Jestr

Journal of Engineering Science and Technology Review 10 (5) (2017) 108 - 115

JOURNAL OF
Engineering Science and Technology Review

Research Article

www.jestr.org

Optimized Adaptive Encoding Based on Visual Attention

Zhao Hong[1], Cao Chang[2], Li Jing Bo[3] and Xiang Yan Zeng[4]

[1] *School of Computer and Communication, Lanzhou University of Technology, Lanzhou, 730050, China*
[2] *School of Computer and Communication, Lanzhou University of Technology, Lanzhou, 730050, China*
[3] *School of Electronics and Information Engineering, Xi'an Jiao tong University, 710049, China*
[4] *Department of Mathematics and Computer Science, Fort Valley State University, Fort Valley, GA 31030, USA*

Received 9 July 2017; Accepted 29 September 2017

Abstract

Visual redundancy, which acts on the low-attention areas of images, can be applied to video encoding due to the selection mechanism of the human eyes, thus improving efficiency. To reduce the impact of visual redundancy on video coding, a novel method of distinguishing the level of image attention was proposed in this study. The method was used to estimate visual attention according to the sensitivity of human eyes to the motion, texture, contrast, and brightness of images. Then, different coding strategies were adopted according to the different visual attention levels of the coding blocks. The structural similarity index algorithm was applied to high-attention coding blocks; the visual attention coefficient was employed to refine the Lagrange multiplier so that the quantizer can adopt a larger quantization step for low-attention coding blocks. Results show that the coding bit rate is reduced by an average of 30.33% when the luminance peak signal-to-noise ratio increments are reduced by merely 0.11 dB and the coding time is increased by only 0.75%. These results indicate that visual redundancy has a considerable influence on video coding efficiency. Thus, the proposed method provides a bright prospect for optimizing the design of encoding strategies.

Keywords: Video coding, Visual redundancy, Visual attention, Lagrange multiplier

1. Introduction

Video coding performance depends on three factors, namely, coding rate, compression distortion, and computational complexity, in which compromise and optimization design are fundamental. Rate-distortion optimization (RDO) determines not only the best balance between the bit rate (BR) and the video image distortion but also the optimal coding parameters for the encoder [1]. The good performance of RDO has made it extensively used for encoding strategies, and its design has attracted significant attention. Therefore, the design of the RDO algorithm is an interesting and significant research subject. Previous studies showed that the correction of the Lagrange multiplier is a popular method for RDO and frequently used to improve the efficiency of coding compression [1]

In the existing literature, the emphasis on RDO design is mainly on the determination of the Lagrange multiplier [2]. This process has become a challenge to coding algorithm designers because of numerous influencing factors, such as the texture features of the image and the complexity of video content. Meanwhile, the computational complexity of this process is also restrictive. In addition, the subjective perception of a video is affected by various factors, such as motion, contrast, color, texture, and brightness, and the masking effects of space, time, and color. Therefore, the influence of subjective visual redundancy on the correction of the Lagrange multiplier cannot be ignored [3].

On the basis of the above analysis, this study investigates a core problem in the design of the RDO algorithm according to visual attention.

2. State of the art

Traditional video coding methods focus mainly on processing redundant information in video signal and encoding optimization algorithms on the basis of video encoding standards. Zhao Hong et al. applied the Canny operator to segment images by taking advantage of the video texture feature, thus effectively reducing the coding complexity and time. Their method predicted the coding depth according to the distribution of the coding unit such that depth judgment was terminated in advance [4]. High-efficiency video coding (HEVC) defines 35 intra prediction modes, thus having a high coding complexity. Duanmu C J et al. proposed an algorithm for effectively shrinking the number of candidate modes to be checked and consequently reducing the complexity of HEVC. They utilized edge detection and Hough transform for the prediction unit with different sizes and statistical analysis for the detected edge line angles to determine the candidate modes to be checked [5].

