

Analysis of the Encoding Efficiency of 3D HEVC

Tian Tian¹, Xiuhua Jiang² and Caihong Wang³

Abstract This paper mainly studies currently developing 3D video coding based on HEVC. HEVC-based 3D video coding mainly focuses on 3DTV and auto-stereoscopic video compression system. A variety of new encoding tools, such as inter-view motion prediction and depth modeling modes, have been added in 3D HEVC. The interest is created to confirm if these new features and tools improve the encoding efficiency of 3D HEVC. The goal of this article is to analyze the encoding and compression efficiency of 3D High Efficiency Video Coding. In our experiments for stereo video without depth maps, 3D HEVC provides 59.44% and 24.32% average bit rate savings with respect to HEVC simulcast for the dependent view and both two views, while MVC provides 12.53% and 5.99% in comparison with H.264/AVC simulcast. The results indicate that 3D HEVC can remove redundancy between the views more effectively than MVC.

Keywords: 3D video coding, 3D HEVC, MVC

1 Introduction

In recent years the interest for 3D television (3DTV) and free viewpoint video (FVV) has been increased constantly. 3D movie, 3DTV and Blu-ray Disc have feasted thousands of consumers' eyes on 3D video. While stereo displays with glasses require 2 views, auto-stereoscopic displays without glasses, which require not only 2 but a multitude of different views, are regarded as a promising technology.

¹ Tian Tian (✉)

Information Engineering School, Communication University of China, Beijing, China
e-mail: tiantian1990210@tom.com

² Xiuhua Jiang

Information Engineering School, Communication University of China, Beijing, China

³ Caihong Wang

Information Engineering School, Communication University of China, Beijing, China

The state-of-the-art standard for multi-view video coding is the MVC extension of H.264/AVC [1]. In MVC, the independent view is coded with conventional H.264/AVC. The dependent views use the same coding tools with the only difference called disparity-compensated prediction (DCP) in purpose of removing redundancy between different views. Due to the multitude of views, the bit rate required for coding multi-view video with MVC, which increases approximately linearly with the number of viewpoints, is too large. MVC is not suitable for compressing and delivering 3D video for auto-stereoscopic displays.

In order to reduce the bitrate, depth image based rendering (DIBR) techniques, which are used to generate additional views for auto-stereoscopic displays based on transmitted video pictures and depth maps, can help to reduce the number of coded views. Therefore, MPEG specified a standard called multi-view video plus depth (MVD) format [2] for efficient compression and transmission of texture video and associated depth maps, in which only a few views need to be coded.

The new video coding standard High Efficiency Video Coding (HEVC) [3] is reported to double the compression efficiency in comparison with the current commercially available video coding standard H.264/AVC [4]. The new 3D video coding is based on HEVC and MVD format is applied to 3D HEVC so that 3D HEVC is likely to meet the demands for auto-stereoscopic displays.

In March 2011, Geneva, MPEG issued MPEG's Call for Proposals on 3D Video Coding Technology [5] with the goal of providing efficient compression and high quality view reconstruction of an arbitrary number of dense views. As a response, the method proposed by German Heinrich Hertz Institute (HHI) has been selected as the Test Model under Consideration (TMuC) [6] after a series of performance and subjective evaluation. The Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) established by ITU-T VCEG and ISO/IEC MPEG has been working on 3D video coding extensions of HEVC [7] and other video coding standards.

At the current stage, the interest is created to analyze the compression efficiency of 3D HEVC. We have designed the experiments to confirm if new tools improve the encoding efficiency of 3D HEVC. The experiments use stereo video (2-view video) without depth maps, for stereo video is in more common cases and MVC doesn't support MVD format. In this paper, we select the PSNR metric and bitrate in terms of video quality and encoding performance.

The paper is organized as follows: Section 2 introduces the new features of 3D HEVC. In Section 3 the encoding conditions of test signals is described. Section 4 presents the results and analysis of the encoding efficiency of 3D HEVC. Finally, the paper concludes with Section 5.

2 Introducing the features of 3D HEVC

The new features of 3D HEVC introduced in this section may be slightly modified and other features may be added in the future.

