Perceptual Distortion Measurement in the Coding Unit Mode Selection for 3D-HEVC

Sima Valizadeh, *Member, IEEE*, Panos Nasiopoulos, *Senior Member, IEEE*, and Rabab Ward, *Fellow, IEEE*

Abstract**—3D-HEVC achieves higher compression efficiency compared with the simulcast HEVC or disparity-compensated multi-view video coding (MVC). Improved compression efficiency is highly desirable in the transmission of 3D video and its storage. The coding efficiency gain is the result of the new coding tools introduced in 3D-HEVC such as inter-view motion prediction and inter-view residual prediction. We propose to integrate a perceptual video quality metric inside the rate distortion optimization process of the 3D-HEVC. Specifically, in the coding unit (CU) mode selection process, PSNR-HVS is used as a measure for distortion. PSNR-HVS is based on the characteristics of the human visual system (HVS). Results show that the proposed approach improves the compression efficiency of the 3D-HEVC and achieves higher video quality compared with 3D-HEVC.**

Index Terms**— Perceptual 3D video coding, rate distortion optimization (RDO), coding unit (CU) mode selection, human visual system (HVS), PSNR-HVS, 3D High Efficiency Video Coding (HEVC)**

I. Introduction

3D video extension of the high efficiency video coding called 3D-HEVC is currently being developed by a joint collaborative team (JCT-3V) from video coding experts group (VCEG) and moving picture experts group (MPEG) [1]. Efficient coding of the 3D video depends on removing the inter-view dependencies as well as the temporal and spatial redundancies. In addition to the disparity compensated prediction (DCP), 3D-HEVC uses new tools such as inter-view motion prediction (IVMP) and the inter-view residual prediction (IVRP) coding tools. In both methods, the dependent view is compressed by taking advantage of the coded data in the base view. In comparison to HEVC simulcast, 3D-HEVC provides 50% bitrate savings. The newly developed tools in 3D-HEVC, provide 20% bitrate savings in comparison to disparity compensated coding [2].

In all ITU-T and ISO/IEC JTC 1 video coding standards prior to HEVC, macroblocks with fixed size of 16*16 were used [3]. In HEVC, the size of the Coding Tree Blocks (CTB) is equal to 16, 32, or 64 based on the needs of encoders in terms of memory and computational requirements. A syntax element in the Sequence Parameter Set (SPS) determines the size of the CTBs. High resolution video content benefits from the larger size of CTBs. CTBs are partitioned into coding blocks (CB) using a tree structure. Based on the characteristics of a region of the video frame, the size of the CB is selected in an iterative process. Based on the intra-picture prediction or inter-picture prediction, the prediction mode for the CU is selected.

In a video coding standard, the values of the syntax elements are determined by minimizing the distortion between the reconstructed and original video sequence subject to some constraints such as constraints for the maximum allowed bitrate or the maximum coding delay. This minimization problem is split into a series of smaller minimization problems to keep the encoding delay within an acceptable range [4].

The coding mode is selected based on the minimization:

$$c^* = \arg\min_{c \in C_k} D_k(c) + \lambda . R_k(c) \qquad (1)$$

where $D_k(c)$ represents the distortion between the original block and its reconstruction. C_k is the set of applicable coding modes for the block and c^* is the selected mode. $R_k(c)$ is the bitrate of the block of samples. Distortion measurement is done either by a mathematical distance or based on perceptual measures. While perceptual metrics have higher correlation with subjective quality of experience, it is challenging to integrate them inside the rate distortion optimization process. In both HEVC and 3D-HEVC, sum of squared differences (SSD) is used for distortion measurement in coding unit mode decision while motion estimation uses sum of absolute differences (SAD).

