCEPT AND CODING SPECIFICS

The novel depth map intra-coding mode is characterized by block-based prediction and residual coding. The prediction consists of four different modes, which are designed to match typical depth map characteristics of piecewise smooth regions bounded by sharp edges. Blocks of constant depth can be perfectly modeled by a DC value. For modeling gradual depth areas, a special gradient mode is used. A third mode divides a block into two segments by a straight line. Each subsegment can then be modeled by a different DC value. Similarly, the fourth DMC prediction mode extracts contour information from the collocated texture block to compute a segmentation of the current depth map block. The resulting segments are again approximated by constant depth values.

This third and fourth modes allow to approximate the typical sharp edges describing object boundaries in depth maps. As all residuals are coded in the pixel-domain and do not undergo any transformation, ringing artifacts, like in traditional transform-based coding, are completely absent when coded with the proposed DMC mode. It is implemented as an alternative mode to the intra coding mode, which is available in the HEVC-based 3DV-HTM reference software [8]. An additional DMC-Flag signals the usage of DMC prediction and coding. Instead of coding quantized transform coefficients DMC-coded blocks need to code the following types of information:

1. The type of segmentation/prediction method of the current block. Possible values are
   (a) DC (no segmentation)
   (b) Gradient (no segmentation)
   (c) Edge (segmentation into two segments by a straight line)
   (d) Texture (segmentation into 2 or 3 segments by thresholding the collocated texture block)
2. For Edge- and Texture-segmentation, some details about the segmentation needs to be coded:
   (a) For Edge-segmentation: Start/End of the straight line of the segmentation
   (b) For Texture-segmentation: Number of segments to split the block into
3. For each segment, a residual value (in the pixel domain) is signaled in the bitstream

Before coding, the residual values are mapped to values, which are present in the original, uncompressed depth map by using a Depth Lookup Table (DLT). Consequently, residual values can be coded by signaling only the index into this lookup table, which reduces the bit depth of residual magnitudes for sequences with reduced depth range due to quantization.

3. AVAILABLE PREDICTION MODES

DMC-coded depth map blocks are predicted by one of four available prediction modes. The optimal mode is selected by the encoder based on the VSO [9] rate-distortion criterion and coded into the bitstream. The most probable mode is predicted from neighboring coding units. A flag codes whether the actual block mode matches the most probable mode. If this is not the case, up to two additional flags are required to signal the actual mode for the DMC-block. All the mentioned flags have their own new context models assigned for the CABAC engine.
The directional intra-prediction modes of HEVC are not available for DMC-coded blocks as most of these can be modeled by the Edge segmentation mode, which will be explained in more detail in the following. The encoder can still decide to not use the proposed DMC mode and code the current depth map block with the HEVC intra-coding tools.

3.1. DC Prediction

The DC prediction mode of DMC is suitable for regions with constant depth. The corresponding DC prediction value is predicted from neighboring blocks by a the mean of all directly adjacent samples of the top and the left tree block. The resulting residual is coded according to Section 4.2.

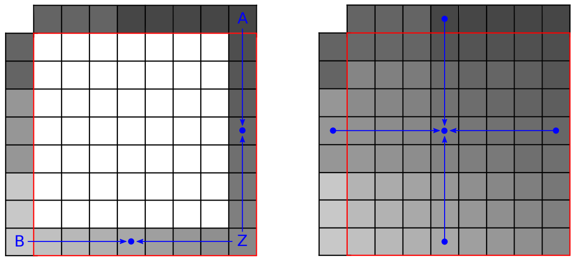


Figure 1: Gradient Prediction: Z value predicted and coded. Bottom and right pixels linearly interpolated, center pixels bilinearly interpolated.

