Fast prediction unit selection method for HEVC intra prediction based on salient regions

FENG Lei (冯磊)[1,2]*, **DAI Ming** (戴明)[1], **ZHAO Chun-lei** (赵春蕾)[1,2], **and XIONG Jing-ying** (熊晶莹)[1,2]

1. *Key Laboratory of Airborne Optical Imaging and Measurement, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China*
2. *University of Chinese Academy of Sciences, Beijing 100039, China*

(Received 24 March 2016)

In order to reduce the computational complexity of the high efficiency video coding (HEVC) standard, a new algorithm for HEVC intra prediction, namely, fast prediction unit (PU) size selection method for HEVC based on salient regions is proposed in this paper. We first build a saliency map for each largest coding unit (LCU) to reduce its texture complexity. Secondly, the optimal PU size is determined via a scheme that implements an information entropy comparison among sub-blocks of saliency maps. Finally, we apply the partitioning result of saliency map on the original LCUs, obtaining the optimal partitioning result. Our algorithm can determine the PU size in advance to the angular prediction in intra coding, reducing computational complexity of HEVC. The experimental results show that our algorithm achieves a 37.9% reduction in encoding time, while producing a negligible loss in Bjontegaard delta bit rate (*BDBR*) of 0.62%.

Document code: A **Article ID:** 1673-1905(2016)04-0316-5
DOI 10.1007/s11801-016-6064-8

With the increasing demand for high definition (HD) video transmission, the joint collaborative team on video coding (JCT-VC) proposed high efficiency video coding (HEVC) standard to develop the video compression technologies, which would provide much higher coding efficiency than the former H.264/AVC standard[1,2]. The HEVC encoder has an increased number of multi-angle intra prediction modes compared with H.264/AVC, so it has high computational complexity due to the rate distortion (RD) cost of running all the possible coding modes in the traversal calculation procedure. The HEVC encoder improves the prediction accuracy, while it has caused significant increase in computational complexity[3,4]. To reduce encoder complexity of HEVC, various methods for improving the standard algorithm have been investigated. Piao et al[5] proposed rough mode decision (RMD), where the calculation of low-complexity RD cost and sum of absolute Hadamard transformed coefficient values may select the most possibility of candidate modes for the intra prediction coding. Then Zhao et al[6] further improved the RMD method. Afterwards, Shen et al[7] proposed a fast coding unit (CU) size decision algorithm utilizing strong correlation between two adjacent tree-blocks. Xiong et al[8] proposed an algorithm based on the theoretical analysis that the non-normalized histogram of oriented gradient can be used to select CU size. Huang et al[9] proposed a block size decision algorithm based on variance computation, which uses a threshold for decision skipping intra prediction modes with sizes of 4×4 and 16×16.

However, few investigations above are focused on texture features in largest coding unit (LCU). In order to further improve the performance of intra prediction, we develop a new algorithm for fast prediction unit (PU) size decision based on the salient region. Our algorithm can determine the PU size in advance to the angular prediction in intra coding, reducing the computation complexity of HEVC.

Many studies have found that when human visual system (HVS) observes videos or pictures, it is not able to perceive all information in the visual field with the same accuracy[10]. When processing the visual information, the HVS has a priority attention mechanism for selecting the region of interest[11]. The areas in frames that can significantly attract the attention of the HVS are called salient regions, while others are the non-salient regions or background regions. When digital video sequence is played, the salient region is perceived by HVS prior to other regions.

Compared with HVS, the HEVC encoder generally uses larger pixel blocks for representing the background regions, whereas using smaller blocks on textured regions for intra prediction[12]. Therefore, if an HEVC encoder extracts texture information ahead in the LCUs, we can obtain proper partitioning results according to different textured regions, avoiding unnecessary recursive

* E-mail: kabakfl@126.com

computation and speeding up the intra prediction process.

Following the above analysis, the saliency map is introduced to extract texture properties[13]. First, take fast Fourier transform (FFT) on the original images to obtain the amplitude spectra. Then take natural logarithm on the amplitude spectra and the logarithm spectra of the amplitudes are obtained, which have similar distributions in different images. Fig.1 shows examples of the transformation from original pictures to the saliency maps.

