SIMPLIFIED DEPTH MAP INTRA CODING WITH AN OPTIONAL DEPTH LOOKUP TABLE

Fabian Jäger

Institut für Nachrichtentechnik
RWTH Aachen University, Germany

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. Therefore, preserving the described signal characteristics of depth maps is a crucial requirement for new depth coding algorithms.
In this paper a novel simplified intra-coding mode is presented, which works as an alternative coding path to the conventional transform-based intra coding scheme. The proposed intra coding mode yields up to 1.3% BD-rate savings in terms of total bitrate, including texture bitrate. An optional Depth Lookup Table can further increase the bitrate reduction to 2.1%.

Index Terms— video coding, depth map coding, residual 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. They are never shown to the user, but used to synthesize new views of the same scene. Due to that, distortions in depth maps only have an indirect impact on the visual quality of the displayed video. 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 depth data compression is using conventional transform-based video coding algorithms as found in H.264/AVC [7]. 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 shown exemplarily in Section 4.
More recent compression algorithms approximate depth map signal characteristics by a partitioning of the frame into triangular meshes [9] or platelets [8] 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 [5]. An SD-coded block is partitioned into two segments, which are described by two constant depth values. This concept can further be extended by incorporating structural information from the corresponding texture component as proposed with Depth Modeling Modes [6].
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 [2]. 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 map intra coding scheme for the 3D extension of High Efficiency Video Coding (HEVC) [1] is presented. For blocks coded with the simplified depth coding (SDC) mode, residual values are coded as DC offsets 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 DC values.
The remainder of this paper is structured as follows: Section 2 introduces the main concepts of the proposed simplification to the HEVC codec for depth map coding. Section 3 presents the concept of the Depth Lookup Table (DLT) and explains how it influences the coding of depth value residuals. Experimental results comparing the proposed approach with a reference implementation are shown in Section 4. Finally Section

5 concludes the paper with a summary and an outlook on future investigations.

2. SIMPLIFIED DEPTH CODING

As stated in the introductory section of this paper, recent coding approaches for depth map videos are still using the same coding structures and tools as for conventional 2D video data. These coding schemes are optimized for signal characteristics of typical texture videos including signal noise, textured areas and higher importance of low frequency components.

In this paper a novel intra coding scheme for depth map videos is proposed, which reduces the amount of signaling in a modern video coder by adapting the coding tree to typical depth map signal characteristics. The latest HEVC-based 3D video coding extension (3D-HEVC) [3], which is currently in its standardization phase, introduces new prediction modes (Wedgelets) [6]. These are designed to model typical depth map signal characteristics like strong edges bounding smooth regions. Wedgelets explicitly model sharp edges by segmenting the current tree block into two distinct regions.

In HEVC there are already directional intra prediction modes available, which are able to model directional structures like segment edges. Consequently the introduction of additional intra prediction modes (Wedgelets) adds redundancy to the prediction stage of the 3D-HEVC as it requires additional signaling bits to define the prediction mode to be used.

Moreover, HEVC allows the current coding tree unit to be split into four prediction units, if the encoder's rate distortion optimization (RDO) decides the splitting to be optimal for coding efficiency. The described splitting of prediction units was introduced for coding conventional 2D video data to allow better prediction of locally varying signal characteristics. This degree of flexibility is typically not necessary for depth map signals, because they are mostly described by piecewise smooth regions bounded by sharp edges along depth discontinuities. Within depth segments, the splitting into smaller prediction units is not necessary. A single prediction unit can already model the signal characteristics. For coding tree units, which cover a depth discontinuity, the quad-splitting into four prediction units does also not result in better prediction signals. The new Wedgelet prediction modes already account for this type of signal. The availability of signaling the splitting of a coding unit into four prediction units is not beneficial for depth map videos and therefore introduces additional bitrate overhead in the coding tree.

Based on this observation, this paper proposes a new shortcut in the intra coding tree of 3D-HEVC, which is adapted to typical depth map signal characteristics. By bypassing the signaling of unnecessary bits, which are only useful for textured 2D videos, a significant number of depth map blocks can be coded with reduced bitrate. Nevertheless, the synthesized views based on the coded depth maps retain the same high visual quality.

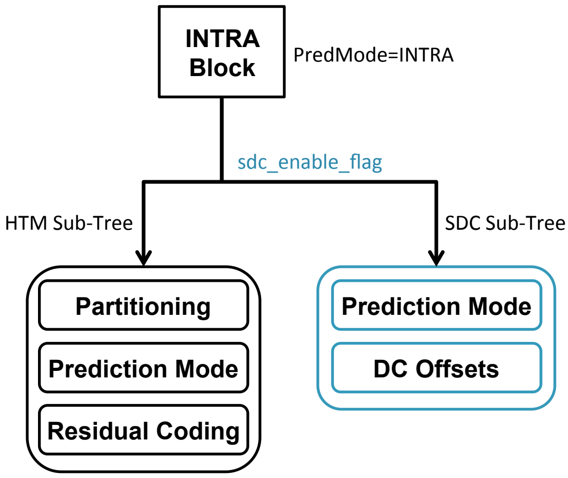

Fig. 1: Coding tree for intra coded depth blocks for conventional 3D-HEVC and for the proposed alternative coding path.

