A New Saliency Based Video Coding Method with HEVC

Xu Zhengrong, Yu Mei, Fang Shuqing, Xu Shengyang

Faculty of Information Science and Engineering

Ningbo University, Ningbo China

xuzhengrong00@sina.cn, Yumei2@126.com, fsq_417@126.com, xsy_417@126.com

Keywords: Saliency; Quantization; HEVC

Abstract. In saliency based video coding, the salient parts of the each frame are encoded with higher quality than non salient areas to improve the signal to noise ratio and get higher compression ratio. In this paper, the background of the saliency based video perceptual coding technology is introduced first. Then the characteristics of motion, color and texture are extracted to form the saliency map. The quantization control in high efficiency video coding (HEVC) is incorporated with the visual saliency map to reduce the bit rate by adaptively changing the quantization parameter. Experimental results show that the proposed method can save the bit rate with improving the perceptual quality of regions-of-interest.

Introduction

Due to the rapid growth of the Internet and wireless technologies, video compression becomes essential for reducing the bandwidth for transmission and storage in many applications. In the case of limited bandwidth and storage resources, however, new requirements have been raised for the current video coding standard, such as higher resolution and image quality [1]. Therefore, higher efficient video compression technology is highly desired. As the new video coding standard after the H.264/AVC, HEVC standard has some new features, such as flexible coding structure and block structure, sample adaptive offset, adaptive loop filter and so on, making it more suitable for the higher resolution video. In addition, compared with H.264/AVC, HEVC doubles the data compression performance by reducing the bit rate by half without dropping the quality just at the cost of increasing computational [2].

With the study of human visual system (HVS) and rapid development of the prediction capacity in human visual perception, perceptual video coding has become one of the hottest research areas in signal and information processing. Currently, video coding techniques focus on taking advantage of the physical and psychological characteristics of the HVS combining the characteristic of the source itself in order to achieve a higher compression ratio and better subjective quality[3].Therefore, there is important theoretical significance in using visual perception principle to improve the quality of video coding, while reducing the computational complexity and the cost of bit resource[4]. Up to now, many scholars have carried out a lot of work on visual perception analysis, object extraction or fast video coding algorithm. Zhang et al [5], Kim et al [6], Chen et al [7] only utilized the texture, color, or contrast information when extract salient objects, but they ignored the high-level information in the real scene, such as movement and depth information. Xue et al [8] used sparse and low-rank matrix factorization to separate foreground from the background, which is also suitable for a variety of complex scenes and low resolution video while ensuring the integrity of the object. Hadi et al [9] proposed a method to reduce the coding artifacts of non salient region and guarantee attention in salient area, meanwhile, allow an increase saliency in high-quality portion and a decrease saliency degree in low-quality portion in each frame. Compared with the conventional video coding, the method improved the visual perceptual quality, however, since each macro-block shared equal value, failing reflecting the differences between significant macro-blocks, it also had limitations. PBAS method was proposed by Deng et al [10] to improve the quality by increasing the bit in salient region, but due to face detection was only used for conference or television interview, thus its applicability was limited. These video coding algorithms based on HVS technology mentioned

above concentrate on the objects extraction or the bit resource allocation but lack consideration on the bit resource allocation optimization under limited bit resources and pay little attention to the extra computational complexity caused by analyzing visual perception [11].

In this paper, we propose a new saliency based HEVC method combing visual perception characteristics, which utilizes the saliency information of video sequences and adaptively determines the quantization parameter based on visual perception characteristics. Compared with other means, there is no need to do accurate object extraction here. The experiment results show that integrating the saliency map into the encoding process can improve the quality of salient regions and save bit resource to a certain extent without affecting the overall perceptual quality.

Characteristics of HEVC Coding Structure

The block-based hybrid coding framework of HEVC, whose inter-frame, intra-frame coding and transformation techniques used to eliminate the redundancy of the time domain and spatial domain and entropy coding eliminate statistical redundancy. Compared to the traditional standards, the flexible quad-tree partition structure used by HEVC increases the maximum size to 64*64, meanwhile, whose varieties of coding tools make HEVC standards greatly improve coding efficiency [12].

