Fast Mode Assignment for Quality Scalable Extension of the High Efficiency Video Coding (HEVC) Standard: A Bayesian Approach

H.R. Tohidypour, H. Bashashati
Dep. of Elec.& Comp. Eng.
Univ. of British Columbia
Vancouver, Canada
{htohidyp, hosseinbs} @ece.ubc.ca

M. T. Pourazad
TELUS Communications
Univ. of British Columbia
Vancouver, Canada
pourazad@ece.ubc.ca

P. Nasiopoulos
Dep. of Elec.& Comp. Eng.
Univ. of British Columbia
Vancouver, Canada
panos@ece.ubc.ca

ABSTRACT
The new compression standard, known as the High Efficiency Video Coding (HEVC), aims at significantly improving the compression efficiency compared to previous standards. There has been significant interest in developing a scalable version of this standard. As expected, the HEVC scalable video version, which is called SHVC, increases the complexity of the codec compared to the non-scalable counterpart. In this paper, we propose an adaptive fast mode assigning method based on a Bayesian classifier that reduces SHVC's coding complexity by up to 68.55%, while maintaining the overall quality and bit-rates.

Categories and Subject Descriptors
I.5 [**Computing Methodologies**]: Pattern Recognition
I.2.10 [**Computing Methodologies**]: Vision and Scene Understanding

General Terms
Algorithms, Measurement, Performance, Standardization

Keywords
Scalable HEVC, Video compression, Low complexity compression, Machine learning.

1. INTRODUCTION
Digital media are becoming interwoven into the fabric of our daily lives at an exponential rate. This has been enabled by the availability of an ever-widening range of playback devices and social media applications. Playback devices come with their own specifications and limitations in terms of screen resolution, frame rate, processing power, battery life, and network requirements. Thus, to support such variety of heterogeneous devices with respect to the bandwidth of network, a video stream needs to be encoded at different quality levels, frame rates, resolution, etc.. This approach is computationally expensive and requires multiple channels and thus overall large amounts of bandwidth. The other approach is to use Scalable Video Coding (SVC), which enables multicast services and video transmission to heterogeneous clients with different capabilities [1]. An SVC stream consists of a base layer (BL) and one or more enhancement layers (ELs). On the decoder side, based on the type of the application and the supported complexity level, an appropriate part of the SVC bit stream will be decoded. A SVC stream may support temporal scalability (frame rate), spatial scalability (resolution), SNR/Quality/Fidelity scalability, or combined scalability (a combination of temporal, spatial and SNR scalabilities). While SVC enables delivery of different versions of the same video content within the same bit stream, it significantly increases coding complexity. The complexity of a scalable video encoder is mainly due to the additional temporal and spatial prediction processes involved when coding the multiple layers.

Recently, the Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) have introduced a new compression standard, known as the High Efficiency Video Coding (HEVC), which has substantially higher compression capabilities than H.264/AVC, the latest widely adopted video coding standard [2]. However, the computational complexity of HEVC is significantly higher due to its advanced features such as increased number of intra modes, and more flexible inter prediction. Due to HEVC's superior compression performance, there has been significant interest in developing a scalable version of this standard. To address the demand of industry, MPEG and ITU are currently working towards standardizing the scalable extension of HEVC, known as SHVC [3]. Considering that HEVC is already highly complex, its scalable implementation is expected to be significantly more complex than previous scalable standards. Thus one of the important factors in widespread adoption of this emerging standard is reducing its complexity so that it can be used for real-time applications.

In SHVC, inter and intra prediction mode-searches are some of the most computationally demanding steps, as the encoder needs to check all the possible/available modes to find the mode with the lowest rate distortion (RD) cost. To facilitate the mode-search process and reduce the coding complexity, in our previous work we proposed an early termination (ET) mode-search scheme for SNR/Quality/Fidelity scalable HEVC [4, 5]. This scheme utilizes the RD information of the base layer to predict the RD cost of the enhancement layer and terminate the mode-search process once the RD cost of the examined mode is equal or smaller than the predicted one. While this scheme is effective in reducing the complexity of coding process, it still requires performing mode-search over a reduced number of inter and intra prediction modes. In this paper, we propose a Bayesian classifier-based method to reduce the complexity of the SNR/Quality/Fidelity scalable HEVC. The proposed scheme instead of performing mode search, predicts the modes in the enhancement layer from the mode information of the already coded blocks. Here, Naive Bayes is used as the classifier and we assumed Dirichlet prior for our

Balkan Conference in Informatics (BCI) '13, Sep 19-21, 2013, Thessaloniki, Greece.

multinomial distributions. An experimental study was performed to confirm the efficiency of the proposed scheme and compare it with the unmodified SNR/Quality/Fidelity scalable HEVC.

