

The Improved Algorithm of Sample Adaptive Offset Based on Visual Saliency

Nana Shan^{a,1}, Wei Zhou^b and Zhemin Duan^b

^a*School of Information Science and Technology, Taishan University*

^b*Department of Electronics and Information, Northwestern Polytechnical University*

Abstract. Sample Adaptive Offset (SAO) in High Efficient Video Coding (HEVC) is a new technic to improve the quality of videos. It categories the pixels and chooses the best way by adding some offsets to the reconstructed video. So, it causes a dramatically increased computational complexity. According to the dependency of sample adaptive offset and visual saliency map, an improved SAO method is proposed in order to minimize the coding time of SAO by skipping some RD cost calculation. Experimental result shows that the proposed method reduces 27.02% SAO encoding time with negligible performance loss.

Keywords. visual saliency, sample adaptive offset, high efficiency video coding, rate distortion optimization

1. Introduction

Sample Adaptive Offset (SAO) is a new coding tool of High Efficiency Video Coding (HEVC). It is an in-loop filter to reduce ringing effect and the distortion after the process of prediction coding and quantization. SAO can produce better coding efficiency and improve the quality of the videos [1]. Extreme correction (EXC) and band correction (BDC) were proposed to reduce the distortion based on the pixel classification in 2010 at the first meeting of JCT-VC [2]. Both techniques are effective to reduce the distortion as well as to improve the subjective quality. MediaTek's proposed each picture can be divided into multi-level quadtree partitions, and each leaf partition can be enhanced by band offset (BO) or edge offset (EO). Both BO and EO classify pixels of a partition into groups, and one offset is derived for each group [3]. SAO is located after deblocking and divides a picture into LCU (Largest Coding Unit)-aligned regions. Each region can be enhanced by either BO or EO with a localized parameter set. The statistical information for deriving offsets is collected, and the decision of all SAO parameters can be made right after the information collection by using fast distortion estimation [4]. Choi explore various SAO encoding policies including complexity reduction in the SAO parameter estimation and efforts to remove inefficiency caused by its implementation in a pipelined manner through numerous experimental tests [5]. In Yang's paper, the human visual characteristics represented by a JND (Just Noticeable Difference) model are introduced into the SAO optimization process for the first time [6].

¹ Shan Nana: School of Information Science and Technology, Taishan University. E-mail: helensnn@hotmail.com.

Most of those methods are based on pixels and lead to inefficient coding problems. This paper proposed a new SAO process by introduce visual saliency map to analysis the region of human visual attention. This proposed method will skip lots of useless calculation and improve coding efficiency with this strategy.

2. Sample Adaptive Offset

Figure 1 shows the ringing effect in the signal processing. The distortion is mainly caused by the loss of high frequencies of the original samples [7].

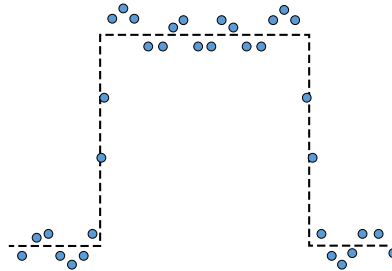


Figure 1. ringing effect

2.1. Types of SAO

SAO adapted in a 64×64 coding units and use different offsets in a region depending on the sample classification with three types including Edge Offset (EO), Band Offset (BO) and Merge [8].

a. edge offset

EO adapts four directional patterns for EO sample classification: 0° horizontal, 90° vertical, 135° diagonal, and 45° diagonal as shown in Figure 2 and named with EO_0, EO_1, EO_2, EO_3 separately.

Every sample will be classified into one of the five categories for the given EO class. And the roles are summarized in Table 1.

The meanings of edge offset are explained in Figure 3. Positive offsets for category 1 and category 2, while negative offsets for category 3 and category 4.

