A NOVEL SATD BASED FAST INTRA PREDICTION FOR HEVC

Jiawen Gu, Minhao Tang, Jiangtao Wen

Tsinghua University
Department of Computer Science
Beijing 100084, China

Hao Zhang

School of Information Science and Engineering
Central South University
Changsha, Hunan 410083, China

ABSTRACT

To better exploit spatial correlations in a video frame, the HEVC video coding standard has introduced many intra prediction modes and a recursive quadtree-based coding unit (CU) structure. As a result, the complexity of Rate-distortion optimized (RDO) HEVC intra mode selection is significantly higher. Many techniques have been proposed to expedite the intra mode selection process to achieve a good overall trade-off between complexity and RD performance. In this paper, we proposed a fast intra decision algorithm consisting of three parts: calculation reduction, adaptive intra candidate selection, and fast depth decision used early termination. Experiments conducted using the HEVC common test conditions show an average of 61.1% (up to 67.7%) time saving with only 1.03% increase in Bjontegaard delta rate (BD-rate) using the proposed algorithm, out-performing existing state-of-the-art algorithms.

***Index Terms*—** HEVC, SATD, Intra Prediction

1. INTRODUCTION

The High Efficiency Video Coding (HEVC) standard jointly developed by IOS/IEC MPEG and ITU-T is a successor to H.264/AVC and achieves about 50% bitrate reduction [1][2] but also lead to tremendous time cost. Many algorithm are proposed to accelerate HEVC such as fast transcoding [3][4], parallel optimization [5]. For intra prediction, HEVC introduced many more spatial prediction modes/tools as compared with H.264/AVC, supporting up to 35 prediction modes and coding uint sizes from 4x4 to 64x64. As a result, the complexity for HEVC intra prediction is significantly higher than for H.264/AVC. To simplify RDO intra prediction mode selection, the HEVC reference software HM adopts several fast algorithms: the rough mode decision (RMD) [6] based on the Sum of Absolute Transform Distortion (SATD) is utilized to eliminate some modes before RDO calculations. Then, considering the high correlation among the neighbors, modes derived from the neighbor CUs, termed the most probable modes (MPMs) [7] are added to the candidate list. After that, RDO quantization (RDOQ) will be applied. Finally the optimal residual quadtree-based TU size (RQT) of the best mode will be selected. However, even with the introduction of such fast algorithms, Intra mode decision in HM still needs to recursively go through all depths, and is still highly time consuming and not suitable for real world applications.

As discussed above, the overall intra prediction process contains best mode selection at each CU depth and CU size decision. In this paper, we proposed fast intra mode decision techniques for both best mode selection and CU size decision. We combined the distribution of SATD values and the correlations between the parent and current CUs for fast mode decision and adaptively chose the mode candidates to reduce RDO calculations. For CU size decision, a fine-grained early termination method based on SATD and RD-Cost estimation were used to avoid unnecessary further splitting and associated calculations. The rest of paper is organized as follows, In Section II, we briefly review existing techniques for fast intra mode decision. The proposed fast intra mode decision algorithm are described in Section III, with experimental results in Section IV and conclusions in Section V.

2. A REVIEW OF FAST INTRA MODE DECISION ALGORITHMS

Fast intra mode decision algorithms for HEVC can be roughly categorized into three main categories, fast intra mode decision, fast RQT and fast CU/TU size decision.

2.1. Fast intra mode decision

Intra mode decision for HEVC essentially consists of four steps, as describe in Section I, each subject to optimization.

Both rough mode decision (RMD) and MPM exploit information from spatial neighbors. Cheng *et al* [8] proposed to use the best mode of the parent CU to reduce the range for rough mode decision and saved 45.2% encoding time in HM2.0 with 0.05dB PSNR loss and a bit rate increases of 0.10%. While in [9], Ismail *et al* used the best and second best modes of the neighbor CU to reduce SATD cost calcualtions and saved 28% of encoding time with 1.7% BD-rate increase.

This work was supported by the Natural Science Foundation of China (Project Number 61521002) and Nanjing Yunyan Technology Inc.