3.2. Gradient Prediction

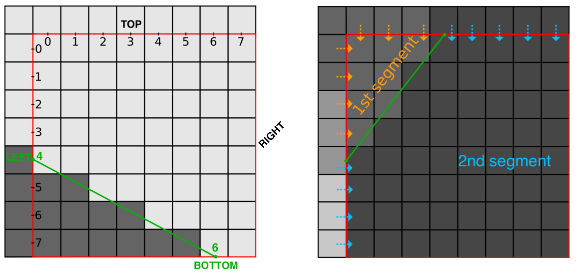
DMC's Gradient Prediction (GP) is introduced to model gradual depth changes in depth maps. Inputs to this mode are the neighboring pixel values, as well as a target value Z in the lower-right corner of the current block. As depicted in Figure 1, the bottom row is linearly interpolated from the values B and Z, the right column respectively from A and Z. In a second step, all remaining pixel values are bilinearly interpolated from the surrounding values.

The Z parameter is again predicted from the directly adjacent depth samples of the upper and left coded tree block. By computing the horizontal gradient at the top and the vertical gradient at the left side of the current block, the Z value can be predicted by adding these two gradients to the depth value at the top-left position. The resulting residual is coded according to Section 4.2.
This gradient prediction mode is similar to HEVC planar prediction, but it explicitly signals the Z value instead of inferring it from neighboring samples. This allows for more flexible approximations of arbitrary gradual depth changes.

3.3. Edge Prediction

The Edge Prediction (EP) divides a block into two segments by a straight line as illustrated in Figure 2. Both segments are modeled by a DC value, which is predicted from the adjacent depth values of top and left coded tree blocks. The prediction of the two DC values is computed by the mean of neighboring depth values, which would belong to the same segment defined by the selected edge division. The resulting residual is coded according to Section 4.2.
Six different start/end border combinations and two corresponding indices define the division of the current block into two segments. Two examples of these combinations are depicted in Figure 2.

To improve continuity of edges through multiple EP-coded blocks and to reduce the required bitrate for edge signaling, DMC predicts the edge information from neighboring blocks. This applies, whenever a neighboring block is also using Edge Prediction and its edge leads into the current block. In this case, the start point of the current block's edge is predicted from the edge endpoint of that block and the endpoint of the current edge is predicted by continuing the slope of the neighboring edge into the current block. The actual border and position offsets relative to the predicted edge are entropy coded.
A very similar prediction mode is found as part of the Depth Modeling Modes (DMM) [5], which uses a different way of signaling



(a) Edge parameters in EM-blocks (b) Prediction of DC-coefficients

Figure 2: (a) Segmentation of blocks with Edge Mode defined by start/end border and index. (b) Prediction of segments' DC value from neighboring blocks.

edge information. It should be noted that the resulting residual of DMM-predicted blocks is still transform coded and may therefore introduces ringing artifacts along depth discontinuities.

3.4. Texture Prediction

When predicting the depth map signal for a certain block it is also possible to incorporate the already coded, collocated block of the texture component of the same view. By applying a simple thresholding of the luminance component of the texture block, a segmentation of the block into two or three segments is computed. The resulting segmentation mask is used to compute the mean depth value of each of these segments. The resulting DC values are again predicted similarly as with Edge or DC Prediction by the mean depth value of directly adjacent samples of the particular segment. The resulting residual is coded according to Section 4.2.

Depth Modeling Modes (DMM), as they are in the current MPEG 3DV reference software, also allow this kind of texture-to-depth prediction, but DMM is more restrictive as it does only allow for two segments.

4. DEPTH LOOKUP TABLE

An analysis of the MPEG 3DV test sequences [10] has shown that many depth maps do not utilize the full available depth range of 2^8. Only a small amount of different depth levels occur in those sequences due to strong quantization. The advantage of using the DLT in these cases is the reduced bit depth of the residual values.

4.1. Analysis Step

In the analysis step the encoder reads a predefined number of frames from the input video sequence to be coded and scans all pixels for available depth values. During this process a mapping table is generated that maps all possible depth values to valid depth values of the original uncompressed depth map. The analysis step results in a Depth Lookup Table $D[d]$ (mapping all possible depth values to valid depth values) and an Index Lookup Table $I[d]$ (mapping valid depth values to their indices). The number of available depth values in the original, uncompressed depth map sequence is $i_{\max} = \max(I[d])$.

