|  |  |
| --- | --- |
| **Joint Collaborative Team on 3D Video Coding Extension Development**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  1st Meeting: Stockholm, SE, 16–20 July 2012 | Document: JCT2-A0010 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **3D-CE6.h related: Model-based Intra Coding for Depth Maps using a Depth Lookup Table** | | |
| *Status:* | Input Document | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Fabian Jäger Institut für Nachrichtentechnik RWTH Aachen University | Tel: Email: | +49 (0) 241 80 27678 jaeger@ient.rwth-aachen.de |
| *Source:* | RWTH Aachen University | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This contribution presents a modified intra-coding scheme for depth map coding in 3D video. As depth maps show unique characteristics like piecewise smooth regions bounded by sharp edges at depth discontinuities, new coding tools are required to approximate these signal characteristics. In the current 3DV-HTM software, there are two kinds of intra prediction modes for depth maps: 1) The directional intra prediction known from HEVC and 2) the depth modeling modes (DMM). The latter can enhance the BD-rate of synthesized views especially when modeling the mentioned depth discontinuities.  
In this contribution a very similar prediction scheme as DMM is used to approximate the depth map’s sharp edges. With DMM the resulting residual is transformed and quantized as with conventional intra prediction modes. At this stage the proposed Depth Mode-based Coding (DMC) does not use transform and quantization to code the resulting residual signal. The DMC prediction stage always results in two to three depth segments per CTB and for each of these segments a single residual DC depth value is coded. By skipping the transform step and coding the residual based on the pixel-domain information, ringing artifacts are eliminated for DMC-coded blocks. Moreover, the number of bits to signal the residual values for each segment can be further reduced by integrating a Depth Lookup Table (DLT), which depth values maps to valid depth values of the original depth map. The DLT is constructed based on an initial analysis of the input depth map and is then coded in the SPS. For sequences with strongly quantized depth maps DLT yields additional gain on top of the gain coming from DMC alone.

**Random Access Coding Performance**

The proposed intra coding scheme for depth maps results in **average BD-Rate savings of 4.02 %** (depth rate) and **up to 8.75 %** (depth rate) for certain sequences, like Newspaper, Kendo and Balloons.  
In terms of the overall coding performance including texture and depth DMC yields 0.76 % BD-Rate savings on average.

**All Intra Coding Performance**

For an all-intra test case, the proposed coding scheme yields **average BD-Rate savings of 24.06 %** (depth rate) and **up to 42.27 %** (depth rate) for the sequence Kendo.  
In terms of the overall intra coding performance including texture and depth DMC yields 1.46 % BD-Rate savings on average.

# Introduction

Depth maps, like in 3D video, show different signal characteristics compared to natural video data. Distortions in these 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, while the depth information is never directly shown to the user. Compressing depth maps with algorithms optimized for natural 2D videos results in strong ringing artifacts along depth discontinuities, which then produce geometric distortions in the synthesized views. Thus, new compression algorithms have to be developed that are adapted to the signal characteristics of depth maps.

Previous work on compression of depth data regarded depth data as gray-colored video and compressed it with conventional transform-based video coding algorithms as found in H.264/AVC [[1](#Platzhalter1)]. It was shown [[2](#Jäg11)] 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 [[2](#Jäg11)].

More recent depth compression algorithms approximate the depth map's signal characteristics by partitioning into triangular meshes [[3](#Sar10)] or platelets [[4](#Mor06)] and modeling each segment by an appropriate 2D function. These purely 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](#Liu10)]. Here, an SD-coded block is partitioned into two segments, which are described by two constant depth values.

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](#Jäg11)]. The disadvantage of this approach is the inability of reaching low bitrates due to the lossless encoding of depth contours.

# General Concept

The proposed DMC coding approach is an extension of the intra coding mode, which is available in the HEVC-based 3DV-HTM reference software. For a DMC-coded block, the prediction mode is still INTRA. An additional DMC-Flag signals the usage of DMC prediction and coding. If a block is coded with DMC, partition size is always 2Nx2N and therefore not signaled in the bitstream. Instead of coding quantized transform coefficients DMC-coded blocks need to code the following types of information:

1. The **type of segmentation/prediction** of the current block. Possible values are
   1. DC (no segmentation)
   2. Gradient (no segmentation)
   3. Edge (segmentation into two segments by a straight line)
   4. 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:
   1. For Edge-segmentation: Start/End of the straight line of the segmentation
   2. 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.