The block partition structure of HEVC is described briefly in the following.

A. Partition of LCU

First of all, each frame will be divided into a series of LCU (Largest Coding Units, the maximum coding unit) in the coding process. Each LCU is consist of a luminance block of size N*N and two chrominance blocks corresponding. The concept of LCU in HEVC is similar to macro-block in H.264/AVC, where N is 64.

B. Coding Unit CU

As the basic encoding unit, the size of CU (Coding Unit, encoding unit) ranges 8*8 to 64*64, it takes quad-tree iterative form to divide. The division process is adaptively, according to the video content, larger block structure is used in the flat region and smaller block structure is used in complex texture or motion intense region.

Due to the larger LCU, HEVC is suitable for high resolution and ultra high definition video, especially in the large area of texture in accordance, bigger block structure can greatly save the coding resource.

C. Prediction Unit PU

As the basic unit of intra and inter prediction, PU is not limited to the size of the square so as to split easily and fit object's edge. Each CU contains one or more PU, the maximum size is equal to the size of CU, and the minimum size is 4*8 or 8*4. Only when the size of corresponding CU is greater than the minimum allowable size, the shape of PU is square.

D. Transform Unit TU

As the basic unit of transform and quantization, the shape of TU is closely linked to division of PU. TU is square when PU is same and its luminance block's size may range from 4*4 to 32*32. TU is rectangular as well as PU and its block's size can be 32*8, 8*32, 14*4 and 4*16. Non-square TU is also applicable to chrominance blocks when the size of the corresponding luminance block is 8*32 and 32*8.

Each frame is divided into a series of CU and the size of CUs is different. Generally, the size of CU is small when the CU belongs to a region with complex texture, whereas the size of CU is big when the CU belongs to a flat region. Choosing an appropriate QP for the CU during the quantization process will directly affect the encoding quality. Traditional methods choose the same QP value for the different CUs, and then the encoding quality of different regions shares the same level. It's not in conformity with the laws of perception. In fact, the human eyes only pay special attention to some regions in the scene, while seldom or never focus on other regions. Those special regions we call them salient regions. The quality of entire region mostly depends on the quality of salient region. So, we can improve the perceptual quality of entire region by enhancing the encoding quality of salient region, namely, adjusting the QP for the CU which belongs to the salient region.

A New Saliency Based HEVC Method using Visual Perception Characteristic

With visual perception characteristics, a new saliency based HEVC method is proposed, its framework is illustrated as Figure 1.

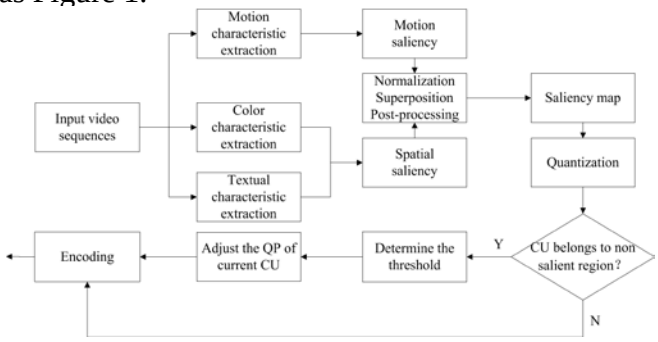


Figure 1. Framework of coding method combing visual perception characteristics

A. Detection of salient regions

Salient regions refer to the region that attracts the highest probability of concerns when the human eyes view a scene. Here we utilize robust temporal alignment and local-global spatial contrast method to get the initial saliency map, and then improve the quality of initial saliency map by post-processing to get final saliency map.

The saliency calculation algorithm consists of three steps:

Step-1: Get the saliency map

Motion information is important to perceptual coding, however, motion detection is critical and directly taking the difference between adjacent frames as salient motion may not satisfy the need. Here, according to [3], we detect salient motion map I_M by robust alignment. We take advantage of robust alignment by sparse and low-rank decomposition to jointly estimate the salient foreground motion and the camera. Consecutive frames are transformed and aligned, and then decomposed to a low-rank matrix representing the background and a sparse matrix indicating the objects with salient motion.