Although traditional coding algorithms perform well, they ignore the key point. That is, the final recipients of videos are human eyes, which implies that the perceived quality of videos is inevitably affected by the human visual system (HVS). Therefore, cues from HVS can be used for further compression optimization in modern hybrid video coding platforms and were an effective means of reducing complexity [6]. Furthermore, the accurate estimation of

*E-mail address: caochang1012@163.com
ISSN: 1791-2377
doi:10.25103/jestr.105.14

visual redundancy in video images is important for the design of video coding algorithms. Currently, most analytical methods for video coding regard visual redundancy segments as video images according to visual features, and then encode different image areas [7]. Kalva H et al. explored and exploited motion-related attentional limitations and developed algorithms for exploiting motion-triggered attention [8]. Guraya F F E et al. used a state-of-the-art visual attention model developed by combining bottom-up, top-down, and motion cues [9]. However, visual attention is developed specifically for surveillance videos. Li F et al. divided the macroblock (MB) into region-of-interest (ROI) MB and non-ROI (NROI) MB according to human visual features; for NROI MB coding, the active MB concealment (AMC) mode in RDO was proposed; AMC trades off the quality of the NROI MBs with the rate, distortion, and improved quality of the ROI MBs at the cost of the quality decreasing of the NROI MBs, thus achieving rate control on the basis of the MB [10]. Hua K L et al. presented a novel block-based image coding algorithm that applied a tree-structured multi-tree dictionary and a perceptual rate distortion optimization scheme [11]. Although multi-tree dictionary is employed to support various tailings, perceptual rate distortion optimization utilizes the structural similarity index (SSIM) metric instead of the popular mean squared error metric to allocate the BR according to HVS.

Furthermore, these methods are limited to 2D videos and exhibit poor performance in 3D animation. Guillotel P et al. proposed a new perceptual coding scheme that considered the HVS. They include perceptual distortion measures in the encoding loop to compute the adaptive local quantization step size and optimize the choice of MB quantization parameters on the basis of HVS [12]. Jin G et al. proposed a coding scheme that jointly applied perceptual quality metrics to prediction, quantization, and RDO within the HEVC framework. They introduced a new prediction approach that used template matching, which employed an SSIM and the just-noticeable distortion model. The matched candidates were linearly filtered to generate a prediction [13]. These methods either fail to adopt exclusive coding schemes for different visual attention areas or fail to realize the adaptive adjustment of the Lagrange multiplier.

The above methods consider the impact of HVS on video coding. However, videos often have sound, and video sound also impact the focus of human eyes on video images. Lee J S et al. proposed an efficient video coding method using audiovisual focus of attention, which was based on the observation that the sound-emitting regions in an audiovisual sequence draw viewer attention. First, an audiovisual source localization algorithm was presented, in which the sound source was identified using the correlation between the sound signal and the visual motion information. The localization result was then used to encode different regions in the scene with different qualities such that the regions closed to the source were encoded with higher quality than those far from the source to reduce redundant high-frequency information and achieve coding efficiency [14–16]. This method increases computational complexity, that is, the accurate estimation of high-visual-attention areas in video images, and makes the synchronization problem of sound and image difficult to consider because the influence of sound is considered in video coding. On the basis of the analysis above, the audiovisual model is not adopted by this study.

The redundant information in the video is divided into video signal redundancy and visual redundancy. These existing methods have been successfully applied to video signal redundancy processing, whereas few studies have taken note of visual redundancy. The color, brightness, contrast, and texture in 3D animation are robust; thus, visual redundancy research is particularly important [17]. This study proposes a novel adaptive encoding algorithm that is based on visual attention. Different coding strategies are adopted according to the different visual attention of coding blocks to reduce BR allocation to low-visual-attention areas and improve coding efficiency.

The remainder of this study is organized as follows. Section 3 establishes the adaptive coding compression model and proposes the method for calculating the visual attention factor. Section 4 discusses and analyzes the experimental results and the performance of this method. Section 5 summarizes the conclusions.

3 Methodology

3.1 Visual Attention Based on HVS

Humans can stare at a natural scene and choose the information of interest. They usually notice the rest of the field while paying attention to their target of interest. This selection mechanism has been widely used in computer vision research [18]. Visual perception is controlled by the mechanism of the selective attention of the brain, which has the option of maintaining and stimulating certain stimuli while ignoring others [19]. Therefore, the key to attention detection is establishing an attention model that simulates the selective attention mechanism of the HVS.

Studies have shown that the HVS has a low degree of attention to flat and non-moving regions in an image and the information in these areas are easily ignored by human eyes. On the contrary, areas with many changes and abundant details are more likely to capture attention in the visual system than flat regions. Meanwhile, people tend to look for obvious target features as influenced by psychological factors [4]. To a certain extent, the selective attention mechanism of obtaining information plays a role in information compression.

