PROBABILISTIC GRAPHICAL MODEL BASED FAST HEVC INTER PREDICTION

*Meiyuan Fang, Jiangtao Wen**

TsingHua University
Beijing, 100084, China

Yuxing Han†

South China Agriculture University
Guangzhou, 510642, China

ABSTRACT

The High Efficiency Video Coding (HEVC) standard achieves 50% improvement in coding efficiency compared with H.264/ AVC by introducing many more video encoding tools achieving different coding performance and complexity tradeoffs [1]. Various techniques have been proposed to reduce the complexity of HEVC encoding. In this paper, an effective Bayesian network based complexity reduction framework for HEVC encoding is proposed. The proposed framework is able to calculate the conditional probabilities of different status in the encoding process, which can be used to speedup the encoders by neglecting small probability events. Experimental results show that the proposed algorithm can save 52.5% of computational complexity with a loss of 1.41% in compression performance on average.

***Index Terms*—** probabilistic graphical model, HEVC, video coding, complexity reduction

1. INTRODUCTION

Compared with H.264/AVC, the HEVC standard developed by ISO-IEC/MPEG and ITU-T/VCEG introduced video tools such as more flexible quadtree-based block partitioning, many more intra/inter prediction modes, sample adaptive offset (SAO), motion merge and etc. It achieves 50% improvement in coding efficiency with a significant increase in computational complexity, up to 500% in comparison with H.264/AVC encoders [2].

In this paper, we propose a probabilistic graphical model based fast HEVC encoding framework. A Bayesian network is parameterized by conditional probabilities calculated using values collected in the encoding process. Based on the probabilistic graphical model, fast HEVC encoding is achieved by skipping small probability events in RD optimized encoding. Experiments using the HEVC test clips and the HM reference software show an average of 52.5% reduction in encoding time with a loss of 1.41% in BD Rate [3].

The rest of the paper is organized as follows. We briefly review related work in Section 2. Section 3 describes the proposed framework. Experimental results using the HM are presented in Section 4. And Section 5 concludes this paper.

2. RELATED WORK

2.1. Complexity Analysis

Several analysis have been published on the complexity of HEVC encoding, by evaluating the relationship between Peak Signal-to-Noise Ratio (PSNR) based Rate-Distortion (RD) performance and encoding time. The RD performance is usually measured using the BD-rate or BD-BR. Vanne $et\ al.$ provided a comprehensive rate-distortion-complexity (RDC) comparison between the HM-6.0 Main Profile (MP) and the AVC reference software (JM-18.0) High Profile (HP)[4], and reported an average bit rate decrements of 23%, 35%, 40% and 35% for the All-Intra (AI), Random Access (RA), Low-delay B (LB) and Low-delay P (LP) encoding settings respectively, with a corresponding average computational complexity ratios of 3.2×, 1.2×, 1.5× and 1.3×. In addition to processing time, theoretical analysis and instruction-level profiling were conducted in [5]. Profiling of the HM encoder show that a significant amount of time (24.4%) was spent by the RD optimized quantization process in the AI configuration, while the calculation of the sum of absolute differences (SAD) and other distortion metrics occupied the most time (38.8%) in the RA configuration.

2.2. Complexity Reduction

The body of work on reducing the computational complexity of HEVC encoding generally focus on Mode Decision (MD) and Motion Estimation (ME), as well as parameter adjustments and simplifications of the RDO process.

A fast coding unit (CU) size decision algorithm for HM was proposed by Shen $et\ al.$ in [6]. It determined a CU depth range and skipped some specific depth levels rarely used in previously encoded frames or neighbor CUs. Experimental results show a reduction of around 40% encoding time with a negligible coding efficiency loss when implemented on HM-2.0 with low complexity profile.

[7] proposed a fast CU depth decision algorithm based on spatial-temporal correlations between neighboring coding tree units (CTUs). The method was able to save only about 25% of encoding time in HM-8.0.

Fast inter CU decision based on latent sum of absolute differences (SAD) estimation was proposed in [8] using a two-layer ME method. It calculates the SAD cost of the CU and the latent SAD cost of the sub-CUs. The algorithm then skipped unnecessary splittings according to a threshold calculated by an exponential model.