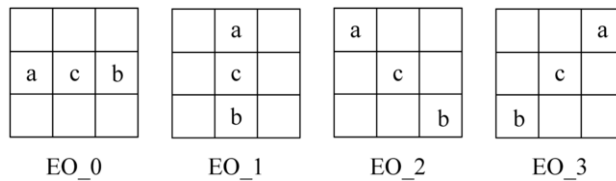


Figure 2. Edge Offset

Table 1. Category of Edge Offset

Category	Condition
1	$c < a \ \& \ c < b$
2	$(c < a \ \& \ c == b) \ \ (c == a \ \& \ c < b)$
3	$(c > a \ \& \ c == b) \ \ (c == a \ \& \ c > b)$
4	$c > a \ \& \ c > b$
0	Others

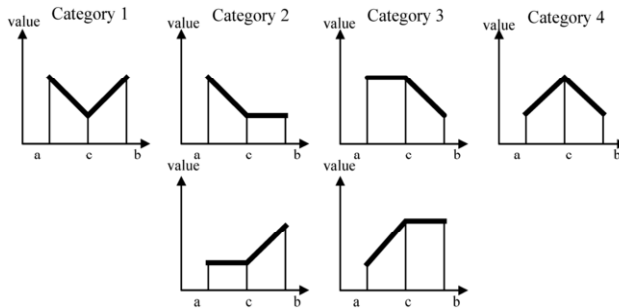


Figure 3. Offsets relations of Edge Offset

b. band offset

The sample value is equally divided into 32 bands in the band offset mode as Figure 4 shows. For 8-bit samples changing from 0 to 255, and the width of a band is 8, so the sample value from $8k$ to $8k+7$ belongs to band k . The value of band offset could be positive or negative. Four consecutive band are signaled to the decoder.

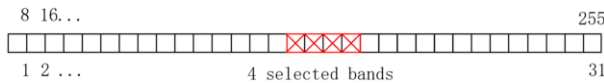


Figure 4. Band Offset

c. Merge

SAO also could use the parameters of their neighbor and transmit it with the syntax element to reduce the coding time. There are three options of them: reusing SAO parameters of the left CTU, reusing SAO parameters of the above CTU, or sending a new SAO parameters [9].

2.2. Process of SAO

As we mentioned before, the types of SAO are EO_0, EO_1, EO_2, EO_3, merge_left, merge_above and OFF. When the SAO processing is starting, the SAO parameter will be chosen among those eight types [10].

(x, y) represents the position of a certain pixel, $s(x, y)$ represents the value of the original pixels, $u(x, y)$ represents the value of the reconstructed pixels, and m represents the value of the offsets, c represents the range of the pixels, the distortion between the original pixels and the reconstructed pixels will be presented as:

$$D_{pre} = \sum_{(x,y) \in c} (s(x, y) - u(x, y))^2 \tag{1}$$

And the distortion between the original pixels and the reconstructed pixels after SAO will be presented as:

$$D_{\text{post}} = \sum_{(x,y) \in c} (s(x,y) - (u(x,y) + m))^2 \tag{2}$$

So, the difference of them is:

$$\Delta D = D_{\text{post}} - D_{\text{pre}} = Nm^2 - 2mE \tag{3}$$

N represent the number of pixels, E is the summation of the difference between the original pixels and the reconstructed pixels.

$$E = \sum_{(x,y) \in c} (s(x,y) - u(x,y)) \tag{4}$$

The delta rate distortion cost is defined in the following equation:

$$\Delta J = \Delta D + \lambda R \tag{5}$$

λ is the Lagrange multiplier, R represents the number of bits to encoding the SAO parameters.

3. Visual Saliency

Visual saliency can describe the most important factors in visual attention of human. An infrared small target detection algorithm based on visual saliency improved by spatial distance is proposed by Yang Linna [11]. The gray value of target is weighted by the ratio of the target pixel block and its surrounding blocks through saliency map. A new algorithm based on multi-scale fusion of the image high frequency information is proposed by Yang Xuye [12].

Seo presents a bottom-up approach and computes local regression kernels. Then the framework results in a saliency map where each pixel indicates the statistical likelihood of saliency of a feature matrix given its surrounding feature matrices [13].

