A LOW COMPLEXITY MODE DECISION APPROACH FOR HEVC-BASED 3D VIDEO CODING USING A BAYESIAN METHOD

Hamid Reza Tohidypour, Mahsa T. Pourazad[1,2], and Panos Nasiopoulos[1]
[1]Department of Electrical & Computer Engineering, University of British Columbia, Canada
[2]TELUS Communications Inc., Canada

ABSTRACT

The 3D extension of High Efficiency Video Coding (HEVC) standard (3D-HEVC) aims at improving coding efficiency by introducing new and unique approaches for utilizing correlations between the different views of a scene. Reported coding efficiency, however, comes at the expense of increased computational complexity. For real-time applications, reducing the computational complexity of 3D-HEVC is very important. In this paper, we propose an adaptive fast mode assigning method based on a Bayesian classifier that reduces 3D-HEVC's coding complexity by up to 51.95%, while maintaining the overall quality and bit-rate.

Index Terms— 3D HEVC, video compression, low complexity compression, Bayesian classifier

1. INTRODUCTION

Multiview video provides more immersive viewing quality of experience compared to the traditional 2D video content. Each multiview video stream contains several video sequences, which are simultaneously captured from the same scene. One the major challenges involved in multiview video applications is compressing and transmitting the resulting very large amount of data.

The latest multiview video coding standard (MVC) is an extension of the H.264/AVC standard [1]. To improve multiview video coding efficiency, MVC is equipped with several inter-view prediction schemes. While these schemes improve coding efficiency, they also increase the overall computational complexity. Several studies were conducted towards reducing MVC coding complexity for real-time applications [2-4]. Researchers in [2] proposed the use of a threshold rate distortion (RD) cost for each to-be-encoded block so that the MVC encoder does not examine all the possible prediction modes. The method presented in [3] examines only a limited number of inter and intra prediction modes for each block, based on the inter and intra prediction modes of the corresponding block in the reference view and eight spatial neighboring blocks. To reduce the number of examined modes by the encoder, the researchers in [4] propose to utilize the depth information in cases where this information is available (multiview plus depth (MVD) content).

Recently, the Joint Video Team (JVT) of the ISO/IEC MPEG and the ITU-T VCEG have introduced a new compression standard, known as the High Efficiency Video Coding (HEVC) [5], which achieves significantly higher compression (up to 45.54% in terms of bit rate) than the H.264/AVC standard [6]. Considering the superior performance of HEVC and the market trends towards the adoption of a multiview system, the JVT of MPEG and VCEG has initiated the development of the 3D extension of HEVC (3D-HEVC). Note that while HEVC's advanced coding features, such as the increased number of intra modes and more flexible inter prediction, improve coding performance, they also result in increased computational cost. In the case of 3D-HEVC, the computational complexity elevates due to inter-view prediction. Reduction of coding complexity is one of the critical issues that need to be addressed in the development of the 3D-HEVC codec. To this end, in our previous work we proposed an adaptive search range adjustment method and an early termination mode-search scheme for 3D-HEVC to decrease the coding complexity [7].

In this study, we propose a fast mode decision scheme based on a Bayesian classifier to predict the mode of the blocks in the dependent view using information of already encoded neighboring blocks. This scheme significantly speeds up the encoding process by preventing the encoder to go through an extensive mode-search process (a 3D-HEVC encoder by default checks all the available modes to find the mode with the lowest rate distortion cost).

The rest of this paper is organized as follows: Section 2 includes a short overview of the 3D-HEVC emerging standard, Section 3 elaborates on our proposed method, performance evaluation of our method is presented in Section 4, and the conclusion is drawn in Section 5.

2. OUR PROPOSED SCHEME

The basic structure of the 3D-HEVC codec is shown in Fig. 1[8]. As it can be observed, in 3D-HEVC one of the views is selected as the base view (BV) and is HEVC encoded independent of the other views. The rest of the views (called dependent views (DVs)) are coded utilizing disparity-compensated prediction (DCP) in addition to the spatial prediction and motion-compensated prediction (MCP) as used in conventional HEVC. In DCP, the already encoded frames of other views in the same time instance are added to

This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant STPGP 447339-13.

the reference list (see arrows in Fig. 1). The main objective of our study is to decrease the computational complexity of the 3D-HEVC encoder by utilizing the correlation between the base view (BV) and the dependent views (DVs). During the inter/intra prediction process, 3D-HEVC computes the RD cost for all of the available modes (based on the size of the to-be-encoded CU). Then, the encoder selects the mode that has the lowest RD cost. In our study, we propose a fast mode assigning (FMA) technique, which uses the mode information of the CUs in BV as well as the mode information of the already-encoded neighboring CUs in DVs to predict the mode of the to-be-encoded CU in each DV. This approach enables the encoder to avoid the extensive computational cost involved in the mode search process.