RDO and RDOQ are two of the most time consuming steps in intra prediction mode decision. Few optimization techniques for RDO and RDOQ have been proposed that can achieve good speedup while avoiding severe overall coding efficiency reduction in the mean time. Zhao *et al*. [10] studied the impact of the number of mode candidates after RMD in HM. They proposed a method for reducing the number of candidates in the full RDOQ process, and achieved 20% time saving with 0.12% average increase of BD-rates.

RQT is applied to further exploit spatial correlation and improve coding efficiency [1] with recursive transform unit (TU) splitting. The computational complexity is high. Tan *et al*. [11] proposed fast RQT algorithms for both intra and inter coding so as to reduce the complexity. Their techniques for fast RQT saved 13% time with 0.1% BD-rate increase.

2.2. Fast CU size decision

Fast CU size decision have been widely investigated. For example, Shen *et al*. [12] used the depth of neighbor CUs to adjust current CU depth and saved 21% time with 1.74% BD-rate increase. Early termination is also an important technique for expediting CU size decision. Lin *et al*. [13] proposed an early termination algorithm based on Sobel edge density of the coding tree block (CTB). The time saving was about 14.2% with 0.02% BD-Rate increase. Hao *et al*. [14] used the Hadamard costs to predict the RD cost of four sub-CUs and sub-TUs. The predicted cost was then compared with the RD cost of its parent CU to decide whether the CU split is necessary. They achieved 32% encoding time reduction with 1.1% BD-rate increase. Zhang *et al*. [15] select SATDs of the first several units as references to analyze the early pruning and early splitting thresholds and achieve by 44% saving on average with only 0.9% BD rate increment.

3. NOVEL SATD BASED FAST INTRA PREDICTION

Among the techniques we proposed in this paper, SATD based early-termination reduces calculations by utilizing the quadtree structure and exploiting the check order in RDO. For the sake of clarity, we introduce two indexing methods for LCU in this paper. One is geometric index as shown in Fig. 1, which contains both depth and location information. The other is z-order index, between 0 to 63 in z-order, each representing a CU of the size of 8×8.

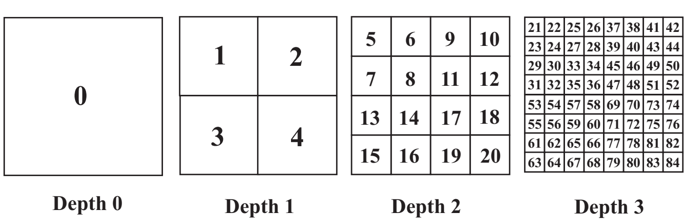

Fig. 1. The geometric index of CU

3.1. A Novel Fast Depth Decision method

Intra prediction searches the best mode using a top-down approach. Conventionally, the RD cost of the current CU without split (referred to as intra 2Nx2N) will be calculated first. Then it is required that encoding for all sub-CUs be finished with the total RD cost found and the split flag set. The RD costs of the intra 2Nx2N and the total RD cost for the sub-CUs are then compared with the encoding option corresponding to the lower of the two costs chosen for subsequent processing. It is easy to see that the complexity of the above process could be reduced if we can estimated the total RD cost before splitting of all sub-CUs. A simple approach to such early termination is to check adjacent depth, such as the algorithm in [16]. In fact, after processing of one CU at a certain depth, all upper depths could be checked. The idea can be summarized in the following formula:

$$\hat{J}_d > \beta_{\Delta d,k} J_d, (\Delta d = 1, 2, ..., 3 - d), \tag{1}$$

where J_d and $\hat{J}_d$ denote the real best and the estimated RD costs at depth d, $\beta_{\Delta d,k}$ is a penalty factor, Δd represents the difference between the current and predicted depths, k denotes the location. It should be noted that $\hat{J}_d$ can be estimated in any depth deeper than d. For instance, the CU of the 67-th geometric index can be used to estimate the RD costs of CUs of the 16-th, the 3-th and the 0-th geometric index at corresponding depth, as shown in Fig. 2.