Histogram statistics is used to extract color information. Specifically, the saliency value of a pixel is defined by the color contrast between the other pixels. In this way, the pixels of the same color will have the same saliency value. Therefore pixels of similar color are gathered to form salient color map I_C.

Finally, we perform textural feature extraction based on gray histogram to form salient textural map I_T.

Step-2: Normalization and Superposition

For each frame, we extract the feature maps and compute a set of initial saliency maps, and then normalization and superposition are taken to form the final saliency map. Different weights are given to three feature maps when the final saliency map is synthesized.

$$I_{\text{sal}} = \mu_M I_M + \mu_C I_C + \mu_T I_T \qquad (1)$$

Here, I_{sal} stands for the final saliency map, I_M, I_C, I_T represent motion map, color map and texture map, and μ_M, μ_C, μ_T denote the corresponding weights. Since the human eye is most sensitive to motion information, here we take μ_M as 0.4, μ_C and μ_T as 0.3.

Step-3: Post-processing

Considering the salient areas are generally distributed in the central part of the scene and the surrounding part is little focused on by the human eye, consequently, in order to reduce complexity and the bit rate of the CU far away from the focus, we set the saliency of these CUs to zero.

B. Adaptive QP selection for salient or non salient regions

Each frame is divided into a series of coding units to process in HEVC, and the size of CU has a relationship with quantization parameter (QP) and quantization step, in addition, the value of quantization parameter and quantization step affect the number of bits allocated to each CU.

Obviously, the more bits allocated to the region, the higher the quality is. According to the saliency map obtained above, a CU will get more attention when it belongs to a salient region and it will be allocated more bits to make encoding quality as high as possible. Conversely, the human eye is not sensitive to a CU when it belongs to a non salient region, it's possible to save unnecessary bits at the cost of lower quality in these regions. While quantization and bit allocation are closely related, we can adaptively adjust the QP of non salient region to reduce the bits. The range of QP is from 0 to 51 in HEVC, and there is an exponent relationship between quantization step and quantization parameter. In addition, the larger value of the QP, the fewer bits allocated to CU, of course, lager distortion of the image. As a result, it is need to increase the QP of non salient region and reduce bit rate without declining the overall perceptual quality.

The process of adaptive QP selection for different regions is explained as follow.

First, computing the saliency of coding units according to the saliency map obtained above:

$$S_{CU} = \frac{1}{16 \times 16} \sum_{x=1}^{16} \sum_{y=1}^{16} S(X) \tag{2}$$

Determine the threshold value based on the distribution of saliency:

$$S_{Level} = \begin{cases} 1, & if \quad S_{CU} > T_1 \\ \ldots & \\ m, & if \quad T_{m-1} < S_{CU} < T_m \\ \ldots & \\ M, & if \quad S_{CU} > T_{M-1} \end{cases} \tag{3}$$

Here, in the formula (3), as the threshold relates to the selection of the quantization parameter, how to select the threshold is critical.

QP plays an important role in the encoding process, adjusting QP value adaptively to optimize the quantization process and guarantee the quality of reconstruction video is main purpose. Selecting an appropriate initial QP for a given frame is crucial,

$$QP_{\mathrm{i}} = round\left(\frac{QP_{frame}}{\sqrt{\omega_i}} \right) \tag{4}$$

Here, QP_{frame} stands for the QP of given frame, QP_{i} stands for the initial QP of CU. ω_i stands for an adjustment factor related to saliency, just taken as [9],

$$\omega_{\mathrm{i}} = a + \frac{b}{1 + \exp\left(-c\left(S_{CU} - \bar{s} \right) / \bar{s} \right)} \tag{5}$$

After the initial QP of each CU is set up, we will do a fine-tuning based on the saliency of CU. The QP is adjusted according to the following formula,

$$QP = \begin{cases} QP_i, & S_{Level} = 1, \\ QP_i + \left(S_{Level} - 1 \right) \times \Delta QP, & S_{Level} \geq 2 \end{cases} \tag{6}$$