In our method for predicting the inter/intra prediction mode of the to-be-encoded CU in a DV, the mode information of the neighboring (top left, top, top right, and left) CUs that are already coded as well as the mode information of the corresponding CU in the BV are used. These CUs are called predictor CUs hereafter. Fig. 2 shows an example of a current to-be-coded CU in the n^{th} view and the predictor CUs, i.e., its four spatial neighbors and its corresponding CU in the BV. Note that the neighboring CUs in the dependent view are similar to the candidates that 3D-HEVC chooses for the inter prediction merge mode.

Our goal here is to approximate a function whose input is the mode information of the predictor CUs and its output is the predicted mode of the current CU. In other words, we would like to estimate the posterior probability of the current CU's mode, given the mode information of the predictor CUs. This problem can be modeled as a supervised learning problem with training and testing processes. To formulate the problem, assume Y is the random variable corresponding to the probability of possible modes for the current to-be-coded CU in a DV, and X is a random vector corresponding to the probabilities of the modes of predictor CUs. If there are M different 3D-HEVC inter and intra modes, the random variable Y has M different values representing the probability of each mode. If there are L predictor CUs, then the length of vector X will be equal to L and each of its components can take M possible values which represent the probability. This results in M^L-1 different possible probability values for the random vector X. The term -1 comes from the fact that the probability values should sum to one. The probability of each mode of the current CU in DV given the probability of the modes of the predictor CUs, i.e., the posterior probability $P(Y|X)$, is calculated using the Bayes rule as follows:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

where $P(Y)$ is the prior probability of the mode of the to-be encoded CU, $P(X|Y)$ is the class-conditional density, which defines the distribution probability of observing a combination of modes for predictor CUs given the probability of the mode of current CU.

To find $P(Y|X)$, the learning algorithms needs to estimate $P(Y)$ and $P(X|Y)$. The former requires estimating M-1values (variable Y can have M different values), and the later requires learning of an exponential number of parameters, which is an intractable problem [9]. In order to estimate $P(X|Y)$,we use the Naive Bayes classifier [9]. The Naive Bayes classifier dramatically reduces the complexity of estimating $P(X|Y)$ by making a conditional independence assumption. This learning algorithm assumes that different components of the X vector are independent with respect to a given Y. Taking into account the conditional independence assumption we have:

$$P(X|Y) = P(X_1, X_2, \dots, X_L|Y) = \prod_{l=1}^{L} P(X_l|Y) \quad (2)$$

Therefore,

$$P(Y|X) = \frac{\prod_{l=1}^{L} P(X_l|Y)\, P(Y)}{P(X)} \quad (3)$$

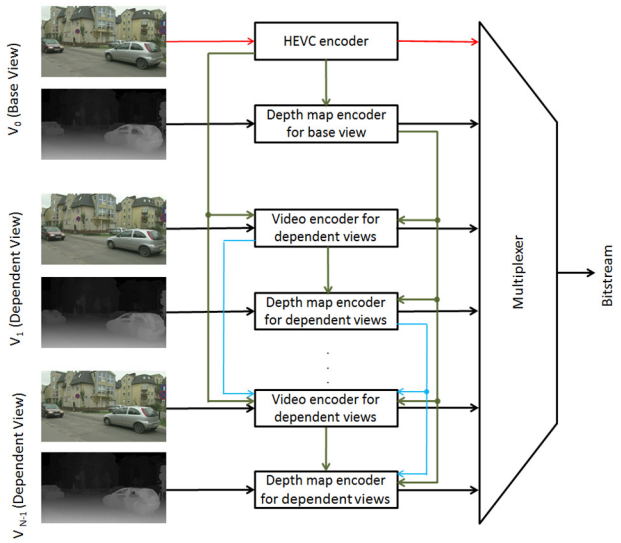

Fig. 1. The structure of 3D-HEVC [7].

