Early SKIP Mode Decision Method in HEVC Based on Perceptual Distortion Measure

Jieon Kim
Queen Mary University of London
London, UK

Ebroul Izquierdo
Queen Mary University of London
London, UK

***Abstract*—An effective fast SKIP mode decision method is proposed for the High Efficiency Video Coding (HEVC) encoding. In order to determine the best Prediction Unit (PU) mode for a given Coding Unit (CU), the reference software for HEVC checks every PU mode candidate. This causes enormous computational complexity, especially in HEVC inter coding, which should be tackled for a fast encoder. An algorithm is proposed which exploits the fact that the SKIP mode can be directly decided as an optimal PU mode when reconstructed CUs of all candidates are similar. This is because bits for the SKIP mode is pretty small compared to other candidates. Thus, the distortion value between reconstructed CUs in 2N $\times$ 2N Merge and SKIP mode is measured in terms of the Human Visual System (HVS). If it is determined as unnoticeable, then a given CU is encoded in the SKIP mode. Experimental results show that the proposed method can save encoding time by 48.11 % and 41.66 % on average with minor objective quality losses under the Random Access (RA) and Low Delay (LD)-B configurations respectively.**

***Index Terms*—HEVC, perceptual quality, SKIP mode, fast encoders, spatial-JND**

I. Introduction

The High Efficiency Video Coding (HEVC) standard is developed by the Joint Collaborative Team on Video Coding (JCT-VC) which cuts bit-rate nearly in half compared to its predecessor, H.264/Advanced Video Coding (AVC), retaining the equivalent visual quality [1]. Despite this splendid outcome, it comes at the price of a huge amount of computationally complex burden, which makes it impossible to implement practical HEVC encoders [2]. For instance, the Rate-Distortion Optimisation (RDO) technique is employed in inter-mode decision and it checks all possible candidates in a brute-force way in order to decide the best Prediction Unit (PU) mode for a given Coding Unit (CU), which results in a considerable increase in computational complexity. Hence, it is essential to develop fast algorithms to reduce the complexity of HEVC encoders.

Several research works have been proposed to tackle the high computational complexity caused by the exhaustive RDO procedure in inter-mode decision. In particular, studies on Early SKIP mode Decision (ESD) can be summarised as follows [3]–[6]. In [3] Inter 2N $\times$ 2N mode is performed prior to SKIP mode in order to exploit its encoding information such as Coded Block Flag (CBF) and Motion Vector Difference (MVD). In [4] 2N $\times$ 2N Merge mode is utilised together with SKIP mode conditions based on a high RD correlation between 2N $\times$ 2N Merge and SKIP mode, which is aimed at improving coding quality of the SKIP mode. In [5] a Unimodal Stopping Model (USM) is designed for ESD. In [6] SKIP mode correlations between the current CU and its neighbouring CUs are exploited for ESD. However, all of these works were developed to find an optimal point between the complexity and RD performance based on objective visual quality measurements, such as Peak Signal-to-Noise Ratio (PSNR), which do not always correspond to the Human Visual System (HVS).

In this paper, a method to reduce the complexity in inter-mode decision of a HEVC encoder based on early SKIP mode decision is presented. The method is based on perceptual quality. The rest of the paper is organised as follow: Section II introduces a motivation of the proposed method and the Just Noticeable Difference (JND) model; the proposed early SKIP mode decision method is described in Section III; experimental results and analysis are presented in Section IV; and lastly, conclusions are drawn in Section V.