As for ME, the MV merging (MVM) heuristic algorithm in [9] reduced the inter-prediction complexity by merging N×N prediction unit (PU) partitions into larger ones. In the worst case, MVM reduced 34% of processing time with a loss of 0.08dB in PSNR and 1.9% increase in bitrate.

Two fast RDO techniques, top skip and early termination, were discussed in [10]. Based on the proposed algorithm, 45% of the encoding time for the HM-2.0 could be saved with a quality loss of 0.1dB in PSNR.

*This work was supported by the State Key Program of National Natural Science of China (Grant No. 61133008), the Natural Science Foundation of China (Grant No. 61521002) and Nanjing Yunyan Technology Limited.

†Yuxing Han is corresponding author.

3. PROPOSED FAST ENCODING FRAMEWORK

3.1. Motivation

Bayesian network is an important probabilistic graphical model for a wide spectrum of applications. It is capable of effectively model and infer the conditional dependencies between variables [11]. In this paper, we introduce Bayesian belief network based mining to HEVC encoding.

A Bayesian network is characterized by the structure of the network (nodes and edges in the graph) and the probabilistic distributions. It can be constructed in three main steps: data collection and pre-processing, followed by learning network structure and finally learning parameters of the probabilistic distributions.

3.2. Data collection and pre-processing

In this step, we prepare training data for a Bayesian Network. We select a subset of possible HEVC encoding parameters to be modeled by the Bayesian Network as shown in Fig. 1. Although introducing more parameters might in general improve the performance of the proposed framework, we included only the subset in the table in our preliminary implementation as these were the parameters that have the most significant impact on HEVC encoding performance and speed tradeoff. Further study of the optimal design of the parameters for the proposed Bayesian network based HEVC framework is warranted.

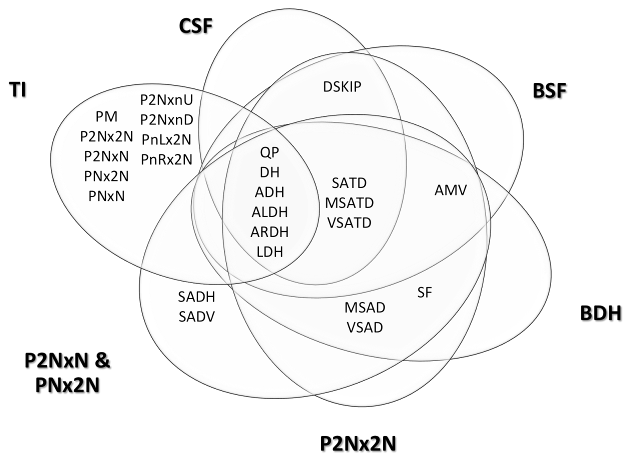

Fig. 1. Attributes for decision parameters

As shown in Fig. 1, the following information used in or produced by HEVC encoding information are also used:

1) QP, ranges from 1 to 51.

2) Current Depth (DH), ranges from 0 to 3.

3) Absolute value of motion vector

$$AMV = |MV_x| + |MV_y|, \tag{1}$$

where MV_x and MV_y are the motion vectors from the merge mode.

4) SAD of SKIP mode

$$DSKIP = Distortion_{SKIP}/(W \times H), \tag{2}$$

where W and H are width and height of the current CU, and

$$SAD = Distortion = \sum_{i=0}^{W}\sum_{j=0}^{H} |Org_{ij} - Rec_{ij}|, \tag{3}$$

with Org and Rec representing the original and reconstructed pixel respectively.

5) Mean SAD of inter $2N \times 2N$ mode $(D2N \times 2N)$

$$MSAD = D2N \times 2N = Distortion_{Inter2N\times 2N}/(W \times H). \tag{4}$$

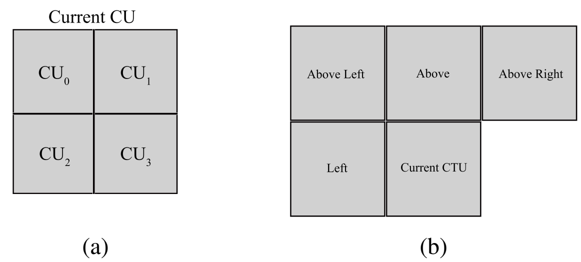

Fig. 2. Sub-CUs and neighbor CUs of the current CU.