According to the saliency map obtained above, we do a fine adjust of the initial QP, here ΔQP has a correlation with salient level of CU and initial QP of the frame. First of all, the variation range of ΔQP is limited within 3 when the initial QP is relatively small, for example, the initial QP is 22. If the initial QP is medium such as 27 and 32, the variation range is limited within 6. And when the initial QP is large, the range of variation is limited within 9. With the limitation of above, the following points are considered. If the types of adjacent coding units are the same, in other words, salient and non salient, their ΔQP must not exceed 3; if the types of adjacent coding units are not same, in order to reduce the block artifact and ensure the quality smoothing between the salient region and non salient region, we adopt the average QP of two coding units.

Experimental results and analyses

The performance of adaptive quantization approach based saliency map is evaluated in this section. Several standard HEVC test sequences with different spatial resolution are carried out in our experiments, whose information is shown in TABLE I. The simulation is implemented with the reference HEVC software HM12.0. Based on the conditions and instructions of the encoder, coding parameters are set up as following, MaxCUWidth =16, MaxCUHeight =16, IntraPeriod = 1, MaxPartitionDepth=1, GOPSize =1, and so on.

TABLE I. INFORMATION OF TEST SEQUENCES

Information of test sequences			
Sequence	***Resolution***	***Frame number***	***Frame rate***
BasketballDrill	832*480	500	50
RaceHorses	416*240	300	30
BQMall	832*480	600	60
PeopleOnStreet	2560*1600	150	30
Johny	1280*720	600	60

The extracted saliency maps are listed in Figure 2.

As shown in Figure 2, (a)(d)(g) express original frame from test video, (b)(e)(h) express the corresponding saliency map and (c)(f)(i) express the extracted salient region. It can be seen basically in line with the law of human visual system. Compared with other means, there is no need to do accurate object extraction here.

In order to reflect the difference between original and modification intuitively, RD curves are drawn and shown in Figure 3, where the curve A represents the result of HM12.0 and the curve B illustrates the result of the proposed method.

The RD curves of encoding algorithm with different test sequences are shown in Figure 3. It is clear that the proposed method can improve the encoding quality and save bit resource to a certain extent when we integrate the saliency map into the quantization process. In particular, we can know that our method does better at low bit rate indicated as Figure 3(a). That is because the overall quality would not be high under limited bit resources, on this occasion, highlighting the salient region can enhance the overall perceptual quality. The encoded results of intermediate frame are listed in Figure 4. QP_{SVA} and QP_{BG} refer to the quantization parameter of salient region and background region respectively, $PSNR\_Y_{SVA}$, $PSNR\_Y_{BG}$ and $PSNR\_Y$ stand for the peak signal to noise ratio of Y component in salient region, background region and entire region respectively. It can be seen from the presented figures that the overall subjective quality doesn't decrease, even though at the cost of a drop in PSNR of overall region. The increasing QP value of the background region leads to decreasing PSNR in these regions when the two sequences after treatment. Furthermore, the increasing QP of background region results in a decline in bit use ratio, more bits are allocated to the salient region so that the PSNR of these regions will increase. The results demonstrated in Figure 4 as expected. It is evident that integrating the saliency map into the encoding process can improve the quality of salient regions and save bit resource to a certain extent without affecting the overall perceptual quality.

Conclusion

In this paper, a new saliency based HEVC method using visual perception characteristics is proposed. First, motion characteristic, color characteristic and textual characteristic are extracted to form the saliency map. Second, we integrate the saliency map into the quantization process, improving the encoding quality by adjusting the QP of coding units. The results indicate that the proposed method is able to reduce the rate by adjusting the QP of coding unit in non salient regions without perceptual quality loss. However, we have ignored depth information of real world during the encoding. Consequently, the framework could be further enhanced by integrating the depth saliency item into the current saliency map.

Acknowledgement

This work is supported by the Natural Science Foundation of China (Grant Nos. 61271270, 61311140262).

