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

ISO/IEC JTC1/SC29/WG11
MPEG2014/N14512
April 2014, Valencia, Spain

Source: **Requirements**
Status: **Approved**
Title: **Description of Exploration Experiments on Free-viewpoint Television (FTV)**
Editors: **Krzysztof Wegner, Takanori Senoh, Gauthier Lafruit**

1 Introduction

The aim of this exploration experiment is to evaluate currently available depth estimation, view synthesis and coding technology in the context of super-multiview and free viewpoint applications as first focus of the FTV activity. This document describes the conditions of the experiment and the evaluation procedures under consideration. Experiments are performed to get more insight into quality of depth estimation algorithms, view synthesis algorithms, and to get judgments about the quality of available test materials.

2 Description of Exploration Experiments

In this experiment, depth estimation, view synthesis and current MPEG multiview coding technology will be evaluated. Several contributions were received about these essential building blocks of FTV systems. The goal is to start experimental evaluation of currently available building blocks and proposed algorithms in a structured, more formal, defined and comparable way.

2.1 Reference Software

Poznan University of Technology kindly agreed to provide enhanced software for depth estimation (DERS 6.1) as well as for view synthesis (VSRS 6.0) to be used as reference in this EE. It is available a MPEG SVN (http://wg11.sc29.org/svn/repos/Explorations/FTV).

For coding efficiency evaluation lasts version of HTM v.10.0 software shall be used.
This is available at https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-10.0.

2.2 Test data

Test data will be selected from the test sequences summarized in Table 1, which contain super-multiview video and if available corresponding depth data.
There are currently four types of the materials available (11 test materials in total, highlighted green in Table 1):

- Linear camera arrangement still images– 2 test material
- Linear camera arrangement sequences – 6 test materials
- Arc camera arrangement sequences – 2 test materials
- 2D parallel camera arrangement sequence – 1 test material

This data can be used for depth estimation experiments and comparison with reference depth data when available. These sequences can also be used directly for encoding/decoding and view synthesis experiments using the available depth data.

Additional test materials will be added as they become available.

Table 1. Summary of the test sequence to be used in the exploration experiments.

No.	Provider's Name	Seq. Name	Number of Views	Resolution (pel)	Frame rate (fps)	Length	Cam Arrangement (1D parall, 1D arc, 2D parall, 2D arc, Sphere, Arbitrary)	Ground Truth Depth Available	Contents	condition to use (if any)	URL of sequence (ID, PW)
1	NICT	Little world (Bee)	185	1920x1088	Still Image	1 frame	1D parallel	Yes	CG: Sun flowers and bee	See M32201	http://fujii.nuee.nagoya-u.ac.jp/NICT/NICT.htm
2	UHasselt	San Miguel	200	1920x1080	Still image	1 frame	1D parallel	Yes	CG, Raytraced	See M33162	https://wg11.sc29.org/content/MPEG-04/Part02-Visual/FTV_AhG/UHasselt_San_Miguel
3	Nagoya University	Champagne Tower	80	1280 x 960	29.4114	10 sec 300 frames	1D parallel	No	Glass tower	See M15378 and provider webpage	http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data no password required
4	Nagoya University	Pantomime	80	1280 x 960	29.4114	10 sec 300 frames	1D parallel	No	Two clowns	See M15378 and provider webpage	
5	Nagoya University	Dog	80	1280 x 960	29.4114	10 sec 300 frames	1D parallel	No	Dog and person	See M15378 and provider webpage	
6	UHasselt	Soccer-linear 1	8	1600x1200	25	22 sec 550 frames	1D parallel	No	Soccer game with cameras ca 1m apart	Provided to MPEG community acknowledgement to Hasselt University.	sftp://mpeg.edm.uhasselt.be port 3333 username: restro password: olezas95
7	UHasselt	Soccer-linear 2	8	1600x1200	25	22 sec 550 frames	1D parallel	No	Soccer game with cameras ca 1m apart		

8	NICT	Shark	185	1920x1088	30	2 sec 60 frames	1D parallel	Yes	CG: See and shark	See M32201	http://fujii.nuee.nagoya-u.ac.jp/NICT/NICT.htm
9	Poznan University of Technology	Poznan Blocks	10	1920x1080	25	40 sec 1000 frames	100deg. on arc around the scene	No	gaming scene	See M32243	ftp://multimedia.edu.pl/ftv Password provided upon request please email kwegner@multimedia.edu.pl
10	UHasselt	Soccer-corner	7	1920x1080	25	22 sec 550 frames	120 deg. corner, 1D arc	No	Soccer game with cameras ca 10m apart	Provided to MPEG community acknowledgement to Hasselt University.	https://wg11.sc29.org/content/MPEG-04/Part02-Visual/FTV_AhG/UHasselt_Soccer
11	Nagoya University	Akko & Kayo	15	640x480	29.4114	10 sec 300 frames	2D parallel	No	Two persons	See M12338 and provider webpage	http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data no password required
12	NHK	CGIP_nhk	Approx. 300	Approx. 200x120 (3840x2160)	TBD		2D lens-array		TBD	TBD	TBD

2.3 Experiments

The main goal of this first round of EE until the 109th meeting is to establish the experimental environment and probably to get first simulations done. Reference software shall be distributed, participants shall get it running, get familiar with it, evaluate it, and should produce results in the form of estimated depth maps, synthesized intermediate views, and produced bitstreams.

