A Practical Foveation-Based Rate-Shaping Mechanism for MPEG Videos

Chia-Chiang Ho, *Member, IEEE*, Ja-Ling Wu, *Senior Member, IEEE*, and Wen-Huang Cheng, *Student Member, IEEE*

***Abstract*—Foveation is one of the nonuniform resolution properties of the human visual system. Recently, different foveation models are proposed and utilized for image and video coding, for the sake of bit-rate saving with no or minor perceptual quality distortion. In the first part of this paper, we propose an efficient and practical DCT-domain foveation model, which is deduced from existing experimental results. In the second part, we present a foveation-based rate-shaping mechanism for MPEG bitstreams, as an application example of the proposed foveation model. The rate shaper is based on eliminating DCT coefficients embedded in MPEG bitstreams. An efficient rate-shaping mechanism is developed to meet various bit-rate requirements. Our simulation confirmed that the proposed foveation model and the rate-shaping mechanism are practical for real-world usage.**

***Index Terms*—Foveated image processing, foveation model, human visual system (HVS), MPEG, rate shaping.**

I. Introduction

MODERN lossy image and video compression techniques try to discard perceptually unimportant or insensible information based on some predefined models of the human visual system (HVS) [1]–[4]. For example, the sensitivity of chromatic signals is found less than that of luminance signals; so typical image or video coding schemes will often compress chromatic signals with a down-sampled resolution. Another example is the design of quantization matrixes for transform domain coefficients. By noticing that the sensitivity of HVS varies for different spatial frequencies, coarser quantizers are applied to coefficients of higher frequency and vice versa. It is also found that, in the HVS, spatial resolution depends on the distribution of photoreceptors on the retina. Foveation, yet another HVS model, is thus proposed to describe this phenomenon. The fovea refers to the region with the densest photoreceptors. Specifically, the sampling density and contrast sensibility decrease dramatically with increasing eccentricity (i.e., the viewing angle with respect to the fovea). With the foveation model in mind, foveated image and video processing preserve the quality of the region that a user gazed at and discards insensible signals outside that region. This helps to reduce the required bit rate and, thus, achieves the purpose of compression. In this paper, we will deduce a DCT-domain foveation model from experimental results presented in previous works and envision its applicability for popular DCT-based image and video coding standards.

With rapid progress of broad-band networks and computation power of general CPUs, more and more applications depend on transporting videos over networks are developed. However, the infrastructure of today's largest network, the Internet, is still insufficient to support guaranteed quality of service (QoS) for real-time multimedia data. While the terminology "video streaming" appears frequently and many researchers are devoted to its realization, there is still a large space for improving its quality and efficiency. The most vital obstacle to achieve real-time video streaming is the heterogeneity problem, which comes from both network and user perspectives. First, different networks have different channel behaviors, and these behaviors are mostly time-varying. Second, different users (or computation environments) also have different QoS requirements. All of these heterogeneities require that video bitstreams be available on a continuum of bit rates such that the available bandwidth can be efficiently utilized. This raises a new challenge for the design of video coding schemes. In recent years, scalable coding schemes have been proposed to deal with such heterogeneous situations, in which one video is coded into base-layer and enhancement-layer bitstreams [5], [6]. However, the usage of scalable coding has its limitations. If one video is coded into a bitstream with a few layers, it cannot satisfy the prementioned bit-rate continuum demand; on the contrary, if many layers are generated, the coding efficiency of the bitstream appears to be a serious issue [7].

An alternative approach other than scalable coding is to encode the video with higher bit rate and provide some rate reduction mechanism for dynamically adapting the video bitstream to channel condition when actual transmission is required. Another reason to develop rate reduction mechanisms is that there are already many compressed videos in the world, coded by traditional nonscalable coding schemes. In the literature, mechanisms proposed for rate reduction can be divided into two categories: *transcoding* and *rate shaping*. Transcoding refers to some kind of reencoding, but with less complexity as compared to a complete encoding process, by wisely making use of available coding information embedded in the original bitstream. Typically, video transcoders reduce video bit rate by requantization, possibly with lowered spatial or temporal resolution. These operations may demand large computation power [8], [9]. Rate shaping, on the other hand, constitutes a lightweight solution for rate reduction. In a nutshell, rate shaping discards some information (generally high-frequency DCT coefficients) residing in the original bitstream and leaves other parts unchanged. When computation complexity is concerned, rate shaping is more suitable for real-time applications.

Manuscript received December 6, 2002; revised May 25, 2005. This work was supported in part by the CIET-NTU (MOE) and the National Science Council of R.O.C. under Contract NSC93-2622-E-002-033, Contract NSC93-2752-E-002-006-PAE, and Contract NSC93-2213-002-006. This paper was recommended by Associate Editor Q. Tian.

C.-C. Ho is with the CyberLink Corporation, Taipei Hsien 231, Taiwan, R.O.C. (e-mail: conrad_ho@gocyberlink.com).

J.-L. Wu and W.-H. Cheng are with the Communication and Multimedia Laboratory and the Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei 10617, Taiwan, R.O.C. (e-mail: wjl@csie.ntu.edu.tw; wisley@cmlab.csie.ntu.edu.tw).

Digital Object Identifier 10.1109/TCSVT.2005.856929

1051-8215/$20.00 © 2005 IEEE

In this paper, we present a foveation-based rate-shaping mechanism for MPEG bitstreams, as an application example of the proposed foveation model.

This paper is organized as follows. In Section II, we detail the DCT-domain foveation model deduced in our work. With such a model in mind, we introduce two foveation-based rate-shaping methods in Section III. Implementation issues of applying the rate-shaping mechanism to MPEG bitstreams are discussed thereafter. Experimental results are then presented in Section IV. Finally, conclusion remarks are given in Section V.

II. Foveation Model

Foveated images can be obtained through pixel-domain approaches [10]–[12]. However, transform-domain approaches have the advantage of applicability for both live-encoded and precompressed bitstreams. In [13], an experimental-proven foveation model was presented associated with a proprietary multiresolution video coding scheme. This model was later applited to wavelet image coding in [14]. Our work adopts the same model but acts on DCT coefficients.