They define saliency at pixel position x_i as a posterior probability as follows:

$$S_i = P_r(y_i = 1 | F) \tag{6}$$

F_i contains a set of feature vectors in a local neighborhood and $F_i = (f_i^1, \dots, f_i^{L_i})$. $F = (F_1, \dots, F_N)$ is a matrix containing features not only from the center, but also a surrounding region. If x_i is salient, y_i equals 1, otherwise y_i equals 0. Using Bayes' theorem, it could be written as:

$$S_i = P_r(y_i = 1 | F) = \frac{p(F | y_i = 1) P_r(y_i = 1)}{p(F)} \tag{7}$$

They adapted a locally data-adaptive kernel density estimator as follows:

$$S_i = \hat{p}(F | y_i = 1) = \frac{G_i(\bar{F}_i - \bar{F}_i)}{\sum_{j=1}^N G_i(\bar{F}_i - \bar{F}_j)} \tag{8}$$

G is weight function and can be computed as follows:

$$G_i(\overline{F}_i - \overline{F}_j) = \exp\left(\frac{-\|\overline{F}_i - \overline{F}_j\|_F^2}{2\sigma^2}\right) = \exp\left(\frac{-1 + \rho(F_i, F_j)}{\sigma^2}\right) \tag{9}$$

The matrix cosine similarity can be written as a weighted sum of the vector cosine similarities.

$$\rho_i = \sum_{l=1}^L \rho(f_i^l, f_j^l) \frac{\|f_i^l\| \|f_j^l\|}{\|F_i\|_F \|F_j\|_F} \tag{10}$$

Then, the saliency map from color channels can be analogously defined as follows:

$$S_i = \frac{1}{\sum_{j=1}^N \exp\left(\frac{-1 + \rho(F_i, F_j)}{\sigma^2}\right)} \tag{11}$$

4. Proposed Method

SAO have several directional modes as we mentioned before. In combination of the perceiving characteristics of human eyes for brightness, chroma, contrast and moving targets, an objective assessment method of video quality based on contrast sensitivity characteristics of human visual system was proposed by Yao Juncai [14]. Li proposed a visual attention-based bit allocation strategy to detect interesting regions in video [15]. There are two bottlenecks for the processing of SAO, one is the huge number of pixels at the statistic stage. The other is the high complexity in calculating the rate-distortion cost. An improved SAO algorithm that based on the visual saliency will be proposed in this paper to save the encoding time of SAO process.

4.1. Determination of SAO with visual saliency

As shown in figure 5, using visual saliency algorithm in [13] to generate visual saliency maps for the sequence BasketballPass (416x240, frame rate 50f/s). For the natural image, the bright region is more salient than the dark region. Red values in saliency map images represent higher saliency, while blue values mean lower saliency as we know.

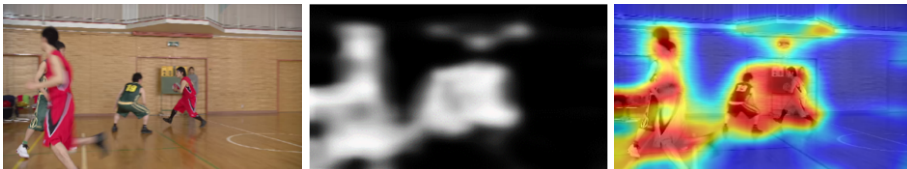


Figure 5. Image in a video and visual saliency map

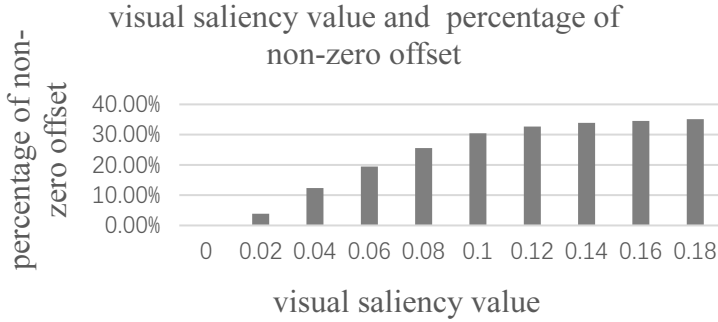


Figure 6. visual saliency value and percentage of non-zero offset in CTU

Figure 6 shows the statistics results of visual saliency average values and the sample adaptive offset which equals zero in a CTU of sequence BasketballDrive (1920x1080, 50f/s). It is important to note that in the regions with lower visual saliency, the sample adaptive offset has a very good chance to be zero, although with a series bitrate optimal calculation. It could be divided into many kinds of regions by their visual saliency values to skip some zero offset.

