|  |  |
| --- | --- |
| **Joint Collaborative Team on 3D Video Coding Extensions**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  9th Meeting: Sapporo, JP, 3 – 9 July 2014 | Document: JCT3V-I0058 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Segment-wise Prediction (SP) Merge Candidate** | | |
| *Status:* | Input Document | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Fabian Jäger Institut für Nachrichtentechnik RWTH Aachen University  Jacek Konieczny, Giovanni Cordara Huawei Technologies Munich Office, European Research Center | Tel: Email: | +49 (0) 241 80 27678 [jaeger@ient.rwth-aachen.de](mailto:jaeger@ient.rwth-aachen.de)  +49 (0) 89 158834 4334 [jacek.konieczny@huawei.com](mailto:jacek.konieczny@huawei.com) |
| *Source:* | RWTH Aachen University | | |

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

# Abstract

At the 6th JCT-3V meeting, a motion/disparity prediction method was introduced in JCT3V-G0106 [1] that uses a depth-derived binary segmentation mask for the derivation of PU partitioning and for merging of two prediction signals. The method is called Depth-based Block Partitioning (DBBP). The segmentation process in JCT3V-G0106 results in two segments, which motion information is coded as a conventional bi-partitioned coding unit with two sets of motion information. This contribution proposed to use the same prediction method as a new Segment-wise Prediction (SP) merge candidate. The corresponding two sets motion information are derived from the already checked merge candidate positions. The proposed method yields 0.12% coding efficiency gain under common test conditions for video PSNR over total bitrate.

# Depth-based Block Partitioning Algorithm

The proposed algorithm for using depth information for block partitioning in texture views consists of three sequential steps, which will be described in more detail in the following section.

## Depth Segmentation

In an initial step the collocated depth block of the current coded tree block (CTB) of the texture component is segmented into two arbitrarily shaped segments. As the depth component is coded after the corresponding texture view in the current CTC, a virtual depth map is derived from the base view’s reconstructed depth and shifted by a disparity vector, which is itself derived from the neighboring blocks (by means of DoNBDV).

The segmentation of the (virtual) depth map is performed based on a very simple thresholding mechanism where the threshold is computed from the mean depth value.

Here, defines the width/height of the current texture block and resembles the already coded, corresponding depth map of the reference view’s texture frame.

Afterwards, a binary segmentation mask is generated based on as follows.

The resulting binary segmentation mask defines the shape of the partitioning of the texture block. While motion or disparity compensation in a modern video coder (e.g. in HEVC) is performed on a block-basis, an arbitrarily shaped block partitioning typically requires pixel-based compensation. This concept is applied in pixel-based view-synthesis prediction (VSP), which warps each pixel position based on the corresponding depth value to the position of the particular reference frame. By this fine-grained compensation approach higher order deformations can be approximated. To achieve this amount of precision for the prediction stage of a video coder, pixel-based VSP introduces very high complexity compared to conventional block-based motion/disparity compensation. This is mainly due to the irregular access to the reference buffer and the pixel-wise conversion between depth and disparity.