# **Available Prediction Modes**

DMC-coded depth map coding units are predicted by one of four available prediction modes. The optimal mode is selected based on the VSO 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 section.

## 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.

## 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 in the lower-right corner of the current block. As depicted in Figure 1, the bottom row is linearly interpolated from the values and , the right column respectively from and . In a second step, all remaining pixel values are bilinearly interpolated from the surrounding values.

Macintosh HD:Users:jaeger:Documents:Publications:VCIP 2012:img:planar_AB.pdf

Figure 1 In Gradient Prediction the bottom and right pixels of the current block are interpolated between already coded samples and the predictively coded Z value. Afterwards, all the remaining pixels are bilinearly interpolated.

The 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 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.

## 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.

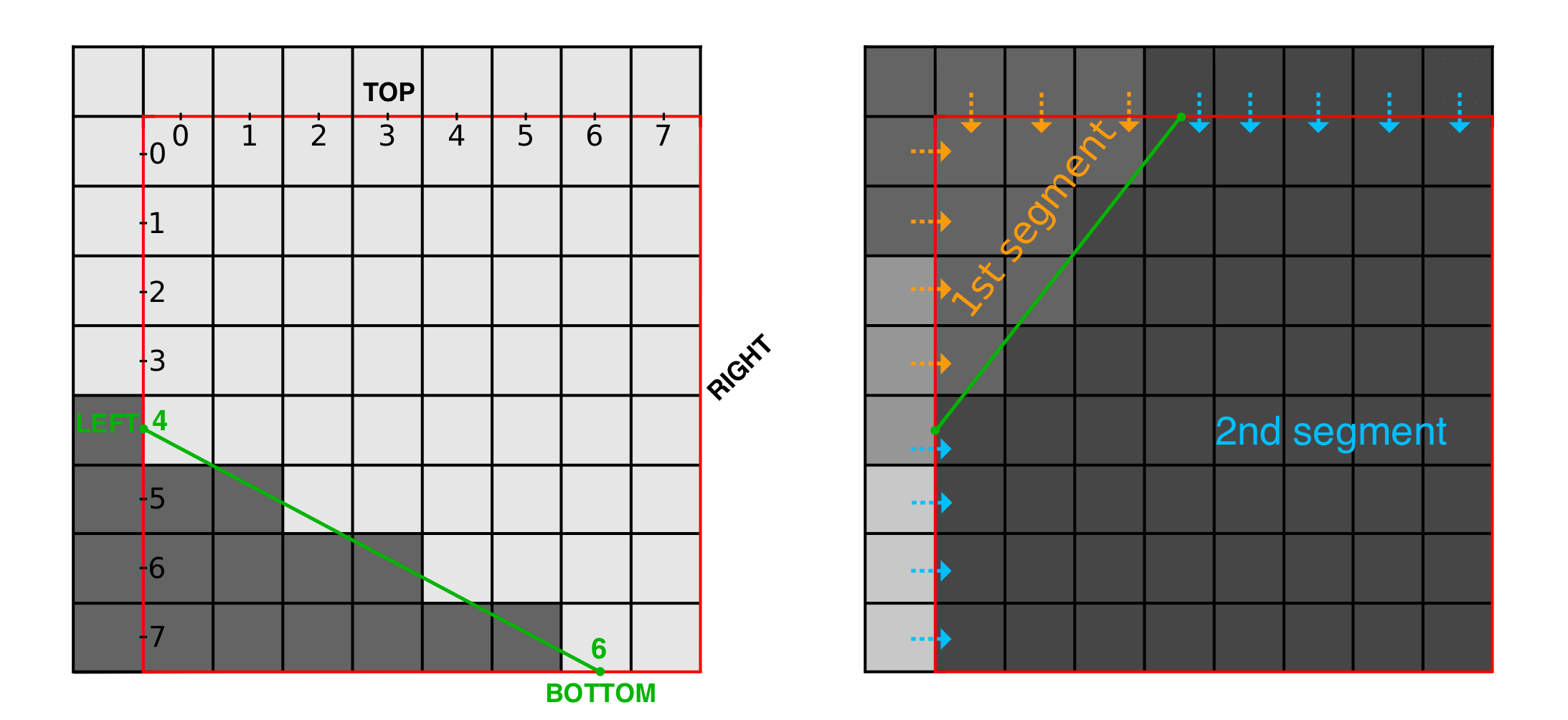


Figure 2 Segmentation of a depth block when using Edge Prediction. Edge information is defined by start/end side (here: LEFT/BOTTOM) and corresponding index (here: 4/6). DC values for each segment are predicted by neighboring depth values.