4.2. The processing of proposed algorithm

In the stand SAO algorithm, every pixel (with one luma value and two chroma value) will go through every kind of mode, including edge offset, band offset and merge. And it will be calculated in different categories. Those method leads to the problem of huge calculation complexity. We notice that the visual saliency statistics can be used as an important measurement for SAO. The proposed algorithm generates the visual saliency value of the pixels according to the visual saliency maps. And the average of visual saliency value will be calculated for every coding units. Then the coding units will be categorized by the average visual saliency values. For the regions with higher saliency, it attracts human visual attentions and is like to be a non-zero offset. On the contrary, for the regions with lower saliency, we can make early decision to stop the SAO process.

As shown in figure 6, there are about 30.42% non-zero offset values when visual saliency reaching 0.10, while the increasing process is getting slower. So, we set 0.10 as the SAO threshold.

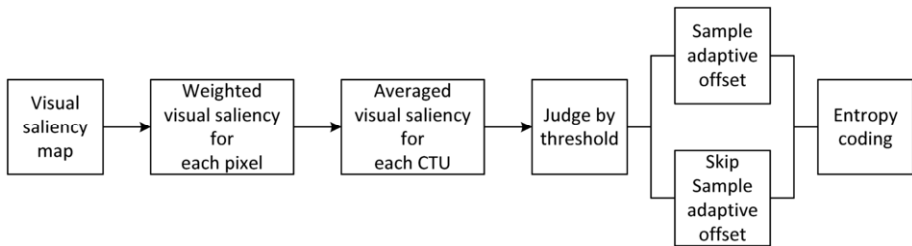


Figure 7. flow chart of proposed algorithm

Figure 7 shows the flowchart of the proposed method, it is explained in several stages as follows:

- Visual saliency map
- Weighted visual saliency for each pixel
- Averaged visual saliency for each CTU
- Judge by threshold: calculate sample adaptive offset or skip it
- Entropy coding

5. Experimental Results

This section presents experimental results to illustrate the benefits of proposed method.

5.1. Test Conditions

Those experiments were conducted using the HEVC test model version 16.0

- Five classes of test sequences
- 12 test sequences
- set quantization parameter with 22, 27, 32, 37
- using all intra configuration

5.2. Simulation Results

Table 2 shows the change of time, BD-rate and PSNR of the proposed SAO algorithm. It achieves about 27.02% encoding time saving only with 0.02 BD_rate and 0.0019 PSNR degradation.

Table 2. the result of proposed algorithm

Class	Size	Name	Δ PSNR/dB	Δ BR/%	Δ TIME/%
Class A	2560×1600	Traffic	-0.0007	0.01	-41.84
		Kimono	-0.0013	0.01	-36.53
Class B	1920×1080	BQTerrace	-0.0011	0.01	-33.44
		BasketballDrive	-0.0010	0.01	-32.09
		RaceHorsesC	-0.0013	0.01	-26.45
Class C	832×480	BQMall	-0.0019	0.01	-29.31
		BasketballDrill	-0.0021	0.01	-26.29
Class D	416×240	BasketballPass	-0.0034	0.03	-21.08
		BQSquare	-0.0025	0.02	-16.29
		Vidyo1	-0.0036	0.03	-21.60
Class E	1280×720	Vidyo3	-0.0011	0.02	-15.79
		Vidyo4	-0.0026	0.02	-23.53
Average			-0.0019	0.02	-27.02

5.3. Types of SAO

Figure 8 shows the pictures of a certain frame for both original HM algorithm and the proposed algorithm. From these pictures, we can see the proposed algorithm does not cause any notable visual distortions.