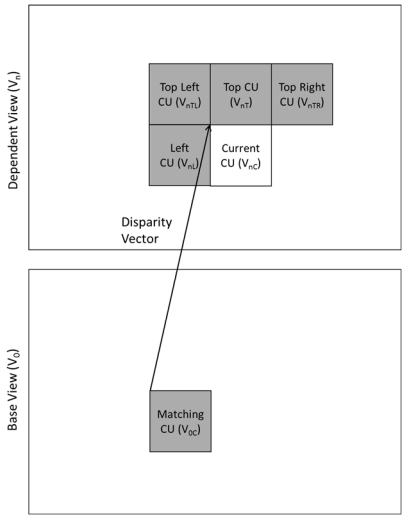

Fig. 2. Current CU and its four spatial neighbors in base view and current view.

According to the optimal Bayes decision rule [9], the mode of the posterior probability distribution (predicted mode of the current CU in DV) is the mode, which has the largest probability among all the modes. Therefore, for classifying a new X, the following formula can be used:

$$y_m = \underset{y_m}{\operatorname{argmax}} P(Y = y_m) \prod_{l=1}^{L} P(X_l | Y = y_m) \quad (4)$$

where y_m is the m^{th} possible value of Y. Note that $P(X)$ is the normalization factor in (3), thus it has been omitted in calculation of y_m (i.e., the value of y_m is independent of $P(X)$). The value of y_m represents the predicted mode for the to-be-encoded CU in DV.

To find the optimal value of y_m in (4), we need to have $P(X|Y)$ and $P(Y)$. These probabilities need to be computed during the training process. A very popular method to estimate these probabilities is the Maximum Likelihood Estimation (MLE) [9]. A major drawback of MLE is that when MLE is used for estimating the probabilities, there are some situations in which we have not seen some states (modes) in the training set. To resolve this problem, we employ Maximum a Posteriori (MAP) estimation [9]. MAP estimation incorporates a prior distribution function over $P(X|Y)$ and $P(Y)$. In order to use MAP estimation, it is required to assign appropriate conjugate prior distribution for the parameters. Since the distribution of the posterior probability is a Multinomial (Categorical) distribution, in our method we use Drichlet distribution as the conjugate prior [9]. Thus, the solution to the MAP estimation for $P(Y)$ is as follows:

$$P(Y = y_k) = \frac{N_k + \alpha_k}{Total\ number\ of\ tries + \sum_{k=1}^{M} \alpha_k} \quad (5)$$

where α_k determines the strength of the prior assumptions relative to the observed data, M is equal to the number of different values which Y can take, and N_k indicates the number of times the modes of the current CU is equal to y_k. Similarly the MAP estimation for $P(X|Y)$ is as follows:

$$P(X_l = x_{lm} | Y = y_k) = \frac{N_{lmk} + \alpha_{lm}}{Total\ number\ of\ tries + \sum_{m=1}^{M} \alpha_{lm}} \quad (6)$$

where α_{lm} determines the strength of the prior assumptions relative to the observed data, M is equal to the number of distinct values which X_l can take, and N_{lmk} indicates the number of times that the mode of the l^{th} predictor is equal to x_{lm}, given the mode of the to-be-encoded CU in DV is equal to y_k. To find the hyper parameters α_k and α_{lm}, which constitute the initial model, four representative video sequences are used in our approach. These video sequences are excluded from the video sets used to test our approach.

In order to improve the efficiency of the initial model and also make sure that it works for all possible encoding configurations, we fine-tune it by using the first few frames of each scene. In our NB-FMA implementation (see Fig. 3), the first second of the video (i.e., 25 frames for 25fps format) is coded using a conventional 3D-HEVC encoder. Note that the number of frames is based on our empirical tests. The mode information of these frames is used for fine-tuning the model. During the training/fine-tuning process, the BV and DVs are encoded using the original 3D-HEVC encoder [8]. For each CU in the DV, the information about the chosen mode is stored and the probability values are updated as the coding process continues. Based on this information, the probability of each mode (for a to-be-coded CU in DV) given the predictor CUs' mode is calculated. The 3D-HEVC modes (inter and intra modes) are labeled by discrete numbers and each mode is considered as a class.

During the training process, the probability values are updated as the coding process continues. In the testing process, first the BV is encoded, and then the encoder encodes the DVs. Unlike the training process, the encoder does not check all of the inter and intra prediction modes. Instead, the modes of the predictor CUs are used for predicting the mode of the current CU. In this study, the three mode candidates with the highest probability among all the available modes are chosen, and the encoder calculates the RD cost for these three candidates and chooses the one with the smallest RD cost. If scene change occurs, the training process is repeated to update the probabilities of the model.

