|  |  |
| --- | --- |
| **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**  2nd Meeting: Shanghai, CN, 13–19 Oct. 2012 | Document: JCT3V-B0054 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **3D-CE1.a: Generalized view synthesis prediction (GVSP) mode** | | |
| *Status:* | Input Document | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Xin Zhao Li Zhang Ying Chen  5775 Morehouse Drive San Diego, CA 92121 USA | Tel: Email: | 86-10-5776-0696 [xinzhao@qti.qualcomm.com](mailto:xinzhao@qti.qualcomm.com)  1-858-651-6660 [lizhang@qti.qualcomm.com](mailto:lizhang@qti.qualcomm.com)  1-858-845-6589 [cheny@qti.qualcomm.com](mailto:cheny@qti.qualcomm.com) |
| *Source:* | Qualcomm Incorporated | | |

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

# Abstract

In current 3DV-ATM, View Synthesis Prediction (VSP) is performed at macroblock (MB) level. To further improve the coding performance, the concept of VSP is extended to sub-MB level with proposed block-level signaling of VSP. This proposal is technically identical to JCT3V-A0103, but integrated into the latest 3D-ATM and incorporated with block-based VSP (B-VSP). Without complexity increase, the proposed method achieves 0.61% BD Rate gain for texture coding under common test conditition (CTC).

# Introduction

In previous 3DV-ATM, in-loop View Synthesis Prediction (VSP), which utilizes a synthesized reference image (VSP frame) generated by a forward warping process, is utilized for enhanced texture coding. Due to the high complexity of generating a synthesized reference picture in VSP, a B-VSP implementation is proposed in [1] and incorporated in the latest 3DV-ATM version 5.1. With B-VSP, texture pixels of a dependent view are not predicted from a synthesized VSP frame, but directly from the base view using a backward warping process. Therefore, the process of generating a VSP reference frame is removed by B-VSP, and the frame buffer for storing a VSP frame is also saved.

In the current WD, VSP is applied at MB level, i.e., VSP skip/direct [2] mode. In this document, generalized VSP (GVSP) mode is proposed by adding a flag per MB partition to indicate whether a MB partition is predicted using B-VSP.

# Proposal

The proposed GVSP mode is summarized as follows.

For a non-skip MB, a flag is introduced to indicate whether the whole MB is predicted using B-VSP or not.

For a MB which is not totally predicted from B-VSP, each of its MB partition (no smaller than 8x8) has a flag to signal whether the MB partition is predicted using B-VSP or not.

If a MB or MB partition is predicted from B-VSP, no reference index is transmitted and residual information such as CBP and coded coefficients are signaled as usual.

As the prediction of current partition by GVSP is derived using B-VSP, the depth map and the texture image in base view are used to perform the prediction in the proposed GVSP mode.

# Experimental results

In this section, objective coding performance of the proposed method is presented and compared with 3DV-ATM v5.1 anchor EHP profile.

The proposed algorithm is implemented based on 3DV-ATM v5.1 software. As specified in the common test condition (CTC) [3], four rate distortion points are generated for each sequence with QP equal to 41, 36, 31 and 26.

## Bitrates vs. decoded texture view PSNRs

In Table 1, the results are reported for coding gain with respect to the anchor where the bitrates represent the bitrates of texture and the PSNR values are the average values of all decoded texture views. The BD bitrate reduction is about 0.61%.

Table 2 lists the coding gain of the proposed algorithm with respect to the anchor where the bitrates represent the total bitrates of bitstream containing both texture and depth of all three views and the PSNR values are the average PSNR values of the three decoded texture views. The BD bitrate reduction is about 0.55%.

**Table 1: Coding gain with respect to anchor for 3-view case   
(total bitrates of texture vs. average PSNR of decoded texture views)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **Coding Gain relative to Anchor**  **(3-view)** | | |
| Resolution | Class | Seq. No. | Seq. Name | BD Bitrate (%) | BD PSNR Gain (dB) |
| 1920x1088 | Class A | S01 | Poznan\_Hall2 | -0.90 | 0.03 |
| S02 | Poznan\_Street | -0.33 | 0.01 |
| S03 | Undo\_Dancer | -0.57 | 0.02 |
| S04 | GT\_Fly | -0.78 | 0.03 |
| 1024x768 | Class C | S05 | Kendo | -0.73 | 0.03 |
| S06 | Balloons | -0.65 | 0.03 |
| S08 | Newspaper | -0.32 | 0.01 |
|  |  |  | **Overall Avg** | -0.61 | 0.02 |

