Foveated Video Compression with Optimal Rate Control

Sanghoon Lee, Marios S. Pattichis, and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*—Recently, foveated video compression algorithms have been proposed which, in certain applications, deliver high-quality video at reduced bit rates by seeking to match the nonuniform sampling of the human retina. We describe such a framework here where foveated video is created by a nonuniform filtering scheme that increases the compressibility of the video stream. We maximize a new foveal visual quality metric, the foveal signal-to-noise ratio (FSNR) to determine the best compression and rate control parameters for a given target bit rate. Specifically, we establish a new optimal rate control algorithm for maximizing the FSNR using a Lagrange multiplier method defined on a curvilinear coordinate system. For optimal rate control, we also develop a piecewise R–D (rate–distortion)/R–Q (rate–quantization) model. A fast algorithm for searching for an optimal Lagrange multiplier λ^* is subsequently presented. For the new models, we show how the reconstructed video quality is affected, where the FPSNR is maximized, and demonstrate the coding performance for H.263,+,++/MPEG-4 video coding. For H.263/MPEG video coding, a suboptimal rate control algorithm is developed for fast, high-performance applications. In the simulations, we compare the reconstructed pictures obtained using optimal rate control methods for foveated and normal video. We show that foveated video coding using the suboptimal rate control algorithm delivers excellent performance under 64 kb/s.**

***Index Terms*—Digital video, foveation, image compression, rate control, video compression.**

I. Introduction

VIDEO standards have always been associated with particular ranges of bit rates. In order to maximize the video compression ratio for a given video standard, it is necessary to use the maximum degree of quantization, typically determined by a quantization parameter (QP) that is provided by the standard. At the maximum compression setting, the compressed bit rate achieves the minimum bound on the number of generated bits, which depends on the codeword density used to represent the discrete cosine transform (DCT) coefficients, i.e., the complexity of the input image sequence. By removing unessential spatial frequency information from a video sequence, the spatial redundancy decreases, due primarily to the reduction or elimination of high-frequency DCT coefficients that are deemed to be visually unimportant. Moreover, motion compensation errors also tend to be reduced. Because of such spatial/temporal redundancy reduction, the coding efficiency is improved, and the minimum bound on the compressed bit rate is reduced. For example, suppose that a CIF image sequence (352×288) is compressed to 50–1000 kb/s for a QP ranging between 31 and 1. If the bit rate is further reduced by 40% by selectively removing some kind of information, then the bit rate range is scaled down to 30–600 kb/s, which is a range of interest for applications where the transmission rate is severely restricted by the channel capacity, as in wireless networks or PSTN.

Naturally, reducing the bit rate in this way has the potential to degrade the visual fidelity in some way, depending on the type of information that is removed. The use of other transforms, such as wavelet methods, offers promise, but even in those domains the limits of additional compression that can be obtained is probably being probed already, and in any case, does not offer the current advantage of standards-compliance. In this paper, we explore the possibility of increasing compression performance, while maintaining or even improving visual fidelity, while also maintaining standards-compliance. We show how this can be done in an effective way by the selective reduction of high-frequency coefficients according to a nonuniform spatial law. The method we will explore is called *foveation.*

The human retina possesses a nonuniform spatial distribution (resolution) of photoreceptors, with highest density on that part of the retina aligned with the visual axis: the fovea. The photoreceptor density rapidly decreases with distance away ("eccentricity") from the fovea, hence the local visual frequency bandwidth also falls away. Subjective image quality can be measured, to some degree, as a function of viewing distance, resolution, picture size, and the contrast sensitivity of the human eye [1], [2].

Recently, very sophisticated commercial eye trackers (head-mounted or desktop) have become available that either track an infrared (IR) reflection of the retina, or directly detect and track the pupil image [3]–[5]. Using an eye tracker, the point of visual fixation can be determined in real-time and delivered over an end-to-end visual communication system. Several real-time/nonreal-time visual communication systems associated with eye trackers have already been proposed and demonstrated in the field of visual communications (wireless video phones, video conferencing systems, web-news, web-advertisement, and personal communication systems) as well as virtual reality (virtual space teleconferencing, virtual three-dimensional games, computer-aided design, remote telepresence,

Manuscript received November 18, 1998; revised March 28, 2001. This work was supported in part by Bell Labs, Lucent Technologies, Texas Instruments Inc., and by the Texas Advanced Technology Program. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Boon-Lock Yeo.