II. Motivation and Just Noticeable Difference Model

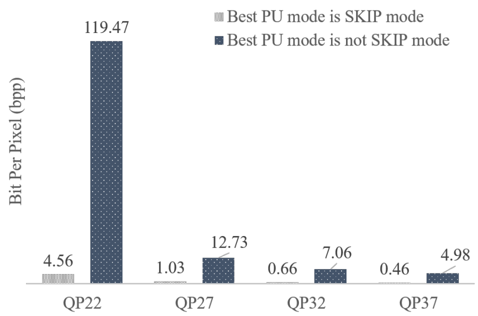
The best PU mode is selected when it refers to the minimisation of the following RD cost

$$J_M = D_M + \lambda_{\text{MODE}} \cdot R_M, \qquad (1)$$

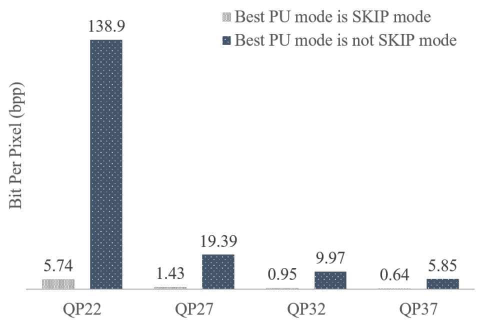
where M indicates a selected PU mode, D_M is the distortion value between the original CU and its reconstruction, R_M is the number of bits corresponding to M, and λ_{MODE} is a Lagrange multiplier that depends on Quantisation Parameter (QP).

In the case of the SKIP mode, R_{SKIP} is comparatively small in value. For example, Fig.1 shows average number of Bits Per Pixel (bpp) values of the two cases: 1) the best PU mode is the SKIP mode; 2) the best PU mode is not the SKIP mode, where bpp is the value of R_M divided by the number of CU samples. As can be seen, the values are generally small in the first case compared to the second case. As for *Traffic* sequence, the average value indicates 4.56 in the first case, whereas it indicates 119.47 in the second case with respect to QP 22. This is because the SKIP mode does not send any residue to decoder side. A flag called *cu_skip_flag*, set to value one, is sent to indicate the associated Coding Blocks (CBs) are encoded in the SKIP mode, and *merge_idx* is transmitted to indicate from which candidate the motion information is inherited. With regard to residual data, it is completely omitted

978-9-0827-9703-9/19/$31.00 ©2019 IEEE



(a) Traffic



(b) ParkScene

Fig. 1: Comparisons average number of bits per pixel (bpp) for a given quantisation parameter (QP) between skipped CU or otherwise.

because transform coefficients of prediction errors are regarded as value zero.

It implies that the SKIP mode can be directly selected as the best PU mode if reconstructions of all PU candidates are determined as identical. That is, if difference values between reconstructed CU in the SKIP mode and other PU modes are small, the current CU can be encoded in the SKIP mode. In order to exploit the aforementioned case for ESD, the differences can be measured by subjective measurement. If the human eye cannot distinguish the differences, the reconstructed CU in the SKIP mode can be considered as a represented CU of the other candidates for a given CU.

As the subjective measurement, JND is one of the widely used models in signal processing to provide the maximum difference value which human eye cannot perceive. Fig.2 shows two images. Fig.2 (a) represents the first frame of *BQMall* sequence and Fig.2 (b) represents JND-injection image to Fig.2 (a). When it comes to objective quality measurement such as PSNR, it provides visual quality of the target image (Fig.2 (b)) as a concrete value compared to the original image, which is 32.57 dB. However, it may require several threshold values to exploit for the aforementioned case. On the other hand, it can be seen that two images have zero visual difference in terms of subjective quality. Thus, the differences among reconstructions of candidates are estimated by the JND model.

(a)

(b)

Fig. 2: PSNR and perceptual visual quality. (a) Original BQMall Sequence (b) JND-Injection to the first frame of BQMall Sequence, PSNR: 32.57 dB.

