MPEG Activities for 3D Video Coding

Yo-Sung Ho and Yunseok Song
Gwangju Institute of Science and Technology (GIST), Gwangju, Rep. of Korea
E-mail: {hoyo, ysong}@gist.ac.kr

Abstract— **In this paper, we introduce the 3D video coding research activities of the moving picture experts group (MPEG), the leading international standardization group for multimedia. Since 2001, MPEG has worked on a number of 3D video coding projects, notably, 3D audio-visual (3DAV), free-viewpoint television (FTV), multi-view video coding (MVC) and 3D video coding (3DVC). Such research works have significantly contributed to the 3D video processing industry. Since multi-view video and depth maps are critical parts of 3D video, the handling techniques of such components have evolved throughout the years. MPEG will continuously develop efficient standards for emerging 3D video applications.**

I. INTRODUCTION

3D video provides natural depth perception that cannot be achieved by 2D video. 3D video applications include films, games, advertisement, and many others. Generally, depth information is vital for 3D video. This is represented by a depth map, a gray-scale image that contains camera-to-object distance information. Multi-view video is a core part as well. One of the most used 3D video format is multi-view video plus depth (MVD) [1]. Using this format, a number of views of texture and depth data can be coded and transmitted. Fig. 1 represents three-view MVD data of a 3D video test sequence.

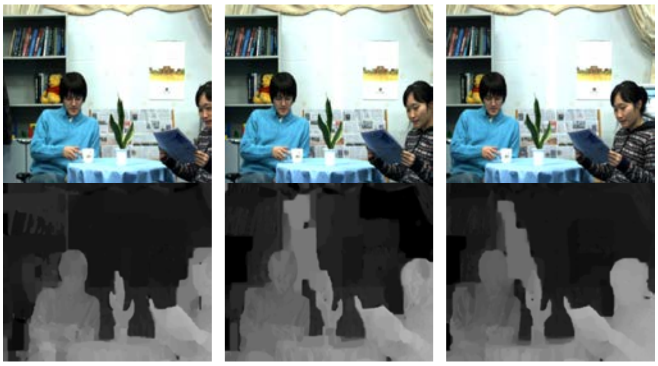

Fig. 1 Three-view MVD data

Moving picture experts group (MPEG) is a working group under ISO/IEC. The 3D video group of MPEG has led the research of 3D video, namely, video compression, depth estimation and view synthesis. In 2001, MPEG started 3D audio-visual (3DAV) project to investigate compression methods of various 3D video information. A more focused study on multi-view video coding (MVC) followed [2]. Recently, the development of 3D video coding based on advanced video coding (3D-AVC) has been concluded. In addition, 3D video coding based on high efficiency video coding (3D-HEVC) is expected to be completed in 2015. Currently, free-viewpoint (FTV) ad hoc group (AHG) of MPEG is exploring depth estimation and view synthesis methods specifically targeted for., super multi-view and free-navigation applications.

II. 3D AUDIO-VISUAL (3DAV)

From 2002 to 2004, four exploration experiments (EE) were proceeded in 3DAV. The main goal is to investigate compression methods beyond 3D mesh coding.

A. EE1: Omni-directional video

Omni-directional video is captured by a single camera. Many internet portal websites provide road view service using this type of video. Fig. 2 represents an omni-directional image captured from above. Picture-to-mesh mapping followed by 3D conversion technique was considered.

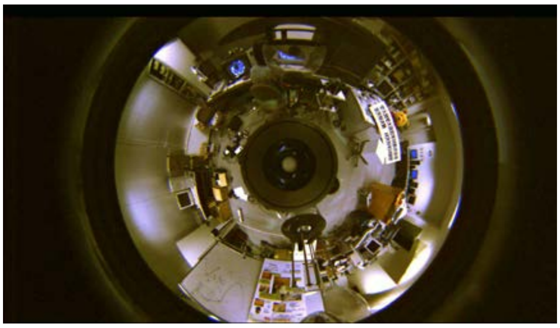

Fig. 2 Omni-directional image

B. EE2: Free-viewpoint TV

Free-viewpoint video allows the user to decide the view. Addition of ray space into MPEG-4 system was proposed. However, this had compatibility issues and was not efficient for the decoder.