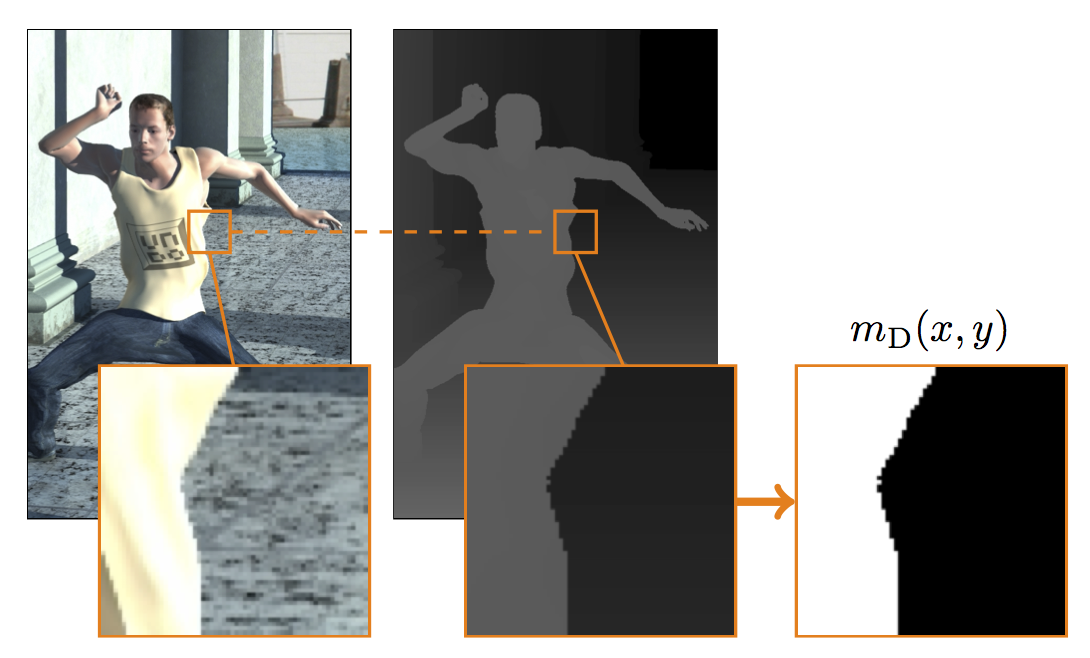


Figure 1: Cropped frame from *Undo\_Dancer* test sequence with magnified component blocks of an example coding unit. The collocated depth block segments the block into foreground and background, which is defined by a binary mask .

To overcome this issue of VSP, the proposed depth-based block-partitioning (DBBP) scheme still uses block-based compensation in the prediction stage, as it is described in the following subsection.

## Block-based Compensation

In the proposed DBBP scheme, the actual motion or disparity compensation is performed on a partitioning, which means that the full CTB is shifted by the coded vector information. This full-size motion/disparity compensation is performed twice, once for each segment, and results in two prediction signals and .

Consequently, two sets of vector information need to be coded for a DBBP block. The assumption behind this approach is that a texture block is typically segmented into foreground and background based on the collocated depth block. These two depth layers can then be compensated independently by their own sets of motion or disparity vectors.

## Merging of Prediction Signals

After having generated two full-size prediction signals and for a DBBP-coded block, the segmentation mask is used to merge these into the final prediction signal for the current texture CTB.

By merging the two prediction signals, shape information from the depth map allows to independently compensate foreground and background objects in the same texture block. At the same time, DBBP does not require pixel-wise motion/disparity compensation. Memory access to the reference buffers is always regular (block-based) for DBBP-coded blocks in contrast to approaches like VSP. Moreover, DBBP always uses full-size blocks for compensation. This is preferable with respect to complexity, because of the higher probability of finding the data in the memory cache.

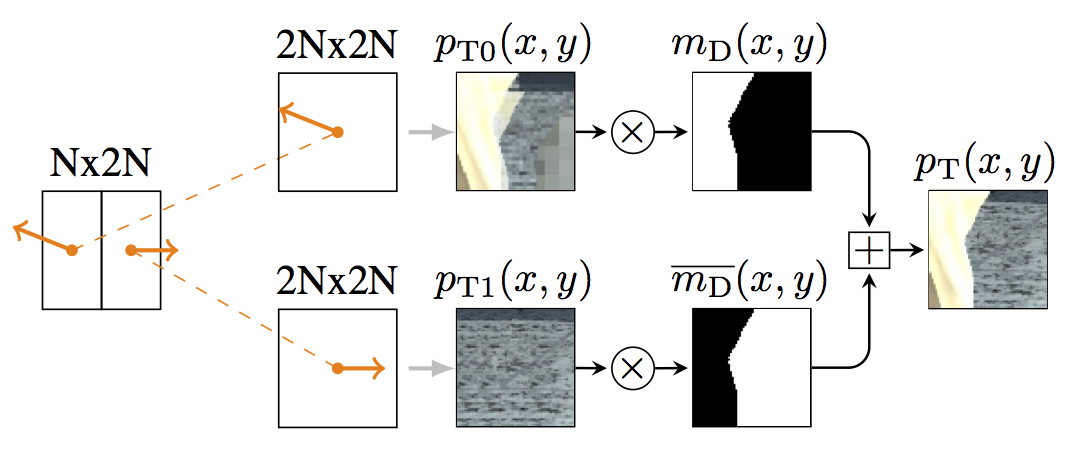


Figure 2: DBBP merging process: For each of the two decoded motion parameters a motion compensation is performed. The resulting prediction signals and are combined using the DBBP mask .