Calculating visual attention is an important step in adaptive compression, and the computational complexity should not be excessively high. In this study, a visual attention model is established by simulating the visual perception process. In the following, four HVS characteristics are studied: motion, brightness, texture complexity, and contrast factor.

3.1.1 Motion Factor Based on Gray Projection Method

Studies have shown that when a region of the video moves relative to the background, human eyes track these motion areas subconsciously. This behavior is called the smooth pursuit eye movement. Observers are more sensitive to areas with moving objects. The human eyes perceive distortions in these areas easily. On the contrary, distortions in non-moving regions have a relatively slight effect on visual perception [20].

In the case of a stationary background, the frame difference-based method has been used to extract motion areas and achieved good results. However, the general method based on frame difference cannot get a good segmentation effect when the foreground and background are moving simultaneously [4]. Therefore, to adapt to the

identification of various occasions by the motion areas, this study uses a low-computational-complexity and robust algorithm to calculate the motion factor (MF).

The gray projection method (GPM) is a simple and effective method that can estimate the global motion vector. This method projects a 2D image into the X and Y directions to obtain the projection curve and then finds the maximum cross-correlation value according to the gray scale projection curve of the current and previous frames. This method is widely used in electronic image stabilization [21].

GPM is used to obtain the motion vector of the coding unit of the current frame to analyze the motion areas in the image. The MF is shown in Eq. (1).

$$MF = \frac{GV_x^2 + GV_y^2}{P \times \max\{GV_x, GV_y\}} \tag{1}$$

where P is the current coding unit size, $\max\{GV_x, GV_y\}$ represents the larger of the two, and MF is the motion factor or the ratio of the moving distance in the current coding unit to the moving distance sin the direction.

GPM has strong robustness to foreground extraction because of the statistical properties of the adjacent frames it uses. Moreover, the GPM has low computational complexity and can be applied easily to real-time systems.

3.1.2 Texture Factor

According to the study of visual psychology, similar to the human desire to explore, human eyes pay more attention to areas of complex texture. Texture complexity is introduced as an important factor in the formula of attention.

Luminance histogram represents the probability of each intensity level in the image, and the distribution range of each brightness level can represent the texture flatness degree of the image [10]. The highest frequency level of brightness ($B_{popular}$) must exist in the luminance histogram. The ratio of a certain range of pixels distributed around $B_{popular}$ to the total number of pixels denotes the texture smoothness of the largest coding unit (LCU), which is the texture factor in Eq. (2).

$$TF = 1 - \frac{1}{M} \times \sum_{i=B_{popular}-8}^{B_{popular}+8} N_i \times 100, i \in 0,1,\mathrm{L}\ ,L-1 \tag{2}$$

where M is the total number of pixels within the LCU, N_i is the total number of pixels at brightness level i in the LCU, and L is the level of total brightness. L is equal to 256 when the 8-bit quantization of histogram statistics is used.

3.1.3 Contrast Factor Based on Four-Neighbor Method

Studies have shown that human vision is more sensitive to color contrast, which is another important factor that captures human visual attention [18]. Therefore, contrast is introduced into the attention calculation as an important influence factor. According to the different areas of contrast calculation, contrast can be divided into two categories: local and global contrast. Local contrast is a measure of the difference in each pixel relative to their neighboring pixel, whereas global contrast is the spatial distribution of the image area color.

Inspired by the contrast algorithm in [22], the four-neighborhood method is used to calculate the regional-level global contrast factor (CF) while ignoring unnecessary texture information.

