|  |  |
| --- | --- |
| **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-B1101 |

|  |  |  |
| --- | --- | --- |
| *Title:* | **Description of Core Experiment 1 (CE1) on View Synthesis Prediction** | |
| *Status:* | Output Document | |
| *Purpose:* | Core Experiment Description | |
| *Author(s)):* | Shinya Shimizu Fabian Jäger | [shimizu.shinya@lab.ntt.co.jp](mailto:shimizu.shinya@lab.ntt.co.jp) [jaeger@ient.rwth-aachen.de](mailto:jaeger@ient.rwth-aachen.de) | |
| *Source:* | JCT-3V CE Coordinators | |

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

# Abstract

This document describes Core Experiment 1 (CE1) of the development of 3D extensions for HEVC and AVC. This Core Experiment investigates coding methods based on view synthesis prediction, which uses texture and depth information of the independent view to construct a synthesized prediction signal for coding the dependent views.

# Participants

|  |  |  |  |
| --- | --- | --- | --- |
| ***Participant*** | ***Contact*** | ***Email*** | ***Proponent*** |
| NTT | Shinya Shimizu | [shimizu.shinya@lab.ntt.co.jp](mailto:shimizu.shinya@lab.ntt.co.jp) | P/C |
| Mitsubishi Electric | Dong Tian | [tian@merl.com](mailto:tian@merl.com) | P/C |
| RWTH Aachen University | Fabian Jäger | [jaeger@ient.rwth-aachen.de](mailto:jaeger@ient.rwth-aachen.de) | C |
| MediaTek | Yu-Lin Chang  Yi-Wen Chen  Yu-Wen Huang | [yulin.chang@mediatek.com](mailto:yulin.chang@mediatek.com)  [yiwen.chen@mediatek.com](mailto:yiwen.chen@mediatek.com)  [yuwen.huang@mediatek.com](mailto:yuwen.huang@mediatek.com) | P/C |
| Qualcomm | Xin Zhao  Jewon Kang | [xinzhao@qualcomm.com](mailto:xinzhao@qualcomm.com)  [jewonk@qti.qualcomm.com](mailto:jewonk@qti.qualcomm.com) | P/C |
| Nokia | Dmytro Rusanovskyy | [Dmytro.Rusanovskyy@nokia.com](mailto:Dmytro.Rusanovskyy@nokia.com) | P/C |
| Zhejiang University | Yin Zhao | [zhaoyin@zju.edu.cn](mailto:zhaoyin@zju.edu.cn) | P/C |
| LG Electronics | Sehoon Yea | [sehoon.yea@lge.com](mailto:sehoon.yea@lge.com) | C |
| Samsung | Jin Young Lee | [jinyoung79.lee@samsung.com](mailto:jinyoung79.lee@samsung.com) | C |
| ETRI | Gun Bang | [gbang@etri.re.kr](mailto:gbang@etri.re.kr) | C |
| KyungHee University | Kyung Yong Kim | [kimky@khu.ac.kr](mailto:kimky@khu.ac.kr) | C |

(P = proponent, C = cross checker)

# Tools under Investigation

The tools to be investigated in this Core Experiment use the decoded texture and depth information from the independent base view to compute a prediction signal for the texture and depth component of the dependent views. This process is done by means of a warping algorithm, which is similar to the non-normative algorithms used in the final view synthesis stage at the receiver.

For the case of view synthesis prediction (VSP), as tested in this CE, the view synthesis algorithm is part of the coding loop and therefore a normative coding tool.

## CE1.a: View Synthesis Prediction based on 3D-ATM

3D-ATM supports a view synthesis prediction using a backward warping, where displacement shift is determined for each 2x2 block and disparity compensation is applied.

The tools to be investigated can be categorized into the following 5 categories. All separable tools should be evaluated separately according to the following categorization. Reports on combined tools are also encouraged. TABLE I summarizes the combinations of the tools to be studied.