C. EE3: Stereoscopic video

EE3 studied disparity estimation between two adjacent images; then, the disparity map and the coded information of one view are transmitted. In many aspects, this method was more effective than coding two views.

D. EE4: Depth/disparity coding

This experiment compared coding performances of sets with different filters for depth and disparity data. In general, with deblocking filter or median filter, bitrates decreased and the quality of the synthesized image improved. Based on such characteristics, various depth/disparity coding techniques were discussed.

978-616-361-823-8 © 2014 APSIPA

III. MULTI-VIEW VIDEO CODING (MVC)

MVC enables compression of multi-view video. In addition to spatial and temporal redundancies exploited in conventional video coding, inter-view redundancy is considered as well. Illumination compensation and motion skip mode techniques are included as well. Fig. 3 represents the framework of MVC.

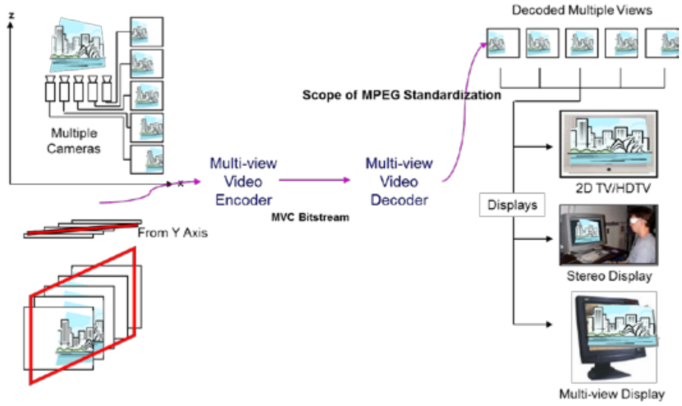

Fig. 3 Framework of MVC

A. *Prediction structure*

MVC uses inter-view prediction where pictures between views at the same time instance are estimated. For view and time axes, IBPBP … and temporal B prediction structure are employed, respectively. Fig. 4 represents inter-view/temporal prediction of MVC. *Sn* and *Tn* denote *n*-th view and *n*-th frame, respectively. *S0* and *S2* are coded first, then, *S1* is coded using the decoded pictures of *S0* and *S2*. Since time correlation is different for all test sequences, the group of picture (GOP) needs to be determined accordingly.

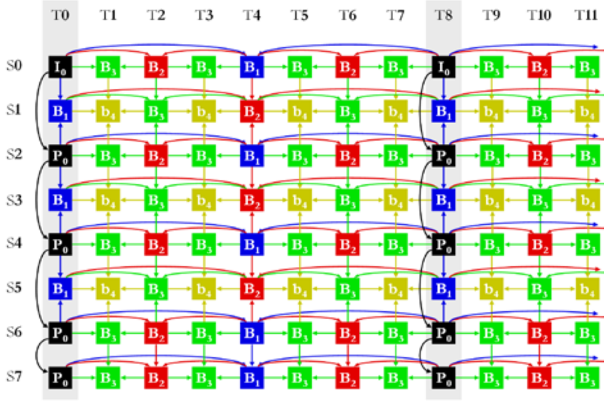

Fig. 4 Inter-view/temporal prediction structure

B. *Illumination compensation*

Multi-view video data are captured by a number of cameras. Even with the same configurations, the captured images can be slightly different due to hardware difference. Fig.5 shows images captured by different cameras. The illumination is noticeably dissimilar. Illumination compensation is used for minimizing the pixel difference in the same object of each view.

Fig. 5 Illumination variance in multi-view video

C. *Motion skip mode*

In multi-view video, since cameras are capturing the same scene, motion information is similar. Motion skip mode allows sharing of motion information. Fig. 6 represents the motion correlation between views. In the neighboring view, acroblock $P_{0'}$ at time *T0* refers to $P_{1'}$ of time *T1*, motion vector is mv_{nbr}. In the current view, if P_0 is coded referring to P_1, the motion vector mv_{cur} corresponds to mv_{nbr}. In addition, disparity vectors *dv* and *dv'* are similar.