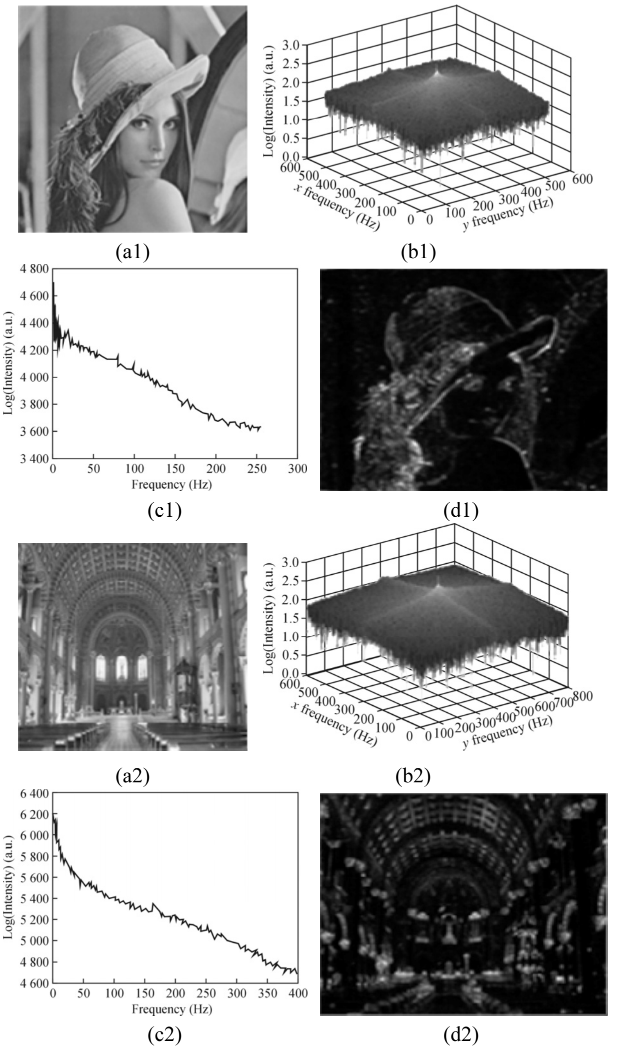


Fig.1 (a1) and (a2) Two natural pictures; (b1) and (b2) Amplitude spectrum distributions obtained via FFT; (c1) and (c2) One-dimensional marginal distributions corresponding to (b1) and (b2); (d1) and (d2) The saliency maps

From Fig.1, we can observe that the logarithm spectra of different images have similar tendencies in (c1) and (c2), both of which display a local approximate proportion relation between frequency and logarithm spectrum of amplitude[14]. After using mean filtering on the logarithm spectra and acquiring the linear component of every image, the images satisfy the following equation:

$$R(\text{residual}) = I(\text{image}) - I(\text{priorknowledge}), \tag{1}$$

where I(image) denotes the entirety of the image components, and I(priorknowledge) denotes the linear sections in (c1) and (c2), i.e., the components of non-salient region of HVS. R(residual) denotes the components visually captured by audiences, namely, the salient map. From Eq.(1), we can obtain R(residual) by removing redundant components from the image. In addition, the scenes in the backgrounds of (d1) and (d2) have texture complexities that are greatly reduced, while the salient region has no significant loss in details.

We can use the above computational model of visual selective attention to extract the texture information in HEVC, utilizing the characteristics of saliency maps to reduce texture complexity of LCUs and boost the speed of intra prediction.

The process of obtaining saliency maps of LCUs is as follows: take FFT over an LCU block $I_i(x)$; then the amplitude spectrum, denoted as $A_i(\omega)$, is obtained from Eq.(4); the phase spectrum $P_i(\omega)$ is given by Eq.(3), and we can compute logarithm spectrum of the amplitude $L_i(\omega)$ by taking natural logarithm on $A_i(\omega)$.

$$A_i(\omega) = (\text{real}\{F[I_i(x)]\}^2 + \text{imag}\{F[I_i(x)]\}^2)^{0.5}, \tag{2}$$

$$P_i(\omega) = \arctan\frac{\text{imag}\{F[I_i(x)]\}}{\text{real}\{F[I_i(x)]\}}, \tag{3}$$

$$L_i(\omega) = \log\left[A_i(\omega)\right]. \tag{4}$$

The residual spectrum $R_i(\omega)$ is:

$$R_i(\omega) = L_i(\omega) - h_n(\omega) * L_i(\omega), \tag{5}$$

where $h_n(\omega)$ is the convolution kernel defined by the following matrix:

$$h_n(\omega) = \frac{1}{n^2}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}, n=3. \tag{6}$$