The rest of this paper is organized as follows: Section 2 provides a short background, Section 3 elaborates on our proposed method, the performance evaluation of our method is presented in Section 4 and the conclusion is drawn in Section 5.

2. BACKGROUND

The High Efficiency Video Coding (HEVC) standard is the most recent joint video project of the ITU-T VCEG and the ISO/IEC MPEG standardization organizations, which has been finalized in January 2013. HEVC offers a substantially higher compression performance compared to the widely adopted H.264/AVC standard. Objective comparison results show that the current HEVC design outperforms H.264/AVC by 29.14% to 45.54% in terms of bit rate or 1.4 dB to 1.87dB in terms of PSNR [2]. Subjective comparisons of the quality of compressed videos – for the same (linearly interpolated) Mean Opinion Score (MOS) points also show that HEVC outperforms H.264/AVC, yielding average bitrate savings of 58% [6].

HEVC utilizes a quad-tree based coding structure with support for coding units of more diverse sizes than the macro-block sizes in H.264/AVC. The basic block in HEVC, known as the Largest Coding Unit (LCU), is 64x64 and can be re-cursively split into smaller Coding Units (CU), which in turn can be split into small Prediction Units (PU) and Transform Units (TU). HEVC employs more complicated intra prediction modes and more flexible motion compensation than H.264/AVC to reduce the spatial and temporal redundancies [2]. For intra prediction, HEVC uses 35 luma intra prediction modes compared to 9 used in H.264/AVC. Furthermore, intra prediction can be done at different block sizes, ranging from 4x4 to 64x64.

In the case of inter-prediction, for every inter-coded PU, the encoder can choose between 1) the motion merge mode, 2) the SKIP mode, or 3) explicit encoding of motion parameters. The motion merge mode involves creating a list of previously coded neighboring (spatially or temporally) PUs (called candidates) for the PU being encoded. The motion information for the current PU is copied from one selected candidate, avoiding the need to encode a motion vector for the PU; instead only the index of a candidate in the motion merge list is encoded as well as the residual data.

In the SKIP mode, the encoder signals the index of a motion merge candidate and the motion parameters for the current PU are copied from the selected candidate, without sending any residual data.

In explicit coding, inter-coded CUs can use Symmetric and Asymmetric Motion Partitions (AMP). AMPs allow for asymmetrical splitting of a CU into smaller PUs. AMP can be used on CUs of size 64x64 down to 16x16, improving coding efficiency since it allows PUs to more accurately conform to the shape of objects, without requiring further splitting [2]. Each inter-prediction coded PU, has a set of motion parameters, which consists of a motion vector, a reference picture index and a reference list flag.

Although HEVC's advanced coding tools such as increased number of intra modes and flexible inter prediction improve its coding performance, they also result in increased computational complexity. In the case of scalable extension of HEVC, where the base layer and multiple enhancement layers are required to be coded as a single stream, the increased computational complexity becomes one of the critical issues that need to be addressed.

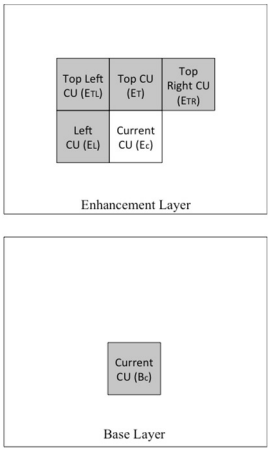

Figure 1. Current CU (white block) and its five predictors (Gray blocks).