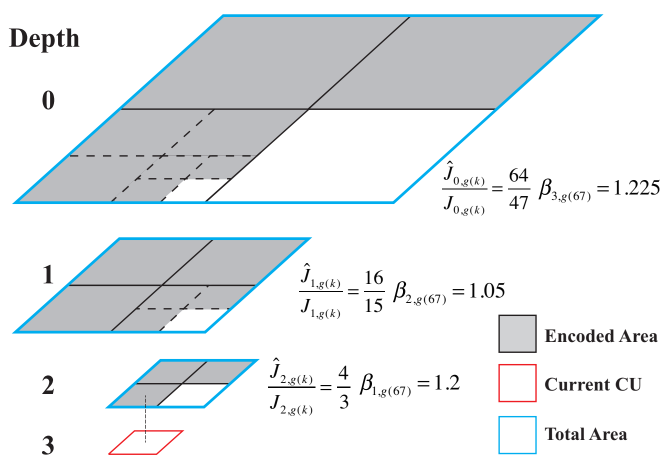

Fig. 2. Illustrative Example for SATD based early termination at 67-th geometric index

Numerous experiments have shown that SATD has a high correlation with RD cost. As is the case for many HEVC encoders, in our algorithm, we also use the SATD values as an estimate for the RD costs in early termination of CU split decisions. Let $\hat{H}_{d,g(i)}$ denote the aggregated SATD cost at depth d at the i-th geometric CU, with the corresponding total SATD at depth d denoted as $H_{d,g(i)}$. Since SATD cost is obtained by summing 8×8 units in the current CU, it can be stored in z-order according to its location. Hence

$$H_{d,g(i)} = \sum_{j=start}^{start-1+64/4^{d(g(i))}} had_{d,z(j)}, \tag{2}$$

and

$$start = \{g(i) - geom_{start}(d[g(i)])\} \times 64/4^{dg(i)}, \tag{3}$$

where $z(j)$ is the $j\text{-}th$ index in z-order, $had_{d,z(j)}$ is the SATD cost of the $j\text{-}th$ z-order index at depth d, $d(g(i))$ is the depth of the $i\text{-}th$ geometric CU, and $geom_{start} = \{0, 1, 5, 21\}$ for each depth.

Similarly, $\hat{S}_{d,g(i)}$ represents the aggregated encoded area at depth d in the $i\text{-}th$ geometric CU, while $S_{d,g(i)}$ denotes the total area at depth d of the $i\text{-}th$ geometric CU. Then the RD cost can be estimated using

$$\hat{J}_d = (min\{\frac{S_{D,g(k)}}{\hat{S}_{D,g(k)}}, \frac{H_{D,g(k)}}{\hat{H}_{D,g(k)}}\}) \cdot J_{D,a}, (D = d, ..., 3), \tag{4}$$

where the $J_{D,a}$ is the aggregated RD cost at its depth. Combining (1) - (Eq.4), we have

$$(min\{\frac{S_{D,g(k)}}{\hat{S}_{D,g(k)}}, \frac{H_{d,g(k)}}{\hat{H}_{d,g(k)}}\}) \cdot J_{D,a} > \beta_{\Delta d,k} J_d, (D = d, ..., 3) \tag{5}$$

where $\Delta d = D - d$. If (5) holds for any k, CU split is terminated with depth d is chosen as the best depth. In the formula above, $\beta_{\Delta d,k}$ is a penalty factor. Because video may have various textures, there is higher risk to skip at the first few CUs than subsequent CUs. These penalty factors were chosen in order of descending empirically as given in Fig.3.

2.2	**1.6**
1.2	**1**

△**d=1**

3.2	**2.8**	**2.0**	**1.8**
2.5	**2.2**	**1.7**	**1.6**
1.5	**1.3**	**1.15**	**1.10**
1.3	**1.2**	**1.05**	**1**

△**d=2**

3.8	**3.6**	**3.1**	**3.0**	**2.15**	**2.10**	**1.95**	**1.90**
3.4	**3.2**	**2.9**	**2.8**	**2.05**	**2.0**	**1.85**	**1.80**
2.725	**2.65**	**2.425**	**2.35**	**1.775**	**1.75**	**1.675**	**1.65**
2.575	**2.5**	**2.275**	**2.2**	**1.725**	**1.70**	**1.625**	**1.6**
1.575	**1.55**	**1.475**	**1.45**	**1.18**	**1.17**	**1.14**	**1.125**
1.525	**1.5**	**1.425**	**1.4**	**1.16**	**1.15**	**1.11**	**1.1**
1.375	**1.35**	**1.275**	**1.25**	**1.085**	**1.07**	**1.03**	**1.02**
1.325	**1.3**	**1.225**	**1.2**	**1.06**	**1.05**	**1.01**	**1**

△**d=3**

Fig. 3. The penalty factors

Fig.2 shows an example of early termination process of the CU with the $67\text{-}th$ geometric index. For this CU, every depth smaller than the current CU will be checked. If Equ. 5 is satisfied at depth d, the remaining CUs will be skipped and intra 2Nx2N at depth d will be chosen as the best mode. Meanwhile, if this formula holds for several depths, the higher depth is chosen.