From the Euler's formula:

$$\exp(r + \mathrm{i}\cdot\varphi) = \exp(r)\cdot\left[\cos(\varphi) + \mathrm{i}\cdot\sin(\varphi)\right] =$$
$$\exp(r)\cdot\cos(\varphi) + \mathrm{i}\cdot\exp(r)\cdot\sin(\varphi), \tag{7}$$

where r denotes $R_i(\omega)$ and φ denotes $P_i(\omega)$, satisfying:

$$\sin(\varphi) = \frac{\text{imag}\{F[I_i(x)]\}}{\left|A_i(\omega)\right|},$$

$$\cos(\varphi) = \frac{\text{real}\{F[I_i(x)]\}}{\left|A_i(\omega)\right|}, \tag{8}$$

we have the saliency map:

$$S_i(x) = g(x)\cdot F^{-1}\left\{\exp[R_i(\omega) + P_i(\omega)]\right\}, \tag{9}$$

where $g(x)$ is the Gaussian filtering function.

The information entropy provides a measure of the uncertainty of random events that occur in an image area, namely, the amount of information contained in an image area[15].

Supposing A_i is a random variable, its information entropy is given by:

$$H(A_i)=-\sum_{i=1}^{n}P(A_i)\cdot\log(A_i)\,, \tag{10}$$

where $P(A_i)$ denotes the probability of event A_i. A digital image contains a series of discrete pixels, and the gray-scale is divided into 256 levels (0—255). P_j denotes the probability of pixels at gray level j emerging in the image. Thus we have the following equation:

$$H(P_j)=-P_j\cdot\log_2(P_j)\,,\ 0\le j\le 255\,. \tag{11}$$

Subsequently, we have the equation for information entropy:

$$E=\sum_{j=0}^{255}H(P_j)=-\sum_{j=0}^{255}P_j\cdot\log_2(P_j)\,. \tag{12}$$

In general, the information entropy implies the richness of gray level information in an image, and the amount of texture in an image is proportional to the information entropy. Therefore, we can utilize image information entropy to determine the flatness of a saliency map, and partition the saliency map efficiently with a proper threshold.

In summary, we first compute the information entropy of saliency map in LCUs, and compare the entropy value with its four 2×2 sub-blocks in the next depth; if the difference of entropy value in four sub-blocks is less than 10%, and each of the sub-blocks has a smaller value than its father-block, we use the size of the parent block in the former depth to make intra prediction and the partition process is terminated; if the entropy value in each of the sub-blocks is larger than that of its father-block, or the difference of entropy value in four sub-blocks is more than 10%, we use the size of the sub-block in the next depth to make intra prediction. We run the above process recursively for each LCU, until the PU size is in 4×4. Once the optimal block size is found, the partitioning process is terminated.

The process of fast PU size selection based on salient regions is described as follows:

(1) Set the LCU size to be 64×64, then count the LCU numbers in the current frame;

(2) Take FFT on every LCU, and obtain the saliency map $S_n(x)$ of every LCU;

(3) Compute information entropy of the current saliency map $S_n(x)$, and recursively compute the information entropy of 2×2 sub-blocks, updating the optimal partitioning result, until the current PU is in 4×4;

(4) Appling the partitioning result of the saliency map to each original LCU, we can obtain the final partitioning result of the intra prediction.

The proposed algorithm is implemented using the HEVC reference software HM11.0. In order to give an objective evaluation of the performance of the algorithm, we use standard test sequences with texture features of multiple sizes (416×240, 832×480, 1 920×1 080 and 2 560×1 600). The encoder is set to be all-intra main configuration. The quantization parameter (QP) values are 22, 27, 32 and 37. The coding performance is evaluated under Bjontegaard delta bit rate ($BDBR$)[16] and time saving (TS) criteria, which are given by

$$\Delta BR=\frac{BR_{\text{new}}-BR_{\text{HM11.0}}}{BR_{\text{HM11.0}}}\times 100\%\ , \tag{13}$$

$$\Delta TS=\frac{T_{\text{new}}-T_{\text{HM11.0}}}{T_{\text{HM11.0}}}\times 100\%\ . \tag{14}$$

The experimental results are shown in Tab.1.