An access unit, which consists of not only all video pictures but also depth maps at the same time instant, is coded one by one. Inside an access unit, independent view is coded first as view 0 and the other views are dependent views. Inside a view, depth map is coded after texture picture using inter-component prediction techniques. The texture picture of independent view is coded with non-modified HEVC coder and the other components are coded with modified HEVC coder. Texture pictures can be predicted from already coded texture pictures of other views using inter-view prediction techniques and the information from already coded depth maps of other views. Depth maps can be predicted from already coded depth maps of other views.

The access unit can be configured as the access unit which consists of only texture pictures. In our experiments, the configuration is like that.

2.1 New Inter-view Prediction

Except for the known concept DCP similar with motion-compensated prediction (MCP) in MVC, inter-view prediction adds new tools, including inter-view motion prediction and inter-view residual prediction. The difference between DCP and MCP is that the reference picture list insert already coded video pictures of the same access unit.

The motion in different views at the same time instant is very similar so that redundancy lies in the motion parameters of different views. Motion skip mode [8] in MVC uses a constant disparity to predict the motion parameters for dependent views. The new inter-view motion prediction [9] employs estimated associated depth map to get the disparity. If texture and depth are coded together, the depth of the dependent view can be obtained by mapping the reconstructed depth map of reference view into the coordinate system of the current view. If texture pictures are coded only, the depth map is estimated by converting the disparity vectors for DCP into depth values or deriving from the corresponding depth block of the already reconstructed depth map in the same view with MCP parameters of the associated texture block. The process is presented on Fig.1. Whether coding depth maps or not, the depth map estimate is necessary. Through estimated depth map, each block of the texture pictures in the dependent view finds the corresponding block in the reference view and the motion parameters of the corresponding block are used as the candidate motion parameters for the block of dependent view.

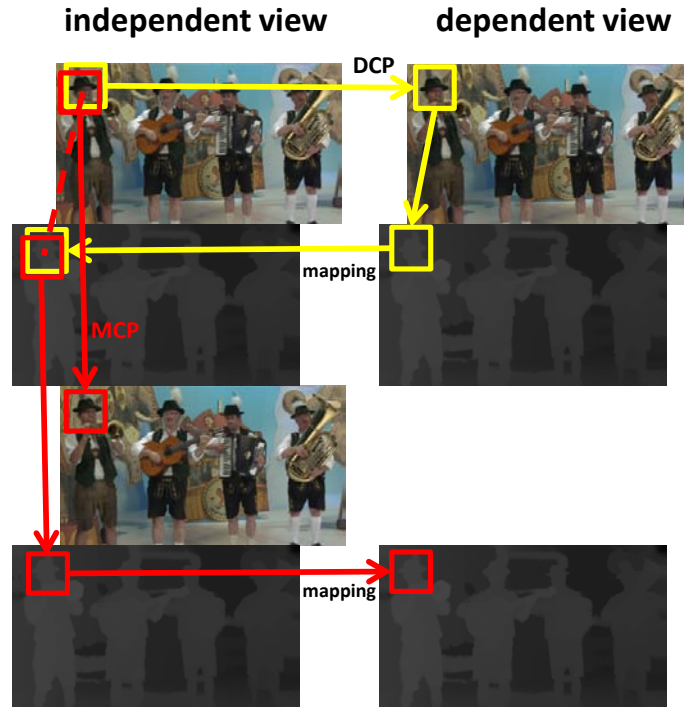


Fig. 1 Depth map estimates for only coding texture pictures

Inter-view residual prediction also uses depth map estimate. The difference between the corresponding residual block in the reference view and the current residual block is transform coded to reduce residual redundancy.

2.2 Coding and Encoder Control for Depth Maps

The characteristics of depth maps differ from texture pictures. Depth maps contain large areas of smoothly changing gray levels and sharp edges of step functions. Four new depth modeling modes, including Explicit Wedgelet, Intra-predicted Wedgelet, Inter-component Wedgelet and Inter-component Contour, allow two types of partitioning: wedgelets and contours. Wedgelets and contours both divide a depth block into two non-rectangular regions of constant values, but use a straight line and an arbitrary curve separately. New modes use new concepts of inter-component prediction [10].