3.2. SATD calculation reduction

Many fast motion estimation algorithms for Inter prediction, such as diamond search, 3-step search, use a rough-search-followed-by-refinement approach. In this paper, we introduce the same concept to intra mode decision in HEVC. HEVC intra prediction methods can be classified into two categories. Angular prediction methods for edges and textures, and PLANAR and DC prediction for smooth areas. HEVC uses 33 angular intra prediction modes, along with DC and PLANAR modes. For convenience, we use mode 0 and mode 1 to refer the PLANAR and DC mode respectively. The remaining 33 angular prediction are indexed from 2 to 34. As mentioned above, the SATD of all 35 modes will be calculated before RDO to screen out unlikely modes, with full RDO for the remaining modes. The size of the subset of candidates for full RDO is $8, 8, 3, 3, 3$ for CU size 4×4, 8×8, 16×16, 32×32 and 64×64 respectively.

Our algorithm for finding local optimum without traversing all modes is based on several observations. The first is the high correlation between the optimal modes of the current and parent CU, which has been used in papers such as [8] and [12], where the best prediction mode of parent CUs in the upper depths are used for narrowing down the range of predict directions. The other observation is about SATD. Several observations are concluded as below: One is that in many cases, the predicted sample for many mode are very close, leading to nearly identical SATD values. Then, while the SATD values for many modes may be similar, the optimal mode is often one that produces a significantly different SATD value. Further analysis found that nearly identical SATD values usually occur for consecutive modes. Based on these observations, we proposed a fast algorithm as follows:

1. Calculate the SATDs of PLANAR and DC mode.
2. If current depth is 0 or the best mode of its parent CU is 0 or 1, then we check $2 + 4 \cdot \delta, 0 \leq \delta \leq 8$. Otherwise we use the best mode i of its parent CU as the search center. The SATD of mode j will be checked if $j > 1$ and $i \equiv j \pmod 4$.
3. Find the best six modes among candidates , then calculate the SATD of two-distance neighbors.
4. Find the best two modes from the previous step, and calculate the SATD of one-distance neighbors. If mode m and mode n share the same SATDs and $|m-n| = 2$, then skip the one-distance mode.
5. Add MPMs of the current CU. Choose best M candidate modes from all the calculated modes, where $M \in (8, 8, 3, 3, 3)$

3.3. Adaptive candidate selection

Experiments shows that the algorithm described in the previous section can reduce close to 55% SATD calculations, as shown in Table. 1. However, compared with full RDO, the time used for SATD calculations is relatively small. As we

Table 1. Reduction of SATD calculations (%)

Video \ Depth	0	1	2	3	4
NebutaFestival	55.93	54.70	54.09	54.21	54.38
ParkScene	55.46	54.25	54.21	54.81	55.11
RaceHorses	55.11	53.72	53.55	54.53	54.74
Bubbles	56.21	54.30	53.36	53.58	53.70
FourPeople	53.51	52.88	53.31	53.87	53.93
Average	55.24	53.97	53.71	54.20	54.37

use the same number of effective candidates as the HM, these mode will be evaluated by further RDO and RQT to arrive at the best mode. The calculations in this process can be adaptively reduced to further reduce the overall complexity using three insights. First, there is a high probability that local optimal mode is the best mode in its 2-neighborhood, i.e. if a mode i has the lowest SATD value among modes from $i-2$ to $i+2$ and is chosen as one of the candidates, the probability that mode i is the best mode is vey high. Second, it has been shown in [10] that there is a high probability that the best mode chosen by the RDO corresponds to a SATD value that is among the lowest 3 for all modes in the candidate list. Finally, as the SATD of the best mode is usually unique, it is possible to skip RDO if repetition is detected. Based on these insights, an adaptive candidate selection algorithm can be designed using the following rules:

1. If current mode shares the same SATD value with the calculated mode, skip.
2. If the index of current mode (i-th) is closed to the calculated mode (j-th), i.e, $|i-j| \leq 2$, and the SATD of current mode is higher, then skip.
3. If the SATD value is more than 1.5 times the best SATD, skip.