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 prediction are residual coded using the CABAC entropy coder.

A very similar prediction mode is found as part of the Depth Modeling Modes (DMM), which uses a different signaling of the edge information. Moreover, the resulting residual of DMM predicted blocks is still transform coded.

## 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 then 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 reference software also allow texture-to-depth prediction, but the DMM is more restrictive as it does only allow for two segments and there is a subsequent transform step instead of coding the depth values directly.

# Depth Lookup Table

An analysis of the 3DV test sequences [[6](#MPE11)] has shown that the estimated depth maps do not utilize the full available depth range of . Only a small amount of different depth levels occur in those sequences due to strong quantization. In an initial analysis step, DMC therefore constructs a dynamic depth lookup-table by analyzing a certain number of frames of the input sequence. This depth lookup-table is used during the coding process to reduce the effective signal bit-depth and consequently the range of residual signal coefficients used by other DMC coding tools.

## Analysis Step

In the analysis step the encoder reads a pre-defined 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 depth values to valid depth values based on the original uncompressed depth map.

**Algorithmic Details**

Input: Depth map of pixels at time instance

Output: Depth Lookup Table

Index Lookup Table

Depth Mapping Table

Number of valid depth values

Algorithm:

1. Initialization

* boolean vector for all depth values
* index counter

1. Process each pixel position in for multiple time instances :

* Set to mark valid depth values

1. Count number of values in 🡪
2. For each with :

* Set
* Set
* Set

1. For each with :

* Find and
* Set

Set

## **Coding Step**

Instead of coding a residual depth value for a given coding unit, the depth value is mapped to its corresponding index in the list of valid depth maps. This mapping table needs to be transmitted to the decoder for the inverse lookup from index to valid depth value.

The advantage of using this lookup table is the reduced bit depth of the residual index for sequences with reduced depth range (e.g. all estimated depth maps in the MPEG 3DV activity [[6](#MPE11)]).

**Algorithmic Details**

Input: Original depth value

Predicted depth value

Index Lookup Table

Number of valid depth values

Output: Residual index to be coded

Algorithm:

The computed residual index is then coded with a significance flag, a sign flag and with bits for the magnitude of the residual index.

# Details on Implementation

The proposed algorithm is implemented into the 3DV-HTM 3.1 reference software. All modifications to the source code are encapsulated by preprocessor statements (RWTH\_DMC and RWTH\_DLT), which make it possible to (de-) activate the coding tools easily.

The algorithm and its implementation do not interfere with other coding tools of the reference software and can therefore be uses in addition to all available tools.

As three of the four DMC prediction modes are highly similar to the Depth Modeling Modes (DMM) that are already in the reference implementation, DMM is disabled for our simulations. In the reference simulation, DMM was enabled.