3. OUR PROPOSED SCHEME

The focus of our study is to reduce the complexity of SNR/Quality scalable HEVC by utilizing the correlation between the base layer and enhancement layer. The HEVC standard utilizes several inter and intra prediction modes to achieve high compression performance. Considering in the inter/intra prediction mode-selection process, the encoder is required to calculate the Rate Distortion (RD) cost for each mode to find the best mode with minimum RD cost, mode decision is one of the most computationally involved procedures in a HEVC-based encoder. In the case of SHVC, since there is a high correlation between the base layer and the enhancement layer, the CU modes of the frames in the base layer can help us to speed up the process of selecting modes for the corresponding enhancement frames. Also, the modes of the already encoded neighboring CUs in the enhancement layer are valuable for predicting the mode of the current CU. Therefore the modes of the neighboring CUs in the enhancement layer and the corresponding CU in the base layer are used to predict the mode of the current CU. These five predictors are called predictor CUs hereafter. Figure 1 shows an example of the current CU, its corresponding base layer CU and its four neighbors. E_C indicates the current CU and E_L(left), E_{TL}(top left), E_T(top) and E_{TR}(top right) are its four spatial neighbors whose information is exploited to predict a mode for EC. The neighboring CUs in the enhancement layer are similar to the candidates that HEVC chooses for the merge mode.

The objective here is to implement a fast mode assigning (FMA) mode-search, so that the encoder does not need to go through all the modes, thus significantly reducing the computational complexity. To this end we approximate a function from the predictor CUs to the current CU or, equivalently, a posterior probability of the mode of current CU given the mode of its predictor CUs. The estimation of this posterior probability (or equivalently this function approximation problem) can be modeled as a supervised learning problem. This supervised learning problem consists of two stages; the first stage is the training process and the second stage is the test process. During the training process, the encoder encodes the BL using the unmodified SHVC [3]. The SHVC encoder, in the inter/intra prediction mode selection process, calculates the RD cost for each

mode and the one with minimum RD cost is selected. Then, the EL is also encoded using the conventional SHVC. In this process all of the available inter-intra modes are checked to find the lowest rate distortion cost. For each CU in the EL, the information about the mode chosen by the encoder is stored. Based on this information, the probability of each mode in the current CU in EL (given the predictors modes) is updated. These conditional probabilities will then be used in the test process. The second stage is the test process. In this stage, the information stored during encoding the BL and the previously encoded CU in the EL is used to estimate a mode for the to-be-encoded CU in the EL, without the encoder being required to check all of the inter-intra prediction modes.

As mentioned above, a probability can be assigned to each mode for the current CU given the modes of its predictor CUs. To define this posterior probability, different numbers are assigned to different HEVC modes (inter and intra modes), and each mode is considered as a class. Assume Y is the random variable corresponding to the mode of the current CU, and X is a random vector corresponding to the modes of its predictor CUs. In the case, where there is M different modes in HEVC, the random variable Y can have M different values. Regarding the predictor vector X, if there are L predictor CUs, the length of vector X will be equal to L and each of its components can take M possible discrete values (M possible modes). This results in M^L-1 different possible values for the random vector X. The posterior probability $P(Y|X)$, which in our case indicates the probability of the modes of the predictor CUs given each mode of the current CU in EL, can be 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 of the data that is expected to be seen in each class. The learning algorithm needs to estimate M-1 different parameters to estimate $P(Y)$, because the probability should sum to one. However, estimating $P(X|Y)$ requires learning of an exponential number of parameters, which is an intractable problem [7]. Thereby, the key to use Bayes rule is to specify a suitable model for $P(X|Y)$. In our study to solve the above-mentioned intractability problem, Naive Bayes classifier [7] has been used. 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 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,

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

According to the optimal Bayes decision rule [7], the mode of the posterior probability distribution is the predicted mode of the current CU. Therefore, for classifying a new X, the following formula can be used:

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

where y_m is the m^{th} possible value of Y. The normalization part, i.e., $P(X)$, of the posterior distribution has been omitted due to the fact that the denominator does not depend on y_m. The resulting y_m is the predicted mode for the current to be encoded CU.