2.3.1 EE1 Depth estimation

Several depth estimation algorithms have been presented. The goal of this first round of EE is to make a first step in the direction of a comparative study, and an evaluation of depth quality attainable for general camera arrangements. The final goal is to improve DERS 6.1 for FTV applications, i.e.:

a) Support the use of more than three input views in estimating high-quality depth maps which are spatio-temporally coherent over adjacent views for providing good view synthesis results in EE2
b) Improve depth estimation performance for spare viewpoints for providing high-quality view synthesis in EE2 with as little input views as possible.
c) Evaluate execution time of depth estimation (capability of real-time operation)

For the super-multiview data set with many input views, proponents should take sparser subsets of the available views and estimate the depth map. It is recommended to take a dyadic reduction of input views for the successive subsets under test, e.g. if 32 input views are originally provided as input data set, the depth estimation should be performed successively on 32, 16, 8 and 4 regularly spaced input views. The depth maps for these different input settings are compared to establish at which reduced number of input views the depth estimation quality is judged as unsatisfactory, giving insights about the baseline specs for the proposed depth estimation.

For the super-multiview data set with many input views, proponents should estimate depth maps for given view several times, each time using different reference views (i.e. different views fed into the software as a left and right views while the center view remains unchanged). The aim of this experiment is to examine the influence of reference view distance used for depth estimation on the quality of the estimated depth. Proponents should provide PSNR values of the estimated data against ground truth depth data when available. When ground truth data is not available, the proponents should provide PSNR values expressing the quality of the view synthesized based on the estimated depth data with respect to the view captured by the real camera at the same spatial position as the synthesized view.

Proponents shall provide the estimated depth maps as results, yielding evidence for (a) and (b). The most promising visual results should be profiled against their execution time (c), clearly mentioning on which platform the profiling has been done for comparison purposes.

2.3.2 EE2 View synthesis

A few view synthesis algorithms have been presented so far as well. Availability of high-quality view synthesis is crucial not only for Free Viewpoint applications, but also in general for further experimentation within the FTV activity.

In this part of the EE available view synthesis algorithms shall be evaluated. The final goal is to improve VSRS 6.0 for FTV applications, i.e.:

a) Make the View Synthesis robust against error prone depth maps (i.e. estimated depth maps that are different from ground truth)
b) Improve the quality of View Synthesis with depth maps for intermediate views

c) Improve the quality of View Synthesis with depth maps outside the camera plane/cylinder/freeform for true Free Navigation (Extrapolation)
d) Perform view synthesis without depth maps
e) Evaluate execution time of view synthesis with and without depth maps (capability of real-time operation)

This EE will be done using the XX sequence. Other data shall be added as soon as depth data becomes available (cfr. EE1).

For (a), proponents shall show View Synthesis results with the depth maps obtain with use of DERS 6.1, and compare them with the improved depth maps from EE1.

For (b), the results to be provided in this round of EE shall be the synthesized intermediate views. Regular subsets of the available views should be used as input and number of other views at spatial position of real views should be synthesized, providing a PSNR quality figure against real views available, e.g. every odd view is synthesized from its surrounding even views, using ground truth depth maps (if available) and DERS 6.1 depth maps (providing EE2(a) results) and improved depth maps from EE1(b) (providing EE2(b) results). This principle should be repeated over all subsets of sparse input views of EE1(b), e.g. if in the original input views (0, 1, 2, …, 7, 8, 9, …, 14) the subset (0, 7, 14) is taken as the sparse input views, then views (1, 2, …, 6) and (8, 9, …, 13) should be synthesized and the PSNR calculated against their original, uncompressed ground truth.

For (c), the procedure of (b) is repeated, but for synthesized views slightly outside the camera plane/cylinder/freeform, with a PSNR calculation based on a scaled ground-truth (equivalent to a zoom-in/zoom-out operation) using bi-cubic resampling of the input images (with the resampling factor providing the highest PSNR for that particular view).

For (d), other methods of view synthesis not using any depth maps, are solicited.

Subjective viewing of the results will be carried out at 109th meeting.
Additionally, it is suggested to synthesize a dense set of intermediate views for subjective quality evaluation on a 3D display made available at the 109th meeting.