References

[1] Muller K, Schwarz H, Marpe D, et al. 3D high-efficiency video coding for multi-view video and depth data, IEEE Transactions on Image Processing, 2013, 22(9): 3366-3378.

[2] Wulf S, Zolzer U. Extension of a visual saliency guided bit allocation approach using laplace distribution of DCT coefficients, 56th International Symposium ELMAR (ELMAR), 2014: 1-4.

[3] Ren Z, Chia L T, Rajan D. Video saliency detection with robust temporal alignment and local-global spatial contrast, the 2nd ACM International Conference on Multimedia Retrieval. ACM, 2012: 47.

[4] Borji A, Itti L. State-of-the-art in visual attention modeling, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 185-207.

[5] Zhang Y, Mao Z, Li J, et al. Salient region detection for complex background images using integrated features, Information Sciences, 2014.

[6] Kim J, Grauman K. Shape sharing for object segmentation, Computer Vision–ECCV 2012, 2012: 444-458.

[7] Chen S, Shi W, Zhang W. Visual saliency detection via multiple background estimation and spatial distribution, Optik-International Journal for Light and Electron Optics, 2014, 125(1): 569-574.

[8] Xue Y, Guo X, Cao X. Motion saliency detection using low-rank and sparse decomposition, International Conference on Acoustics, Speech and Signal Processing, 2012: 1485-1488.

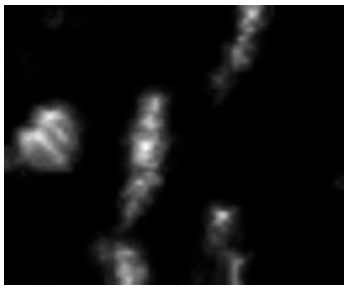
[9] Hadi Hadizadeh, Ivan V.Bajic. Saliency-Aware video compression, IEEE Transactions on Image Processing. 2014, 23(1):19-31

[10]Deng X, Xu M, Wang Z. A ROI-based bit allocation scheme for HEVC towards perceptual conversational video coding, Int. Conf. on Advanced Computational Intelligence, 2013: 206-211.

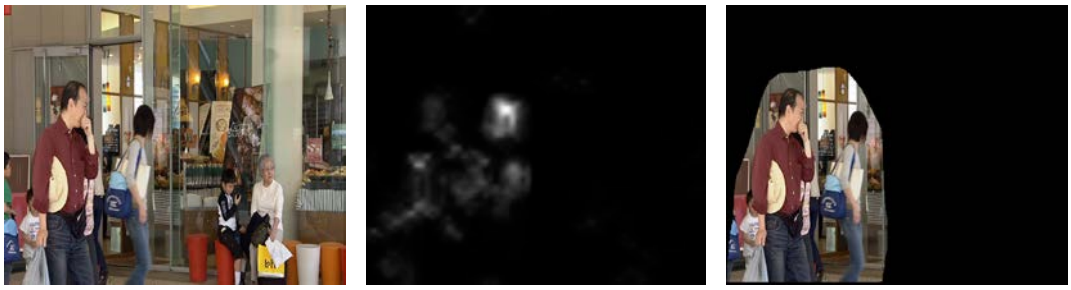
[11]Seo C W, Moon J H, Han J K. Rate control for consistent objective quality in high efficiency video coding, IEEE Transactions on Image Processing, 2013, 22(6): 2442-2454.

[12]Sullivan G J, Ohm J, Han W J, et al. Overview of the high efficiency video coding (HEVC) standard, IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668.