1. Block size for warping (JCT3V-B0094)
2. The disparity vector derivation method from a depth block
3. Removal of VSP picture from the reference picture list (JCT3V-B0054)
4. Skipping motion vector coding (JCT3V-A0107)
5. Block-based VSP using forward warping (JCT3V-B0121)

TABLE : Combinations of the tools under investigation.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Block size | Disparity vector derivation | VSP picture in the reference picture list | MV coding | Warping direction | Proponent |
| ATM  (JCT2-A0107) | 2x2 | N/A | Insert | Skipped | Backward | Anchor |
| Test 1 | 4x4 | Max among 4 corners | Insert | Skipped | Backward | MediaTek |
| Test 2 | 4x4 | Other methods | Insert | Skipped | Backward |  |
| Test 3 | 2x2 | N/A | Remove | N/A | Backward | Qualcomm |
| Test 4 | 2x2 | N/A | Insert | Coded | Backward | NTT/MERL |
| Test 5 (1+3) | 4x4 | Max among 4 corners | Remove | N/A | Backward | MediaTek |
| Test 6 (1+4) | 4x4 | Max among 4 corners | Insert | Coded | Backward |  |
| Test 7 (2+3) | 4x4 | Other methods | Remove | N/A | Backward |  |
| Test 8 (2+4) | 4x4 | Other methods | Insert | Coded | Backward |  |
| Test 9 | 2x2 | N/A | Insert | Skipped | Forward | Zhejiang Univ. |
| Test 10 | 4x4 | Max among 4 corners | Insert | Skipped | Forward | Zhejiang Univ. |

In addition, it is also encouraged to investigate the use of larger block size than 4x4 and the restriction on the amount of 4x4 blocks per picture, similar as motion compensation in AVC.

### JCT3V-B0094: Interview Skip mode with sub-partition scheme

The inter-view skip mode is proposed to eliminate full frame view synthesis for reducing the computational complexity, memory bandwidth, and buffer. In contrast to the current VSP design, the inter-view skip mode is to utilize the depth from the current view and pick up sample predictors from a reference view by using a backward warping. No sample level vector is used, but a maximum depth value in the 4x4 is used for the whole 4x4 block.

**🡪This contribution should be evaluated in terms of the tools in categories (a).**

### JCT3V-B0054: Generalized view synthesis prediction (GVSP) mode

GVSP is a way to enable sub-MB skip/direct using VSP reference pictures. In addition to a vsp\_flag at MB level, a smaller partition, up to 8x8 blocks, can be signaled whether it is predicted from a VSP reference picture. It is also proposed to remove VSP picture from the reference picture list.

**🡪This contribution should be evaluated in terms of the tools in categories (c).**

### JCT3V-A0107: Block-based View Synthesis Prediction for 3DV-ATM

A block-based VSP implementation (B-VSP), which is based on a backward warping process, is utilized in 3D-ATM. With the B-VSP, sample predictors are derived directly from the texture pixels of the reference view, not the synthesized VSP-frame. The displacement vectors required for this process are produced for each pixel from the depth map data of the current view. No motion information is coded when the reference picture index, which corresponds to the VSP, is coded. Zero motion vectors are assumed for such blocks.

**🡪This contribution should be evaluated in terms of the tools in categories (d).**

### JCT3V-B0121: Block-based VSP using Forward Warping

Forward warping Block-based VSP (FBVSP) is proposed to generate VSP prediction samples for a target block in the current coding texture image, which utilizes depth map of a reference view (e.g., base view) instead of that of the current dependent view. For each macroblock using VSP in the target view (called a target block), the synthesis samples are obtained by forward warping of pixels in a proper window in the reference view. The window, i.e., a reference block, is decided by a three-step refinement method based on estimating the correspondences of the four corners in the target block. FBVSP enables block-based VSP for both texture-first and depth-first coding order of the dependent views.

**🡪 This contribution should be evaluated in terms of the tools in categories (e).**

## CE1.h: View Synthesis Prediction based on 3D-HTM

### JCT3V-B0102: View Synthesis Prediction for 3D-HTM

In this contribution, a basic view synthesis prediction which utilizes the merging candidate list is proposed. The proposed VSP scheme is

1. Append a synthetic picture to the reference picture lists, both LIST\_0 and LIST\_1. The synthetic pictures are generated from a decoded texture and depth on the reference view using a simplified VSRS-1D-Fast.
2. Add a VSP candidate to the merging candidate list.

This contribution also investigates whether to transmit non-zero motion vector for the general inter modes. Transmission of non-zero motion vector was recommended.

**🡪This VSP scheme should be further investigated to find out to what extent the proposed method can reduce the bitrate of the depending views.**

### ****JCT3V-B0103: Block-based VSP using synthesized depth map****

The proposed scheme is also based on two-step backward (or reverse) warping.