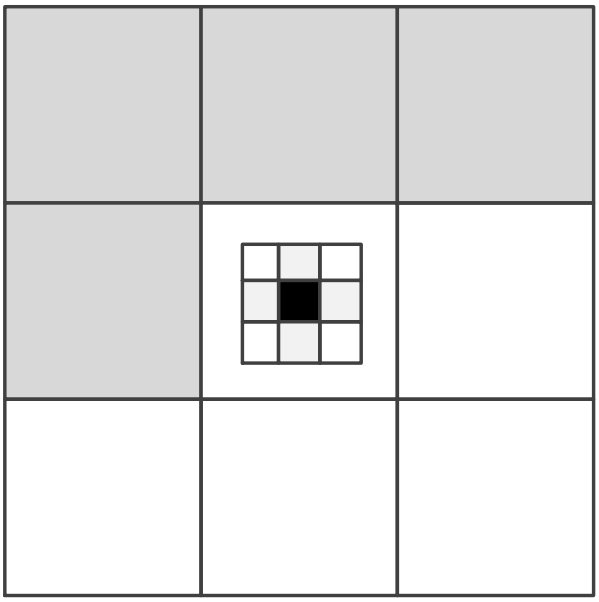

Fig. 1. Illustration of CFN and PFN

Fig. 1 presents a schematic of the coding unit four-neighbor method (CFN) and the pixel four-neighbor method (PFN), in which the black center part represents the current pixel of the PFN, with the same size on the upper, lower, and left sides. In addition, the right gray part represents the four-neighbor pixels. The upper left, upper, upper right, and the large left gray portions represent the four-neighbor coding units (CU) of the CFN.

First, the PFN is used to calculate the difference (COT) between the pixel and the neighboring pixels, as shown in Eq. (3).

$$COT = \frac{\sum \left| C_{(i,j)} - C_{(i',j')} \right|}{P_{cur\_cu}} \tag{3}$$

where $C_{(i,j)}$ represents the chroma value of the current pixel, $C_{(i',j')}$ stands for the chroma value within PFN, and P_{cur_cu} indicates the total number of pixels in the current CU.

Then, CFN is used to calculate the chromaticity change rate of the four-nearest-neighbor contrast of the CU, and the CF of the color is fully sensed, as shown by Eq. (4). Compared with the local contrast, the algorithm has higher reliability and better contrast calculation ability.

$$CF = 4 \times \frac{COT_{cur\_cu}}{COT_{left\_up} + COT_{up} + COT_{right\_up} + COT_{left}} \tag{4}$$

where COT_{cur_cu} represents the chroma contrast of the current CU, COT_{left_up} stands for the contrast of the upper left side CU, COT_{up} means the contrast of the upper side CU, COT_{right_up} is the contrast of the upper right side CU, and COT_{left} represents the contrast of the left side CU.

3.1.4 Luminance Factor

Studies have revealed that brightness affects visual perception. When the gray value of the scene is between 56 and 108 in the region and the image changes are relatively slow in time and space, the human eyes have the strongest resolution and the visual sensitivity is relatively high. Under high or low brightness, the capability of human eyes to distinguish is subject to a certain impact and visual sensitivity is relatively low [23]. Meanwhile, the subjective sense of brightness depends not only on the actual brightness of the scene but also on the average brightness of the surrounding environment, which means that the subjective sense of the same brightness is different in different environments.

On the basis of this theory, the CFN algorithm is used to calculate the luminance factor (LF), as shown in Eq. (5).

$$LF_{avg} = 4 \times \frac{\sum_{(i,j) \subset cur\_cu} L_{(i,j)}}{\sum_{(i,j) \subset CFN} L_{(i,j)}} \tag{5}$$

where LF_{avg} indicates the average value of LF, $\sum_{(i,j) \subset cur\_cu} L(i,j)$ represents the total luminance of the luminance pixels in the current CU, and $\sum_{(i,j) \subset CFN} L(i,j)$ indicates the total pixel brightness in the four neighborhoods of the CU.

3.1.5 Calculation of Attention

By obtaining the weighted average of the above key factors, the saliency factor (SF) is obtained as shown by Eq. (6).

$$\begin{cases} SF = \alpha \times MF + \beta \times TF + \delta \times LF + \gamma \times CF \\ \alpha + \beta + \delta + \gamma = 1 \end{cases} \tag{6}$$

Numerous experimental results show that the best visual SF can be obtained when $\alpha = 0.41$, $\beta = 0.25$, $\delta = 0.22$, and $\gamma = 0.12$.

When the calculated SF is within the thresholds TH_l and TH_h and it is a high-attention CU, the Attention is 1. By contrast, when the SF is not in the interval [TH_l, TH_h], it is low-attention CU, as shown in Eq. (7).

$$Attention_i = \begin{cases} 1.04 \times SF + 0.13 & SF \le TH_l \\ 1 & TH_l < SF \le TH_h \\ -0.52 \times SF + 0.09 & SF > TH_h \end{cases} \tag{7}$$

where $Attention_i$ represents the attention of the i-th coding unit. The best attention division can be obtained when $TH_l = 0.83$ and $TH_h = 2.1$.