Concerning about the contrast sensitivity of human eyes, reference [13] proposed a model that fits psychological experiment data, as shown in

$$CT(f,e) = CT_0 \exp\left(\alpha f \frac{e_2 + e}{e}\right) \tag{1}$$

where f is the spatial frequency (in cycles/degree), e is the eccentricity (in degrees), CT_0 is the minimal contrast threshold, α is the spatial frequency decay constant, and e_2 is the half-resolution eccentricity constant (in degrees). The best fitting parameters reported in [13] are $\alpha = 0.106$, $e_2 = 2.3$, and $CT_0 = 1/64$. It was also reported that the same α and e_2 provide a good fit to the data in [15] and a proper fit to the data in [16], where CT_0 equals 1/75 and 1/76, respectively.

The *Foveation point* refers to the point at which a human observer gazes in an image. For any given points $\overrightarrow{p} = (x, y)$ in that image, the corresponding eccentricity with respect to the foveation point $\overrightarrow{p}_f = (x_f, y_f)$ can be calculated as

$$e(\overrightarrow{p}) = \tan^{-1}\left(\frac{\sqrt{(x-x_f)^2 + (y-y_f)^2}}{WD}\right) \tag{2}$$

where D is the viewing distance measured in the unit of the image width W (pixels). Note that this calculation is still valid even when the image is displayed with scaling, provided that the scaling ratio is the same for both horizontal and vertical directions.

Now let us consider the visibility of the DCT basis functions. For an $N \times N$ DCT kernel, the (m, n)th basis function can be written as

$$B_{m,n}(j,k) = C_{m,n} \cos\left(\frac{(2j+1)\pi m}{2N}\right) \cos\left(\frac{(2k+1)\pi n}{2N}\right), \quad j,k = 0, \cdots, N-1 \tag{3}$$

where $C_{m,n}$ is the normalization constant.

For single-oriented $B_{m,n}$ ($m = 0$ or $n = 0$, but not both), the corresponding spatial frequency $f_{m,n}$ is [17]

$$f_{m,0} = \frac{m}{2Nw_x} \quad \text{and} \quad f_{0,n} = \frac{n}{2Nw_y} \tag{4}$$

where w_x and w_y are the horizontal width and the vertical height of one pixel in degrees of visual angle, respectively. In this paper, we approximate w_x and w_y by using

$$w = w_x = w_y = \frac{\tan^{-1}\left(\frac{\frac{1}{2}W}{DW}\right)}{\frac{W}{2}} = \frac{\tan^{-1}\left(\frac{1}{2D}\right)}{\frac{W}{2}} \tag{5}$$

assuming that pixels are horizontally and vertically displayed with the same spacing distance. Note that $\tan^{-1}(1/2D)$ represents the visual angle of half the image width $W/2$.

Double-orientated $B_{m,n}$ can be viewed as a sum of two frequency components with the same spatial frequency [17]

$$f_{m,n} = \sqrt{f_{m,0}^2 + f_{0,n}^2} = \frac{1}{2Nw}\sqrt{m^2 + n^2} \tag{6}$$

but with different orientations. Also, the angle between these two components is

$$\theta_{m,n} = \sin^{-1} \frac{2f_{m,0}f_{0,n}}{f_{m,n}^2}. \tag{7}$$

Moreover, a multiplicative factor $1/(r + (1-r)\cos^2\theta_{m,n})$ should be applied to the minimum contrast threshold CT_0 [17] to account for the imperfect summation of the two frequency components and the reduced sensitivity due to the obliqueness of the two components [18], [19]. Note that $\cos\theta_{m,n}$ can be derived from (4), (6), and (7) as

$$\cos\theta_{m,n} = \frac{|m^2 - n^2|}{m^2 + n^2}. \tag{8}$$

The value of r was suggested to be set as 0.6, based on a fourth power summation rule for the two frequency components [17].

Introducing the multiplicative factor and integrating (6) into (1) yields the following equation:

$$CT(f_{m,n}, e) = \frac{CT_0}{r + (1-r)\cos^2\theta_{m,n}} \times \exp\left(\frac{\alpha(e_2 + e)\sqrt{m^2+n^2}}{2e_2Nw}\right). \tag{9}$$

The *critical amplitude* $A_c(f_{m,n}, e)$ is found by multiplying the maximal value of the (m, n)th coefficient $A_{m,n}^*$ to the right-hand side of (9), that is,

$$A_c(f_{m,n}, e) = \frac{A_{m,n}^* CT_0}{r + (1-r)\cos^2\theta_{m,n}} \times \exp\left(\frac{\alpha(e_2 + e)\sqrt{m^2+n^2}}{2e_2Nw}\right). \tag{10}$$

This means that, if the (m, n)th DCT coefficient of one block with eccentricity e is smaller than $A_c(f_{m,n}, e)$, for human eyes, it is indistinguishable from zero.

Taking another point of view, we can deduce the *critical eccentricity* $e_c(m,n)$ for a fixed $f_{m,n}$ by setting the left side of (9) to 1.0 (the maximum contrast) and solving for e

$$e_c(m,n) = \frac{2e_2Nw}{\alpha\sqrt{m^2+n^2}} \ln\left(\frac{r+(1-r)\cos^2\theta_{m,n}}{CT_0}\right) - e_2. \tag{11}$$

For one block with center point (x,y), we can thus have the following critical condition for each (m,n)th DCT coefficient:

$$e(x,y) > e_c(m,n). \tag{12}$$

That is, if (12) is true, no matter how large the (m,n)th DCT coefficient is, for human eyes, it is indistinguishable from zero.

A final note is that the contrast sensitivity of dc values is not included in the above model. In fact, we do not consider altering dc values in the following rate-shaping mechanism, because doing that demands much more computation since dc is predicted-coded and the corresponding coded block pattern may change.