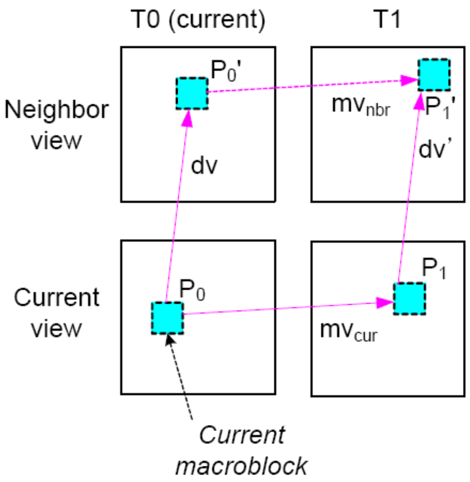

Fig. 6 Motion correlation between current and neighbor views

After acquiring the camera parameters, we choose the fourth camera as the reference camera since this is closest to the middle. Other color cameras use the intrinsic parameters and rotation matrices of the fourth camera. Camera centers are adjusted as well. For the non-reference cameras, since the extrinsic matrices have changed, pixels of certain areas, i.e., top/bottom, left/right sides may not exist. We add the same focal offset to each camera in order to make the frames contain only captured data. Due to the offset, rectified images become slightly

IV. 3D-AVC

In 2011, MPEG issued a call for proposals (CfP) on 3D video coding technology [3, 4]. Test categories were AVC-compatible and HEVC-compatible. From numerous proposals, the best performing test model for each category was selected and tools were enhanced throughout MPEG meetings. 3D-AVC and MVC are both compatible with AVC. While MVC codes texture and depth views independently, AVC exploits redundancy between texture and depth videos [5]. Thus, texture and depth coding are dependent.

A. Reduced-resolution depth coding

Depth coding is less important than texture coding. Reduced-resolution depth coding may decrease the quality of coded depth maps but can save many bits. Depth video is down-sampled prior to coding; both height and width are reduced by half.

B. Non-linear depth representation

Non-linear depth representation (NDR) maps depth values to represents closer objects more accurately than distance objects. After decoding, inverse mapping is performed. Fig. 7 represents the piecewise linear segments for mapping in NDR.

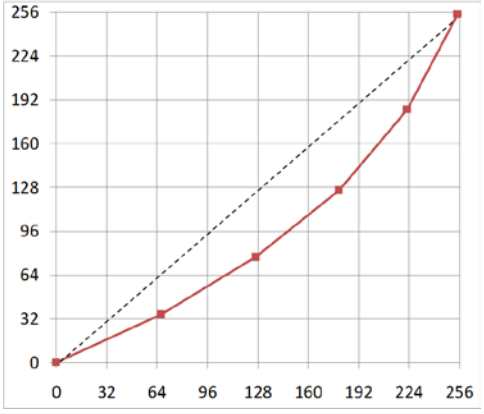

Fig. 7 Piecewise linear segments for mapping in NDR

C. View synthesis prediction (VSP)

View synthesis prediction (VSP) is supported for enhanced texture coding. VSP generates an additional prediction signal. Rectified data are assumed, i.e., vertical displacement is zero does not exist. For the current block *Cb*, the synthesized reference block *R(Cb)* is retrieved from base view *T0* with disparity vector *D*. This backward warping based VSP is shown in Fig. 8.

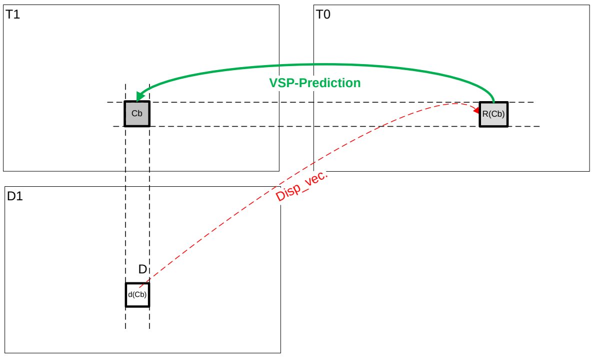

Fig. 8 Backward warping based VSP

V. 3D-HEVC

HEVC is the newest video coding standard which outpaces AVC. 3D-HEVC also performs better than 3D-AVC. The coding structure of 3D-HEVC is shown in Fig. 3. For each view depth coding follows texture coding [6].