To find the optimal value of y_m in equation 4, we need to have $P(X|Y)$ and $P(Y)$. These probabilities are computed during the training process. A very popular method to estimate these probabilities is the Maximum Likelihood Estimation (MLE) [7]. 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. Therefore, in this case the classifier overfits and will have problem during the test process [7]. To resolve this problem Maximum a Posteriori (MAP) [8] estimation is employed in this study. MAP estimate resolves the above-mentioned problem by incorporating a prior distribution over the parameter that we want to approximate. In order to facilitate MAP estimation, it is required to assign appropriate conjugate prior distribution for the parameters. Drichlet distribution [7] is chosen as the conjugate prior since the distribution of the posterior probability is a categorical (Multinomial) distribution,. As a result, the solution to the MAP estimate for $P(Y)$ is:

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

where α_k determines the strength of the prior assumptions relative to the observed data and M is equal to the number of different values which Y can take. The element N_k is the number of observed instances of class y_k. That is, N_k indicates the number of times the modes of the current CU is equal to y_k. On the other hand, the estimate 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 and M is equal to the number of distinct values which X_l can take. The element N_{lmk} denotes the number of times X_l=x_{lm} has been observed in the instances of class y_k [7]. That is, N_{lmk} indicates the number of times the modes of the l^{th} predictor is equal to x_{lm}, while the mode of the current CU is equal to y_k. To find the hyper parameters α_k and α_{lm}, five representative video sequences are used in our approach. These video sequences are different from the video sets used to test our approach.

To implement NB-FMA, the first second (exp., 30 frames if the frame rate is 30fps) of the video is coded based on non-modified SHVC. The coding information (modes) of these frames is used for the training process. During the training process the program updates the probabilities. Then, for finding the modes of the rest of the frames the fast mode assigning method is applied. In this stage, first the program encodes the BL. Then, the encoder starts encoding the EL. Unlike the training process, the encoder does not check all of the inter-intra prediction modes. Instead the information of the predictor CUs is used for predicting the mode of the to-be-encoded CU. In this study, the two mode candidates

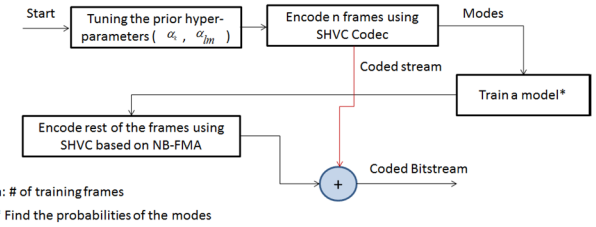

Figure 2. Block diagram of the proposed method.

Table 1. Impact of the proposed methods on Bitrate, PSNR and Complexity

Name	Resolution, Frame Rate (fps)	NB-FMA Method			ET-Method [4]		
		Average PSNR Degrade	Average Bitrate Increase	Average Complexity Reduction	Average PSNR Degrade	Average Bitrate Increase	Average Complexity Reduction
BasketballDrill	832x480, 50	0.081dB	0.98%	61.68%	.051dB	0.56%	33.73%
China Speed	1024x768, 30	0.173dB	0.95%	68.45%	0.165dB	0.72%	38.75%
Race Horse	832x480, 30	0.0189dB	0.94%	68.55%	.0097dB	0.70%	48.125%

with the highest probability among all modes are chosen, and the encoder calculates the RD cost for these two candidates and chooses the one with the smallest RD cost is selected. The block diagram of the overall method is shown in Figure 2.

4. RESULTS & DISCUSSION

In our experiment we used three test videos from the data set provided by MPEG for HEVC Call for proposals [6] (see Table 1). Our method was implemented into the SHVC software (SHM 1.0 [3]). The Random Access High Efficiency (RA-HE) configuration (hierarchical B pictures, GOP length 8, ALF, SAO, and RDOQ were enabled) was used in our study.

In our implementation we used one base layer (BL) and one enhancement layer (EL). The quantization parameters of BL and EL were set to four different values (QP_{B1}, QP_{E1})=(22,20), (QP_{B2}, QP_{E2})=(32,28), (QP_{B3}, QP_{E3})=(36,32) and (QP_{B4}, QP_{E4})=(40,36).