III. Foveation-Based Rate Shaping

Rate shaping was first proposed in [20] and [21], with two schemes for DCT coefficient elimination, and was known as the constrained and general dynamic rate shaping (DRS). An iterative approach, based on Lagrange multipliers, was proposed to optimize the corresponding rate-distortion tradeoff. It is intuitive that low-frequency coefficients are considered to be more important so high-frequency coefficients are dropped first. Later on, combined with a TCP congestion control algorithm, an online implementation was proposed and tested in [23]. Integration of rate shaping and error concealment can improve the quality of rate-shaped bitstream, and this idea was addressed in [24], where selective block dropping is treated as another rate-shaping choice on the condition that the adopted error concealment scheme can interpolate dropped blocks well. This work was later extended by considering perceptual-oriented image features, specifically, removing the blocking effect by preserving coefficients of edge blocks [25]. Recently, discarding slices (group of blocks) of bidirectional predicted pictures was proposed, together with the constrained DRS, for transmission MPEG-2 videos over satellite channels [26]. The concept of joint source-channel coding was brought up for rate shaping in [27], in which channel error conditions were taken into consideration. However, the work in [27] is more likely a selection scheme for protection and transmission of layer-coded videos and thus is far from previous works mentioned here.

In general, the prescribed schemes try to minimize the overall distortion for a specific video unit, e.g., a frame or a group of pictures (GOP). However, in some applications, users focus more on some regions of images and expect better quality in those regions. For example, in a videoconference environment, more user attention is paid to the face region of the speaker than other regions [28]. Yet another example is in the remote education applications, where students focus mostly on the teacher or some specific region of the blackboard or the lecture slide. These scenarios call for developing a new rate-shaping scheme adaptive to the content and user preference. This motivates us to propose a foveation-based rate-shaping mechanism.

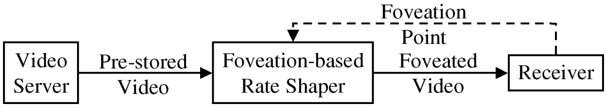

Fig. 1. System architecture of the foveated rate-shaping scheme.

The work presented in [29] explored a similar idea. Each image was divided into the fovea (foreground) region and the periphery (background) region, and reduction of both temporal and spatial resolutions was considered. An associated rate control scheme was presented in [30] to make the proposed architecture more practical. However, this work relied on heuristic rules which are determined empirically, without any religious model for foveation.

Once the relationship between the foveation model and the sensitivity of DCT coefficients has been established, as described in the previous section, we are now ready to apply it to shape the rate of MPEG videos. Fig. 1 shows the system model adopted in this paper. When one precompressed video is requested for transmission, it is sent to the foveation-based rate shaper for downsizing to meet the available bit-rate constraint. The rate shaper may be located in the bottleneck of the network, or be colocated with the video source. The available bandwidth can be estimated either by sender-based, receiver-based, or hybrid methods [31], and the choice among them is beyond the scope of this paper. However, we do propose a rate control algorithm to meet the estimated available bandwidth. A back channel is assumed in between the rate shaper and the receiver to transmit control signal (most important of all, the foveation point) between them. In our experiments, the foveation point is specified explicitly by the mouse click activated by the user. Collaboration with other kind of user interfaces, such as an eye tracker, has no contradiction to the fundamental idea proposed here.

A. Foveation-Based Rate Shaping by Coefficient Elimination

Since our foveation filter is block-based, the actual foveation point used is the center of the block at which the user-specified point is located, and we denote the block as the *foveation block*. We then consider two rate reduction methods in this paper, and both methods are based on the elimination of DCT coefficients.

Method 1—Amplitude Threshold-Based Elimination: This method is based on (10). Each DCT coefficients (except dc) in each coded block is compared to its corresponding critical amplitude calculated from (10). Those coefficients smaller than their corresponding critical amplitudes are eliminated, i.e., set to zero. This method is somewhat similar to the concept of general DRS [20], [21]. It was observed that one block of more zero coefficients requires fewer bits to code statistically [32], [33], so this method does perform the rate reduction task, though not intuitively.

Method 2—Breakpoint-Based Elimination: This method is based on the critical condition (12) and borrows the concept of breakpoint proposed for the constrained DRS problem [20], [21]. The bit rate of one coded block is reduced by eliminating a series of DCT coefficients at the end of that block, in zigzag scanning order. The number of DCT coefficients kept is called the *breakpoint*, corresponding to the first (m,n) pair (in reverse zigzag order) that does not satisfy the critical condition (12).

Computation complexity is always an issue for real-time rate reduction mechanisms. To ease heavy computation incurred by (10) or (11), we can calculate critical amplitudes (or critical eccentricities) in advance and generate *amplitude threshold maps* (or *breakpoint maps*) for different viewing distances. In this way, the required computation at runtime can be reduced largely.

1) Amplitude Threshold Maps: The image width W is derived from parsing the sequence header of the input bitstream. The viewing distance D is restricted to be of integer values, for example, $D = \{d | d \in N, 1 \le d \le 6\}$ for normal viewing distances. The foveation block is set to be the most upper-left block in one frame. Based on these settings, we can calculate critical amplitudes of all coefficients in all blocks, for all viewing distances considered. These values are stored in array form, i.e., $AT[d][b_x][b_y][c_i]$, where (b_x, b_y) represents the location of one particular block and c_i $(i = 1, \dots, 63)$ is the index of one particular coefficient, in zigzag order. We call critical amplitude values for different viewing distances as different *amplitude threshold maps*.

By this way, at runtime, the required amplitude thresholds can be retrieved simply by a look-up table, with some index shifting. It can be observed that the critical amplitudes are symmetric horizontally and vertically with respect to the foveation point. Thus, assuming that the foveation block is specified at location (B_x, B_y) in the runtime, the critical amplitude of one coefficient c_i in the block located at (b_x, b_y) is just $AT[d][|b_x - B_x|][|b_y - B_y|][c_i]$.

2) Breakpoint Maps: The idea used in previous subsection can also be used for calculating breakpoints. First, $e_c(m,n)$'s of all (m,n) pairs are calculated, for each viewing distance d. Then, by setting the foveation block as the top-left block, the breakpoints of every block are found by comparing the corresponding eccentricity to $e_c(m,n)$'s. These breakpoints are also stored in array form, i.e., $BK[d][b_x][b_y]$, and we call breakpoints for different viewing distances as different *breakpoint maps*. At runtime, assuming that the foveation block is specified at (B_x, B_y), the breakpoint of one block at (b_x, b_y) is $BK[d][|b_x - B_x|][|b_y - B_y|]$.