The most promising visual results should be profiled against their execution time (e), clearly mentioning on which platform the profiling has been done for comparison purposes.

2.3.3 EE3 Compression

The second phase of FTV MPEG developed a number of multiview coding technologies, i.e MV-HEVC and 3D-HEVC.

The goal is to investigate the coding efficiency of the best of currently available coding technologies in the context of FTV third phase video.

a) Proponents shall compress the input views of EE1 and EE2 with MV-HEVC and 3D-HEVC and report the PSNR of all views (input and synthesized views) vs. the total bitrate over all input views. The coding condition should follow the JCT-3V Common Test Condition (JCT3V-H1100), except for the number of views. The number of views is TBD after the selection of test materials.
b) Proponents shall investigate the number and density of the input views to be codded that provides highest and equally distributed quality of the render views (under EE2) at given bitrate. Proponent shall include total bitrate, quality of coded views, average and the worst quality of synthesized views under EE2 conditions. Results should be compared with the results of view synthesized from uncompressed data at the same spatial positions.

2.3.4 EE4 Registration

In FTV third phase it is expected that many input views will be used and one cannot guarantee that all input views can be correctly registered within subpixel accuracy.

The goal of this first round of EE is to investigate the influence of misregistration (a couple of pixels between the left-most and right-most input camera) on the View Synthesis performance.

EE1 and EE2 will be conducted with and without slight registration errors in the input image set.

For data sequences with a guaranteed very accurate registration, small registration errors of a couple of pixels will be artificially introduced (linearly increasing between the left-most and right-most input camera) and the PSNR evaluated as in EE2 and EE3.

For data sequences with unknown registration error, experiments EE2 and EE3 will have to be flagged as “intrinsic registration error” for proper interpretation of the obtained results between different data sequences.

2.3.5 EE5 Free Navigation

This category includes EE on performance evaluation of free navigation by an image-based approach and possible improvement by using model information, and EE on the possibility of free navigation by new approaches.

A relation between an existing MPEG standard and free navigation application has been mentioned. The goal is to investigate the representation capability of the existing MPEG standards like a binary format for scenes (BIFS) in MPEG-4 Part 11 in the context of FTV third phase application for free navigation. The final goal is to clarify the required elements to be standardized in the FTV third phase for free navigation from natural video sequences.

2.4 Evaluation

Evaluation in this first round of EE shall be mainly based on subjective assessment of generated depth and synthesized views at the 109th meeting. Quality and artifacts shall be judged by experts viewing the results.

It is not expected that very strong conclusions can be drawn at that stage, however, first initial insights can be obtained already in several ways

- Monoscopic viewing of a synthesized view.
- PSNR comparisons shall be made for view synthesis experiments with real view at the same spatial position.
- Results of depth estimation for other data shall be inspected visually and compared.
- Results of depth estimation shall be evaluated by view synthesis and visual inspection.

It is expected that by the next meeting the group will have established a running experimental framework for further FTV testing, including depth estimation, view synthesis and multiview compression. Then this framework can be improved and refined meeting by meeting.

Discussion of the results at the 109th meeting will provide insights about how to improve, extend, and refine experiments and algorithms during the next period.

3 Participants

NISRI.	Masayuki Tanimoto, tanimototentative3@yahoo.co.jp

NTT	Shinya Shimizu, shimizu.shinya@lab.ntt.co.jp
Poznan University	Krzysztof Wegner, kwegner@multimedia.edu.pl,
of Technology	Olgierd Stankiewicz, ostank@multimedia.edu.pl
Uhasselt	Gauthier Lafruit, **,** gauthier.lafruit@uhasselt.be
NICT	Takanori Senoh, senoh@nict.go.jp
NHK	Jun Arai, arai.j-gy@nhk.or.jp
LZ Associates	Lazar Bivoarsky, lazarmb@yahoo.com
KDDI	Kei Kawamura, ki-kawamura@kddi.com
Zhejiang University	Yichen Zhang, felixzyc@gmail.com;
	Qing Wang wqingxmu@gmail.com;
	Lu Yu, yul@zju.edu.cn

Experiment	Participants	
EE1 - Poznan Blocks EE1 - Soccer	Poznan University of Technology	Krzysztof Wegner, kwegner@multimedia.edu.pl, Olgierd Stankiewicz, ostank@multimedia.edu.pl
EE1 and EE2 Soccer EE1 and EE2 San Miguel	Uhasselt	Gauthier Lafruit, gauthier.lafruit@uhasselt.be
EE1 and EE2 ?????	NICT	Takanori Senoh, senoh@nict.go.jp
EE1 and EE2 Soccer	LZ Associates	Lazar Bivoarsky, lazarmb@yahoo.com

Coordination:
Krzysztof Wegner, kwegner@multimedia.edu.pl
Taka Senoh, senoh@nict.go.jp