III. Proposed Method

In HEVC, the SKIP mode does not incorporate transform and quantisation process. Thus, the spatial JND model [7] has been adopted in the proposed method. Based on the JND model, the total number of visible distortion in a CU (N) is defined as follows:

$$N = \sum_{y=0}^{H-1} \sum_{x=0}^{W-1} \mathrm{v}_{x,y}, \tag{2}$$

where W and H are the width and height of the CU, and $\mathrm{v}_{x,\,y}$ represents whether the distortion value at position x and y is visible or invisible which is calculated by

$$\mathrm{v}_{x,y} = \begin{cases} 1, & |\mathbf{d}^{\mathrm{SKIP}}_{x,y}| > \mathbf{s}_{x,y} \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

where $\mathbf{d}^{\mathrm{SKIP}}$ is defined as

$$\mathbf{d}^{\mathrm{SKIP}} = \hat{\mathbf{y}}^{\mathrm{m}} - \hat{\mathbf{y}}^{\mathrm{SKIP}}, \tag{4}$$

where $\hat{\mathbf{y}}^{\mathrm{m}}$ and $\hat{\mathbf{y}}^{\mathrm{SKIP}}$ denote reconstructed CU in PU candidate except the SKIP mode and in SKIP mode respectively. The $\mathbf{s}$ is defined in [7] as follows:

$$\mathbf{s}_{x,y} = \max\{\mathbf{l}_{x,y}, \mathbf{c}_{x,y}\}, \tag{5}$$

where $\mathbf{l}_{x,y}$ and $\mathbf{c}_{x,y}$ are pixel values to refer to Luminance Adaptation (LA) which represents the making effect and

Contrast Masking (CM) which denotes the spatial masking respectively [7]. $\mathbf{l}_{x,y}$ and $\mathbf{c}_{x,y}$ are denoted respectively, by:

$$\mathbf{l}_{x,y} = \begin{cases} T_0 \times (1 - \sqrt{\mathbf{b}_{x,y}/127}) + \varepsilon, \text{if } \mathbf{b}_{x,y} \leq 127 \\ \gamma \times (\mathbf{b}_{x,y} - 127) + \varepsilon, \text{if } \mathbf{b}_{x,y} > 127 \end{cases}, \quad (6)$$

$$\mathbf{c}_{x,y} = \mathbf{m}_{x,y} \times \alpha(\mathbf{b}_{x,y}) + \beta(\mathbf{b}_{x,y}), \quad (7)$$

where $\mathbf{b}_{x,y}$ is the average background luminance in a CU by a weighted low-pass filter *B* [7].

$$\mathbf{b}_{x,y} = \frac{1}{32} \sum_{j=1}^{5} \sum_{i=1}^{5} \mathbf{y}_{x-3+i,y-3+j} \times B_{i,j}, \quad (8)$$

and ε represents the Minimum Visibility Threshold (MVT). $\mathbf{m}_{x,y}$ is the maximum weighted average of luminance differences around the pixel at position x and y using the operator *G* [7].

$$\mathbf{m}_{x,y} = \max_{k=1,2,3,4} |\text{grad}^k_{x,y}|, \quad (9)$$

where

$$\text{grad}^k_{x,y} = \frac{1}{16} \sum_{j=1}^{5} \sum_{i=1}^{5} \mathbf{y}_{x-3+i,y-3+j} \times G_{i,j}. \quad (10)$$

In addition, $\alpha(\mathbf{b}_{x,y})$ and $\beta(\mathbf{b}_{x,y})$ are defined as

$$\alpha(\mathbf{b}_{x,y}) = \mathbf{b}_{x,y} \times 0.0001 + 0.115 \quad (11)$$

and

$$\beta(\mathbf{b}_{x,y}) = \frac{1}{4} - \mathbf{b}_{x,y} \times 0.01. \quad (12)$$

Note that T_0, γ and ε are set to value 17, 3/128 and 3, respectively.

In [5] RD cost relationships among PU modes were exploited for ESD. When the RD cost of SKIP mode is less than its 2N $\times$ 2N Merge mode and Inter 2N $\times$ 2N, the SKIP mode is selected as the best PU mode. Similar to this approach the reconstructed CU in 2N $\times$ 2N Merge mode is compared with the reconstructed CU in the SKIP mode in the proposed method, and comparisons with the other PU modes are skipped. In addition, the reconstructed CU in 2N $\times$ 2N Merge mode serves as an anchor for calculating the visual quality threshold using the JND model.