6) Variance of SAD

$$VSAD = \sum_{i=0}^{3} (SAD_i/(W \times H) - MSAD)^2/4, \tag{5}$$

where i is the index of sub-CUs as shown in Fig. 2(a).

7) Sum of absolute transformed difference (SATD). Instead of the 16×16 CU based pre-compression analysis technique in [12], in this paper, the sum of the absolute values of the Hadamard coefficients h_{ij} of 8×8 blocks

$$SATD = \sum_{n=1}^{N}\sum_{i=0}^{7}\sum_{j=0}^{7} |h_{ij}| \tag{6}$$

is used, where N is the number of 8×8 CUs in the current CU.

8) Mean SATD

$$MSATD = \sum_{i=0}^{3} SATD_i/4, \tag{7}$$

where i is the index of sub-CUs in the current CU as shown in Fig. 2(a).

9) Variance of SATD

$$VSATD = \sum_{i=0}^{3} (SATD_i - MSATD)^2/4. \tag{8}$$

10) SAD horizontal

$$SADH = \begin{cases} \frac{SAD_0+SAD_1}{SAD_2+SAD_3} & if\ SAD_0 + SAD_1 < SAD_2 + SAD_3 \\ \frac{SAD_2+SAD_3}{SAD_0+SAD_1} & Otherwise \end{cases} \tag{9}$$

11) SAD vertical

$$SADV = \begin{cases} \frac{SAD_0+SAD_2}{SAD_1+SAD_3} & if\ SAD_0 + SAD_2 < SAD_1 + SAD_3 \\ \frac{SAD_1+SAD_3}{SAD_0+SAD_2} & Otherwise \end{cases} \tag{10}$$

12) Prediction mode (PM) of the current CU, can be either inter (0) and intra (1).

13) Transform index (TI) for the partition of transform unit (TU), ranges from 0 to 2.

14) SKIP flag (SF), indicating whether current CU is skipped or not.

15) Best is SKIP flag (BSF) indicating if the optimal partition mode is SKIP.

16) Best depth (BDH) of the current CU, ranges from 0 to 3.

17) CU split flag (CSF) indicating whether the current CU is further split.

18) Average depth of the neighboring CTUs is calculated using 4×4 CUs,

$$Avg_{Depth} = \sum_{i=1}^{N} Depth_i / N, \tag{11}$$

where N is the number of 4×4 CUs in the CTU. As shown in Fig. 2(b), in this paper, we use the depth information of the CTU above (ADH), above left (ALDH), above right (ARDH) and left (LDH).

19) Partition size information. P2N×2N, P2N×N, PN×2N, PN×N, P2N×nU, P2N×nD, PnL×2N, PnR×2N are binary variables representing whether corresponding partition is the optimal of the current CU.

The above data were collected using the first 150 frames of the HEVC Class D sequences, including BQSquare, RaceHorses, BasketballPass, and BlowingBubbles. The Main Profile (MP) of the HM 16.12 with the Random Access (RA) coding structure was used in our experiment.

3.3. Learning the Structure of the Bayesian Network

Bayesian network is a directed acyclic graph (DAG) that shows the relationship between interdependent variables in the network and a collection of conditional probabilistic distributions for each of the network variables. The model structure is trained so that it is consistent with the conditional dependencies from the observations. Since some of the variables used in our model as listed in the previous section are continuous-valued, a function converting continuous data to discrete values is used when establishing the structure of the network. In this paper, we quantized any continuous-valued variable to one of 10 equal-sized bins between its max and the min values. After pre-processing, we conducted training of the Bayesian network with a constraint-based approach. As described in [11], the base hypothesis denoted by H_0 states that the data were sampled from a distribution of random variables X, Y that are independent.

$$P^*(X, Y) \; = \; P^*(X) P^*(Y). \tag{12}$$

Since we have $P^*(X)$ and $P^*(Y)$, we use the best approximation $\hat{P}(X)$ and $\hat{P}(Y)$, thus

$$H_0 \; : \; P^*(X, Y) \; = \; \hat{P}(X)\hat{P}(Y). \tag{13}$$

We can devise a rule for testing whether we want to accept the hypothesis using:

$$R_{d,t}(D) = \begin{cases} Accept & d(D) \le t \\ \\ Reject & d(D) > t, \end{cases} \tag{14}$$