3.2 Adaptive Coding Compression

Different coding schemes are adopted to optimize the HEVC for different attention CUs after obtaining the attention value of the CU. For the high-attention CU, the SSIM rate distortion optimization algorithm is used as a measure factor of distortion and video clarity to reduce the coding computational complexity [12]. For the low-attention CU, the Lagrange multiplier is adjusted to correct the size of the quantization step and achieve a larger quantization step for the coefficients of the visual insignificant region and filter the partial high-frequency components to reduce the coding rate.

3.2.1 High-attention CU

In SSIM, the average of the pixels is used as the estimation of brightness, the standard deviation as the estimation of contrast, the covariance as a measure of the structural similarity, and the distortion as a combination of brightness, contrast, and structural similarity, as shown in Eq. (8) [24].

$$SSIM = \left(\frac{2\mu_x \mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \right) \left(\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \right) \tag{8}$$

where *SSIM* is the structural similarity evaluation function, μ_x and μ_y denote the two sequence pixel mean values, σ_x and σ_y are the mean standard deviations of unbiased estimation, σ_{xy} is the covariance of *x* and *y*, and c_1 and c_2 are the regulatory parameters. In addition, the value range of SSIM is [0, 1]. The closer the value is to 1, the better the quality of the CU block can be.

For a high-attention CU, the Lagrange multiplier of the rate distortion is obtained using the dynamic rate distortion model mentioned in [13].

First, the relationship between the distortion and quantization parameters is derived from many experimental statistics, as shown in Eq. (9).

$$D_{SSIM} = 10^{-3} \times 2.95 \times e^{QP/10.20} \tag{9}$$

where D_{SSIM} represents the calculated distortion when the SSIM is the distortion metric and *QP* is the quantization parameter. The relationship between the structural distortion and the quantization parameter is hence obtained.

To adjust the RDO model dynamically on the basis of the input frame as in the research of [18], the dynamic adjustment factor σ_{sd} is introduced, which is the standard deviation of the discrete cosine transform (DCT) residual integer transform, as shown in Eq. (9). For a series of video frames, the DCT residual integer transform is uncertain, but the standard deviation of the residual integer transform is a relatively stable eigenvalue. When the contents of the input sequences are similar, the change in σ_{sd} is small. When the change in the contents of the latter frame is relative large, σ_{sd} can represent its conversion range. Therefore, σ_{sd} can stably represent the transform degree of the video texture, as shown in Eq. (10).

$$\sigma_{sd} = \sqrt{E(x^2) - [E(x)]^2} \tag{10}$$

where *x* is the coefficient of the current frame residual transform and $E(x)$ is the expectation of *x*.

Then, the relationship between the coding rate *R* and the quantization parameter *QP* based on the residual transform standard deviation is obtained, as shown in Eq. (19).

$$R = 1.205 \times 10^3 \times e^{-QP/10.08} \times (\sigma_{sd}^{\tau} + b) \tag{11}$$

The Lagrange multiplier is obtained according to the traditional HEVC rate distortion formula, as shown in Eq. (12).

$$\lambda_{SSIM} = -\frac{dD_{SSIM}}{dR} = -\frac{\frac{\partial D_{SSIM}}{\partial QP}}{\frac{\partial R}{\partial QP}} \tag{12}$$

where λ_{SSIM} represents the Lagrange multiplier, which is calculated in the SSIM. Substituting Eqs. (9) and (11) into Eq. (12) yields Eq. (13).

$$\lambda_{SSIM} = \frac{10^{-6} \times 2.42}{\sigma_{sd}^{\tau} - 11.50} \times e^{0.02QP} \tag{13}$$

where the value range of τ is [0.6, 1], and its value is 0.8 in [18]. However, an animation video is selected as the main test object in this study because scenes of artificial design are richer and colors are more vibrant in animation than in other videos. Furthermore, the HEVC HM16.6 experimental platform is used in such videos, thus making a value of 0.92 reasonable. When the change in the current background frame is small and the possibility that the HEVC coding unit is selected as the "SKIP" mode increases, the high-attention CU is unacceptable because the picture quality is affected. To avoid this issue, the algorithm sets the same constraint. Given that the residual transform standard deviation of the frames cannot be obtained before encoding, this algorithm estimates the current frame (σ_{sd}) by using the simple arithmetic mean of the first five frames (σ_{sd}).