In order to investigate the relationship between skipped CU and invisible distortion in a CU, proportion of the invisible distortion (*p*) has been collected. The skipped CU refers to the CU in the SKIP mode by HM software, and the *p* is defined as follows:

$$p = 1 - \frac{N}{W \times H}. \quad (13)$$

In the experiment, eighteen test sequences in Table I are tested under random access (RA) configuration.

Fig.3 shows experimental results with respect to QP22 and QP37 respectively. The test results are categorised into two cases, Case A and Case B. Case A refers to skipped CUs and Case B refers to non-skipped CUs. As can be seen, the distribution of *p* values are different in Case A and Case B. In the case of Case A, *p* values are close to value one, while in the case of Case B, *p* have much lower values than Case A, i.e., from 0.46 to 0.91 with QP22. In respect of QP 37, the values of p are decreased overall but it can be still seen the same pattern in QP 22. CUs in Case A have p values from 0.92 to 1, whereas CUs in Case B have values from 0.2 to 0.6. From it, Eq.2 can be considered an appropriate element for ESD.

TABLE I: Test Sequences for Experiments and Its Labels

Class	Label	Sequence	Resolution
Class A	a	Traffic	2560 × 1600
	b	PeopleOnStreet	
Class B	c	Kimono1	1920 × 1080
	d	ParkScene	
	e	Cactus	
	f	BQTerrace	
	g	BasketballDrive	
Class C	h	RaceHorsesC	832 × 480
	i	BQMall	
	j	PartyScene	
	k	BasketballDrill	
Class D	l	RaceHorses	416 × 240
	m	BQSquare	
	n	BlowingBubbles	
	o	BasketballPass	
Class E	p	FourPeople	1280 × 720
	q	Johnny	
	r	KristenAndSara	

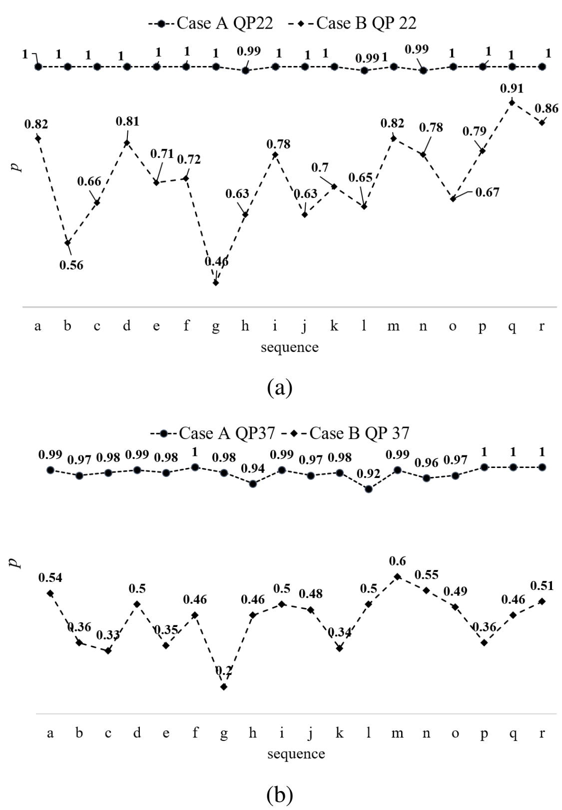


Fig. 3: Average proportion of invisible distortion in SKIP mode measured on reconstructed CU. (a) QP = 22 (b) QP = 37

TABLE II: Average Results of the Probabilities P($\mathbf{R}|\mathbf{H}$)