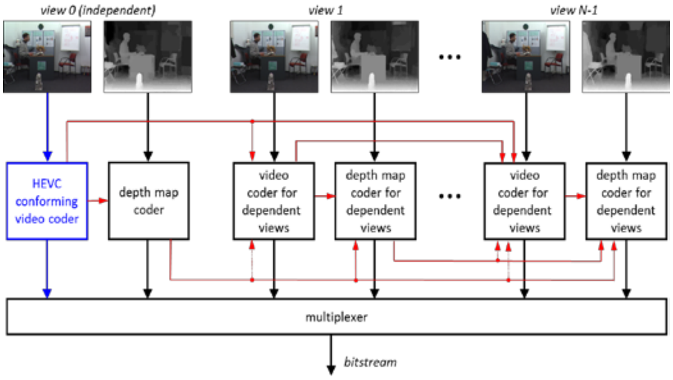

Fig. 9 Coding structure of 3D-HEVC

A. Disparity-compensated prediction

Disparity-compensated prediction (DCP) is similar to motion-compensated prediction (MCP). MCP refers to temporally-different pictures in the same view. DCP refers to different view pictures in the same access unit. This allows additional candidates for optimal prediction. Fig. 10 represents DCP and MCP for dependent view coding.

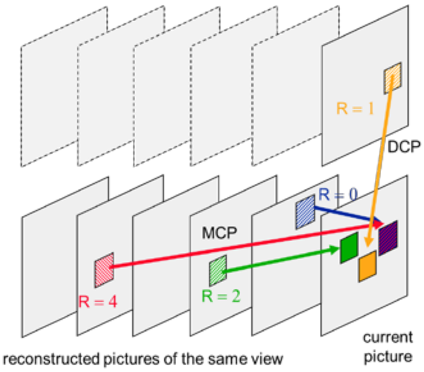

Fig. 10 DCP and MCP for dependent view coding

B. Depth-based block partitioning

Depth-based block partitioning uses depth information to partition the texture block. The collocated block can be binary-segmented. This is applied to the texture block. MCP is applied to each partition, later they are merged. Such a process takes regional characteristics into consideration. Fig. 11 displays partitioning by this method.

C. Motion parameter inheritance

Motion parameter inheritance (MPI) assumes similar motion in texture and depth data at the same view. Thus, motion information can be inherited. Fig. 12 exhibits the MPI process.

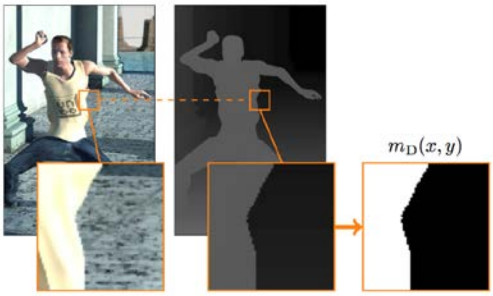

Fig. 11 Binary segmentation for DBBP

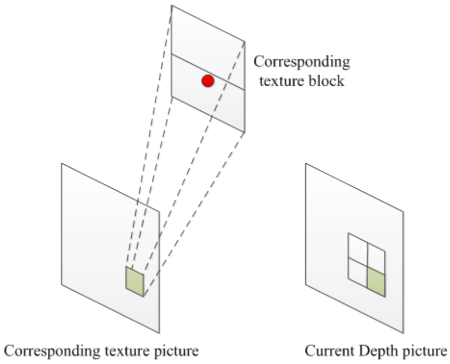

Fig. 12 Depth coding inherits motion information from texture coding results

VI. FREE-VIEWPOINT TELEVISION (FTV)

In 2013, MPEG started the third phase of FTV activity. The first and second phases are MVC and 3D video coding (3D-AVC and 3D-HEVC) [7]. The goal is to establish a new FTV framework. With a 3D space input, after coding and transmission, either a single view, multi-view or super multi-view application can display the 3D data. This framework is shown in Fig. 13. Super-multi-view typically means more than 100 views. In the previous phases, only 1D linear camera arrangement was considered. The third phase assumes arbitrary camera arrangements, including arc and 2D parallel settings. In the latest MPEG meeting, three EEs have been examined [8].

