|  |  |
| --- | --- |
| **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-A0011 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **3D-CE1.a Summary Report: View Synthesis and Inter-view Prediction** | | |
| *Status:* | Input Document | | |
| *Purpose:* | Report | | |
| *Author(s) or Contact(s):* | Dong Tian Mitsubishi Electric Research Labs | Email: | [tian@merl.com](mailto:tian@merl.com) |
| *Source:* | Summary Report | | |

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

# Abstract

This document is the summary report of Core Experiment 1 in Advanced Video Coding (AVC) based 3D Video Coding (CE1.a). The Core Experiment was established at the 100th MPEG Meeting in Geneva, Switzerland to find out the best solution for view synthesis prediction and inter-view prediction [1].

The methods to be investigated needed to be implemented into the 3DV-ATM 0.4 reference software and evaluated according to the common test conditions [2].

# Participants

During the 100th MPEG Meeting the following participants registered for this Core Experiment:

* Samsung
* Zhejiang Univ
* Mitsubishi
* Qualcomm
* NTT
* Nokia
* ETRI

Besides, MediaTek took part in the CE discussion. There are totally 24 documents, including 11 proposals and 13 cross checking documents.

The following CE proposal/cross checker combinations have been assigned:

|  |  |
| --- | --- |
| Proponent | Cross Checker |
| Nokia (m26069, JCT2-A0107) | Mitsubishi (m26086, JCT2-A0019),  ETRI (m26163, JCT2-A0145) |
| Samsung (m25851, JCT2-A0015) | NTT (m26095, JCT2-A0024) |
| Samsung (m25879, JCT2-A0035) | Zhejiang Univ (m25937, JCT2-A0058) |
| Zhejiang Univ (m25932, JCT2-A0055) | Qualcomm (m26118, JCT2-A0130) |
| Qualcomm (m26058, JCT2-A0103) | Nokia (m26136, JCT2-A0138),  ETRI (m26164, JCT2-A0146) |
| NTT (m26093, JCT2-A0022) | Samsung (m26172, JCT2-A0152) |

The following CE related proposal/cross checker combinations have been suggested. Though it is optional, all the four related contributions are cross checked.

|  |  |
| --- | --- |
| Proponent | Cross Checker |
| NTT (m26094, JCT2-A0023) | Samsung (m25984, JCT2-A0061) |
| Samsung (m25882, JCT2-A0036) | NTT (m26096, JCT2-A0025) |
| Zhejiang Univ (m25933, JCT2-A0056) | Samsung (m25985, JCT2-A0062) |
| MediaTek (m25921, JCT2-A0050) | Samsung (m25987, JCT2-A0063) |
| Samsung /NTT (m26198, JCT2-A0158) | Qualcomm (mxxxxx, JCT2-Axxxx) |

# Tool Description

The tools to be investigated in this Core Experiment use the decoded texture and depth information to compute a prediction signal for a dependent view. In concept, 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.

Briefly, the CE study items established for this meeting cycle are given below [1]. The proposals are grouped according to the study items.

* Study block-based view synthesis;
* Study benefit of depth up-sampling in combination with VSP;
* Inter-view and view synthesis prediction with adaptive luminance compensation;
* Evaluate benefit of not transmitting MV info and building context;
* Evaluate benefit of sub-MB skip/direct;
* Evaluate benefit of reference picture refinement;
* Further evaluation of variable precision synthesis.

## Block-based view synthesis

|  |  |
| --- | --- |
| Proponent | Cross Checker |
| Nokia (m26069, JCT2-A0107) | Mitsubishi (m26086, JCT2-A0019),  ETRI (m26163, JCT2-A0145) |
| MediaTek (m25921, JCT2-A0050) | Samsung (m25987, JCT2-A0063) |

The current implementation of view synthesis prediction (VSP) in 3DV-ATM utilizes a forward warping to produce a full VSP reference image. An asserted drawback of such frame-level implementation is its high complexity at the decoder side.

Nokia proposes a block-based VSP implementation (B-VSP) which is based on a backward warping process. As the displacement vectors required for this process are produced from the depth map data of the current view, the texture component would reply on the depth component from the same view (in addition to other reference views, if any).