where we measure the deviance of the data from H_0 using χ^2 statistics

$$d(D) \; = \; \sum_{x,y} \frac{(M[x, y] \; - \; M \cdot \hat{P}(x) \cdot \hat{P}(y))^2}{M \cdot \hat{P}(x) \cdot \hat{P}(y)}, \tag{15}$$

where M is the number of samples, and we expect that $M[x, y]$ from the dataset to be close to $M \cdot \hat{P}(x) \cdot \hat{P}(y)$. We use the p-value to determines the false rejection probability of the decision rule:

$$p\text{-}value(t) \; = \; P(D : d(D) \; > \; t \mid H_0, M). \tag{16}$$

Setting the p-value at 0.05 in our experiments led to the Bayesian network as shown in Fig. 3.

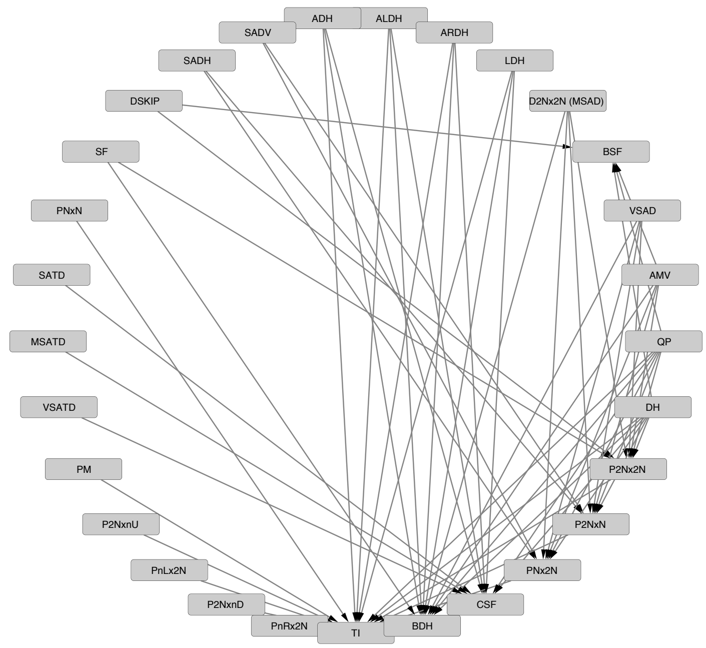

Fig. 3. Bayesian network proposal

3.4. Parameters Learning of the Bayesian Network

To properly model both discrete and continuous valued variables, we use a Gaussian Bayesian network, where the joint probabilitic distributions of variables in the network are multivariate Gaussian distributions. We denote nodes as $\mathbf{X} = \{X_1, X_2, ..., X_{29}\}$, which correspond to variables in Section 3.2. Assume that the set of parent nodes of node i is $\mathbf{P}_i \subseteq \mathbf{X}$, then we have parameters $\beta_0, ..., \beta_k$ and σ^2 such that:

$$p(X_i | \boldsymbol{P}_i) = N(\beta_0 + \sum_{y_j \in \mathbf{P}_i} \beta_j y_j ; \sigma^2), \quad j = 1, 2, ..., k. \tag{17}$$

The values $\beta_0, ..., \beta_k$ and the variance σ^2 for each variable X_i were the parameters to be estimated using Maximum likelihood Estimation (MLE) in our experiments. Let $\boldsymbol{\theta}$ be the set of all parameters to be estimated for each X_i,

$$\boldsymbol{\theta} = < \beta_0, ..., \beta_k, \sigma > . \tag{18}$$

To find the MLE values of θ, and based on the definition of the Gaussian distribution, we have:

$$\begin{aligned} \ell(\boldsymbol{\theta} : D) = \log L(\boldsymbol{\theta} : D) = \sum_m [-\frac{1}{2} \log{(2\pi\sigma^2)} \\ - \frac{1}{2\sigma^2} (\beta_0 + \beta_1 y_1[m] + \cdots + \beta_k y_k[m] - x[m])^2], \end{aligned} \tag{19}$$

where m denotes the indices of the instances in our experiments. We obtain $\boldsymbol{\theta}$ by solving

$$\frac{\partial}{\partial \beta_i} \ell(\boldsymbol{\theta} : D) = 0, \qquad i = 0, 1, 2, ..., k. \tag{20}$$

3.5. Fast HEVC Encoding based on Bayesian Network

Based on the Bayesian network and parameters from Section 3.2 and 3.3, we can calculate the conditional probabilities for different status in the encoding process using (17) and neglect the events with probabilities smaller than a threshold. Algorithm 1 is a pseudo code for the proposed algorithm, and the thresholds used in our experiment are given in Table 1.