B. Foveation Mismatch Problem

The foveation methods described above make no difference between blocks with different coding types. Thus, for predicted blocks (i.e., those blocks in P- or B-type macroblocks), we are actually foveating the prediction error. This possibly leads to the foveation mismatch problem discussed as follows. In the original video, assume one block B with location (b_x, b_y) is used to predict another block $B' = B + E$ with location (b'_x, b'_y), where E represents the prediction error. For simplicity, we assume that B is coded in the intramode and the breakpoint-based method is applied. In our work, the block B will be foveation filtered with its corresponding breakpoint value BK_a. Let us denote the filtered data as B_a. Now consider foveation filtering of the block B'. For an ideal foveation, we expect to reconstruct $B_b + E_b$, that is, B' foveation filtered with its breakpoint value BK_b. However, in our work, we actually reconstruct $B_a + E_b$. If BK_a is not the same as BK_b, the foveation mismatch problem occurs. Fortunately, we found that this is not a big issue according to two observations of MPEG compressed videos. First, most motion vectors are very small [22]—this is confirmed by many researches and is due to the signal nature of typical videos. Second, due to the complexity issue of typical video encoders, the motion vector is limited within a small region, for example, not exceeding two macroblocks wide (±32 pixel). Thus, for most predicted blocks, the difference between BK_a and BK_b is zero or negligible.

C. Rate Shaping

Rate shaping is required to generate a bitstream with suitable bit rate to fit the estimated available bandwidth. The rate-shaping scheme needed here is different from typical rate control schemes proposed for raw video compression in two ways. First, the input of rate control is an already compressed bitstream, whose intrinsic information may reveal useful clues for good bit reallocation. Second, owing to limited computation power or the need of serving many video streams, the rate-shaping scheme should be as simple as possible to ease the burden of the device.

In our work, rate shaping is achieved by properly increasing the minimum contrast threshold CT_0. We restrict the modified CT_0, denoted as CT_1, to be some fixed values, that is,

$$CT_1(k) = CT_0 + k\mathrm{S}, \quad k = 0, 1, \dots, K \tag{13}$$

where S is a fixed step size. Adding an additional dimension to CT_1, amplitude threshold maps and breakpoint maps can be represented in the forms of $AT[d][k][b_x][b_y][c_i]$ and $BK[d][k][b_x][b_y]$, respectively.

Successful rate shaping requires an underlying rate model. Here we choose the ρ-domain rate model proposed in [32] and [33], for its simplicity, efficiency, and ease of integration with the rate-shaping mechanism proposed here. Typical rate models try to model the relationship between the coding bit rate R and the value of quantization scale q [34], [35]. Taking another viewpoint, based on the insight that the number of zeros plays an important role in transform coding of images and videos, it was observed in [32], [33] that a linear relation existed between R and ρ (the percentage of zeros among the quantized transform coefficients). This linear relation can be modeled as

$$R(\rho) = \theta(1 - \rho) \tag{14}$$

where θ is a frame-dependent constant. Note that the bit rate R discussed here excludes all header information other than DCT coefficients. In our work, since the input is a compressed MPEG bitstream, the value of θ for each coded frame can be easily found by partial decoding.

Let us come back to our foveation model. For a particular viewing distance and a specified foveation point, we can figure out a one-to-one relationship between CT_1 and ρ. In this way, we relate the foveation model with the rate model. Before presenting the rate-shaping scheme used in our work, we define necessary variables and list them in Table I.

The proposed mechanism of applying foveation-based rate shaping to one frame F_t is briefed as follows. (Notice that this mechanism is general enough for applying to the prescribed two rate shaping algorithms.)

Step 0) (Initialization): Set $r_S(t)$, Z_t and all $z_t(k)$'s to zero.

Step 1) Decode one coded block into DCT coefficients and increase $r_S(t)$, Z_t, and each $z_t(k)$ accordingly (with amplitude threshold maps or breakpoint

TABLE I
VARIABLES FOR THE PROPOSED RATE-SHAPING SCHEME

Notation	Meaning (unit)
R_S	The bitrate of the source bitstream. (bps)
R_T	The target bitrate of the new bitstream. (bps)
$r_S(t)$	The bitcount of the frame F_t in the source bitstream. The header information is excluded. (bits)
$r_T(t)$	The target bitcount of the frame F_t in the foveated bitstream. The header information is excluded. (bits)
$r_h(t)$	The bitcount of the header information of the frame F_t. (bits)
$r_r(t)$	The actual bitcount of the foveated frame F_t. The header information is excluded. (bits)
$r_l(t)$	The difference between $r_T(t)$ and $r_r(t)$. (bits)
B	The size of the encoding buffer. (bits)
B_{t+1}	The number of bits in the buffer after encoding the frame F_t. (bits)
θ_t	The frame dependent constant of the rate model, for the frame F_t.
Z_t	The count of zero coefficients of the frame F_t in the source bitstream.
$z_t(k)$	The count of zero coefficients when applying the foveation model with $CT_1(k)$ to the frame F_t.
M_{total}	The number of blocks in one frame.

maps). Iteratively perform this step until all blocks are processed.

Step 2) The frame-dependent constant θ_t is found as $r_S(t)/(1-Z_t/(64\cdot M_{total})$, and the target bitcount $r_T(t)$ is calculated proportionally to the ratio of the target bit rate to the source bit rate, with adjustment according to the current buffer fullness

$$r_T(t) = \left(\frac{R_T}{R_S} r_S(t) + r_l(t-1)\right) \frac{2(B-B_t)+B_t}{(B-B_t)+2B_t}. \quad (15)$$

Step 3) The target percentage of zero is derived as

$$\rho = 1 - \frac{r_T(t)}{\theta_t} \quad (16)$$

and the intended value of CT_1 is found by

$$CT_1^* = \underset{CT_1(k)}{\arg\min} \left|64\cdot\rho\cdot M_{total} - z_t(k)\right|. \quad (17)$$

Step 4) The frame is foveation-rate-shaped with the amplitude threshold map (or the breakpoint map) corresponding to CT_1^*, and the bitcount of all coefficients are collected as $r_r(t)$.