	QP	a	b	c	d	e	f	g	Avg.
P($\mathbf{R}|\mathbf{H}$)	22	0.98	0.96	1.01	0.99	0.94	0.92	0.94	0.96
	27	0.97	0.94	0.97	0.97	0.96	0.93	0.97	0.96
	32	0.97	0.92	0.95	0.96	0.96	0.97	0.96	0.96
	37	0.98	0.91	0.94	0.97	0.96	0.99	0.97	0.96

The proposed early SKIP mode decision in HEVC is summarised in Algorithm 1. First of all, the 2N $\times$ 2N Merge and SKIP PU mode predictions are performed. If the RD cost of SKIP mode is less than the 2N $\times$ 2N Merge mode, N in Eq.(2) is calculated, where $\mathbf{d}^{\text{SKIP}} = \mathbf{y}^{\text{2N} \times \text{2N Merge}} - \mathbf{y}^{\text{SKIP}}$. If the N is equal to value zero, then the SKIP mode is selected as an optimal mode and RDO procedures for the rest of candidates are skipped. Otherwise, the conventional way in HM software is followed.

Algorithm 1 Proposed Early SKIP Mode Decision Algorithm

```
Input:
    CU size / depth / total-depth : 64 / 4 / 4

for CU Depth = 0 to 3 do
    Calculate J_SKIP and J_2Nx2N Merge
    if J_SKIP < J_2Nx2N Merge then
        calculate N in Eq.(2)
        if N == 0 then
            the best PU mode ← SKIP Mode
        else
            Encode the current CU with other PU modes
        end if
    else
        Encode the current CU with other PU modes
    end if
    Output:
    The best PU mode of the current CU
end for
```

IV. Experimental Results

A. Test Conditions

Proposed method was implemented into HM software 16.13 [8]. Test sequences, which are listed in Table I, were encoded with four different QPs 22, 27, 32 and 37 under the common test conditions (CTCs) [9] and two configurations. All the experiments were performed on the computer equipped with Intel Xeon E5645 CPU @ 2.40 GHz, 24GB memory.

B. Accuracy of the Proposed Method

First of all, an experiment has been conducted to evaluate the conditional probabilities P($\mathbf{R}|\mathbf{H}$) under Random Access (RA) configuration. $\mathbf{H}$ refers to the event that the current CU is encoded in the SKIP mode by HM software. $\mathbf{R}$ represents events that the SKIP mode is selected as the best PU mode for the current CU using reconstructed CU in 2N $\times$ 2N Merge mode. Thus, P($\mathbf{R}|\mathbf{H}$) represents the probability of event that the SKIP mode is chosen by the proposed method when the current CU is encoded in the SKIP mode by HM software.

The experimental results of P($\mathbf{R}|\mathbf{H}$) is tabulated in Table II. From it, most of skipped CUs are predicted with high accuracy.

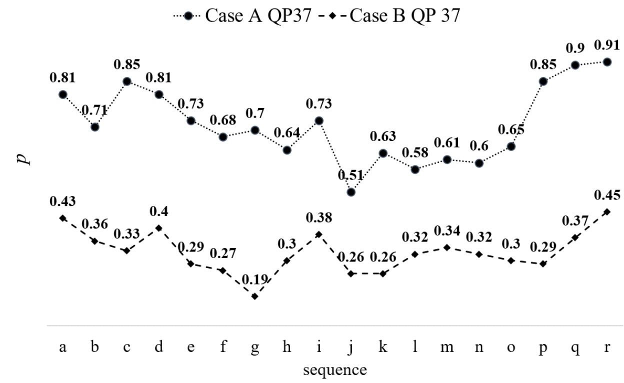


Fig. 4: Average proportion of invisible distortion in SKIP mode measured on original CU. QP = 37

The average values of P($\mathbf{R}|\mathbf{H}$) are 0.96, 0.96, 0.96 and 0.96 with QP 22, 27, 32 and 37 respectively.