Tab.1 Results of the proposed algorithm and a comparison with the standard

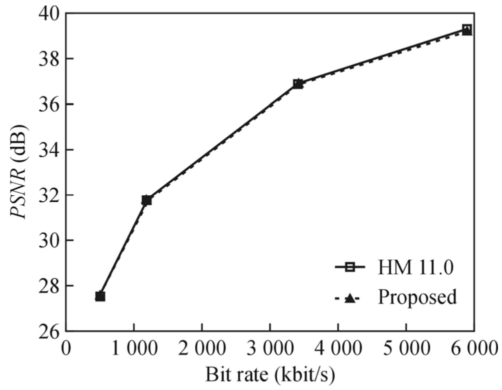
Test sequences	Image size	ΔBR (%)	ΔTS (%)
Traffic	2 560×1 600	−0.97	−33.6
People On Street	2 560×1 600	−0.88	−37.8
BQ Terrace	1 920×1 080	−0.91	−34.4
Kimonol	1 920×1 080	−0.61	−40.2
Basketball Drive	1 920×1 080	−0.79	−33.2
Park Scene	1 920×1 080	−0.84	−38.3
Cactus	1 920×1 080	−0.72	−35.5
Basketball Drill	832×480	−0.57	−31.7
Slide Show	832×480	−0.38	−39.7
Party Scene	832×480	−0.54	−41.3
BQ Mall	832×480	−0.86	−43.1
Blowing Bubbles	416×240	−0.16	−45.1
BQ Square	416×240	−0.25	−40.4
Basketball Pass	416×240	−0.38	−37.9
Race Horses	416×240	−0.46	−36.7
Average		−0.62	−37.9

The results in Tab.1 indicate that we have achieved higher acceleration on the given sequences whose scene is relatively smooth, or the sequences where background region dominates the frames, the moving objects are fewer in number and occupy more relatively large PUs in the frames, whereby the encoder skips the partitioning judgment cycle earlier. On the contrary, in the images containing more moving objects or complicated scenes, more of smaller PUs are allocated to the textured areas, and the encoder consumes more time in the computation procedure, resulting in lower time reduction. Besides, our algorithm compares the entropy of the 2×2 sub-blocks with that of their father-blocks for every coding depth. Compared with the standard algorithm, HEVC encoder avoids the RD summation process, therefore the coding time has been effectively reduced. For example, according to the experimental results in Tab.1, some sequences contain objects with slow movement, and the image resolution is not very high. The proposed algorithm uses more larger-PUs for intra prediction coding, therefore the coding speed increases significantly.

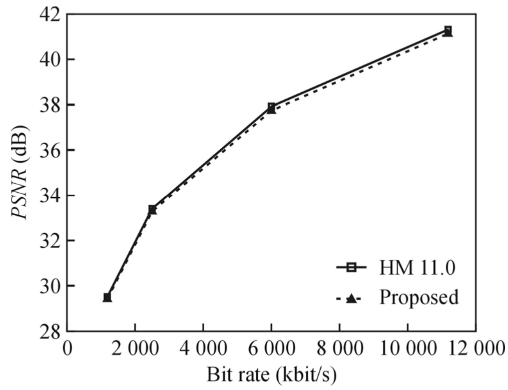
Tab.2 compares the time saving performance of our algorithm and that of existing popular algorithms. We can see that our algorithm performs better on the test sequences containing smooth textures. For instance, Fig.2 shows RD curves of two typical sequences, and the highest proportion of time is saved, which is 45.1%, in sequence "Blowing Bubbles"; and we still have the lowest coding time reduction of 33.2% in sequence "Basketball Drill". Fig.3 shows the saliency map and partitioning results in LCUs of sequence "Blowing Bubbles".