Many methods have been proposed to integrate the properties of the human visual system into some aspects of the H.264 video coding standard [5]-[6]. Inter frame prediction has been modified by using structural similarity (SSIM) in the rate distortion optimization process [7]-[9]. SSIM has also been integrated in the intra mode selection process of H.264 [10]-[11]. Integration of SSIM as a perceptual video quality metric inside the rate distortion optimization process of

S. Valizadeh is with the Electrical and Computer Engineering Department, University of British Columbia, Canada (email:simav@ece.ubc.ca)

P. Nasiopoulos is a professor at the Electrical and Computer Engineering Department, University of British Columbia, Canada (email:panosn@ece.ubc.ca)

R. Ward is a professor at the Electrical and Computer Engineering Department, University of British Columbia, Canada (email:rababw@ece.ubc.ca)

This work was supported by the NPRP grant # NPRP 4-463-2-172 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

the H.264 and HEVC standards has been explored in the literature [12]-[14]. For 3D video coding standard, a perceptual video quality metric is incorporated inside the rate distortion optimization process of multi-view video coding (MVC) which is the multi-view extension of H.264 [15]. The study in [15] focuses on modifying prediction mode selection in the rate distortion optimization process. In addition to modifying the intra prediction mode, they also study the integration of the perceptual video quality metric in the motion estimation process.

In the HEVC video coding, prior to the intra mode selection and motion vector estimation, intra/inter mode is selected for the coding unit [4]. The size of the coding unit is determined iteratively by a quad-tree structure while macroblock size was fixed in the H.264 standard and MVC. Also, new tools have been introduced in the 3D-HEVC standard including the inter-view motion prediction and the inter-view residual prediction. Introduction of the new tools in the 3D-HEVC increases its compression efficiency beyond what disparity compensation can provide. In this paper, we investigate improvement of the coding efficiency of the 3D-HEVC by integrating a perceptual video quality metric in the coding unit (CU) mode selection process.

II. Proposed Approach

In 3D-HEVC, the base view is coded using an unmodified HEVC. The dependent view uses all the tools of an HEVC encoder as well as additional tools such as inter-view motion prediction and the inter-view residual prediction. Fig. 1 shows the basic structure of the 3D video codec for compressing two views.

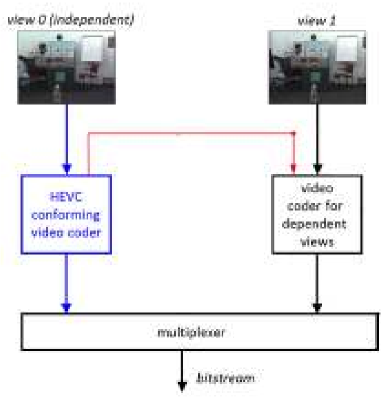

Fig. 1 Basic codec structure for compression of the base view and the dependent view

For coding unit mode decision, the smallest cost is selected among all mode candidates [4]. Lagrangian cost function, $D + \lambda.R$, weights the distortion and the bitrate. D is the distortion caused by coding the considered block in a given mode and R is the associated number of the required bits. λ is the Lagrangian multiplier that balances the trade-off between bitrate and distortion. In HEVC and 3D-HEVC mode selection process, λ is selected as:

$$\lambda = 0.85.2^{\frac{QP-12}{3}} \tag{2}$$

where QP is the quantization parameter. This dependency on the quantization parameter has been derived empirically [16]-[17]. As measure for the distortion, the sum of squared differences (SSD) is used in the 3D-HEVC reference software.

In our proposed approach, PSNR-HVS is used for measuring distortion in the coding unit mode selection. PSNR-HVS is a full-reference quality metric that is based on characteristics of the human visual system and shows higher correlation with the subjective video quality evaluations than PSNR [18]. PSNR-HVS is defined as:

$$PSNR - HVS = 10\log\left(\frac{255^2}{MSE_{HVS}}\right) \tag{3}$$

$$MSE_{HVS} = K\sum_{i=1}^{I-7}\sum_{j=1}^{J-7}\sum_{m=1}^{8}\sum_{n=1}^{8}((X[m,n]_{ij} - X[m,n]_{ij}^{e})T_C[m,n])^2$$

where K=1/[(I-7)(J-7)64]. I, J are the image width and height. X_{ij} is the DCT coefficient of an 8×8 image block with its upper left corner at (i,j). X_{ij}^{e} is the DCT coefficient of the corresponding block in the original image. T_c is a matrix adopted from the JPEG quantization table proposed in the JPEG [19]. PSNR-HVS is adaptable to block sizes with different sizes and is not too computationally complex. These characteristics allows its application inside the rate distortion optimization process of 3D-HEVC.