# Complexity Discussion

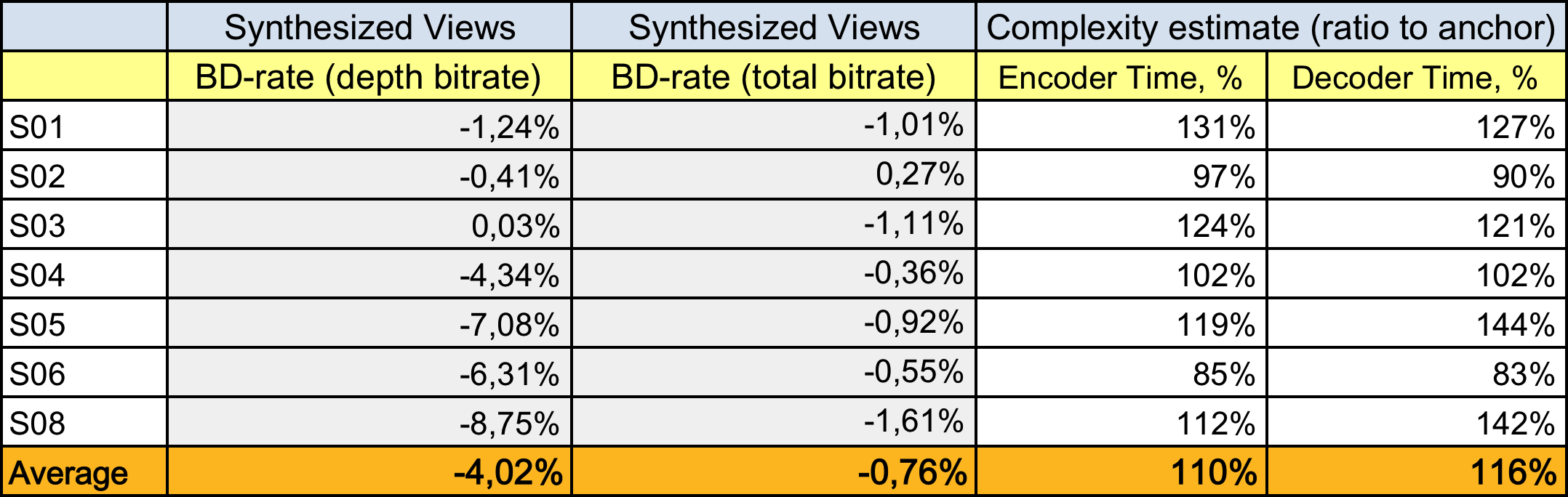
DMC does not introduce significant additional computational complexity compared to the current 3DV-HTM 3.1 reference software. Only the Edge Prediction mode is relatively complex at the encoder side, as it needs to test all possible segmentations to find the optimum. This behavior is very similar to the DMM Wedgelet prediction that can be found in the current HTM reference software.  
All other DMC components and especially the decoder-side algorithms are very low complex as there is neither a de-quantization nor an inverse transform for DMC-coded blocks involved.

# Simulation Results

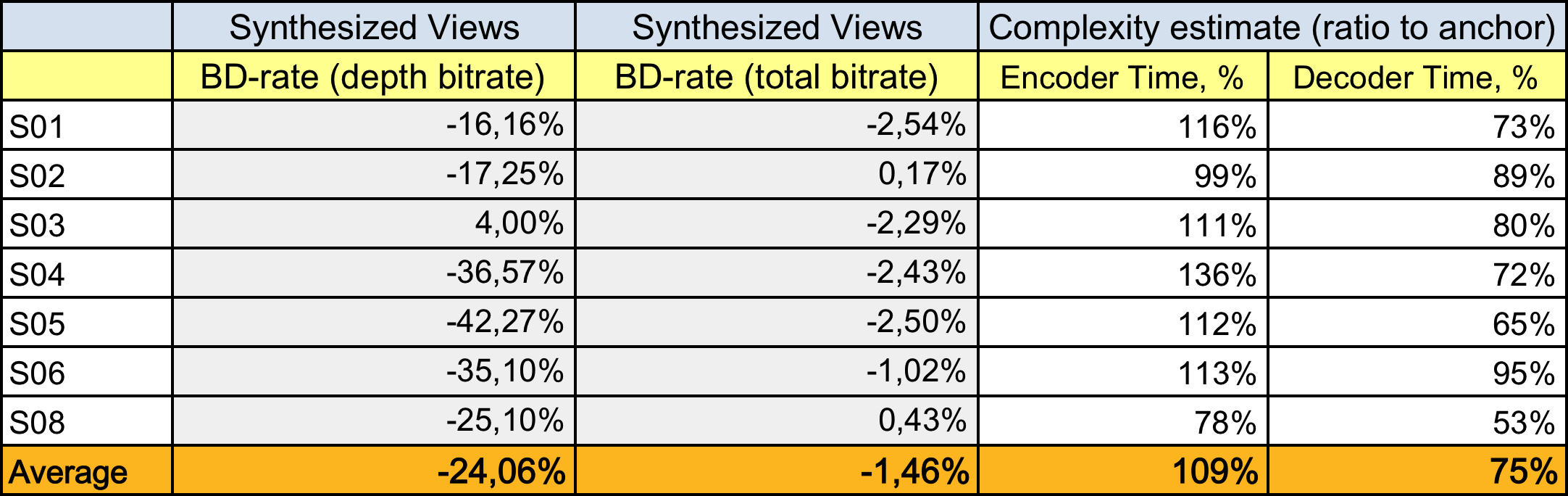
The DMC simulations were performed according the common test conditions [[7](#Hei11)]. For the All-Intra coder configuration configuration files from the Core Experiment 6.h on Depth Map Intra Coding Tools were used. The only deviation from the common test conditions was the disabling of DMM for the DMC simulations, as DMM prediction modes are very similar to the prediction used in DMC and the complexity increase would be much larger with potentially little additional gain on bitrate savings.