Step 5) (Housekeeping): The difference between $r_T(t)$ and $r_l(t)$ is calculated as

$$r_l(t) = r_T(t) - r_r(t). \quad (18)$$

The encoding buffer fullness is initially set to half the buffer size and updated as

$$B_{t+1} = B_t + (r_r(t) + r_h(t)) - (r_r(t-1) + r_h(t-1)). \quad (19)$$

It should be noted that this rate-shaping scheme achieves a minimal one-frame delay, which is irrelative to most applications.

IV. EXPERIMENTAL RESULTS

Two foveation-based methods have been proposed in the previous section. For simplicity and without loss of generality, experiments focused on the breakpoint-based method only. The amplitude threshold-based method may not be easily used for intercoded frames and macroblocks, and this will be illustrated in the following. For one coefficient of value c_i in one intercoded block BI and its amplitude threshold AT, we should check if $c_i + c_i' < AT$, where c_i' is the value of the same coefficient in the predicted block of BI (denoted as BI'). However, usually c_i' is not readily available, unless BI' is intracoded. If BI' is intercoded, we can get c_i' by performing DCT on the reconstructed BI'; however, this requires much more computation and hinders the amplitude threshold method from being applied to shape the rate of predictive coded videos.

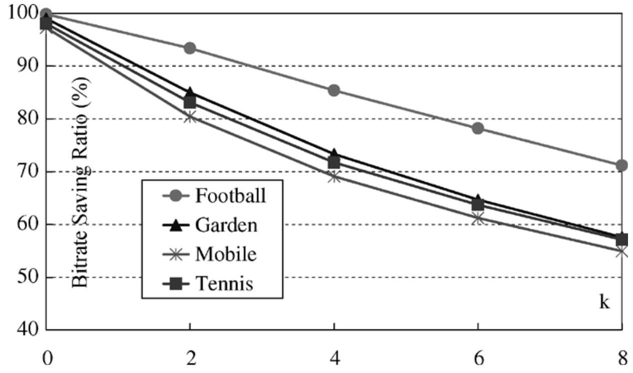

Fig. 2. Relationships between CT_1 and BSR's, for the four different video sequences. The stepsize S used here is 0.03, i.e., $CT_1 = CT_0 + 0.03k$. The bit rate is 800 kb/s, and the viewing distance D is set to 1.

In our experiments, the first 60 frames of four well-known test sequences, namely *football*, *garden*, *tennis*, and *mobile*, were compressed in MPEG-1 format. The frame size is 352 × 240 and the frame rate is 24. The size of the GOP is 12, and the distance between two anchor frames is 3 (i.e., two B-frames in between). Two different bit rates are considered: 800 and 1125 kb/s. The adopted parameters of the foveation model are $\alpha = 0.106$, $e_2 = 2.3$, and $CT_0 = 1/64$. For the sake of fair comparison, the foveation block was set as the center block of one frame.

Let us define the bit-rate saving ratio BSR (percentage) as

$$\text{BSR} = \frac{\text{bitcount\_of\_foveated\_video} * 100}{\text{bitcount\_of\_original\_video}}. \quad (20)$$

Figs. 2–5 show the relationships between CT_1 and BSR under the combinations of two source bit rates and two normal viewing distances ($D = 1$ and $D = 6$). It is found that the bit-rate saving of the sequence *football* is lower than that of other sequences. This is because *football* is a sequence of large motion, and most bits are spent on coefficients of lower frequency when coding intercoded blocks. These coefficients are harder to be foveation-eliminated. In fact, we can expect that BSR will

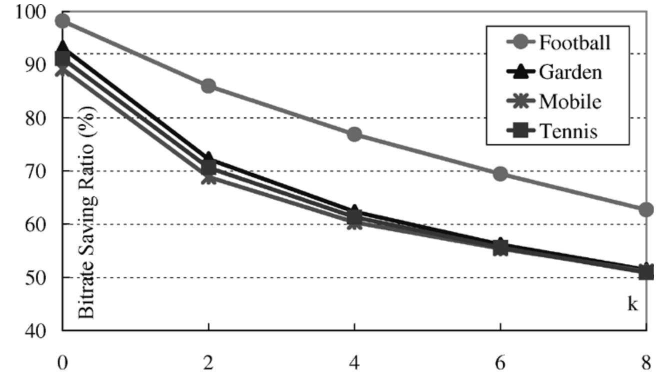

Fig. 3. Same settings as that of Fig. 2, but with $D = 6$.

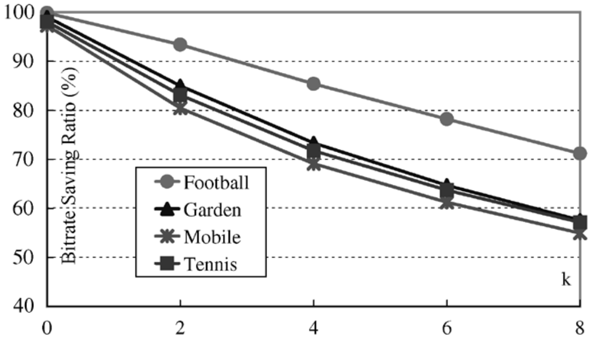

Fig. 4. Same settings as that of Fig. 2, but the bit rate is 1125 kb/s.