In the rate distortion optimization process, λ balances the trade-off in decreasing the bitrate or distortion. In selecting the coding modes, Lower bitrate comes at the expense of more distortion. Thus, effective choice of λ plays an important role in reaching higher compression efficiency. In this paper, modified λ, is the product of a scalar coefficient in the traditional λ used in 3D-HEVC reference software. This scalar coefficient is determined by experimental results in the next section.

III. Experimental Results

To test the performance of the proposed approach, PSNR-HVS is integrated into the 3D-HEVC reference software HTM10. Test video sequences are selected from the standard MPEG test sequences as summarized in Table 1. The test video sequences are in the YCbCr format. The length of each sequence is 10 seconds.

Table 1 Test video sequences

VIDEO SEQUENCES	RESOLUTION	TOTAL FRAMES
POZNANSTREET	1920x1088	250
NEWSPAPERCC	1024x768	300
GHOSTTOWNFLY	1920x1088	250
POZNANHALL	1920x1088	250

In our approach, the Lagrangian multiplier is modified by a scaling factor, c, as:

$$\lambda_{proposed} = c.\lambda_{mode} \tag{4}$$

We run our tests for four different QPs: 25, 30, 35 and 40.

These values are used in the standard MPEG 3D-HEVC tests [20]. By running the reference 3D-HEVC software for these QP values, four points in the rate distortion plot are obtained. To compare the proposed approach with the reference 3D-HEVC software, Bjontegaard's Delta (BD) Rate [21] is used. BD rate measures the average difference along the bitrate axis, between the proposed and reference curves. Fig. 2 shows the BD rate savings of the proposed approach depending on the scaling factor.

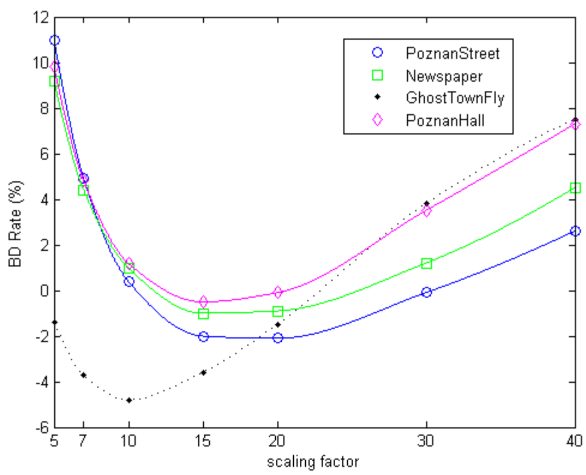

Fig. 2 Bitrate savings versus the scaling factor for different video sequences

BD savings show a similar trend in three videos of PoznanStreet, Newspaper and PoznanHall. We are interested in the highest amount of BD rate saving. For these three test videos the extremum of the plot happen around the same value of scaling factor. However, GhostTownFly video has a different trend and its minimum happen on a different scaling factor. GhostTownFly is the only test sequence that is computer generated. The other three video sequences are natural scenes. The optimal scaling factor for the natural scene videos happen at scaling factor 15. Based on Fig. 2, the scaling factor value of 15 leads to the most amount of BD rate savings in the proposed approach compared with the reference software for the natural scene videos. This scaling factor is selected for our tests. In Table 3, we summarized the results (bitrate and quality) of the proposed approach and the reference software for the video. The quality is measured by PSNR-HVS. PSNR-HVS shows higher correlation with the subjective results compared with the PSNR [18]. Video0 is the base view while video1 refers to the dependent view. Video (2v) contains both base view (video0) and dependent view (video1). The quality of the video(2v) is the average of the quality of video0 and video1. The bitrate of the video(2v) is equal to summation of the bitrate of video0 and video1. Video(2v) is reported in all MPEG 3D-HEVC tests.