C. JND measure on Original CU

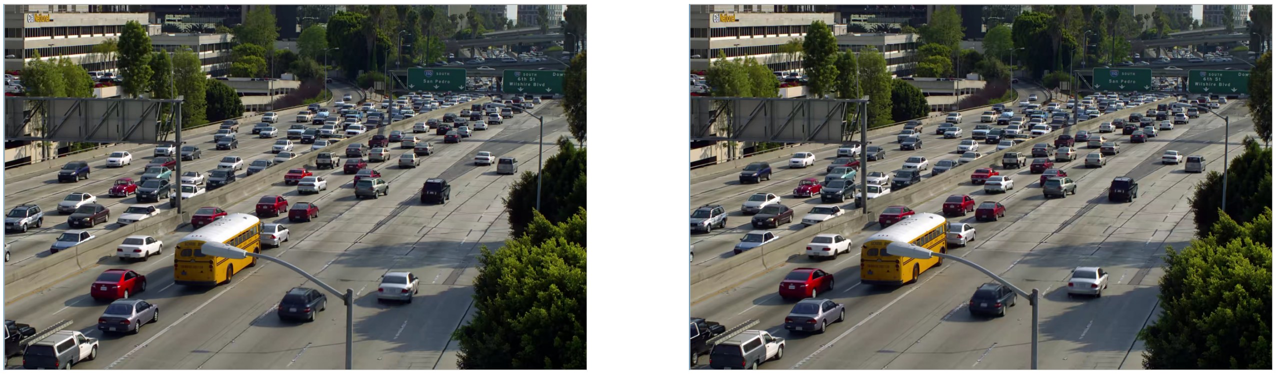
The RD cost refers the distortion between the original CU and its reconstruction. Therefore, an experiment has been conducted to collect the p values when the original CU served as an anchor instead of the reconstructed CU in 2N $\times$ 2N Merge. Fig.4 shows the results with respect to QP 37. As shown, a variation of the p values seems to be high with skipped CUs. It ranges in value from 0.91 to 0.51. It is because the maximum threshold values for the distortions in the SKIP mode are estimated on current CUs, and the visible errors include quantisation errors together with prediction errors.

D. Results of Encoding Time Compared with HM software and the state-of-the art Algorithms

Time saving (TS) was calculated to measure the performance of the proposed method in terms of encoding time which is defined as TS $= \frac{T_{\text{A}} - T_{\text{P}}}{T_{\text{A}}} \times 100$, where T_{A} and T_{P} are the total encoding time of the original HM software and the proposed method respectively. The other two state-of-art approaches to ESD were selected in order to compare TS of the proposed method. All three approaches were tested with the HM 16.13 software. Table III and Table IV summarise the experimental results of [3], [6] and the proposed algorithm under the RA and Low Delay (LD) B configurations respectively. As observed in the table, the proposed method reduces the encoding time by 48.11% and 38.59% on average under RA, LD B conditions respectively, outperforming the 38.59% and 34.02% in [3] and the 41.31% and 34.16% in [6].

E. Quality Comparisons

The visual quality comparisons are shown in Fig.5. The decoded sequences by HM software were compared as a ground truth with the decoded ones by the proposed method. As shown, subjective detail of contents information is well preserved. Considering that the proposed method substantially reduced the encoder complexity, reconstruction videos in the



(a) HM16.13: QP 37 (b) Proposed: QP 37

Fig. 5: The subjective quality comparisons. Traffic sequence (Class A).

TABLE III: Time Saving Comparisons Among the Proposed Method and Previous Works under RA Configuration

Class	Proposed	[3]	[6]
Average (Class A)	44.51%	36.16%	37.90%
Average (Class B)	51.76%	40.74%	44.67%
Average (Class C)	38.62%	32.24%	31.85%
Average (Class D)	36.36%	31.33%	30.20%
Average (Class E)	72.77%	54.77%	65.43%
Average	48.11%	38.59%	41.31%

TABLE IV: Time Saving Comparisons Among the Proposed Method and Previous Works Under LD-B Configuration