4. EXPERIMENTS

We implemented our techniques in HEVC reference software HM16.12. Tests were conducted using standard test clips (classes A to E) in the HEVC common test conditions. All intra (AI) configuration of the HEVC Main Profile was used with 64x64 LCU for $QP \in \{22, 27, 32, 37\}$. The BD-Rate and BD-PSNR results [17] are given in Table 1. In the table, overall refers to the overall time saving of the algorithm, where the time saving is calculated by $Speedup = \sum_{qp\in\{22,27,32,37\}} \frac{T_{HM}-T_{proposed}}{4\cdot T_{HM}}$. where the T_{HM} and $T_{proposed}$ are denoted as the encoding time of HM and proposed method respectively. Table.2 shows the overall results of proposed algorithms, which are SATD based Early termination (refers to ET), SATD reduction and Adaptive candidate selection (refer to RDO). On average, our proposed algorithm achieves 60.23% encoding time reduction for all intra coding with less than 1.03% BD-Rate increase and 0.07dB BD-PSNR decrease.

Table 2. The performance of proposed algorithm

Class	Video	ΔP(dB)	ΔR(%)	ΔT
A	NebutaFestival	-0.016	0.252	61.9%
A	PeopleOnStreet	-0.053	0.970	63.2%
A	SteamLocomotiveTrain	-0.026	0.847	68.4%
A	Traffic	-0.047	0.960	62.1%
B	BasketballDrive	-0.036	1.269	64.8%
B	BQTerrace	-0.049	1.015	61.9%
B	Cactus	-0.034	0.994	60.9%
B	Kimono	-0.023	0.682	66.7%
B	ParkScene	-0.024	0.565	59.8%
C	BasketballDrill	-0.071	1.512	58.2%
C	BQMall	-0.059	1.100	60.5%
C	PartyScene	-0.051	0.742	54.0%
C	RaceHorses	-0.046	0.829	59.4%
D	BasketballPass	-0.066	1.128	59.8%
D	BlowingBubbles	-0.055	0.850	52.4%
D	BQSquare	-0.069	0.928	53.6%
D	RaceHorses	-0.057	0.993	56.6%
E	FourPeople	-0.067	1.209	64.5%
E	Johnny	-0.080	1.974	67.2%
E	KristenAndSara	-0.087	1.761	67.0%
	Average	-0.051	1.029	61.1%

Table 3 compares the proposed work with state-of-the-arts. Compared with the works [18] - [19], this work reduces more computation time with the better coding performance. In addition, the parameter of penalty factor can be further refined to achieve the further improvement.

Table 3. The performance of proposed algorithm

	ICIP'14 [18]	ICIP'15 [20]	TIP'14 [21]	ICME'16 [19]	proposed
BD-BR(%)	1	0.66	1.08	1.4	1.03
BD-PR(dB)	-	0.03	-	-	0.05
ΔTime(%)	54	37.91	47	60.98	61.1

5. CONCLUSION

In this paper, we proposed a fast intra mode selection algorithm that utilizes SATD cost distribution to adaptively skip SATD calculations and expedite the RDO process. We further used the SATD for estimating the RD cost for early termination. Two main contribution are described in this paper. The first is SATD based early termination. It utilizes features of the quadtree structure and SATD values to estimate the RD cost for early termination decision. The second is based on using the distribution of SATD cost and the correlation of parent and child CUs. The distribution provided a high-accuracy prediction for best mode selection, an observation not widely used in other algorithms. Compared with the HM16.12, the proposed algorithms achieved a good tradeoff between compression performance and encoding time compare to the other intra encoding algorithms. Experimental results show that the proposed algorithms achieved up to 67.7 % time saving with only 1.03% increase in BD-Rate compared to the HM16.12.

6. REFERENCES

[1] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas 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, 2012.