Table 2 summarizes the BD-rate savings for the three natural scene video sequences. The BD rates are reported for the base view (video0), the dependent view (video1) and also for both views together, video (2v). The average bitrate savings of the proposed approach compared with the reference 3D-HEVC software for the stereoscopic video (two views together) is 1.17%. Our proposed approach needs on average 1.17% less bitrate to transmit/store the stereoscopic video compared to the reference 3D-HEVC.

Table 2 Bitrate saving of the proposed approach compared to the 3D-HEVC reference software. Video 0 is the base view and video 1 refers to the dependent view. Video (2v) contains both base view (video0) and dependent view (video1).

VIDEO SEQUENCES	VIDEO 0 BD-RATE	VIDEO 1 BD-RATE	VIDEO (2V) BD-RATE
POZNANSTREET	-1.8%	-4.3%	-2.0%
NEWSPAPERCC	-0.4%	-3.1%	-1.0%
POZNANHALL	-0.5%	-0.8%	-0.5%
AVERAGE	-0.9%	-2.73%	-1.17%

Integration of the quality metric inside the encoder achieves higher compression efficiency at the expense of more complexity. The encoding time of the proposed approach along with the reference 3D-HEVC software are reported in Table 4. By taking the geomean of the encoding times of 3D-HEVC and the proposed approach, the ratio for the two geomeans is 1.27 on average for the two test video sequences. This approach is used by MPEG, in their comparisons as they introduce new tools or propose complexity reduction approaches.

Table 3 Bitrate and quality of the reference 3D-HEVC and the proposed approach for the PoznanStreet test video sequence

PoznanStreet	QP	Reference 3D-HEVC		Proposed Approach		BD-rate (piecewise cubic)	BD-rate (cubic)
		Kbps (kbps)	Quality (dB)	Kbps (kbps)	Quality (dB)		
Video0	25	2438.8984	39.5482	2264.8912	39.4768	-1.8%	-1.8%
	30	956.9616	36.2907	942.2680	36.3013		
	35	461.1040	33.2217	458.1376	33.2359		
	40	241.1088	30.2582	240.9576	30.2788		
Video1	25	702.7456	38.2428	612.4696	38.1710	-4.3%	-4.3%
	30	196.1960	35.2568	183.1904	35.2264		
	35	75.4288	32.3799	73.6064	32.3647		
	40	33.5168	29.5985	33.1120	29.5919		
Video(2v)	25	3141.6440	38.8955	2877.3608	38.8239	-2.0%	-2.0%
	30	1153.1576	35.7738	1125.4584	35.7638		
	35	536.5328	32.8008	531.7440	32.8003		
	40	274.6256	29.9283	274.0696	29.9353		

Table 4 Comparison between encoding time of the proposed approach and 3D-HEVC reference software

		Encoding Time (s)		
	QP	Reference 3D-HEVC	Proposed Approach	Geomean Ratio
PoznanStreet	25	53388.89	69857.61	1.20
	30	43990.69	48030.12	
	35	36790.86	44754.23	
	40	34106.29	40342.94	
Newspaper	25	31445.84	38807.02	1.30
	30	24842.85	35325.01	
	35	21256.16	27610.71	
	40	19075.64	24279.65	
PoznanHall	25	35367.33	45285.11	1.27
	30	28091.54	36219.36	
	35	26095.15	33573.13	
	40	23595.67	29284	
average		1.27		

IV. CONCLUSION

In this paper, we improved the compression efficiency of 3D-HEVC by integrating a perceptual video quality metric inside the encoder. PSNR-HVS is a full-reference quality metric based on the characteristics of the human visual system. We used PSNR-HVS as a measure of distortion in the rate distortion optimization process for coding unit mode selection. The proposed approach was tested on different standard video sequences. The results show that our proposed scheme provides on average 1.17% bitrate saving compared to the 3D-HEVC for the natural scene video.