Class	Proposed	[3]	[6]
Average (Class A)	37.62%	31.23%	30.19%
Average (Class B)	44.58%	36.24%	37.17%
Average (Class C)	31.05%	27.19%	24.36%
Average (Class D)	29.50%	26.11%	21.80%
Average (Class E)	69.87%	51.84%	61.32%
Average	41.66%	34.02%	34.16%

TABLE V: Objective Quality Evaluation of the Proposed Method over HM 16.13

Class	Random Access			Low Delay B		
	ΔR	ΔPSNR [dB]	BD-BR	ΔR	ΔPSNR [dB]	BD-BR
Average (A)	-1.0%	0.11	1.8%	0.08	-0.8%	1.5%
Average (B)	-0.9%	0.07	2.2%	0.06	-0.6%	1.7%
Average (C)	-1.0%	0.11	1.7%	0.08	-0.7%	1.5%
Average (D)	-1.3%	0.13	1.9%	0.12	-1.0%	2.0%
Average (E)	-1.2%	0.09	2.1%	0.09	-1.0%	2.3%
Average	-1.1%	0.10	1.9%	0.09	-0.8%	1.8%

proposed method does not suffer by comparison with the reconstructions by HM software.

In addition, performance was evaluated on the PSNR difference ΔPSNR and the bit-rate difference ΔR which are defined as $\Delta\text{PSNR} = \text{PSNR}_\text{A} - \text{PSNR}_\text{P}$, where PSNR_A is the original PSNR for HM 16.13. PSNR_P is for the proposed method. The bit-rate increase ΔR is defined as $\Delta\text{R} = \frac{R_\text{P} - R_\text{A}}{R_\text{A}} \times 100$, where R_A is the original bit-rate for HM 16.13. R_P is for the proposed method. Table V shows PSNR and bit-rate differences between the proposed method and HM 16.13. From it, the proposed method achieves 0.10 dB loss in delta PSNR and 1.1% bit-rate reduction in delta bit-rate under RA configuration and 0.09 dB loss in delta PSNR 0.6% bit-rate reduction in delta bit-rate under LD-B configuration.

V. Conclusions

This paper proposed early skip mode decision in order to reduce the complexity of inter coding in HEVC. The proposed method was based on the fact that the SKIP mode can be decided as the best PU mode without exhaustive rate-cost comparisons of all PU candidates when reconstructions of all candidates were similar in terms of HVS. In addition, the similarity of the reconstructions were measured based on the JND model. Experimental results show that the proposed approach to the early SKIP mode can save by 48.11 % and 41.66 % on average encoding time of HM software under RA and LD-B configurations respectively.

References

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[2] J. Ohm, G. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards including high efficiency video coding (hevc)," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[3] J. Kim, J. Yang, K. Won, and B. Jeon, "Early determination of mode decision for hevc," in *Proc. IEEE Pict. Coding Symp.*, May 2012, pp. 449–452.

[4] H. Lee, H. J. Shim, Y. Park, and B. Jeon, "Early skip mode decision for hevc encoder with emphasis on coding quality," *IEEE Trans. Broadcast*, vol. 61, no. 3, pp. 388–397, Sep. 2015.

[5] Y. Li, G. B. Yang, Y. P. Zhu, X. L. Ding, and X. M. Sun, "Unimodal stopping model based early skip mode decision for high efficiency video coding," *IEEE Trans. Multimedia*, vol. 19, no. 7, pp. 1431–1441, Jul. 2017.

[6] J. Kim and E.Izquierdo, "An effective strategy for early skip mode decision in hevc," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct. 2018.

[7] C.-H. Chou and Y.-C. Li, "A perceptual tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol*, vol. 5, no. 6, pp. 467–476, Dec. 1995.

[8] ITU, "HM reference software." [Online]. Available: https://hevc.hhi.fraunhofer.de/HM-doc/

[9] F. Bossen, "Common test conditions and software reference configurations."