Algorithm 1 The proposed fast encoding algorithm

```
1: for Each CU do
2:   if p(CSF=true|QP,DH,ADH,ALDH,ARDH,LDH,SATD,MSATD,
     VSATD)-p(CSF=false|QP,DH,ADH,ALDH,ARDH,LDH,SATD,
     MSATD,VSATD)>TH_SA && Current Depth < MAX_DEPTH then
3:       Go to 20;
4:   Check SKIP mode;
5:   if p(BSF=true|QP,DH,AMV,DSKIP)>TH_SF && SF=true then
6:       Go to 17;
7:   Check inter 2N×2N mode;
8:   if p(BDH=Current depth|QP,DH,ADH,ALDH,ARDH,LDH,SF,
     MSAD,VSAD,AMV)>TH_BDH && Current Depth < MAX_DEPTH
     then
9:       Go to 20;
10:  if p(P2N×2N=false|QP,DH,SF,MSAD,VSAD,AMV,DSKIP)
     ≤TH_PS then
11:      Go to 17;
12:  if p(P2N×N=true|QP,DH,MSAD,VSAD,AMV,SADV,SADH)
     >TH_PS_H then
13:      Check inter 2N×N mode;
14:  if p(PN×2N=true|QP,DH,MSAD,VSAD,AMV,SADV,SADH)
     >TH_PS_V then
15:      Check inter N×2N mode;
16:  Check inter AMP mode;
17:  Check intra mode;
18:  if Current Depth=MAX_DEPTH then
19:      Go to 21;
20:  Split the CU into four sub-CUs, Depth++, go to 1;
21:  return
```

In addition, for each TI used in RDO process, we will stop TU partitioning when the following condition is satisfied.

$$p(TI \mid QP, DH, PM, ADH, ALDH, ARDH, LDH,\\ P2N\times2N, P2N\times N, PN\times2N, PN\times N, P2N\times nU,\\ P2N\times nD, PnL\times2N, PnR\times2N) \; > \; TH\_TI \qquad (21)$$

As the overall RD performance of HEVC encoding is impacted by a large number of factors, some suboptimal decisions do not affect R-D performance significantly. This is shown as the the incorrect rate in Table 1.

Table 1. Characteristics of the decision

Decision	Threshold	Value	Incorrect Rate
CSF	TH_SA	0.8	0.4%
BSF	TH_SF	0.96	5.3%
BDH	TH_BDH	0.9	8.3%
$P2N\times2N$	TH_PS	0.01	0.1%
$P2N\times N$	TH_PS_H	0.1	1.9%
$PN\times2N$	TH_PS_V	0.3	0.1%
TI	TH_TI	0.99	12.9%

4. EXPERIMENTAL RESULTS

We tested our proposed algorithm using a server with a 12-core Intel Xeon E5-2697 v2 2.70GHz processor and 64GB of memory. We implemented our algorithm in the HEVC reference software HM-16.12. Video sequences specified in [13] were used with the RA MP configuration. We used fixed QP mode with QP values of 22, 27, 32 and 37. To measure the reduction in total encoding time, time saving (TS) is defined as:

$$TS = \frac{Enc.time(Anchor) - Enc.time(Prop.)}{Enc.time(Anchor)} \times 100\%. \quad (22)$$

BD-rate was used as the metric for RD performance.

Table 2. Performance comparison of proposed method with recent work (RA MP)

