

Title: **3D-CE1.h Summary Report: View Synthesis and Inter-view Prediction**
Status: Input Document
Purpose: Report
Author(s) or Contact(s): Fabian Jäger¹, Krzysztof Wegner²
¹RWTH Aachen University Email: jaeger@ient.rwth-aachen.de
²Poznan University of Technology kwegner@multimedia.edu.pl
Source: Summary Report

Abstract

This document is the summary report of Core Experiment 1 in High Efficiency Video Coding (HEVC) based 3D Video Coding. The Core Experiment was established at the 100th MPEG Meeting in Geneva, Switzerland to investigate the impact of view synthesis based and inter-view prediction on texture and depth coding for the dependent views.

The methods to be investigated needed to be implemented into the 3DV-HTM 3.1 reference software [1] and evaluated according to the common test conditions. [2]

1 Participants

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

- RWTH Aachen University
- Poznan University of Technology
- Samsung
- LG Electronics
- Mitsubishi Electric
- NTT

After the meeting the coordination of the experiments to be conducted was done via email.

RWTH Aachen University, LG Electronics and Samsung decided to not have a proposal tested in this Core Experiment during this meeting cycle. Poznan University of Technology also wanted to have a proposal during this meeting cycle, but their software was not ready for being cross-checked. Samsung offered to be a cross checker for the CE. The following proposal/cross checker combination has been assigned:

Proponent	Cross Checker
NTT/Mitsubishi Electric	Samsung

2 Tool Description

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.

As the view synthesis, which is required to form the described prediction signal, is part of the coding loop, this view synthesis algorithm is a normative tool and may improve the coding efficiency of a 3D video codec.

2.1 Description of Proposed Tool by NTT/Mitsubishi Electric

View synthesis prediction (VSP) is a technique to remove inter-view redundancies when coding multiview video information, which uses a synthetic picture as a reference picture to predict the current picture. This contribution first proposes inserting synthetic pictures to reference picture lists, which shows an average bitrate saving of -2.6% and -0.4% for the dependent views, with maximum gains of -9.9% and -5.8% for these views. This contribution further proposes a VSP skip mode, which enables the skip mode relative to the synthetic reference picture. However, this mode has not been fully implemented and requires further study. The benefit of VSP for depth coding has also been investigated, but the result is not yet conclusive. Lastly, this contribution proposes to simplify the rendering process in HTM, which would reduce the decoding time by 9% when VSP is enabled.

The proponent recommends to adopt the basic VSP scheme by adding the synthetic pictures to the reference picture lists and to adopt the simplified view synthesis procedure into HTM. It is also suggested to continue study VSP skip mode and simplify synthetic picture generation in the context of this CE.

3 Simulation Results

The core experiment investigates the effects of view synthesis based prediction in terms of bitrate reduction for the dependent views. The proposed methods for view synthesis based prediction are implemented into the 3DV-HTM 3.1 software and evaluated according to the common test conditions. [2]

3.1 Results of VSP Tool by NTT/Mitsubishi Electric

The joint contribution from NTT and Mitsubishi Electric includes seven different configurations, which use different variants of view synthesis prediction: Basic VSP scheme by adding VSP reference picture to the reference picture lists; VSP skip mode; VSP for depth component; and simplifications of the rendering process. From the results of the current simulation, adding the synthesized pictures for texture (i.e. test case 2 in the contribution) yields the best performance with an average bitrate saving of -2.6% and -0.4% for the dependent views. Maximum gains can be observed for the sequence Undo_Dancer with -9.9% and -5.8%, as shown in the following table.

More study is needed for VSP skip mode and VSP for depth components, as the implementations are still uncompleted. In addition, the simplification of the rendering process in HTM leads to about 9% less decoding time.

3.1.1 Basic VSP for texture

The following table shows the results for using basic view synthesis prediction by putting a synthesized view into the reference picture list of the texture view to be coded.

	video 0	video 1	video 2	video only	synthesized only	coded & synthesized	enc time	dec time
Balloons	0,0%	1,1%	2,9%	1,0%	1,0%	1,1%	118,2%	177,3%
Kendo	0,0%	2,1%	2,8%	1,2%	1,3%	1,4%	108,6%	190,8%
Newspaper_CC	0,0%	0,3%	0,5%	0,2%	0,1%	0,2%	109,2%	188,1%
GT_Fly	0,0%	-5,7%	-2,7%	-0,8%	0,0%	-0,1%	106,4%	178,7%
Poznan_Hall2	0,0%	-2,1%	2,3%	0,3%	0,6%	0,6%	108,0%	187,8%
Poznan_Street	0,0%	-4,2%	-2,7%	-1,0%	-0,1%	-0,2%	105,6%	186,2%
Undo_Dancer	0,0%	-9,9%	-5,8%	-2,0%	-1,1%	-1,2%	107,3%	187,1%
1024x768	0,0%	1,2%	2,1%	0,8%	0,8%	0,9%	111,9%	185,3%
1920x1088	0,0%	-5,5%	-2,2%	-0,9%	-0,2%	-0,2%	106,8%	184,9%
average	0,0%	-2,6%	-0,4%	-0,2%	0,3%	0,3%	109,0%	185,1%

3.1.2 Simplified rendering process

The decoding time is reduced by the proposed simplification of the rendering process. Test cases 4, 5 and 6 correspond to test cases 1, 2 and 3, respectively, but with a simplified rendering. From the table below, it could be found that the decoding time is reduced by about 9%.

	Case 0	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Balloons	263.2%	178.4%	177.3%	270.0%	180.1%	172.4%	207.1%
Kendo	255.2%	200.6%	190.8%	275.7%	180.4%	174.6%	206.9%
Newspaper	283.0%	191.5%	188.1%	276.4%	190.4%	193.9%	217.6%
GT_Fly	238.4%	179.7%	178.7%	263.0%	172.4%	173.9%	200.7%
Poznan_Hall2	263.8%	192.8%	187.8%	266.7%	174.4%	171.9%	201.7%
Poznan_Street	282.3%	192.2%	186.2%	275.2%	179.4%	176.5%	211.8%
Undo_Dancer	268.8%	191.2%	187.1%	278.2%	186.9%	179.3%	209.5%
1024x768	266.9%	189.9%	185.3%	274.0%	183.6%	180.1%	210.5%
1920x1088	262.8%	188.9%	184.9%	270.7%	178.2%	175.4%	205.8%
average	264.6%	189.4%	185.1%	272.1%	180.5%	177.4%	207.8%

3.2 Cross Check of NTT/MERL by Samsung

The results of Samsung's cross check shows a perfect match of the results presented above and with the additional configurations of the contribution document. At the time of submission of this summary report the crosscheck was only finished for a subset of all test sequences, as the other simulations were still ongoing.

4 References

- [1] Krzysztof Wegner Heiko Schwarz, "Test Model under Consideration for HEVC based 3D video coding v3.0," ISO/IEC JTC1/SC29/WG11 MPEG, N12744 2012.
- [2] Dmytro Rusanovsky Heiko Schwarz, "Common Test Conditions for 3DV experimentation," ISO/IEC JTC1/SC29/WG11 MPEG, N12745 2012.