It is asserted that B-VSP reduces the complexity of view synthesis prediction considerably and avoids the need to allocate a frame buffer for the VSP reference frame. It is claimed the decoding time for B-VSP is ~20% lower on average than the decoding time of the 3DV-ATM v4 anchor. Moreover, the proposed B-VSP scheme outperforms the current 3DV-ATM design of VSP by -0.4% bitrate saving for coded texture views and by -0.7% bitrate saving for synthesized views.

The proposal JCT2-A0050 from MediaTek share a similar idea to use the depth map from the current view to pick up sample predictors from reference views, though more details are described in section 2.4. MediaTek reports some minor loss in texture coding with decoding time reduced by about 10%.

The differences between JCT2-A0107 and JCT2-A0050 might be studied during the meeting.

Both proposals (JCT2-A0107, JCT2-0050) were cross verified by cross checkers.

## Benefit of depth up-sampling in combination with VSP

|  |  |
| --- | --- |
| Proponent | Cross Checker |
| Samsung (m25879, JCT2-A0035) | Zhejiang Univ (m25937, JCT2-A0058) |
| Samsung (m25882, JCT2-A0036) | NTT (m26096, JCT2-A0025) |

Samsung made two contributions on this topic. Both proposals utilize an in-loop dilation filter to process the depth before being used for view synthesis prediction. The first one (JCT2-A0035) includes a linear interpolation for depth upsampling; while the second one (JCT2-A0036) does not involve depth upsampling.

When depth upsampling is performed (JCT2-A0035), the asymmetric VSP need be turned off, as it is to avoid depth upsampling. The proponents claim -0.23% bitrate saving with respect to decoded texture, and -0.26% bitrate saving in terms of synthesized views. The decoding time is increased by 52.6% due to the proposed upsampling, dilation filtering and also due to VSP reference generation at full resolution.

When no depth upsampling is performed (JCT2-A0036), the asymmetric VSP is turned on as defined in CTC. The proponents claim -0.18% bitrate saving with respect to decoded texture, and -0.22% bitrate saving in terms of synthesized view. The decoding time is increased by 1.34% compared to anchor.

Both contributions (JCT2-A0035, JCT2-A0036) were cross verified.

## Inter-view and view synthesis prediction with adaptive luminance compensation

|  |  |
| --- | --- |
| Proponent | Cross Checker |
| Samsung (m25851, JCT2-A0015) | NTT (m26095, JCT2-A0024) |
| NTT (m26093, JCT2-A0022) | Samsung (m26172, JCT2-A0152) |

As multiview sequences (natural or synthetic) often shows luminance variations among different views.

Samsung (JCT2-A0015) proposes an auxiliary mode, which is signaled by a 1-bit flag right after mb\_type at the macroblock level. It is proposed that this auxiliary mode is applied for view synthesis prediction only, and not for inter-view prediction based on the simulation results observed. Texture only tool. Not proposed for depth. Total bitrate saving reported is -1.03%.

NTT (JCT2-A0022) proposes an adaptive illumination compensation for the same purpose. The parameters for the illumination compensation are derived locally at the decoder side by minimizing the differences between the decoded picture and the reference picture. The average bitrate saving is -0.25% with the linear function, and -0.04% with a 2D filter with offset.

It is remarked that NTT shares the position as Samsung that illumination compensation is more efficient for VSP prediction than inter-view prediction.

Both contributions (JCT2-A0015, JCT2-0022) were cross verified.

## Benefit of not transmitting MV info and building context

|  |  |
| --- | --- |
| Proponent | Cross Checker |
| Qualcomm (m26058, JCT2-A0103) | Nokia (m26136, JCT2-A0138),  ETRI (m26164, JCT2-A0146) |

During the last meeting, Qualcomm proposed a technique that did not require (a) transmission of motion information for VSP prediction and (b) inclusion of the VSP picture in the reference picture lists, as part of sub-MB VSP signaling. Based on the discussion, the CE aimed to study the benefits of not transmitting motion information, as a separate item in the CE.

The results reported in JCT2-A0103 do provide a separate evaluation of the benefit of not transmitting the MV info. In the absence of such results, no conclusions on this potential benefit could be drawn.

The contribution (JCT2-A0103) was cross verified.