The motion characteristics of the depth map are similar with the texture picture of the same view. Inside a view, the partitioning and motion parameters of the depth block can be inherited from associated texture block or be newly generated.

This mode is named as Motion Vector Inheritance (MVI) [11], which is inserted into merge mode.

Depth maps can't be watched by viewers so that coding artifacts in depth maps indirectly affect the quality of the synthesized video which is watched by viewers. Therefore, the distortion D in Lagrangian cost measure $J = D + \lambda \bullet R$ should be changed. In the View Synthesis Optimization, D is measured exactly with synthesized views distortion change (SVDC) [12] related to distortion of depth maps by fast rendering mechanism called renderer model.

In encoder control for renderable regions, regions in dependent views that can be rendered based on coded independent view don't need high coding quality, that is to say, these regions can have no residual, for those can be replaced by rendered versions after decoding. But the mechanism [13] leads to the difficulties in objective measures.

However, coding and encoder control tools for depth maps are not used in our experiments.

3 Encoding Conditions of Test Signals

For generating evaluation results, we used a set of two 1080p25 stereo sequences (BMX and Musicians) depicted on Fig. 2, which represent different spatial and temporal activity levels. The test sequences have spatial resolution 1920x1080 and each has both left and right view. BMX has fast camera movements and fast moving objects, while Musicians has slowly moving camera, moving objects and details in background.



Fig. 2 The snapshots of BMX (left) and Musicians (right)

The test input signals were encoded from original uncompressed YUV format to High Profile. The reference software HTM_DEV_0.2 and HM_10.1 were used for 3D HEVC and HEVC encoding process. JMVM_8.5 was chosen for MVC.

For achieving an ideal analysis of 3D HEVC, the configuration files are necessary to have very similar parameter values. The GOP structure of two encoding methods consists of the same I, P and B sequence. We used hierarchical B pictures with a GOP structure of 8 pictures and a random access period of 24 pictures

(roughly one second). The basic codec configuration for 3D HEVC is similar to the Random Access configuration of the common test conditions for HEVC. The basic codec configuration for MVC is also similar to the common test conditions for H.264/AVC.

The number of frames to be coded is set to be 250 (10s).

The Quantization Parameter (QP) impacts greatly on visual quality and the bit rate. The basic QP was set to the value 25, 30, 35 and 40 to calculate the average bit rate savings.

4 Encoding Efficiency of 3D HEVC

We have analyzed the encoding efficiency of 3D HEVC for the independent view, the dependent view and both two views in comparison to HEVC simulcast and MVC. The average bit rate savings is calculated using the Bjøntegaard delta rate [14] for the QP values 25, 30, 35 and 40.

4.1 3D HEVC vs. HEVC simulcast

In this section, we compare the encoding efficiency of 3D HEVC and HEVC simulcast for the independent view, the dependent view and two views in very similar encoding conditions.

Table 1 Average BD rate savings (in percentage)

Sequence	3D HEVC vs. HEVC simulcast		
	independent view	dependent view	two views
BMX	0	61.83	25.27
Musicians	0	57.05	23.37
Average	0	59.44	24.32

From the Table 1, we can find that 3D HEVC provides 59.44% and 24.32% average bit rate savings in comparison to HEVC simulcast for the dependent view and two views separately. Encoding stereo video sequences with 3D HEVC can obtain better compression efficiency than encoding those sequences with HEVC separately. 3D HEVC can remove the redundancy between the views quite effectively.

4.2 MVC vs. H.264/AVC Simulcast

In this section, the compression efficiency of MVC is compared with H.264/AVC simulcast for the independent view, the dependent view and two views. The encoding conditions are very similar.