Tab.2 Comparison of coding time saved in other algorithms for text sequences

Test sequences	Image size	ΔTS (%)		
		Ref.[7]	Ref.[8]	Proposed
Traffic	2 560×1 600	−22.3	−33.1	−33.6
People On Street	2 560×1 600	−21.6	−32.3	−37.8
BQ Terrace	1 920×1 080	−25.5	−34.7	−34.4
Kimonol	1 920×1 080	−36.0	−37.8	−40.2
Basketball Drive	1 920×1 080	−31.8	−38.2	−33.2
Park Scene	1 920×1 080	−26.1	−35.9	−38.3
Cactus	1 920×1 080	−23.5	−33.4	−35.5
Basketball Drill	832×480	−17.9	−31.9	−31.7
Slide Show	832×480	−16.9	−36.3	−39.7
Party Scene	832×480	−17.8	−27.3	−41.3
BQ Mall	832×480	−18.6	−27.5	−43.1
Blowing Bubbles	416×240	−15.1	−28.1	−45.1
BQ Square	416×240	−14.9	−29.5	−40.4
Basketball Pass	416×240	−15.1	−30.4	−37.9
Race Horses	416×240	−12.9	−33.2	−36.7
Average		−21.1	−32.6	−37.9



(a)



(b)

Fig.2 (a) RD curve of sequence "Blowing Bubbles"; (b) RD curve of sequence "Basketball Drill"

Fig.3 The saliency map and partitioning result in LCUs for sequence "Blowing Bubbles"

In this paper, we propose a fast prediction unit size selection method for HEVC intra prediction based on salient regions, which is specified using saliency map to reduce texture-complexity in every LCU, and the optimal PU size is determined efficiently through the information entropy comparison in sub-blocks of LCUs. Along with the technical details, our method provides a new analytical perspective for HEVC intra coding. Experimental results reveal that our algorithm can judge the optimal PU size before the angular prediction in HEVC intra prediction process, avoiding unnecessary calculation steps. The new algorithm can save the coding time by 37.9% effectively with negligible loss of 0.62% in BD-rate.

References

[1] Gary J. Sullivan, Jens-Rainer Ohm and Woo-Jin Han, IEEE Transactions on Circuits and Systems for Video Technology **22**, 1649 (2012).

[2] SUN Le, DAI Ming, LI Gang, CHEN Xiao-lu, WANG Zi-chen and SONG Yue-ming, Journal of Optoelectronics·Laser **24**, 2205 (2013). (in Chinese)

[3] Frank Bossen, Benjamin Bross, Karsten Suhring and David Flynn, IEEE Transactions on Circuits and Sys-

tems for Video Technology **22**, 1 (2012).
[4] ZHANG Qiang and YUAN Chun, Journal of Frontiers of Computer Science and Technology **8**, 537 (2014). (in Chinese)
[5] Y. Piao, J. H. Min and J. Chen, Encoder Improvement of Unified Intra Prediction, JCTVC-C207, Guangzhou, 166 (2010).
[6] L. Zhao, L. Zhang, S. Ma and D. Chen, IEEE Visual Communications and Image Process **5**, 1 (2011).
[7] L. Shen, Z. Zhang and P. An, IEEE Transactions on Consumer Electronics **59**, 207 (2013).
[8] XIONG Jian, Optical Engineering **52**, 071504 (2013).
[9] Y. H. Huang, T. S. Ou and H. H. Chen, IEEE Transactions on Circuits and Systems for Video Technology **20**, 1122 (2010).
[10] Sang Ho Kim and J. P. Allebach, IEEE Transactions on Image Processing **11**, 258 (2002).
[11] M. Xu, X. Deng, S. X. Li and Z. L. Wang, IEEE Journal of Selected Topics in Signal Processing **8**, 475 (2014).
[12] S. Cho and M. Kim, IEEE Transactions on Circuits and Systems for Video Technology **23**, 1555 (2013).
[13] Z. Qi and K. Christof, Journal of Vision **11**, 74 (2011).
[14] X. D. Hou, J. Harel and C. Koch, IEEE Transactions on Pattern Analysis and Machine Intelligence **34**, 194 (2012).
[15] J.Y. Liang and Z. Z. Shi, Fuzziness and Knowledge-Based Systems **12**, 37 (2011).
[16] G. Bjontegaard, Calculation of Average PSNR Differences between RD-curves, document ITU-T SC16/Q6, VCEG-M33, Texas, 1 (2001).