4.2. Coding Step

Instead of coding the actual residual depth value for a given coding unit, the predicted depth value d_{pred} and the original depth value d_{orig} are first mapped to their corresponding indices in the list of valid depth maps to get the residual index as follows.

$$i_{\text{resi}} = I\left[D[d_{\text{orig}}]\right] - I\left[D[d_{\text{pred}}]\right] \qquad (1)$$

The computed residual index i_{resi} is then coded with a significance flag, a sign flag and with $\lceil \log_2(i_{\max}) \rceil$ bits for the magnitude of the residual index. During encoding, no quantization of the residual depth value is applied.
The corresponding mapping information needs to be transmitted to the decoder for the inverse lookup from indices to valid depth values. In the current implementation, the DLT is transmitted as part of the sequence parameter set (SPS), but could also be updated more often, e.g. in the adaptation parameter set (APS).

5. EXPERIMENTAL RESULTS

The proposed Depth Model Coding (DMC) approach was integrated into the MPEG 3DV HEVC Test Model (HTM 3.1) as an additional intra coding mode. As the proposed method is an intra-coding tool, the intra-only performance is investigated separately from the random access coding configuration with a GOP size of 8 and 0.5 seconds random access rate. The simulations follow the common test conditions of the MPEG 3DV standardization activity [10]. For the HTM+DMC simulations DMM was turned off. The reference still uses DMM and follows the mentioned test conditions.

5.1. Objective Evaluation

As depth maps are typically not displayed themselves and can be regarded as supplementary data to the texture videos, RD curves are computed based on the PSNR values of synthesized views. The reference for this type of evaluation is a virtual viewpoint synthesized with the uncompressed texture and uncompressed depth. For a particular coding method (here: DMC), the reconstructed texture and depth is used to compute the virtual viewpoint at the same position. PSNR is then computed between the synthesized view based on the uncompressed data and the synthesized view based on the coded and reconstructed data. This evaluation procedure is also used within MPEG and follows the common test conditions of the MPEG 3DV standardization activity [10].

Table 1: BD-Rate savings on total (incl. texture) and on depth bitrate only for HTM+DMC in Random Access Configuration.

Sequence	Synthesis BD-Rate (total bitrate)	Synthesis BD-Rate (depth bitrate)
Poznan_Hall2	-1,01 %	-1.24 %
Poznan_Street	+0.27 %	-0.41 %
UNDO_Dancer	-1.11 %	+0.03 %
GT_Fly	-0.36 %	-4.34 %
Kendo	-0.92 %	-7.08 %
Balloons	-0.55 %	-6.31 %
Newspaper_CC	-1.61 %	-8.75 %
AVERAGE	**-0.76 %**	**-4.02 %**

Table 2: BD-Rate savings on total (incl. texture) and on depth bitrate only for HTM+DMC in All-Intra Configuration.

Sequence	Synthesis BD-Rate (total bitrate)	Synthesis BD-Rate (depth bitrate)
Poznan_Hall2	-2.54 %	-16.16 %
Poznan_Street	+0.17 %	-17.25 %
UNDO_Dancer	-2.29 %	+4.00 %
GT_Fly	-2.43 %	-36.57 %
Kendo	-2.50 %	-42.27 %
Balloons	-1.02 %	-35.10 %
Newspaper_CC	+0.43 %	-25.10 %
AVERAGE	**-1.46 %**	**-24.06 %**

5.2. Subjective Evaluation

As already stated in Section 1, the most prominent advantage of a model-based approach compared to a conventional transform-based coder is the improved preservation of depth discontinuities due to the removal of ringing artifacts along object boundaries. In the following, an exemplary synthesized viewpoint based on compressed depth maps is illustrated to show the improvements in visual quality for those virtual views. For this experiment, texture data remained uncompressed to concentrate on depth coding artifacts. The view synthesis in this experiment is based on a single viewpoint (extrapolation scenario) and disoccluded regions are not inpainted.