The first step is a forward warping of depth map. On this step, virtual depth map corresponds to the target picture is generated from already decoded depth map on reference picture. Since actual spatial resolution of depth map is not so high, the forward warping is performed at integer-pixel precision although depth precision is kept. Different from the conventional depth warping, we propose to perform hole-filling with considering picture quality on the final synthesized texture. The generated virtual depth map is utilized only for block-based VSP on texture coding. The virtual depth map for VSP on depth map coding is generated in the conventional method. Note that most of the processes are the same in both cases, so that such processes can be shared in order to minimize overall computational cost.

The second step is generating VSP picture by a backward warping which utilizes the virtual depth map generated at the first step. As proposed in A0107, VSP picture is generated only for the blocks where VSP is used.

**🡪The benefits of using block-based VSP in terms of coding efficiency and especially in terms of complexity are to be investigated.**

### ****JCT3V-B0090: Depth-oriented Neighboring Block Disparity Vector (DoNBDV) with virtual depth****

In this contribution, a method to enhance the NBDV by extracting a more accurate disparity vector from the depth map is proposed. A depth block from coded depth map in the same access unit is first retrieved and used as a virtual depth of the current block. While coding the texture in view 1 and view 2 with the common test condition, the depth map in view 0 is already available. So the coding of texture in view 1 and view 2 can be benefited from the depth map in view 0. The overall flow is as following:

1. Use an estimated disparity vector, which is the NBDV in current 3D-HTM, to locate the corresponding block in the coded texture view
2. Use the collocated depth in the coded view for current block (coding unit) as virtual depth.
3. Extract a disparity vector for inter-view motion prediction from the maximum value in the virtual depth retrieved in the previous step.

**🡪The benefits of using depth data to predict the disparity vectors for texture data in terms of coding efficiency are to be investigated.**

### ****JCT3V-B0121: Block-based VSP using Forward Warping****

Forward warping Block-based VSP (FBVSP) is proposed to generate VSP prediction samples for a target block in the current coding texture image, which utilizes depth map of a reference view (e.g., base view) instead of that of the current dependent view. For each macroblock using VSP in the target view (called a target block), the synthesis samples are obtained by forward warping of pixels in a proper window in the reference view. The window, i.e., a reference block, is decided by a three-step refinement method based on estimating the correspondences of the four corners in the target block. FBVSP enables block-based VSP for both texture-first and depth-first coding order of the dependent views.

**🡪The benefits of using forward warping for block-based VSP in terms of coding efficiency and especially in terms of complexity are to be investigated.**

# Core Experiment Conditions

## Software

The proposed methods for view synthesis prediction (VSP) will be implemented into 3D-ATM 6.1, 3D-HTM 5.1, or HTM-5.1-VSP. Proponents are requested to provide software that can be compiled under Windows and Linux platforms.

## Test Sequences

Test sequences will be used according to common test conditions [1].

## Coding Conditions

### CE1.a

The experiments are conducted under the common test conditions [1].

### CE1.h

The experiments are conducted under the common test conditions [1].

## Evaluation Criteria

Evaluation will be done according to the common test conditions [1].

For CE1.h, the results shall be compared to HTM-VSP in addition to the anchor.

**Coding Performance Measurements:** PSNR values shall be computed for the decoded texture views relative to original texture views and for the synthesized views relative to the synthesized views based on uncompressed texture and depth. The 4-point BD-rate according to common test conditions is used to report the overall simulation results.

* As view synthesis prediction requires depth data to encode texture views, the total bitrate (including depth rate) should be compared against all reported PSNR measures (for decoded and synthesized views).   
  A corresponding Excel sheet for reporting the appropriate simulation results will be provided by the CE coordinators.

**Complexity measurements:** For the complexity measurement, the reference software and the reference software with the proposed method implemented will be executed on the same machine with the same configuration and the computational time will be measured. A time ratio will then be calculated between the reference software and the reference software with the proposed method implemented.

Proponents are encouraged to report details about decoder complexity in terms of number of operations and memory consumption.

## Timeline

2012/11/12 Release 3D-ATM 6.1

2012/12/03 Release 3D-HTM 5.1

2012/12/17 Release HTM-5.1-VSP based on 3D-HTM 5.1

2013/01/03 Make source code and simulation results available for all CE1 participants, not only for the cross-checkers

2013/01/10 Register documents for the 3rd JCT-3V meeting

2013/01/10 Upload simulation and cross check results to the JCT-3V document repository

# Reference

1. D. Rusanovksyy, A. Vetro, and K. Müller, "Common Test Conditions for 3D Video Extensions Development," Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) of ITU-T VCEG and ISO/IEC MPEG JCT3V-B1100, Shanghai, China, 2012.