# Segment-wise Prediction Merge Candidate

## Derivation of Motion Information

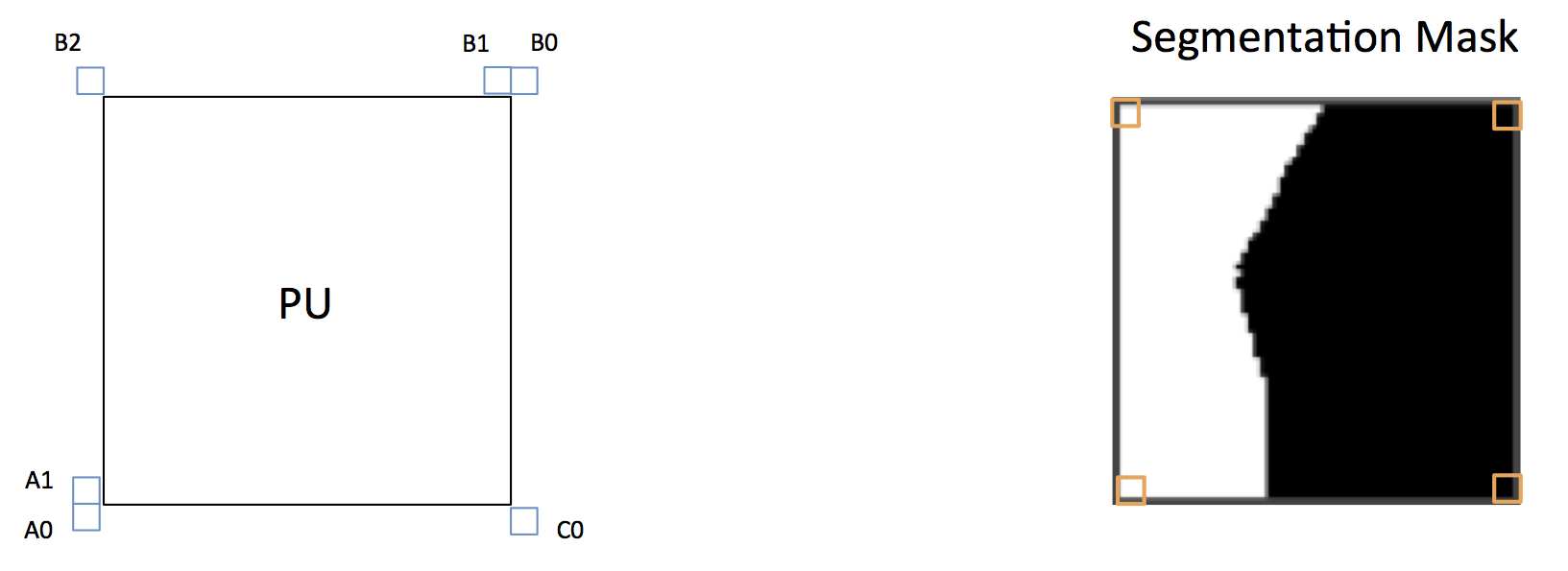


Figure 3: For the SP merge candidate, the conventional merge candidate positions are checked and assigned to one of the two segments, based on the depth-derived segmentation mask.

When the encoder selects the SP merge candidate, the HEVC merge candidates are checked for availability and assigned to one of the two segments, based on the depth-derived segmentation mask. If both segments can be assigned a set of motion information, SP is available as a merge candidate. The prediction mechanism is the same as for DBBP, using the two sets of motion information (first available candidate of each segment).

* Check conventional neighboring merge candidates for availability and assign them to either segment 0 or segment 1, depending on closest segmentation mask corner value.
* If motion info for both segments is found, add segment prediction (SP) merge candidate to the merge candidate list.
* If encoder selects SP merge candidate, apply DBBP-like prediction mechanism to PU
* Only merge index is required to do segment-wise prediction for a PU
* Motion information for both segments is derived from neighboring motion info.