[2] Jens-Rainer Ohm, Gary J Sullivan, Heiko Schwarz, Thiow Keng Tan, and Thomas Wiegand, "Comparison of the coding efficiency of video coding standardsincluding high efficiency video coding (hevc)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1669–1684, 2012.

[3] Tong Shen, Yao Lu, Ziyu Wen, Linxi Zou, Yucong Chen, and Jiangtao Wen, "Ultra fast h. 264/avc to hevc transcoder," in *Data Compression Conference (DCC), 2013*. IEEE, 2013, pp. 241–250.

[4] Yucong Chen, Ziyu Wen, Jiangtao Wen, Minhao Tang, and Pin Tao, "Efficient software h. 264/avc to hevc transcoding on distributed multicore processors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 8, pp. 1423–1434, 2015.

[5] Jiawen Gu, Yuxing Han, and Jiangtao Wen, "A novel low delay in-loop filtering wpp process for parallel hevc encoding," in *Visual Communications and Image Processing (VCIP), 2016*. IEEE, 2016, pp. 1–4.

[6] Detlev Marpe, Thomas Wiegand, and Gary J Sullivan, "The h. 264/mpeg4 advanced video coding standard and its applications," *IEEE communications magazine*, vol. 44, no. 8, pp. 134–143, 2006.

[7] Jani Lainema, Frank Bossen, Woo-Jin Han, Junghye Min, and Kemal Ugur, "Intra coding of the hevc standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1792–1801, 2012.

[8] Yilong Cheng, Guowei Teng, Xuli Shi, and Huosheng Li, "A fast intra prediction algorithm for hevc," in *Advances on Digital Television and Wireless Multimedia Communications*, pp. 292–298. Springer, 2012.

[9] M Ismail, Hyunho Jo, and Donggyu Sim, "Fast intra mode decision for hevc intra coding," in *The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014)*. IEEE, 2014, pp. 1–2.

[10] Liang Zhao, Li Zhang, Siwei Ma, and Debin Zhao, "Fast mode decision algorithm for intra prediction in hevc," in *Visual Communications and Image Processing (VCIP), 2011 IEEE*. IEEE, 2011, pp. 1–4.

[11] Yih Han Tan, Chuohao Yeo, Hui Li Tan, and Zhengguo Li, "On residual quad-tree coding in hevc," in *Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on*. IEEE, 2011, pp. 1–4.

[12] Liquan Shen, Zhaoyang Zhang, and Ping An, "Fast cu size decision and mode decision algorithm for hevc intra coding," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 1, pp. 207–213, 2013.

[13] Yih Chuan Lin and Jian Cheng Lai, "Edge density early termination algorithm for hevc coding tree block," in *Computer, Consumer and Control (IS3C), 2014 International Symposium on*. IEEE, 2014, pp. 39–42.

[14] Hao Zhang and Zhan Ma, "Early termination schemes for fast intra mode decision in high efficiency video coding," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*. IEEE, 2013, pp. 45–48.

[15] Mengmeng Zhang, Yu Liu, and Zhi Liu, "A new fast algorithm based on satd for hevc intra prediction," in *Visual Communications and Image Processing (VCIP), 2016*. IEEE, 2016, pp. 1–4.

[16] Hao Zhang and Zhan Ma, "Fast intra mode decision for high efficiency video coding (hevc)," *IEEE Transactions on circuits and systems for video technology*, vol. 24, no. 4, pp. 660–668, 2014.

[17] Q ITU-T SG16, "6: An excel add-in for computing bjontegaard metric and its evolution," *VCEG-AE07*, 2007.

[18] Yang Wang, Xiaopeng Fan, Liang Zhao, Siwei Ma, Debin Zhao, and Wen Gao, "A fast intra coding algorithm for hevc," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 4117–4121.

[19] Jiawei Chen and Lu Yu, "Effective hevc intra coding unit size decision based on online progressive bayesian classification," in *Multimedia and Expo (ICME), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–6.

[20] Xiwu Shang, Guozhong Wang, Tao Fan, and Yan Li, "Fast cu size decision and pu mode decision algorithm in hevc intra coding," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1593–1597.

[21] Liquan Shen, Zhaoyang Zhang, and Zhi Liu, "Effective cu size decision for hevc intracoding," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4232–4241, 2014.