3. RESULTS

In our experiment, four test videos from the data set provided by MPEG for the 3D Video Coding Call for proposals [10] were used (see Table I). Our method was implemented in the 3D-HEVC software (HTM-DEV-2.0). The performance of our proposed scheme is compared with the performance of the proposed complexity reduction method proposed in [7]. Note that to have a fair comparison the adaptive search range scheme proposed in [7] was not used in our implementation. The "baseCfg_2view+depth" configuration is used (hierarchal B pictures and GOP length 8) [11]. By using this configuration, the 3D-HEVC encodes two views and their corresponding depths [11]. The QPs used for the views and the Depth map (QPV, QPD) are as follows: (25, 34), (30, 39), (35, 42) and (40, 45). Fig. 4 shows the RD curves of four test video sequences (reported PSNR values belong to the video content of the DV and not

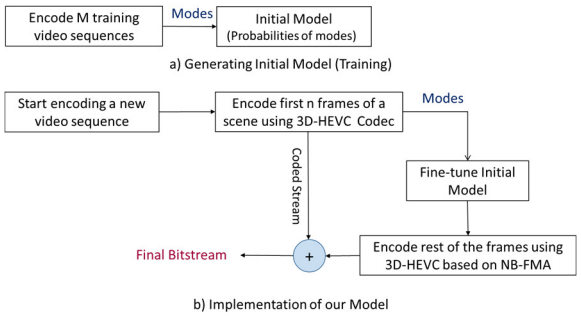

Fig. 3. Block diagram of the proposed method.

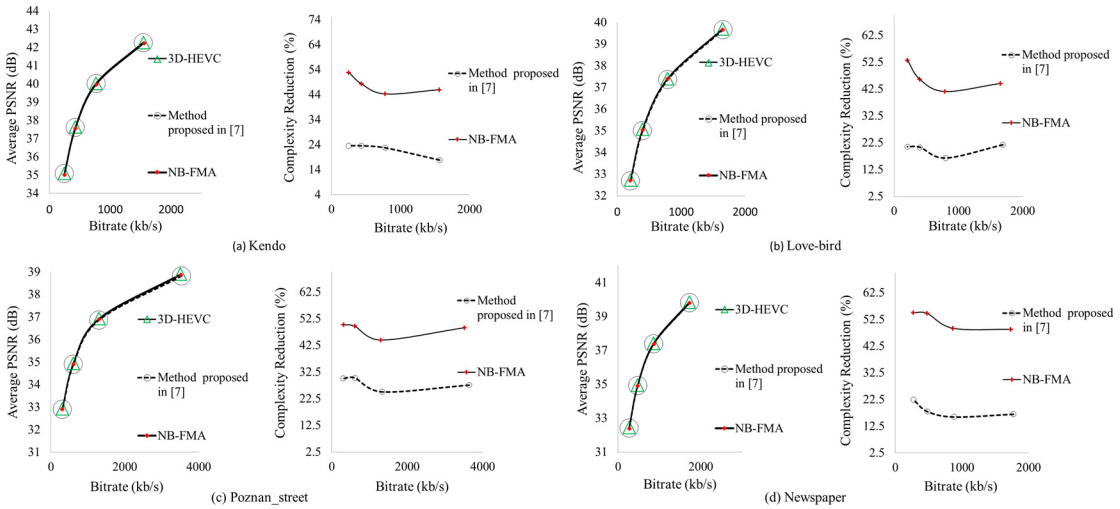

Fig. 4. Rate distortion and the complexity reduction comparison for four video streams

the depth map). Note that the reported bitrate is the bitrate of the 3D video stream, which includes the base view plus depth map and the dependent view plus depth map. As it can be observed from Fig. 4, our proposed scheme (NB-FMA) and the complexity reduction scheme proposed in [7] barely affect the bitrate of the streams.

Fig. 4 also illustrates the percentage of mode-search complexity reduction for each stream. In our study the complexity is computed based on the number of times the encoder searches for the best mode [7]. For example, for inter prediction, for every search point the complexity measure is equal to 1. As can be seen from the complexity curves in Fig. 4, our proposed scheme substantially reduces the computational cost without hampering the total bitrate.