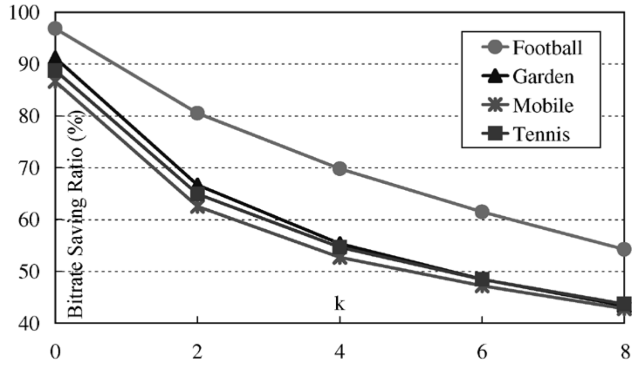

Fig. 5. Same settings as that of Fig. 3, but with $D = 6$.

be influenced by factors of both the video source and the encoder (specifically, encoding parameters) used. Another interesting insight is the near linear relationship between CT_1 and BSR. We can foresee that the computation complexity of the proposed rate-shaping method can be further reduced, if fewer CT_1 values are tested and the most suitable value is generated through proper interpolation.

Figs. 6 and 7 show some example snapshots of foveated videos. For low-value CT_1's, the distortion of the foveated (with respect to the original) frame is almost invisible. This is illustrated by the two second pictures of both figures. In these two figures, the foveated *football* bitstream (with $k = 2$) reduces 7% of the bit rate of the original 1125-kb/s bitstream, while the foveated *mobile* bitstream (with $k = 0$) reduces 11% of the bit rate of the original 800-kb/s bitstream. When CT_1 increases, periphery regions are gradually blurred and more visible block effects appear, while the quality of the center area (near the foveation block) is comparably preserved, as shown in the third and the fourth pictures of both figures.

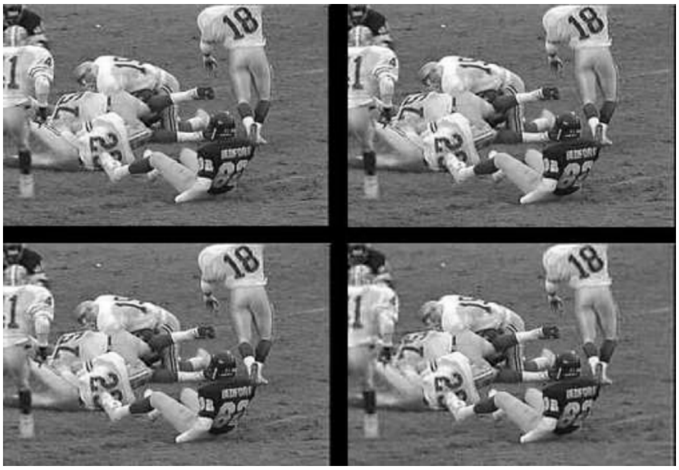

Fig. 6. Snapshots of the foveated *football* video. The bit rate is 1125 kb/s. The first (upper-left) image comes from the 35th frame (an I-frame) of the original bitstream. The second, third, and fourth (upper-right, bottom-left, and bottom-right) images come from the same frame of foveated bitstreams with $k = 2$, 4 and 8, respectively, where $CT_1 = CT_0 + 0.03k$.

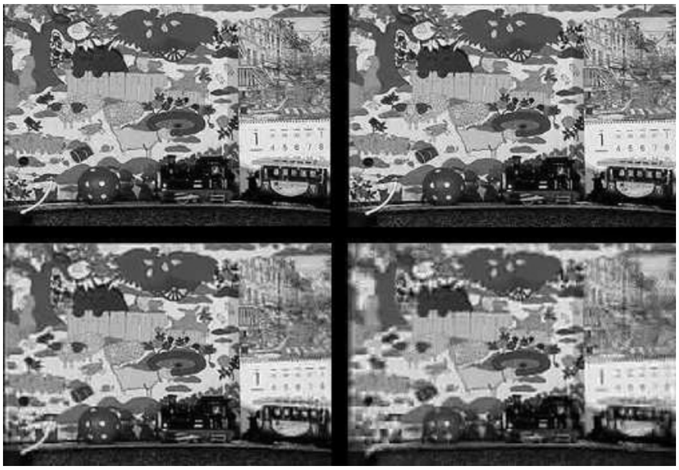

Fig. 7. Snapshots of the foveated *mobile* video. The bit rate is 800 kb/s. The upper-left image comes from the 30th frame (an B-frame) of the original bitstream. The second, third, and fourth (upper-right, bottom-left, and bottom-right) images come from the same frame of foveated bitstreams with $k = 0$, 4 and 8, respectively, where $CT_1 = CT_0 + 0.03k$.

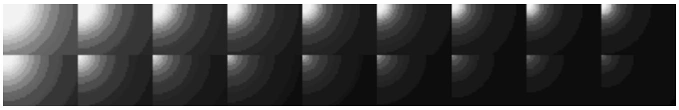

Fig. 8. Sample breakpoint maps for images of size 352 × 240. The value of D is set to 1 and 6 for the upper and the lower rows, respectively. The value of CT_1 increases from left to right. Note that breakpoint values are linearly scaled up into the range [0,255] for display purposes.

As mentioned previously, the proposed rate shaper is of low complexity because breakpoints for each block are calculated in advance. Fig. 8 shows examples of calculated breakpoint maps with the foveation center in the most upper-left block. It is shown that the breakpoint values of $D = 6$ are larger than those of $D = 1$. This illustrates that a larger rate reduction can be obtained with $D = 6$, as compared with that of $D = 1$, as shown in Figs. 2–5.

V. Discussion and Conclusion

In this paper, a DCT-domain foveation model is derived based on existed experimental results. With such a model in mind, we present a foveation-based rate-shaping mechanism for MPEG bitstreams. Two different rate-shaping methods, the critical amplitude-based method and the breakpoint-based method, are then proposed. By restricting ranges of model parameters, we can greatly reduce the computation requirement in the runtime by precalculating and storing critical amplitudes or breakpoints for different viewing settings. To meet various bit-rate requirement, an efficient rate-shaping mechanism has been developed. All of these new developments together with their implementation and experimental results show the applicability of the proposed foveation model and rate-shaping mechanism for real-world usage. While more attention is paid to user-oriented applications nowadays, our work reveals the value of considering user needs.