REFERENCES

[1] G. J. Sullivan, J. M. Boyce, Y. Chen, J. Ohm, C. A. Segall and A. Vetro, "Standardized extensions of high efficiency video coding (HEVC)," *Selected Topics in Signal Processing, IEEE Journal of,* vol. 7, pp. 1001-1016, 2013.

[2] K. Muller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle and F. H. Rhee, "3D high-efficiency video coding for multi-view video and depth data," *Image Processing, IEEE Transactions on,* vol. 22, pp. 3366-3378, 2013.

[3] G. J. Sullivan, J. Ohm, W. Han and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 22, pp. 1649-1668, 2012.

[4] V. Sze, M. Budagavi and G. J. Sullivan, *High Efficiency Video Coding (HEVC).* Springer, 2014.

[5] Z. Chen and C. Guillemot, "Perceptually-Friendly H.264/AVC Video Coding Based on Foveated Just-Noticeable-Distortion Model," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 20, pp. 806-819, 2010.

[6] J. Chen, J. Zheng and Y. He, "Macroblock-Level Adaptive Frequency Weighting for Perceptual Video Coding," *Consumer Electronics, IEEE Transactions on,* vol. 53, pp. 775-781, 2007.

[7] Z. Mai, C. Yang, K. Kuang and L. Po, "A novel motion estimation method based on structural similarity for H.264 inter prediction," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, vol. 2, Feb. 2006, pp. 913-916, 2006.

[8] C. Yang, R.Leung, L. Po and Z. Mai, "An SSIM-optimal H.264/AVC inter frame encoder," in Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on, 2009, pp. 291-295.

[9] C. Yang, H. Wang and L. Po, "Improved inter predic-tion based on structural similarity in H.264," in Signal Pro-cessing and Communications, 2007. ICSPC 2007. IEEE In-ternational Conference on, 2007, pp. 340-343.

[10] Z. Mai, C. Yang, L. Po, and S. Xie, "A new rate-distortion optimization using structural information in H.264 I-frame encoder," *in Proc. ACIVS* 2005, pp. 435–441.

[11] B. H. K. Aswathappa and K. R. Rao, "Rate-distortion optimization using structural information in H.264 strictly intra-frame encoder," in *System Theory (SSST), 2010 42nd Southeastern Symposium on,* 2010, pp. 367-370.

[12] T. Ou, Y. Huang and H. H. Chen, "SSIM-Based Perceptual Rate Control for Video Coding," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 21, pp. 682-691, 2011.

[13] S. Wang, A. Rehman, Z. Wang, S. Ma and W. Gao, "Perceptual Video Coding Based on SSIM-Inspired Divisive Normalization," *Image Processing, IEEE Transactions on,* vol. 22, pp. 1418-1429, 2013.

[14] A. Rehman and Z. Wang, "SSIM-inspired perceptual video coding for HEVC," in *Multimedia and Expo (ICME), 2012 IEEE International Conference on,* 2012, pp. 497-502.

[15] G. Tech, K. Müller, P. Merkle, H. Brust and L. Jin, "Final report on coding algorithms for mobile 3DTV," *Mobile 3DTV, Techical Report D2.6*, February 2011.

[16] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *Signal Processing Magazine, IEEE,* vol. 15, pp. 74-90, 1998.

[17] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Image Processing, 2001. Proceedings. 2001 International Conference on,* 2001, pp. 542-545.

[18] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, New full-reference quality metrics based on HVS, *Proc. of the Second International Workshop on Video Processing and Quality Metrics*, Scottsdale, USA, 2006, 4 p.

[19] G. Wallace, "The JPEG still picture compression standard," *Comm. of the ACM*, vol. 34, No.4, 1991.

[20] Y. Chen, G. Tech, K. Wegner, S. Yea, "Test Model 11 of 3D-HEVC and MV-HEVC." *JCT-3V of ITU-T SG* 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Document JCT3V-K1003, Geneva, February 2015.

[21] G. Bjontegaard, "Calculation of average PSNR difference between RD curves," in Proc. 13th Meeting ITU-T Q.6/SG16 VCEG , Austin, TX, Apr. 2001.