3.2.2 Low-attention CU

For the low-attention CUs, the Lagrange multiplier is adjusted on the basis of the coefficient of visual attention. In the DCT quantization process, a large quantization step is used for the low-attention CU. That is, coarse quantification is used for the high-frequency coefficient.

In the encoding process, frame I is referenced by the subsequent frames. If the coding distortion of the frame I image is large, then the quality of the subsequent coding frames will be affected seriously. Therefore, if the current algorithm is working on frame I, the correction factor (P) will be 1.0. On the basis of visual attention, the Lagrange perception correction factor (Perception Factor) is shown in Eq. (14).

$$P_i = \begin{cases} 1.0 & type = I\_Slice \\ a \times Attention + b & other \end{cases} \tag{14}$$

where type indicates the type of the current encoding frame, *I_Slice* represents frame I, *a* and *b* are the adjustment parameters. Tests show that the rate distortion is optimal when $a = -1.6$ and $b = 2.6$.

Therefore, the rate-distortion function based on visual attention is corrected as follows:

$$J = D_{SSE} + P_i \times \lambda_{SSE} \times R \tag{15}$$

$$\lambda_{SSE} = \beta \times 2^{(QP-12)/3} \tag{16}$$

where J represents the rate distortion cost, D_{SSE} stands for the distortion of the image, λ_{SSE} is the Lagrange multiplier in the original algorithm, and R indicates the BR of the current CU. The Lagrange multiplier is defined as a function of the quantization parameter, and β has a constant value of 0.85.

$$\lambda_{new} = P_i \times \lambda_{SSE} \tag{17}$$

Eq. (17) is the Lagrange multiplier of the proposed algorithm. The changed QP value for the low-attention coding block is shown in Eq. (18).

$$\Delta QP = QP_{new} - QP_{org} = 3 \times \log_2 P_i \tag{18}$$

where ΔQP indicates the change value of the quantization parameter, QP_{new} represents the quantization parameter, which is corrected by the proposed algorithm, and QP_{org} is the original quantization parameter. In the study, we use the high-quantization parameter to sample the low-attention coding block, filter part of the high-frequency components, and adjust the BR resource allocation. On the premise of guaranteeing subjective quality, the bit stream is reduced greatly.

4. Result Analysis and Discussion

The optimized algorithm was integrated into HM16.6 to verify its validity. The performance of the algorithm was compared with that of the original algorithm at low-delay mode, which was configured in the "encoder_intra_main" file. The two algorithms were compared in terms of coding rate and time. The platform configurations for the experimental test were as follows: Intel Core i5 processors, 2.30GHZ CPU clock speed, 6GB memory, 64-bit Windows 7 operating system, and Microsoft Visual Studio 2010IDE. The experiment adopted four standard-resolution YUV test sequences, namely, ElephantsDream_704×576, BigBuckBunny_1024×768, ElephantsDream_1920×1080, and BigBuckBunny_352×288, which were provided by the Joint Collaborative Team on Video Coding (JCT-VC). The coding performance was measured in terms of BR and peak signal-to-noise ratio (PSNR). The encoding complexity was measured in terms of encoding time (ET). The evaluation metrics that the algorithm contrasted with the original coding algorithm for HM16.6 were as follows: the increment of PSNR ($\Delta PSNR(Y)$), increment of coding BR (ΔBR), and increment of ET (ΔET) [24]. These metrics are expressed as follows:

$$\Delta PSNR(Y) = PSNR(Y)_P - PSNR(Y)_{HM16.6} \tag{19}$$

$$\Delta B = \frac{B_P - BR_{HM16.6}}{BR_{HM16.6}} \times 100\% \tag{20}$$

$$\Delta ET = \frac{ET_P - ET_{HM16.6}}{ET_{HM16.6}} \times 100\% \tag{21}$$

where $PSNR(Y)_P$, BR_p, and ET_p are the PSNR, BR, and ET, respectively, of the luminance component of the fast-speed division algorithm, which is proposed by this study. $PSNR(Y)_{HM16.6}$, $BR_{HM16.6}$, and $ET_{HM16.6}$ are the PSNR, BR, and ET, respectively, of the original HM16.6 algorithm.