It is worth remarking that, with respect to the specific problem of rate shaping, the foveation block is fixedly set as the center block as described in Section IV. To obtain better perceived visual quality, the foveation block should be carefully selected. In the literature, a visual attention model [36] has been proven to be affective and accurate to detect the viewer's visual focus for still images. Therefore, we have investigated the extension of visual attention model to video sequences [36] and its possibility to be integrated with the proposed rate-shaping method [37]. However, we think this subject is beyond the scope of this paper and needs further investigation in the future. On the other hand, The authors of [38] have done a complete survey regarding the statistical distribution of visual focus and confirmed that human eyes are mostly attracted to the central portion of an observed image. Therefore, current settings of the foveation block may be inappropriate but would be highly acceptable for rate-shaping purposes.

Besides, to resolve the difficulty of using the amplitude threshold-based method, as mentioned in Section IV, we think the following two possible research directions may be useful.

1) Utilize all kinds of available information. For example, the relationship between temporal information (such as motion vectors) and spatial foveation coefficients can be investigated to improve the coding performance, as shown in [39].
2) Explore the visual attention model [36] further to enhance the effectiveness of the amplitude threshold-based method. As prescribed, the fixed setting scheme of foveation blocks is reasonable but lacks flexibility; therefore, the knowledge of accurate visual focuses is expected to improve the overall performance and be beneficial for reducing the needed computational efforts.

Consideration of multiple foveation points may be another remarkable future work for the rate-shaping mechanism. This is useful for some application scenarios. For example, when more than one user are interacting with the same video through a multicast network environment, they may have different focuses on the video content. Rate shaping with multiple foveation points yields a bitstream which compromising the needs of different users.

As far as MPEG-4 video is concerned, the new standardized fine-granularity-scalable (FGS) coding scheme has drawn considerable attentions. The FGS coding scheme has the advantage of providing a continuous spectrum of applicable bit rate. Performing the proposed foveation-based rate-shaping mechanism on FGS-coded bitstreams gives another dimension of scalability, without any foveation mismatch problem since the enhancement layer data are coded frame by frame.

References

[1] S. Lee, M. S. Pattichis, and A. C. Bovik, "Rate control for foveated MPEG/H.263 video," in *Proc. Int. Conf. Image Processing*, vol. 2, 1998, pp. 365–369.

[2] S. Lee and A. C. Bovik, "Very low bit rate foveated video coding for H.263," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 6, 1999, pp. 3113–3116.

[3] Z. Wang, L. Lu, and A. C. Bovik, "Rate scalable video coding using a foveation-based human visual system model," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, 2001, pp. 1785–1788.

[4] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 4, pp. 129–132, Mar. 2002.

[5] *Information Technology—Generic Coding of Moving Pictures and Associated Audio. Part 2: Video*, 1995.

[6] *Information Technology—Coding of Audio-Visual Objects. Part 2: Visual*, 1999.

[7] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, pp. 269–275, 1991.

[8] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 977–992, Jul. 2001.

[9] S. Lee and A. C. Bovik, "Fast algorithms for foveated video processing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 2, pp. 149–162, Feb. 2003.

[10] R. S. Wallace, P.-W. Ong, B. B. Bederson, and E. L. Schwartz, "Space-variant image processing," *Int. J. Comput. Vis.*, vol. 13, no. 1, pp. 71–90, 1994.

[11] N. Tsumura, C. Endo, H. Haneishi, and Y. Miyake, "Image compression and decomposition based on gazing area," *Proc. SPIE*, vol. 2657, pp. 361–367, 1996.

[12] T. Kuyel, W. Geisler, and J. Ghosh, "Retinally reconstructed images: digital images having a resolution match with the human eyes," *IEEE Trans. Syst, Man, Cybern. A*, vol. 29, no. 3, pp. 235–243, Mar. 1999.

[13] W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," in *Proc. SPIE*, vol. 3299, Human Vision and Electronic Imaging, 1998, pp. 294–305.

[14] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *IEEE Trans. Image Process.*, vol. 10, pp. 1397–1410, Oct. 2001.

[15] W. S. Geisler, "Visual detection following retinal damage: predictions of an inhomogeneous retino-cortical model," in *Proc. SPIE*, vol. 2674, Human Vision and Electronic Imaging, 1996, pp. 119–130.

[16] M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: limits imposed by optics, photoreceptors and receptor pooling," *J. Opt. Soc. Amer. A*, vol. 8, pp. 1775–1787, 1991.

[17] J. Ahumada, Jr. and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," in *Proc. SPIE*, vol. 1666, Human Vision, Visual Processing, and Digital Display III, 1992, pp. 365–374.

[18] G. C. Phillips and H. R. Wilson, "Orientation bandwidths of spatial mechanisms measured by masking," *J. Opt. Soc. Amer. A*, vol. 1, pp. 226–232, 1984.

[19] B. Watson, "Detection and recognition of simple spatial forms," in *Physical and Biological Processing of Images*, O. J. Braddick and A. C. Sleigh, Eds. Berlin, Germany: Springer-Verlag, 1983, pp. 100–114.

[20] A. C. Eleftheriadis and D. Anastassiou, "Constrained and general dynamic rate shaping of compressed digital video," in *Proc. Int. Conf. Image Proc.*, vol. 3, Oct. 1995, pp. 396–399.

[21] ——, "Meeting arbitrary QoS constraints using dynamic rate shaping of coded digital video," in *Proc. NOSDAV*, 1995, pp. 95–106.

[22] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video Processing and Communications*. Upper Saddle River, NJ: Prentice-Hall, 2002.

[23] S. Jacobs and A. Eleftheriadis, "Real-time video on the web using dynamic rate shaping," in *Proc. Int. Conf. Image Proc.*, vol. 2, 1997, pp. 250–253.

[24] W. Zeng and B. Liu, "Rate shaping by block dropping for transmission of MPEG-precoded video over channel of dynamic bandwidth," in *Proc. ACM Multimedia*, Boston, MA, Nov. 1996, pp. 385–393.

[25] W.-J. Zeng, B. Guo, and B. Liu, “Feature-oriented rate shaping of precompressed image/video,” in *Proc. Int. Conf. Image Proc.*, vol. 2, 1997, pp. 772–775.