Table 2 Average BD rate savings(in percentage)

Sequence	MVC vs. H.264/AVC simulcast		
	independent view	dependent view	two views
BMX	0	9.22	4.39
Musicians	0	15.84	7.60
Average	0	12.53	5.99

In Table 2, the experimental results show that 12.53% and 5.99% average bit rate savings are provided by MVC in comparison to H.264/AVC simulcast for the dependent view and both two views. The encoding efficiency for stereo video in MVC is improved just a little in comparison with H.264/AVC simulcast. This suggests that MVC can only remove partial redundancy between different views.

4.3 Analysis

59.44% average bit rate savings is provided by 3D HEVC for the dependent view while MVC provides 12.53% average bit rate savings. Similarly, for both two views, 3D HEVC provides 24.32% average bit rate savings in comparison with MVC's 5.99% average bit rate savings. From the data, we can observe that 3D HEVC's new inter-view prediction tools can remove the redundancy between the views more effectively so that 3D HEVC makes bigger savings on the bit rate with respect to simulcast for the dependent view or two views than MVC.

Additionally, BMX video clip, which has a much richer motion-wise theme than the Musicians one, has more average bit rate savings for the dependent view (BMX 61.83% vs. Musicians 57.05%) or two views (BMX 25.27% vs. Musicians 23.37%) for 3D HEVC. Therefore, new inter-view motion prediction tools can play a significant role in the encoding process.

5 Conclusions

This paper has studied the draft and the encoding efficiency of the developing 3D video coding 3D HEVC. Various new features and encoding tools added in 3D HEVC have been described. Experiments are designed to analyze the encoding efficiency of 3D HEVC. Our experimental results about 3D HEVC show that 59.44% and 24.32% average bit rate savings is provided by 3D HEVC in comparison to HEVC simulcast for the dependent view and both two views while MVC provides 12.53% and 5.99% average bit rate savings with respect to H.264/AVC simulcast. The results indicate that 3D HEVC can remove redundancy between the views more effectively than MVC.

Acknowledgments This work is supported by the National Sci-tech Support Plan of China: The Research of Stereoscopic Content Transmission Technology and Systems. (No. 2012BAH39F02).

References

1. ITU-T, ISO/IEC (2010) Advanced video coding for generic audiovisual services. ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC).
2. P. Merkle, A. Smolic, K. M^uller, T. Wiegand (2007) Multi-view video plus depth representation and coding. In: Proc. of ICIP.
3. JCT-VC (2013) High Efficiency Video Coding (HEVC) text specification draft 10 (for FDIS & Last Call). Doc. JCTVC-L1003.
4. H. Koumaras, M. Kourtis, D. Martakos (2012) Benchmarking the encoding efficiency of H.265/HEVC and H.264/AVC. In: IIMC.
5. ISO/IEC MPEG (2011) Call for Proposals on 3D Video Coding Technology. MPEG N12036.
6. ISO/IEC MPEG (2011) Test Model under Consideration for HEVC based 3D video coding. MPEG N12350.
7. JCT-3V (2013) 3D-HEVC Test Model 4. Doc. JCT3V-D1005.
8. H.-S. Koo, Y.-J. Jeon, B.-M. Jeon (2006) Motion Skip Mode for MVC. JVT-U091.
9. H. Schwarz, T. Wiegand (2012) Inter-View Prediction of Motion Data in Multiview Video Coding. In: Proc. of PCS.
10. P. Merkle, C. Bartnik, K. M^uller, D. Marpe, T. Wiegand (2012) 3D Video: Depth Coding Based on Inter-component Prediction of Block Partitions. In: Proc. of PCS.
11. M. Winken, H. Schwarz, T. Wiegand (2012) Motion Vector Inheritance for High Efficiency 3D Video plus Depth Coding. In: Proc. of PCS.
12. G. Tech, H. Schwarz, K. M^uller, T. Wiegand (2012) 3D Video Coding using the Synthesized View Distortion Change. In: Proc. of PCS.
13. S. Bosse, H. Schwarz, T. Wiegand (2012) Encoder Control for Renderable Regions in High Efficiency 3D Video plus Depth Coding. In: Proc. of PCS.
14. G. Bjøntegaard (2001) Calculation of average PSNR differences between RD-curves. VCEG-M33.