(a) Original Depth (b) HTM (c) HTM+DMC

Figure 3: Synthesis results for sequence *Undo_Dancer* based on (a) uncompressed, (b) HTM 3.1 and (c) HTM+DMC-based coded depth map at 0.0043 bit per pixel.

Figure 3 shows synthesis results based on depth maps coded at the same bitrate. The HTM 3.1 encoding in Figure 3b shows geometrical deformations (person's head) while the DMC method yields convincing synthesis quality, relatively close to the synthesis based on the uncompressed depth map.

5.3. Computational Complexity

The proposed DMC coding mode is relatively complex at the encoder side as the described segmentations need to be tested. The DMM prediction of the HTM reference software comes with a very similar complexity. On the other side, the decoding process of DMC-coded blocks is very low complex and highly efficient as the whole de-quantization and inverse transform steps are not used.
Overall, the current implementation of HTM+DMC has about 109% encoder and 75% decoder complexity, relative to HTM 3.1. These numbers are based on execution times and should therefore be taken as rough estimates of the real complexity of the algorithms.

6. CONCLUSION

In this paper a novel model-based intra-coding mode for depth maps is proposed. It is demonstrated that DMC significantly reduces the depth map bitrate while retaining the same synthesis quality compared to conventional coding methods. At the same bitrate the synthesis quality is visually improved by reducing geometrical distortions in the synthesis coming from ringing artifacts along depth discontinuities.

Further research is needed on whether it is possible to introduce a general pixel-domain coding of depth residual values as an alternative to transform-based coding. In this case, the proposed Depth Lookup Table may be used to further reduce the required bitrate for the depth component in 3D video. Moreover, it needs to be investigated whether conventional intra-coding and model-based approaches can be combined by predicting mode parameters from neighboring blocks.

7. REFERENCES

[1] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proceedings of IEEE International Conference on Image Processing*, 2007, vol. 1, pp. 201–204.

[2] Michel Sarkis, Waqar Zia, and Klaus Diepold, "Fast depth map compression and meshing with compressed tritree," in *Computer Vision – ACCV 2009*, vol. 5995 of *Lecture Notes in Computer Science*, pp. 44–55. Springer Berlin / Heidelberg, 2010.

[3] Y. Morvan, P.H.N. de With, and D. Farin, "Platelet-based coding of depth maps for the transmission of multiview images," in *Proceedings of SPIE Stereoscopic Displays and Applications*, 2006, vol. 6055, pp. 93–100.

[4] S. Liu, P.L. Lai, D. Tian, C. Gomila, and C.W. Chen, "Sparse dyadic mode for depth map compression," in *IEEE International Conference on Image Processing*. IEEE, 2010, pp. 3421–3424.

[5] P. Merkle, C. Bartnik, K. Muller, D. Marpe, and T. Wiegand, "3D video: Depth coding based on inter-component prediction of block partitions," in *Proceedings of IEEE Picture Coding Symposium*, May 2012, pp. 149–152.

[6] F. Jager, "Contour-based segmentation and coding for depth map compression," in *Proceedings of IEEE Visual Communications and Image Processing*, 2011, pp. 1–4.

[7] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, and T. Wiegand, "High Efficiency Video Coding (HEVC text specification draft 7)," Tech. Rep., JCTVC-I1003, May 2012.

[8] MPEG Video Group, "Test model under consideration for HEVC based 3D video coding v3.0," Tech. Rep., MPEG N12744, April 2012.

[9] G. Tech, H. Schwarz, K. Muller, and T. Wiegand, "3D video coding using the synthesized view distortion change," in *Proceedings of IEEE Picture Coding Symposium*, May 2012, pp. 25–28.

[10] MPEG Video Group, "Common test conditions for 3DV experimentation," Tech. Rep., MPEG N12745, April 2012.