[26] N. Celandroni, E. Ferro, F. Potorti, A. Chimienti, and M. Lucenteforte, “DRS compression applied to MPEG-2 video data transmission over a satellite channel,” in *Proc. 5th IEEE Symp. Computers and Communications (ISCC)*, Antibes-Juan Les Pins, France, Jul. 2000, pp. 259–266.

[27] T. P. Chen and T. Chen, “Adaptive joint source-channel coding using rate shaping,” in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, Orlando, FL, May 2002, pp. 1985–1988.

[28] H. R. Sheikh, S. Liu, Z. Wang, and A. C. Bovik, “Foveated multipoint videoconferencing at low bit rates,” in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, Orlando, FL, May 2002, pp. 2069–2072.

[29] T. H. Reeves and J. A. Robinson, “Adaptive foveation of MPEG video,” in *Proc. ACM Multimedia*, Boston, MA, Nov. 1996, pp. 231–241.

[30] ——, “Rate control of foveated MPEG video,” in *Proc. IEEE Canadian Conf. Electrical and Computer Engineering*, vol. 1, 1997, pp. 379–382.

[31] D.-P. Wu, Y.-W. T. Hou, and Y.-Q. Zhang, “Transporting real-time video over the Internet: challenges and approaches,” *Proc. IEEE*, vol. 88, no. 12, pp. 1855–1875, Dec. 2000.

[32] Y. K. Kim, Z. He, and S. K. Mitra, “A novel linear source model and a unified rate control algorithm for H.263/MPEG-2/MPEG-4,” in *Proc. Intl. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, Salt Lake City, UT, May 2001, pp. 1777–1780.

[33] Z. He and S. K. Mitra, “A unified rate-distortion analysis framework for tranform coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 12, pp. 1221–1236, Dec. 2001.

[34] H.-J. Lee, T.-H. Chiang, and Y.-Q. Zhang, “Scalable rate control for MPEG-4 video,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 878–894, Sep. 2000.

[35] B. Tao, B. W. Dickinson, and H. A. Peterson, “Adaptive model-driven bit allocation for MPEG video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 1, pp. 147–157, Feb. 2000.

[36] C.-C. Ho, W.-H. Cheng, T.-J. Pan, and J.-L. Wu, “A user-attention based focus detection framework and its applications,” in *Proc. 4th IEEE Pacific-Rim Conf. Multimedia*, Singapore, Dec. 2003.

[37] C.-C. Ho, “A study of effective techniques for user oriented video streaming,” Ph.D. dissertation, Dept. Comp. Sci. Inf. Eng., National Taiwan Univ., Taipei, May 2003.

[38] L. Itti, “Automatic foveation for video compression using a neurobiological model of visual attention,” *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304–1318, Oct. 2004.

[39] C.-C. Ho, W.-T. Chu, C.-H. Huang, and J.-L. Wu, “User-oriented approach in spatial and temporal domain video coding,” in *Proc. 4th IEEE Pacific-Rim Conf. Multimedia*, Singapore, Dec. 2003, pp. 1341–1345.

Chia-Chiang Ho (M’03) received the B.S. and Ph.D. degrees in computer science and information engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1996 and 2003, respectively.

Since 2004, he has been with Cyberlink Corporation, Taipei, Taiwan, R.O.C., where he is presently a Senior Engineer. His current research interests include video coding and video processing.

Ja-Ling Wu (SM’98) received the B.S. degree in electronic engineering from Tamkang University, Tamshoei, Taiwan, R.O.C., in 1979, and the M.S. and Ph.D. degrees in electrical engineering from Tatung Institute of Technology, Taipei, Taiwan, in 1981 and 1986, respectively.

From 1986 to 1987, he was an Associate Professor with the Electrical Engineering Department, Tatung Institute of Technology, Taipei, Taiwan, R.O.C. In 1987, he transferred to the Department of Computer Science and Information Engineering, National Taiwan University (NTU), Taipei, where he is presently a Professor and the Director of the Communications and Multimedia Laboratory. From 1996 to 1998, he was the first Head of the Department of Information Engineering, National Chi Nan University, Puli, Taiwan. During his sabbatical leave (from 1998 to 1999), he was invited to be the Chief Technology Officer of the Cyberlink Corporation, Taipei. In this one-year term, he was involved with the developments of some well-known audio-video softwares, such as the PowerDVD. Since August 2004, he has been appointed the head the newest research institutes of NTU, the Graduate Institute of Networking and Multimedia. hH has published more than 200 technique and conference papers. His research interests include digital signal processing, image and video compression, digital content analysis, multimedia systems, digital watermarking, and digital right management systems.

Prof. Wu was the recipient of the Outstanding Young Medal of the Republic of China in 1987 and the Outstanding Research Award three times of the National Science Council, Republic of China, in 1998, 2000 and 2004, respectively. He was the recipient of the Award for Distinguished Information People in 1993, the Special Long-Term Award for Collaboratory Research in 1994, the Best Long-Term paper Award in 1995, and the Long-Term Medal for Distinguished Researchers in 1996, all sponsored by the Acer Corporation. In 2001, his paper “Hidden Digital Watermark in Images” (coauthored with Prof. Chiou-Ting Hsu), published in IEEE TRANSACTIONS ON IMAGE PROCESSING, was selected to be one of the winners of the “Honoring Excellence in Taiwanese Research Award” offered by ISI Thomson Scientific. He started his Industrial-Academic Collaborative researches, which are sponsored by both the National Science Council of Taiwan and local industries, in 1992. During the last 12 years, he has conducted four three-year term Industrial-Academic Collaborative projects covering various kinds of multimedia related applications, such as multimedia office, multimedia classroom, multimedia home, etc. Due to his continuous contributions on promoting the cooperation between Industry and Academia, he was the recipient of the Special Award for Collaboratory Research, offered by of the Ministry of Education, Taiwan, R.O.C., in 1997.

Wen-Huang Cheng (S’05) received the B.S. and M.S. degrees in computer science and information engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 2002 and 2004, respectively, where he is currently working toward the Ph.D. degree in the Graduate Institute of Networking and Multimedia.

His research interests include multimedia data management and analysis.