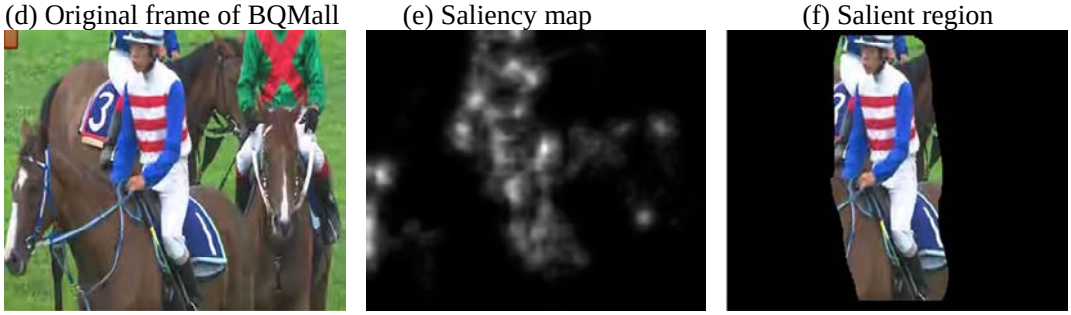

(a)Original frame of Basketball-Drill


(b) Saliency map


(c) Salient region



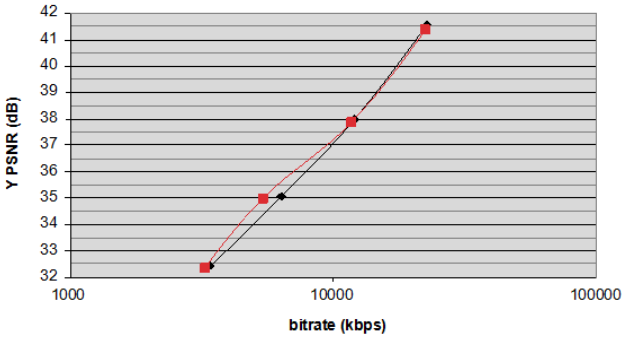
(d) Original frame of BQMall (e) Saliency map (f) Salient region



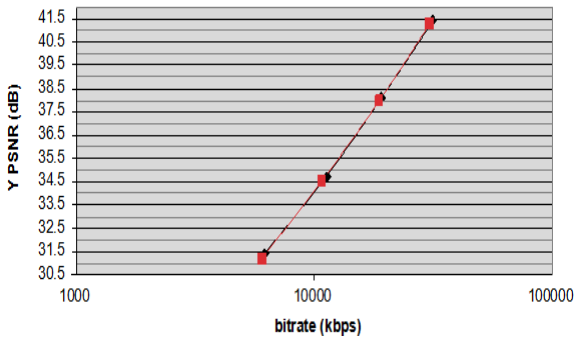
(g) Original frame of RaceHorses (h) Saliency map (i) Salient region

Figure 2. Original frame and corresponding saliency map, (a)(d)(g) stand for one frame of the original sequences, (b)(e)(h)

denote the corresponding saliency map and (c)(f)(i) illustrate the extracted salient region.



(a)RD curve of BasketballDrill



(b)RD curve of BQMall

Figure 3 RD curves of test sequences

HTM12.0
(QP_{SVA}: 27, QP_{BG}:27
PSNR_Y_{SVA}: 37 .93 dB,
PSNR_Y_{BG}: 38.25dB,
PSNR_Y: 37.98dB)

proposed
(QP_{SVA}: 27, QP_{BG}:33
PSNR_Y_{SVA}: 38.02dB, ΔPSNR_Y_{SVA}:**0.09**dB
PSNR_Y_{BG}: 37.88dB, ΔPSNR_Y_{BG} :**-0.37**dB
PSNR_Y:37.92dB)

(a) BasketballDrill

HTM12.0
(QP_{SVA}: 27, QP_{BG}:27
$PSNR\_Y_{SVA}$: 37.68 dB,
$PSNR\_Y_{BG}$:38.34dB,
PSNR_Y: 38.13dB)

proposed
(QP_{SVA}: 27, QP_{BG}:33
$PSNR\_Y_{SVA}$: 38.14dB, $\Delta PSNR\_Y_{SVA}$:**0.46**dB
$PSNR\_Y_{BG}$: 37.75dB, $\Delta PSNR\_Y_{BG}$:**-0.59**dB
PSNR_Y: 38.08dB)

(b) BQMall

Figure 4.Results comparison between the HTM12.0 and the proposed method