Fig. 2 shows the overall comparison chart of the 2282th frame in the video sequences of BigBuckBunny, in which the red box contains the high-attention area and the black box contains the low-attention area. Fig. (a) shows the result of the frame that was encoded by the original HEVC codec, whereas Fig. (b) shows the result of the frame that applied the proposed algorithm. Tests show that the visual perception quality using the algorithm was not significantly changed.

Fig. 3 shows the details of the high-attention area, where Fig. (a) shows the result of the frame that used the original HEVC algorithm and Fig. (b) shows the result of the frame that used the proposed algorithm. Results show that the image quality was almost unchanged in the high-attention area.

Fig. 4 shows the details of the low-attention area, where Fig. (a) shows the result of the frame that used the original HEVC algorithm and Fig. (b) shows the result of the frame that used the proposed algorithm. By contrast, more distortions are observed in the low-attention areas when we used the proposed algorithm. Notably, this finding did not affect the overall perception quality, and the BR was reduced effectively.

Table 1 compares the performances of the proposed algorithm and the HM16.6 algorithm. Compared with the original HM16.6 encoding algorithm in terms of overall coding efficiency, the encoding BR was increased by an average of 30.33%, the ET was increased by an average of 0.75%, and the fluctuation range of the PSNR was reduced by only 0.11 dB by the proposed algorithm. In this study, an adaptive coding algorithm based on visual attention degree was used to adjust the coding BR resource allocation in different-attention areas by introducing an attention algorithm with low computational complexity and can effectively decrease coding BR

(a) (b)

(a) Original coding results. (b) Proposed coding results

Fig. 2. The overall comparison chart of the 2282th BigBuckBunny frame.

(a) (b)

(a) Original coding results. (b) Proposed coding results

Fig. 3. Comparison of details of high attention area.

(a) (b)

(a) Original coding results. (b) Proposed coding results

Fig. 4. Comparison of details of low attention area.

Table 1. Comparison of the performance

Test sequence	Resolution	$\Delta PSNR(Y)$ (dB)	ΔET (%)	ΔBR (%)
ElephantsDream	1920×1080	-0.09	0.61	-32.63
	704×576	-0.12	0.78	-29.17
BigBuckBunny	1024×768	-0.09	0.72	-30.58
	352×288	-0.13	0.87	-28.94
Average	-	-0.11	0.75	-30.33

5. Conclusions

A novel method based on visual attention was developed to adjust the coding rate allocation of different-visual-attention CUs and encode animated 3D videos effectively. A series of video sequences were analyzed to compare the coding efficiency of the traditional and proposed algorithms. The following conclusions could be drawn:

(1) The GPM can be used to calculate the motion vector of the image. In addition, the extraction of the foreground demonstrates strong robustness because the characteristics of adjacent frames are considered.

(2) CFN ignores the texture features of the image when calculating the CF to obtain fine effects. Furthermore, tests show that among the many factors that affect human visual attention, MF is the most prominent but contrast is the least.

(3) The standard deviation of the DCT residual integer transform can stably represent the transform degree of the video texture that can adjust the relationship between the coding rate and the quantization parameter in high-attention CUs. The coefficient of attention can correct the Lagrange multiplier, and the correction factor of frame I is 1 because frame I will be referred to by subsequent frames in the low-attention CU.

This study focuses on the visual characteristics of human eyes, which are applied to the proposed algorithm, thus improving coding performance. However, the HVS is a complex system, and the attention model does not fully consider the factors that affect visual perception, such as masking effects, user preference, and other visual–psychological factors, which affect the accuracy of attention calculation. These influencing factors should be considered in the future.

Acknowledgements
This work was supported by the Natural Science Foundation of China under Grant Nos. 51668043 and 61262016 and the CERNET Innovation Project under Grant Nos. NGII20160311 and NGII20160112.