Table I summarizes the effect of our scheme in terms of bitrate, PSNR, complexity, and the execution time for each stream. In our study, we used the Bugaboo Dell Xeon cluster from WestGrid, a high performance computing consortium in Western Canada [12]. A blade with an Intel Xeon X5650 6-core processor, running at 2.66GHz, and 8-GB RAM was used for the simulations. The execution time reductions reported here are for the dependent view. As it can be observed, our scheme outperforms the method proposed in [7], reducing the complexity by up to 51.95% (24.61% for [7]), or decreasing the execution time by up 44.28% (22.13% for [7]) while increasing the bitrate by a mere 2.06% (4.31% for [7]. In summary, these results show the superiority of our scheme over the method proposed in [7].

4. CONCLUSIONS

In this paper, we proposed a content adaptive complexity reduction scheme for 3D-HEVC. In our approach a fast mode decision scheme is employed using a Bayesian classifier to predict the block mode in the dependent view using information of already encoded neighboring blocks in the base and dependent views. Performance evaluations show that our approach significantly reduces the coding complexity of 3D-HEVC (up to 51.95%) while minimally hampering the overall bitrate.

5. REFERENCES

Table I. Impact of the proposed scheme on Bitrate, PSNR and Complexity

			Complexity reduction method proposed in [7]				NB-FMA			
Name	**Resolution, Frame Rate (fps)**	**Base & Dependent Views**	**Average PSNR Degrade (dB)**	**Average Bitrate Increase**	**Average Complexity Reduction**	**Execution Time Reduction**	**Average PSNR Degrade (dB)**	**Average Bitrate Increase**	**Average Complexity Reduction**	**Execution Time Reduction**
Kendo	1024x768, 30	V_3, V_5	0.005	2.01%	17.75%	15.11%	0.104	2.69%	47.40%	41.11%
Love-bird	1024x768, 30	V_6, V_8	0.003	2.21%	20.25%	18.84%	0.033	0.98%	46.43%	40.45%
Poznan street	1920x1088, 25	V_3, V_4	0.0002	4.36%	24.61%	22.93%	0.042	1.71%	48.38%	42.54%
Newspaper	1024x768, 30	V_4, V_6	0.019	3.21%	16.17%	14.56%	0.084	2.06%	51.95%	44.28%

[1] A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y.-K. Wang, Joint draft 8 of multiview video coding, Hannover, Germany, Joint Video Team (JVT) Doc. JVT-AB204, Jul. 2008.

[2] L. Shen, Z., Liu, P. An, R. Ma and Z. Zhang, ''Low-Complexity Mode Descion for MVC,'' IEEE Trans. on Circuite Systems and Video Technology, Vol. 21, No. 6,pp. 837-843, June 2011.

[3] Shen, Z. Liu, Tao Yan, Z. Zhang and P. An, ''View-Adaptive Motion Estimation and Disparity Estimation for Low Complexity Multiview Video Coding,'' IEEE Trans on Circuits System and Video Technology, Vol. 20(6), pp. 925-930, 2010..

[4] Q. Zhang, P. An, Y. Zhang, L. Shen, and Z. Zhang, ''Low Complexity Multiview Video Plus Depth Coding,'' IEEE Trans. on Consumer Electronics, Vol. 54 (4), pp. 1857 – 1865, Jan. 2012.

[5] G.J. Sullivan, J. Ohm, J; H. Woo-Jin, and T. Wiegand, ''Overview of the High Efficiency Video Coding (HEVC) Standard,'' IEEE Trans. on Circuits and Systems for Video Technology, Vol. 22(12), pp. 1649 – 1668, Dec. 2012.

[6] M. T. Pourazad, C. Doutre, M. Azimi, and P. Nasiopoulos, "HEVC: The New Gold Standard for Video Compression," IEEE Consumer Electronic Magazine, vol.1 , issue 3, pp. 36-46, July 2012.

[7] H.R. Tohidypour, M. T. Pourazad, P. Nasiopoulos, and V. Leung, ''A Content Adaptive Complexity Reduction Scheme for HEVC-Based 3D Video Coding,'' in Proc. of 18th International Conference on Digital Signal Processing (DSP), pp. 1-5, Santorini, Greece, 2013.

[8] ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11,"3D-HEVC Test Model 5," E1005, August 2013.

[9] K. P. Murphy, "Machine Learning: A Probabilistic Perspective," MIT Press, 2012.

[10] ISO/IEC JTC1/SC29/WG11,"Call for Proposals on 3D Video Coding Technology,'' N12036, March 2011.

[11] https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-DEV-2.0/cfg/3D-HEVC/

[12] https://www.westgrid.ca/