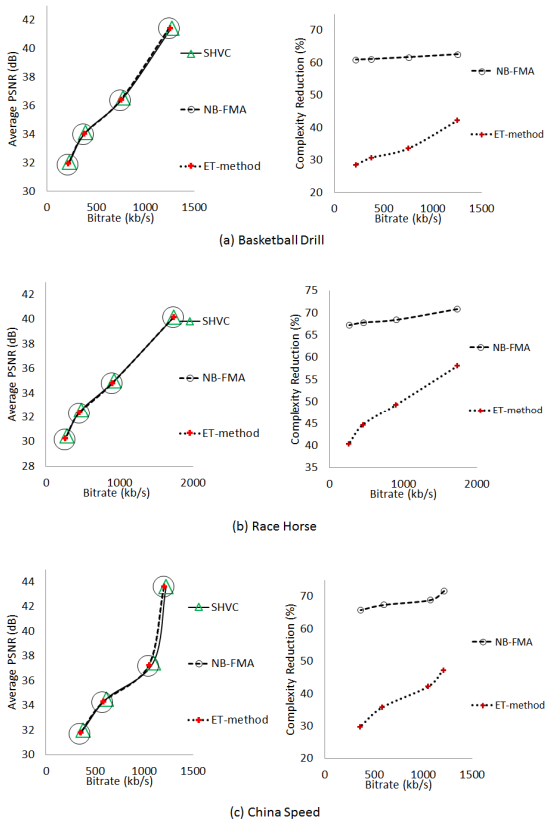

Figure 3. Rate Distortion curves and complexity reduction percentage for different video sequences (for NB-FMA (with two candidates) and ET-method)

We compute the complexity based on the number of times the encoder searches for the best mode. For example, for inter prediction, for every search point the complexity measure is equal to 1. For the skip mode, the complexity measure is equal to 1. For the Merge mode the complexity is up to 5, depending on the available candidates. By adding up these complexity values when coding the enhancement layer, we find the total complexity measure. For intra modes, we also compute the number of candidates which the encoder checks to encode a CU.

The performance of our method is compared with the performance of the presented ET method in [4]. Note that the adaptive search range method proposed in [4] was not applied for a fair comparison. Figure 3 shows the comparison of rate distortion (LHS column) using our NB-FMA approach, the original SHVC and the ET method proposed in [4] and the percentage of complexity reduction (RHS column) using our NB-FMA approach and the ET method compared to SHVC. As it can be seen from the complexity curves in this figure, our proposed scheme substantially reduces the computational complexity. Comparing the bitrate performance in Figure 3, we observe that all the RD curves overlap, an indication that our scheme does not affect the bitrate.

Table 1 summarizes the effect of our scheme in terms of bitrate, PSNR and complexity for each stream. As it can be observed, our scheme reduces the complexity up to 68.55% at a maximum cost of 0.95% bitrate increase. Comparison results show that our NB-FMA method outperforms the ET-method by more than 20% in terms of complexity reduction without affecting the PSNR and bit-rates significantly.

5. CONCLUSION

In this paper, we have proposed an adaptive Bayesian classifier-based mode-assigning method for reducing the complexity of scalable HEVC (SHVC). The results show that our method decreases the complexity of mode decision process by more than 61.68% without affecting the PSNR and bit-rates significantly. Our proposed method can be used in any HEVC SNR/Quality/Fidelity scalability implementation.

6. ACKNOWLEDGMENTS

This work was partly supported by Natural Sciences and Engineering Research Council of Canada (NSERC) and the Institute for Computing Information and Cognitive Systems (ICICS) at UBC.

7. REFERENCES

[1] Schwarz, A.H., Marpe, D., and Wiegand, T. 2007. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 17 (9), pp. 1103-1120, September 2007.

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

[3] Test Model for Scalable Extensions of High Efficiency Video Coding (HEVC). *ISO/IECJTC1/SC29/WG11*, m28348, January 2013.

[4] Tohidypour, H.R., Pourazad, M. T., Nasiopoulos, P. 2013. Content Adaptive Complexity Reduction Scheme for Quality/Fidelity Scalable HEVC. *International Conference on Acoustics, Speech, and Signal Processing*, May 2013.

[5] Tohidypour, H.R., Pourazad, M. T., Nasiopoulos, P. 2013. Content Adaptive Complexity Reduction Scheme For Quality/Fidelity Scalable HEVC. ISO/IECJTC1/SC29/WG11 L0042, January 2013.

[6] Joint Call for Proposals on Video Compression Technology. ISO/IECJTC1/SC29/WG11, N11113, January 2010.

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

[8] Bishop, C. 2006. Pattern Recognition and Machine Learning, Springer.