The measurement of the computational complexity is to be taken with care as it is based on a heterogeneous cluster with different hardware configurations.

## Random Access Configuration



## All Intra Configuration



# Cross Check

A cross check of the proposed method was done by Huawei Technologies. They did not find any issues with the implementation. At the time of submission of this document, cross check simulations were still ongoing.

# Conclusion

In this contribution a model-based algorithm for intra-coding of depth maps is proposed. The proposed DMC method removes ringing artifacts as known from transform-based coding by directly signaling the pixel-domain information in the bitstream. The piecewise smooth regions of depth data are predicted by either constant depth values or by a planar mode, which is able to model depth gradients. In the subsequent residual coding step the proposed Depth Lookup Table maps the residual DC values for each segment to residual indices, which are then entropy coded.

As this contribution shows significant coding gain for intra coding of depth maps, this method should be further investigated in a corresponding Core Experiment.

# References

x

|  |  |
| --- | --- |
| [1] | P. and Smolic, A. and Muller, K. and Wiegand, T. Merkle, "Multi-view video plus depth representation and coding," *IEEE International Conference on Image Processing*, 2007. |
| [2] | F. Jäger, "Contour-based segmentation and coding for depth map compression," *Proceedings of IEEE Visual Communications and Image Processing*, 2011. |
| [3] | M. and Zia, W. and Diepold, K. Sarkis, "Fast Depth Map Compression and Meshing with Compressed Tritree," *Computer Vision*, 2010. |
| [4] | Y. and de With, P.H.N. and Farin, D. Morvan, "Platelet-based coding of depth maps for the transmission of multiview images," *Proceedings of SPIE Stereoscopic Displays and Applications*, 2006. |
| [5] | S. and Lai, P.L. and Tian, D. and Gomila, C. and Chen, C.W. Liu, "Sparse dyadic mode for depth map compression," *Proceedings of IEEE International Conference on Image Processing*, 2010. |
| [6] | MPEG Video and Requirement Groups, "Call for Proposals on 3D Video Coding Technology," *MPEG output document N12036*, March 2011. |
| [7] | Dmytro Rusanovskyy Heiko Schwarz, "Common Test Conditions for HEVC- and AVC-based 3DV," ISO/IEC JTC1/SC29/WG11 MPEG, N12560 2012. |
| [8] | P. and Smolic, A. and Muller, K. and Wiegand, T. Merkle, "Multi-view video plus depth representation and codin," *IEEE International Conference on Image Processing*, 2007. |
| [9] | D. and Schwarz, H. and Wiegand, T. Marpe, "Context-based adaptive binary arithmetic coding in the H. 264/AVC video compression standard," *IEEE Transactions on Circuits and Systems for Video Technology*, 2003. |
| [10] | Dmytro Rusanovskyy Heiko Schwarz, "Common Test Conditions for HEVC- and AVC-based 3DV," ISO/IEC JTC1/SC29/WG11 MPEG, N12352 2011. |
| [11] | Dmytro Rusanovskyy Heiko Schwarz, "Common Test Conditions for 3DV experimentation," ISO/IEC JTC1/SC29/WG11 MPEG, N12745 2012. |
| [12] | Krzysztof Wegner Heiko Schwarz, "Test Model under Consideration for HEVC based 3D video coding v3.0," ISO/IEC JTC1/SC29/WG11 MPEG, N12744 2012. |

x

# Patent rights declaration(s)

**RWTH Aachen University may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**