# Simulation Results

## HTM-11.0+SP-Merge VS. HTM-11.0

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | video 0 | video 1 | video 2 | video PSNR / video bitrate | video PSNR / total bitrate | synth PSNR / total bitrate | enc time | dec time |
| Balloons | 0,00% | -0,37% | -0,09% | -0,08% | -0,05% | -0,09% | 96,4% | 103,4% |
| Kendo | 0,00% | -0,36% | -0,32% | -0,11% | -0,12% | -0,10% | 107,2% | 105,2% |
| Newspaper\_CC | 0,00% | -0,45% | -0,30% | -0,15% | -0,09% | -0,14% | 108,1% | 102,7% |
| GT\_Fly | 0,00% | -0,29% | -0,16% | -0,04% | -0,02% | -0,05% | 106,6% | 105,9% |
| Poznan\_Hall2 | 0,00% | -0,27% | -0,45% | -0,14% | -0,11% | -0,01% | 103,1% | 109,9% |
| Poznan\_Street | 0,00% | -0,20% | -0,36% | -0,11% | -0,08% | -0,02% | 109,9% | 108,2% |
| Undo\_Dancer | 0,00% | -1,58% | -1,60% | -0,43% | -0,37% | -0,28% | 101,1% | 108,5% |
| Shark | 0,00% | -1,14% | -0,57% | -0,18% | -0,13% | -0,08% | 108,0% | 107,1% |
| 1024x768 | 0,00% | -0,39% | -0,24% | -0,12% | -0,09% | -0,11% | 103,9% | 103,8% |
| 1920x1088 | 0,00% | -0,70% | -0,63% | -0,18% | -0,14% | -0,09% | 105,7% | 107,9% |
| **average** | **0,00%** | **-0,58%** | **-0,48%** | **-0,16%** | **-0,12%** | **-0,10%** | **105,1%** | **106,4%** |

## HTM-11.0+SP-Merge VS. HTM-11.0 without DBBP

The following (informal) results compare the combination of both segment-wise prediction tools (DBBP and SP) with an HTM version without these two tools, which are essentially using the same prediction process.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | video 0 | video 1 | video 2 | video PSNR / video bitrate | video PSNR / total bitrate | synth PSNR / total bitrate |
| Balloons | 0,0% | -0,7% | -0,4% | -0,2% | -0,2% | -0,1% |
| Kendo | 0,0% | -0,5% | -0,4% | -0,2% | -0,1% | 0,0% |
| Newspaper\_CC | 0,0% | -0,7% | -0,6% | -0,3% | -0,2% | -0,2% |
| GT\_Fly | 0,0% | -0,8% | -0,6% | -0,2% | -0,2% | -0,1% |
| Poznan\_Hall2 | 0,0% | -0,7% | -0,3% | -0,2% | -0,2% | -0,1% |
| Poznan\_Street | 0,0% | -0,8% | -1,1% | -0,3% | -0,3% | -0,2% |
| Undo\_Dancer | 0,0% | -3,1% | -3,1% | -0,9% | -0,8% | -0,6% |
| Shark | 0,0% | -2,0% | -1,6% | -0,4% | -0,3% | -0,2% |
| 1024x768 | 0,0% | -0,6% | -0,5% | -0,2% | -0,2% | -0,1% |
| 1920x1088 | 0,0% | -1,5% | -1,4% | -0,4% | -0,4% | -0,2% |
| **average** | **0,0%** | **-1,2%** | **-1,0%** | **-0,3%** | **-0,3%** | **-0,2%** |

# Cross Check

The cross check was performed by Qualcomm. They investigated the proposed changes to the software and the specification text and ran the same simulation configuration to confirm the presented simulation results.

# Conclusion

In this contribution we propose to add a new Segment-wise Prediction (SP) merge candidate to 3D-HEVC, which allows using segment-wise motion or disparity prediction by means of a special merge candidate. This candidate derives two sets of motion information to do the prediction, which is itself the same as proposed for DBBP.

# Patent rights declaration

**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).**

**Huawei Technologies 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).**

# References

[1] F. Jäger, J. Konieczny, and G. Cordara, “CE3: Results on Depth-based Block Partitioning (DBBP),” 7th Meeting, San Jose, USA, 2014.

[2] J. L. Lin, Y. W. Chen, T. D. Chuang, X. Zhang, Y. W. Huang, and S. Lei, “Improvement on the signaling of DBBP,” Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) of ITU-T VCEG and ISO/IEC MPEG, Valencia, JCT3V-H0094, Apr. 2014.