**Table 2: Coding gain with respect to anchor for 3-view case   
(total bitrates vs. average PSNR of decoded texture views)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **Coding Gain relative to Anchor**  **(3-view)** | | |
| Resolution | Class | Seq. No. | Seq. Name | BD Bitrate (%) | BD PSNR Gain (dB) |
| 1920x1088 | Class A | S01 | Poznan\_Hall2 | -0.81 | 0.03 |
| S02 | Poznan\_Street | -0.31 | 0.01 |
| S03 | Undo\_Dancer | -0.54 | 0.02 |
| S04 | GT\_Fly | -0.74 | 0.03 |
| 1024x768 | Class C | S05 | Kendo | -0.62 | 0.03 |
| S06 | Balloons | -0.59 | 0.03 |
| S08 | Newspaper | -0.28 | 0.01 |
|  |  |  | **Overall Avg** | -0.55 | 0.02 |

## Bitrates vs. synthesized texture view PSNRs

As specified in the CTC document [3], intermediate views are synthesized between decoded views for each sequence. Table 3 lists the coding gain with respect to the anchor where the bitrates represent the total bitrates of texture and depth and the PSNR values are the average synthesized view PSNRs. The BD rate reduction is about 0.52% for 3-view case.

**Table 3: Coding gain with respect to anchor for 3-view case   
(total bitrates vs. average PSNR of synthesized views)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **Coding Gain relative to Anchor**  **(3-view)** | | |
| Resolution | Class | Seq. No. | Seq. Name | BD Bitrate (%) | BD PSNR Gain (dB) |
| 1920x1088 | Class A | S01 | Poznan\_Hall2 | -0.85 | 0.03 |
| S02 | Poznan\_Street | -0.24 | 0.01 |
| S03 | Undo\_Dancer | -0.41 | 0.01 |
| S04 | GT\_Fly | -0.68 | 0.02 |
| 1024x768 | Class C | S05 | Kendo | -0.63 | 0.03 |
| S06 | Balloons | -0.57 | 0.02 |
| S08 | Newspaper | -0.24 | 0.01 |
|  |  |  | **Overall Avg** | -0.52 | 0.02 |

## Complexity analysis

To analyze the complexity, runtimes of encoding and decoding process were collected from log files, and Table 4 summarizes the runtime ratio of the proposed method to the anchor. Compared to anchor, the proposed encoder has about 9.9% increased runtime for 3-view case which is mainly due to the additional sub-MB partition mode decision with GVSP. The decoding times do not include the I/O times (for output YUV file generation) for both anchor and proposed GVSP. Compared to anchor, the proposed decoder is about 98.05% of the anchor runtime for 3-view case.

**Table 4: Computational time of the proposed method**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Complexity estimate (ratio to anchor) | | |
|  | Encoder Time, % | Decoder Time, % |
| S01 | 109.13% | 97.72% |
| S02 | 108.77% | 96.00% |
| S03 | 107.83% | 97.29% |
| S04 | 108.78% | 96.25% |
| S05 | 115.58% | 100.28% |
| S06 | 110.65% | 100.14% |
| S08 | 108.58% | 98.67% |
| Average | 109.90% | 98.05% |

# Conclusion

The proposed GVSP method is able to bring additional 0.61% gain for texture coding under common test condition without complexity increase. Therefore, it is recommended to adopt this proposal for the next ATM and working draft.

# References

1. W. Su, D. Rusanovskyy, M. M. Hannuksela, “3DV-CE1.a: Block-based View Synthesis Prediction for 3DV-ATM,” JCT3V-A0107, Stockholm, SE, 16–20 July 2012.
2. “Description of 3D video coding technology proposal by ETRI and Kwangwoon University,” ISO/IEC JTC1/SC29/WG11/M23915, San Jose, USA, Feb. 2012.
3. H. Schwarz and D. Rusanovskyy, “Common Test Conditions for 3DV experimentation,” JCT3V-A1100, Stockholm, SE, 16–20 July 2012.

# Patent rights declaration(s)

**Qualcomm Incorporated 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).**