Sequences	Shen[14] BDBR (%)	Shen[14] TS (%)	Xiong[8] BDBR (%)	Xiong[8] TS (%)	Proposed BDBR (%)	Proposed TS (%)
Traffic	2.05	51.4	2.20	69.5	2.05	58.7
PeopleOnStreet	2.47	39.0	1.47	46.3	-1.38	43.7
BasketballDrive	6.62	48.7	3.82	61.3	2.13	47.7
Cactus	5.87	47.6	2.4	62.4	1.55	56.3
Kimono	4.67	39.1	4.93	57.8	2.49	52.5
ParkScene	1.86	44.2	2.05	64.8	2.17	57.0
BQMall	6.11	45.5	1.88	58.3	1.26	53.9
BasketballDrill	4.76	41.5	1.40	52.1	1.13	51.0
PartyScene	1.58	36.4	2.06	44.2	1.48	56.7
RaceHorsesC	3.2	32.1	2.09	47.0	1.57	46.3
BQSquare*	0.51	37.1	1.58	58.8	1.70	60.6
BasketballPass*	2.99	38.9	0.60	59.7	1.14	48.0
BlowingBubbles*	0.58	27.6	2.67	50.1	1.31	54.5
RaceHorses*	1.02	23.3	1.56	34.2	1.20	47.5
Average	3.16	39.5	2.19	54.8	1.41	52.5

* used as training data in our experiments
Results for comparison are all copied from [8].

Data used in Section 3.2 and 3.3 were obtained through offline encodings of following 4 video sequences: BQSquare, RaceHorses, BasketballPass, and BlowingBubbles. Table 2 shows the results including training data, which are mentioned as *. As can be seen in the table, time saving was 52.5% on average along with loss of 1.41% in BD-rate by skipping events with low probability based on the Bayesian network. For the PeopleOnStreet sequence, we achieved 43.7% time saving with a gain of 1.38% in BD-BR. In addition, for sequences excluding training data, 52.4% of complexity reduction with a compression efficiency loss of 1.45% was achieved.

5. CONCLUSIONS

In this paper, we propose a novel fast encoding framework for the HEVC standard. The proposed framework uses a Bayesian network trained offline to model the conditional probabilities between events, parameters and decisions in the encoding process. Experiments with the HEVC HM reference software and the HEVC common test condition clips showed an average saving of 52.5% encoding time with about 1.41% performance loss in comparison with the HM-16.12 anchor.

6. REFERENCES

[1] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec 2012.

[2] G. Correa, P. Assuncao, L. Agostini, and L. A. da Silva Cruz, "Performance and Computational Complexity Assessment of High-Efficiency Video Encoders," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1899–1909, Dec 2012.

[3] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," Document VCEG-M33, Apr. 2001.

[4] J. Vanne, M. Viitanen, T. D. Hamalainen, and A. Hallapuro, "Comparative Rate-Distortion-Complexity Analysis of HEVC and AVC Video Codecs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1885–1898, Dec 2012.

[5] F. Bossen, B. Bross, K. Suhring, and D. Flynn, "HEVC Complexity and Implementation Analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1685–1696, Dec 2012.

[6] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, "An Effective CU Size Decision Method for HEVC Encoders," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 465–470, Feb 2013.

[7] Y. Zhang, H. Wang, and Z. Li, "Fast Coding Unit Depth Decision Algorithm for Interframe Coding in HEVC," in *Data Compression Conference (DCC), 2013*, March 2013, pp. 53–62.

[8] J. Xiong, H. Li, F. Meng, Q. Wu, and K. N. Ngan, "Fast HEVC Inter CU Decision Based on Latent SAD Estimation," *IEEE Transactions on Multimedia*, vol. 17, no. 12, pp. 2147–2159, Dec 2015.

[9] F. Sampaio, S. Bampi, M. Grellert, L. Agostini, and J. Mattos, "Motion Vectors Merging: Low Complexity Prediction Unit Decision Heuristic for the Inter-prediction of HEVC Encoders," in *2012 IEEE International Conference on Multimedia and Expo*, July 2012, pp. 657–662.

[10] M. B. Cassa, M. Naccari, and F. Pereira, "Fast rate distortion optimization for the emerging HEVC standard," in *Picture Coding Symposium (PCS), 2012*, May 2012, pp. 493–496.

[11] Daphne Koller and Nir Friedman, *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*, The MIT Press, 2009.

[12] J. Wen, M. Fang, M. Tang, and K. Wu, "R-(lambda) Model Based Improved Rate Control for HEVC with Pre-Encoding," in *2015 Data Compression Conference*, April 2015, pp. 53–62.

[13] F. Bossen, "Common test conditions and software reference configurations," Document JCTVC-L1100, Jan. 2013.

[14] L. Shen, Z. Zhang, and Z. Liu, "Adaptive Inter-Mode Decision for HEVC Jointly Utilizing Inter-Level and Spatiotemporal Correlations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 10, pp. 1709–1722, Oct 2014.