INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2010/M17900
July 2010, Geneva, Switzerland

Source **Poznan University of Technology,**
Chair of Multimedia Telecommunications and Microelectronics, Poznań, Poland
Status **Proposal**
Title **Input on "Call for Proposals on 3D Video Coding" document**
Author **Krzysztof Wegner** (kwegner@multimedia.edu.pl)
Olgierd Stankiewicz (ostank@multimedia.edu.pl)
Marek Domański (domanski@multimedia.edu.pl)

1. Introduction

In this document, we would like to draw attention of the group to some issues that should be addressed in Call for Proposals (CfP) on 3D Video Coding. Or remarks focus mainly on the backwards compatibility towards MVC, and the evaluation of the CfP responses.

2. Data format

Obviously the CfP should be clear with respect to what input data may be used and what is expected as output.

We propose that the input to the coder should be **texture and depth** video data. Similarly as in classical case of video (texture) coding where a coder is not aware of the source of the content (natural, synthetic), also depth maps should be considered 'as is'. In particular, in real scenarios depth maps may be produced by any means outside the coder e.g by ToF camera.

3. Test Material

Test material should comprise of **texture data and the best depth data** that are currently available. However it is desirable that the future 3DV standard supports many different scene representations (like, e.g. LDV), and allow efficient transmission of various types of input data. Moreover, test material should include all required metadata, like:

- camera parameters (intrinsic and extrinsic),
- information about which view is the 'mono view', (see section 3.2.4 in requirement document [1])
- which views are 'stereo pair'.
- rendering range

We propose the following text:

4.1 Test Material

Test material is a set of texture video sequences with corresponding per pixel depth maps. All data sets correspond to a linear camera arrangement, and moreover source video data

are rectified to avoid misalignment of camera geometry. Camera parameters for each set are given.

All video texture test materials are progressively scanned with 4:2:0 color sampling with 8 bits per sample. All video texture are properly color corrected

All depth maps are progressively scanned and use 4:0:0 color sampling with 8 bits per sample. Each sample represents a normalized depth value for the corresponding pixel in texture video. Each depth map is provided with znear and zfar values that define the range of stereoscopic depth values.

The data sets from Annex I will be used for 3D Video Evaluation

4. Test Scenarios

Test Scenarios should reflect future application in particular those described in requirements [1]. Therefore we propose two scenarios:

- ***Advanced stereoscopic display with varying baseline distance***.
  Stereoscopic pair as an input with respective depth maps, on the output: one view from original pair and one virtual.
- ***Auto-stereoscopic N-view display***.
  Two or three views with respective depth maps as an input, N virtual views as an output.

5. Backwards compatibility

As it was shown in our analysis of MPEG-4 MVC vs expected HEVC performance [2], MPEG-4 MVC cannot be considered as state-of-the-art coding technology. Therefore backwards compatibility with MPEG-4 MVC should be removed.

Transparency to single view coding technology (e.g. HEVC, MPEG-4 AVC), should be considered as advantage of proposed technology..

6. Coding Conditions

6.1 Input data

Video texture data and respective depth data should be input to the encoder. The number of views used for transmitting shall depend on the data format used. Thus, the exact number of views coded directly in the stream should not be specified in CfP.

6.2 Pre-processing

Preprocessing on texture data should be forbidden.

Nevertheless, depth data can be a subject to change of representation that should not be considered as preprocessing.

6.3 Post-processing

All processes leading to creation of a virtual view should not be considered as post-processing.

6.4 Tuning of entropy coding tables

Due to limited number of multiview test sequences available, all coding tables must be determined from the same sequences as used for evaluation. Therefore, it should be more intensively stressed that this applies to ‘large’ coding tables.

7. Rendering Conditions

7.1 Camera Parameters

We propose to use camera parameters in the form used for exploration experiments. It should be possible to give renderer a list of virtual view camera parameters and it should be able to synthesize as many virtual views as requested.

7.2 View range

CfP should not specify any particular depth range of specific positions of virtual views to be synthesized. Instead of that, a range of virtual view positions should be defined. This should meet the following constraints:

- virtual views lay along rectification line,
- no 'out of plane' synthesis is allowed.

8. Submission Requirements

Each proponent should provide bit streams, binary decoder and view renderer. Renderer and decoder can be merged in to one binary file, or they can be two separate binaries. In the latter case, interexchange data format between decoder and the renderer should allow easy cooperation. Renderer should be able to provide requested number of views, based on input decoded data format, and virtual views camera parameters.

9. Testing Procedure

Because of the uncertainty about how to compare different data format (LDV, MultiView+MultiDepth), **only the synthesized virtual views quality should be assessed**.

Moreover exact positions of virtual views to be evaluated should be unknown. We propose that only the range of virtual views positions should be defined by CfP. The testing institute/company should pick random view positions from that range.

On the other hand such a procedure can lead to overburdening the testing institute/company. Another solution could be to specify large number (100) of virtual camera positions from which a testing institute selects only a limited number (3-5) for the tests.

In evaluation of virtual view no specific position should be favored, because in real scenario, the user can potentially view any virtual view.

10. References

[1] Video and Requirements Group, "Applications and Requirements on 3D Video Coding," ISO/IEC JTC1/SC29/WG11 N11061, Xian, China, October 2009.

[2] K. Wegner, O. Stankiewicz, K. Klimaszewski, M. Domański, "Comparison of multiview compression performance using MPEG-4 MVC and prospective HVC technology" ISO/IEC JTC1/SC29/WG11 M17913, Geneva, Switzerland, July 2010.