## Sub-MB VSP skip/direct

|  |  |
| --- | --- |
| Proponent | Cross Checker |
| Qualcomm (m26058, JCT2-A0103) | Nokia (m26136, JCT2-A0138),  ETRI (m26164, JCT2-A0146) |
| MediaTek (m25921, JCT2-A0050) | Samsung (m25987, JCT2-A0063) |

Sub-MB VSP skip/direct is to support VSP skip/direct at block levels smaller than 16x16 MBs.

Qualcomm proposes a way to enable sub-MB skip/direct using VSP reference pictures, known as GVSP in the proposal. 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. Qualcomm further proposes to remove VSP pictures from reference picture lists. The reported bitrate saving claimed in the document is -0.4% as per CTC, and -0.8% if QC\_VSP\_Asymmetric\_Resolution is OFF. The relation and relative benefits between the CE study items in this section and section 2.4 need to be clarified.

MediaTek proposes a way to do inter-view skip for up to 4x4 blocks. In some aspect, the idea is similar to Nokia proposal (JCT2-A0107) on block based VSP. Both proposals utilize the depth from the current view and pick up sample predictors from a reference view. With MediaTek proposal, each 4x4 block has a corresponding reference based on the depth value from the current view. No sample level vector is used as in Nokia proposal (JCT2-A0107), but a maximum depth value in the 4x4 is used for the whole 4x4 block. No syntax change is proposed. Minor bitrate increase (0.07%) in texture coding and small bitrate saving (-0.21%) in terms of synthesis quality are reported, while the reported decoding time is reduced by 10%.

The contributions (JCT2-A0103, JCT2-A0050) were cross verified.

## VSP reference picture refinement

|  |  |
| --- | --- |
| Proponent | Cross Checker |
| - | - |

There were no contributions related to this study item. As a result, the item may be dropped if there is no further interest.

## Variable precision synthesis

|  |  |
| --- | --- |
| Proponent | Cross Checker |
| Zhejiang Univ (m25932, JCT2-A0055) | Qualcomm (m26118, JCT2-A0130) |
| Zhejiang Univ (m25933, JCT2-A0056) | Samsung (m25985, JCT2-A0062) |
| NTT (m26094, JCT2-A0023) | Samsung (m25984, JCT2-A0061) |

In JCT2-A0055, Zhejiang Univ presents a frame level option for flexible rendering precision in VSP generation. By proposing three look-up tables, integer, half-pel and quarter-pel rendering could be realized. The selection of different precision is made based on the frame level RD optimization at the encoder. On average, the proposed method can keep the similar RD performance while reducing decoding time by 5%.

In JCT2-A0056, Zhejiang propose a method to derive the sub-pel disparities by using special ways depending on whether a sample falls along depth edges and within a non-edge area. It is reported that the method would bring -0.38% bitrate saving with respect to texture coding only while maintaining similar decoding time as anchor.

From last MPEG meeting, there were two adoptions related to VSP picture synthesis. One is sub-pel warping without fully interpolating the reference view (NTT). Second is to do warping just based on the depth map at half resolution (Qualcomm). Some issues would arise when combining the two adoptions. In JCT2-A0023, NTT proposes to harmonize the two adoptions by flipping the processing direction of the sample warping without requiring a z-buffer. In addition, the hole-filling and inpainting is also simplified by filling the hole samples between the current warped sample and the last warped sample using the current warped sample. An average bitrate saving of -0.53% is reported in terms of texture coding only.

The contributions (JCT2-A0055, JCT2-A0056, and JCT2-A0023) were cross verified.

## Misc

|  |  |
| --- | --- |
| Proponent | Cross Checker |
| Samsung /NTT (m26198, JCT2-A0158) | Qualcomm (mxxxxx, JCT2-Axxxx) |

Samsung and NTT made a joint contribution by combining their individual proposals JCT2-A0036 and JCT2-A0023. From the results, it shows that the two techniques can be added to each other. With respect to texture coding, the bitrate saving claimed is -0.18% in JCT2-A0036 and -0.53% in JCT2-A0023, which becomes -0.62% in the joint proposal.

The contributions (JCT2-A0158) was cross verified.

# References

1. Anthony Vetro, Karsten Müller, "Description of Core Experiments in 3D Video Coding ," MPEG output document N12746, May 2012, Geneva, CH
2. ISO/IEC JTC1/SC29/WG11, "Common test conditions for 3DV experimentation", ISO/IEC JTC1/SC29/WG11 N12745, May 2012, Geneva, Switzerland.