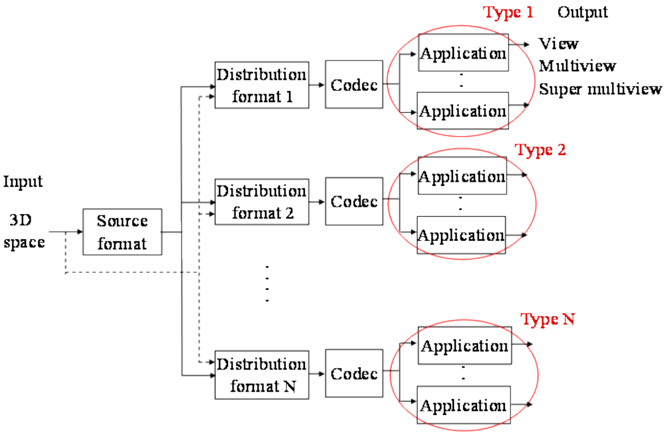

Fig. 13 Framework of FTV

A. EE1: Depth estimation

High quality depth data are essential in FTV applications. The existing depth estimation reference software (DERS) has been improved and extended to consider arbitrary camera arrangements.

B. EE2: View synthesis

Virtual view data can be generated from existing views. For a given display, the number of views needed to reconstruct all views is tested. The synthesized view data are compared with reference view data for evaluation.

C. EE3: Compression

This EE evaluates the compression method for tens of views. HEVC (simulcast), multi-view HEVC (MV-HEVC) and 3D-HEVC are evaluated. MV-HEVC is a subset of 3D-HEVC in which depth coding tools are not included. Direct transmission and depth based transmission tests are conducted.

VII. CONCLUSION

In this paper, we introduced the accomplishments of MPEG on 3D video coding: 3DAV, MVC, 3D-AVC, 3D-HEVC and the latest FTV activities. Plentiful proposals by MPEG participants have contributed to the development of coding standards that are necessary in the 3D video system. Since 3D video hardware technologies will keep growing, research on 3D video handling by MPEG will advance as well.

ACKNOWLEDGMENT

This research was supported by the ‘Cross-Ministry Giga KOREA Project’ of the Ministry of Science, ICT and Future Planning, Republic of Korea(ROK). [GK13C0100, Development of Interactive and Realistic Massive Giga-Content Technology]

REFERENCES

[1] A. Smolic and P. Kauff, “Interactive 3-D Video Representation and Coding Technologies,” *Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery*, vol. 93, pp.98-110, Jan. 2005.

[2] Video and Test, “Updated Call for Proposals on Multi-view Video Coding,” ISO/IEC JTC1/SC29/WG11 N7567, Nice, FR, Oct. 2005.

[3] Y.S. Ho and Y. Song, “Overview of the State-of-the-Art 3D Video Coding Technologies,” *Global 3D Tech Forum*, vol. 3, pp. 28-29, Oct. 2014.

[4] Video and Requirement, “Call for Proposals on 3D Video Coding Technology,” ISO/IEC JTC1/SC29/WG11 N12036, Geneva, CH, March 2011.

[5] D. Rusanovskyy, F.C. Chen, L. Zhang, and T. Suzuki, “3D-AVC Test Model 9,” ITU-T/ISO/IEC JCT3V-G1003, San Jose, US, Jan. 2014.

[6] Y. Chen, G. Tech, K. Wegner, and S. Yea, “Test Model 9 of 3D-HEVC and MV-HEVC,” ITU-T/ISO/IEC JCT3V-I1003, Sapporo, JP, July 2014.

[7] M. Tanimoto, T. Senoh, S. Naito, S. Shimizu, H. Horimai, M. Domanski, A. Vetro, M. Preda, and K. Mueller, “Proposal on a New Activity for the Third Phase of FTV,” ITU-T/ISO/IEC JCT3V-E0258, Vienna, AT, July 2013.

[8] K. Wegner, T. Senoh, and G. Lafruit, “Description of Exploration Experiments on Free-viewpoint Television (FTV),” ISO/IEC JTC1/SC29/WG11 N14551, Sapporo, JP, July 2014.