References

1. Yang K H., "Methods and systems for rate-distortion optimized quantization of transform blocks in block transform video coding". USA. Patent 7957600, 2007.05.08/2011.06.07.
2. Valizadeh S., Nasiopoulos P., Ward R., "Perceptually-friendly rate distortion optimization in high efficiency video coding". In: *European Signal Processing Conference*, Nice, France:IEEE, 2015, pp. 115-119.
3. Chen Z., Guillemot C., "Perceptually-friendly H. 264/AVC video coding based on foveated just-noticeable-distortion model". *IEEE Transactions on Circuits and Systems for Video Technology*, 2010, 20(6), pp. 806-819.
4. Hong Z., Jing-Bo L I., Xiang-Yan Z., "Algorithm for Coding Unit Partition in 3D Animation Using High Efficiency Video Coding Based on Canny Operator Segment". *Journal of Digital Information Management*, 2016, 14(4), pp.237-245.
5. Duanmu C J., Dong D., Yang X Q., "A New Fast Algorithm for Intra Prediction Mode Selection in the High Efficiency Video Coding (HEVC) Standard". *Advanced Materials Research*, 2014, 926(4), pp. 3342-3345.
6. Naccari M., Pereira F., "Advanced H. 264/AVC-based perceptual video coding: architecture, tools, and assessment". *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, 21(6), pp. 766-782.
7. Vetro A., Tourapis A M., Muller K., et al., "3D-TV content storage and transmission". *IEEE Transactions on Broadcasting*, 2011, 57(2), pp.384-394.
8. Kalva H., Adzic V., Cheok L T., "Adapting video delivery based on motion triggered visual attention". 2012.
9. Guraya F F E., Medina V., Cheikh F A., "Visual attention based surveillance videos compression". In: *Color and Imaging Conference. Society for Imaging Science and Technology*, California, USA:CIC, 2012, pp. 2-8.
10. Li F., Li N., "Region-of-interest based rate control algorithm for H. 264/AVC video coding". *Multimedia Tools and Applications*, 2016, 75(8), pp. 4163-4186.
11. Hua K L., Ahmadiyah A S., Anistyasari Y., "A Novel Image Compression Algorithm Based on Multitree Dictionary and Perceptual-based Rate-Distortion Optimization". *Journal of Information Science & Engineering*, 2015, 31(2), pp.475-489.
12. Guillotel P., Aribuki A., Olivier Y., et al., "Perceptual video coding based on MB classification and rate–distortion optimization". *Signal Processing: Image Communication*, 2013, 28(8), pp. 832-842.
13. Jin G., Cohen R., Vetro A., et al., "Joint perceptually-based Intra prediction and quantization for HEVC". In: *Signal & Information Processing Association Annual Summit and Conference*, Taiwan, China: IEEE, 2012, pp. 1-10.
14. Lee J S., De Simone F., Ebrahimi T., "Efficient video coding based on audio-visual focus of attention". *Journal of Visual Communication and Image Representation*, 2011, 22(8), pp. 704-711.
15. Lee J S., De Simone F., Ebrahimi T., "Video coding based on audio-visual attention". In: *Multimedia and Expo*, New York, USA:IEEE, 2009, pp. 57-60.
16. Lee J S., De Simone F., Ebrahimi T., "Subjective quality evaluation of foveated video coding using audio-visual focus of attention". *IEEE Journal of Selected Topics in Signal Processing*, 2011, 5(7), pp. 1322-1331.
17. Banitalebi-Dehkordi A., Pourazad M T., Nasiopoulos P., "An efficient human visual system based quality metric for 3D video". *Multimedia Tools and Applications*, 2016, 75(8), pp. 4187-4215.
18. Fang Y., Lin W., Lee B S., et al., "Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum". *IEEE Transactions on Multimedia*, 2012, 14(1), pp. 187-198.
19. Spering M., Carrasco M., "Similar effects of feature-based attention on motion perception and pursuit eye movements at different levels of awareness". *Journal of Neuroscience*, 2012, 32(22), pp. 7594-7601.
20. Zhao M., Gersch T M., Schnitzer B S., et al., "Eye movements and attention: The role of pre-saccadic shifts of attention in perception, memory and the control of saccades". *Vision research*, 2012, 74, pp.40-60.

21. GU XH W L P., GU G H., "Electronic image stabilization based on improved gray projection". *Journal of Applied Optics*, 2013, 6, pp. 957-963.
22. Chang L., Su-Mei L., "Measurement of the range of contrast parameter influencing the comfort of stereoscopic images". *Journal of Optoelectronics Laser*, 2014, 25(4), pp. 748-755.
23. Shen L., Zhang Z., An P., "Fast CU size decision and mode decision algorithm for HEVC intra coding". *IEEE Transactions on Consumer Electronics*, 2013, 59(1), pp. 207-213.
24. Kanmani M., Narasimhan V., "Swarm intelligent based contrast enhancement algorithm with improved visual perception for color images". *Multimedia Tools and Applications*, 2017, 64(8), pp. 1-24.