The simplified depth mode is implemented as an alternative coding sub-tree to the intra coding sub-tree, which is available in the HEVC-based 3DV-HTM reference software [3]. The general coding tree structure for intra coded blocks is illustrated in Figure 1. An additional SDC-flag, which comes right after the signaling of the prediction mode of the current block, signals the usage of SDC coding. For 3D-HEVC, the partitioning into prediction units, the prediction mode (out of 40) and the transform coefficients are coded. For SDC-coded blocks only the following information needs to be coded:

1. The optimal prediction mode for the current block. Possible values are
   (a) DC Prediction (1 segment)
   (b) Planar Prediction (1 segment)
   (c) DMM Mode 1 - Explicit Wedgelet (2 segments)
   (d) DMM Mode 2 - Intra-Predicted Wedgelet (2 segments)
2. For each segment, a residual DC value (in the pixel domain) is signaled in the bitstream

The resulting residual DC values are optionally 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. This reduces the bit depth of residual magnitudes for sequences with reduced depth range due to quantization. This technique is further described in Section 3 of this paper.

3. DEPTH LOOKUP TABLE

An analysis of the JCT-3V 3DV test sequences [4] has shown that many depth maps do not utilize the full available depth

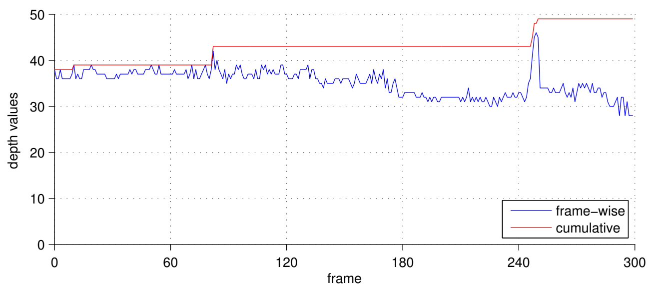

Fig. 2: Number of different depth values in original, uncompressed depth map of sequence 'Balloons'.

range of 2^8. Only a small amount of different depth levels occur in those sequences due to strong quantization. An example of such a depth histogram is illustrated in Figure 2. The advantage of using the DLT in these cases is the reduced bit depth required for signaling the residual values.

3.1. 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. 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 $i_{\max} = \max(I[d])$.

3.2. 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.

$$i_{\text{resi}} = I\big[D[d_{\text{orig}}]\big] - I\big[D[d_{\text{pred}}]\big] \quad (1)$$

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 slice header.

4. EXPERIMENTAL RESULTS

The proposed Simplified Depth Coding (SDC) approach was integrated into the JCT-3V HEVC Test Model (HTM 4.0.1) [3] as an alternative intra coding mode. As the proposed method is an intra-coding tool, the performance with respect to the intra-only coding configuration is investigated. The simulations follow the common test conditions of the JCT-3V standardization activity [4].

4.1. 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: SDC), 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 JCT-3V and complies with the common test conditions of the JCT-3V standardization activity [4].

Table 1: BD-Rate savings on total (incl. texture) bitrate and synthesized PSNR for all-intra coding configuration.

Sequence	HTM+SDC	HTM+SDC+DLT
Balloons	-0.5 %	-2.0 %
Kendo	-1.5 %	-2.7 %
Newspaper_CC	-0.6 %	-2.5 %
GT_Fly	-1.8 %	-2.0 %
Poznan_Hall2	-1.7 %	-3.3 %
Poznan_Street	-0.6 %	-0.7 %
Undo_Dancer	-2.4 %	-2.7 %
1024x768	-0.9 %	-2.4 %
1920x1088	-1.6 %	-2.2 %
AVERAGE	**-1.3 %**	**-2.3 %**

The presented bitrate savings are always measured with respect to the total bitrate including texture bitrate, which accounts for about 90% of the total rate in a typical 3D video stream. As the proposed coding scheme does not touch the coding of the texture component, the bitrate savings on the depth component alone is approximately 15-20%.

4.2. Subjective Evaluation

As already stated in Section 1, the most prominent advantage of the proposed method compared to a conventional transform-based coding scheme is the improved preservation of depth discontinuities due to the removal of ringing artifacts along object boundaries. 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+SDC+DLT

Fig. 3: Synthesis results for sequence *Undo_Dancer* based on (a) uncompressed, (b) HTM 4.0.1 and (c) HTM+SDC-based coded depth map at the same bitrate.

Figure 3 shows synthesis results based on depth maps coded at the same bitrate. The HTM 4.0.1 encoding in Figure 3b shows geometrical deformations (person's head) while the SDC method yields convincing synthesis quality, relatively close to the synthesis based on the uncompressed depth map.

4.3. Computational Complexity

At the encoder the proposed SDC coding mode is slightly more complex than the reference software, because of the additional testing of the alternative coding path. The decoding process of SDC-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+SDC has about 114% encoder and only 70% decoder complexity, relative to HTM 4.0.1. The usage of the Depth Lookup Table does not have an impact on the complexity, because there are no additional computations necessary for the mapping of depth values to table indices.

5. CONCLUSION

In this paper a novel alternative coding path for depth maps is proposed. It is demonstrated that SDC 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 even improved by reducing geometrical distortions coming from ringing artifacts along depth discontinuities.

Further research is needed on the possibility 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 has to be investigated whether the simplified coding tree for depth maps can also be applied to motion predicted blocks.

6. REFERENCES

[1] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, and T. Wiegand. High Efficiency Video Coding (HEVC) text specification draft 9. Technical report, JCTVC-K1003, October 2012.

[2] F. Jäger. Contour-based segmentation and coding for depth map compression. In *Proceedings of IEEE Visual Communications and Image Processing*, pages 1–4, 2011.

[3] Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) of ITU-T VCEG and ISO/IEC MPEG. 3D-HEVC Test Model 1. Technical report, JCT3V-A1005, July 2012.

[4] Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) of ITU-T VCEG and ISO/IEC MPEG. Common test conditions of 3DV core experiments. Technical report, JCT3V-B1100, October 2012.

[5] 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*, pages 3421–3424. IEEE, 2010.

[6] 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*, pages 149–152, May 2012.

[7] 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*, volume 1, pages 201–204, 2007.

[8] 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*, volume 6055, pages 93–100, 2006.

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