Figure 8. result of HM and proposed algorithm

6. Conclusion

In this paper, a relationship between SAO and visual saliency feature of the coding units is established by analyzing the SAO processing. Instead of calculate the SAO parameter one by one, the necessity will be attained by their visual saliency maps. The proposed method reduces SAO encoding time to average 27.02% only with trivial BD_rate and PSNR loss.

References

- [1] Sullivan G J, Ohm J, Han W J, et al. Overview of the High Efficiency Video Coding(HEVC) Standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668.
- [2] McCann K, Han W J, Kim I K, et al. Samsung's Response to the Call for Proposals on Video Compression Technology[C]//Proc. of ITU-T SG16 WP3 and ISO/ IEC JTC1/ SC29/ WG11, Dresden, German: JCT-VC, 2010: 1-3.
- [3] Fu Chihming, Chen Ching-Yeh, Huang Yuwen, et al. Picture Quadtree Adaptive Offset[C]// Proc. of ITU-T SG16 WP3 and ISO/ IEC JTC1/ SC29/ WG11, Daegu, Korea: JCT-VC, 2011: 1-10.
- [4] Fu ChihMing, Chen ChingYeh, C.Y. Tsai, Y.W. Huang, et al. Sample Adaptive Offset with LCU-Independent Decoding[C]// Proc. of ITU-T SG16 WP3 and ISO/ IEC JTC1/ SC29/ WG11, Geneva, Switzerland: JCT-VC, 2011: 1-6.
- [5] Choi Y, Joo J. Exploration of Practical HEVC/H.265 Sample Adaptive Offset Encoding Policies[J]. IEEE Signal Processing Letters, 2015, 22(4): 465-468.
- [6] Yang Kaifang, Wan Shuai, Gong Yanchao, et al. Perceptual Based SAO Rate-distortion Optimization Method with a Simplified JND Model for H.265/HEVC[J]. Signal Processing Image Communication, 2015, 31:10-24.
- [7] Fu Chih-Ming, Alshina E, Alshin A, et al. Sample Adaptive Offset in the HEVC Standard[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2012, 22(12): 1755-1764.
- [8] Benjamin B, Han Woo-Jin, Ohm J R, et al. High Efficiency Video Coding Test Model 10 (HM10) Encoder Descriptions[C]// Proc. of ITU-T SG16 WP3 and ISO/ IEC JTC1/ SC29/ WG11, Geneva, Switzerland: JCT-VC, 2013: 1-310.
- [9] Joo J, Choi Y, Lee K. Fast Sample Adaptive Offset Encoding Algorithm for HEVC Based on Intra Prediction Mode[C]//Proceedings of the 3rd IEEE International Conference on Consumer Electronics. Washington D. C., USA: IEEE Press, 2013: 50-53.
- [10] Eun-kyung R, Jung-hak N, Lee S, et al. Sample Adaptive Offset Parallelism in HEVC[J]. Multimedia and Ubiquitous Engineering, 2013, 240: 1113-1119.

- [11] Yang Linna, An Wei, Lin Zaiping, *et al.*. Small Target Detection Based on Visual Saliency Improved by Spatial Distance[J]. *Acta Optica Sinica*, 2015(7):246-251.
- [12] Yang Xuye, Li Xuewei, Zhang Libao, *et al.*. Saliency Region Detection of Remote Sensing Image Based on Multi-Scale Frequency Analyses[J]. *Acta Optica Sinica*, 2014(13):111-115.
- [13] Seo H J, Milanfar P. Static and space-time visual saliency detection by self-resemblance[J]. *Journal of Vision*, 2009, 9(12):1-27.
- [14] Yao Juncai, Liu Guizhong. Video quality objective assessment combined contrast sensitivity characteristics of human visual system[J]. *Optics and Precision Engineering*, 2016, 24(03):659-667.
- [15] Li Z, Qin S, Itti L. Visual attention guided bit allocation in video compression[J]. *Image & Vision Computing*, 2011, 29(1):1-14.