S. Lee is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA.

M. S. Pattichis is with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131 USA.

A. C. Bovik is with the Center for Vision and Image Sciences, Department of Electrical and Computer Engineering, The University of Texas, Austin, TX 78712-1084 USA (e-mail: bovik@ece.utexas.edu).

Publisher Item Identifier S 1057-7149(01)05446-X.

remote training systems, and remote surgery) [6]–[10]. In addition to the eye tracker, an interactive approach using a mouse is also feasible. It is also possible that an automatic algorithm could be used which selects a "best" or "likely" fixation point [11].

We define the fixation point of a displayed digital image to be that point which intersects the visual axis, viz., to which a human observer is directing his/her visual attention. We will use the fixation point as a reference point for calculating local spatial image bandwidths. We will also create a modified image, from which undetectable high visual frequencies are removed (given a fixation point) which will be termed the *foveated image*. Fig. 2(a) shows an original image, and Fig. 2(b) shows a foveated version of the original image. The fixation point is also indicated on the image; given an appropriate viewing distance, and assuming that the viewer fixates at the indicated point, the image appears normal. By removing such invisible information, it becomes possible to increase compression performance without sacrificing visual quality, provided that the fixation point can be discovered and tracked.

The first foveated true video compression scheme that we are aware of was reported in [12], using an effective but nonstandard-compliant basis coding algorithm. The practical use of eye-tracking hardware in concert with a foveated compression algorithm is demonstrated in [4]. An MPEG-compliant foveated compression scheme was reported in [13]. MPEG and H.263-compliant video compression algorithms were demonstrated in [14] and used in the development of an automated algorithm for assessing the quality of foveated video streams using a human visual model [1], [15], [16]. In [14], [17], and [18], we demonstrated reduction of the computational overhead in implementations of real-time video processing algorithms such as foveation filtering, motion estimation, motion compensation, and video rate control. In addition, we have presented a prototype end-to-end video communications system suitable for human interactive multimedia applications over wireless channels [19]–[21].

In this paper, we exploit foveation as a tool for exploring methods for optimal rate control of foveated video, since it exhibits the advantages of standards-compliance, low coding complexity, algorithmic simplicity, and, when utilized properly, high-fidelity performance. Most so-called optimal rate control algorithms attempt to maximize the SNR under a rate constraint by using a Lagrange multiplier method. However, current rate control methods do not necessarily provide for the best resource allocation in terms of subjective video quality. In [6], [22], and [23], we defined the foveal signal-to-noise ratio (FSNR) as an objective video quality criterion that matches the nonuniform spatial resolution of the human visual system. The FSNR is an objective way to measure subjective image quality in the sense that it exploits more known information about the receiver. Maximizing the FSNR instead of the SNR for a given target bit rate provides for perceptually better-quality video (or reduced bit rate at an equivalent visual quality), as demonstrated in [24].

The main contributions of the paper are as follows. We explore the coordinate transformation approach to foveated imaging. In this approach, a foveated image can be mapped into an image that has been uniformly sampled, allowing analysis to proceed without considering the superimposed variable local bandwidth. We develop a new optimal rate control algorithm for maximizing the FSNR using a Lagrange multiplier technique cast in a curvilinear coordinate system. For efficient algorithm implementations, a piecewise R–D (rate–distortion)/R–Q (rate–quantization) model is described. Based on these models, a fast iterative algorithm for searching for an optimal Lagrange multiplier λ^* is presented. For applications requiring very low bit rate video coding (e.g., under 64 kb/s), we present a suboptimal rate control algorithm which is able to adapt to the normal/modified quantization mode defined in H.263. In order to measure the performance gain, we define compression gains due to foveation filtering and nonuniform quantization. In the simulations, we obtain compression gains ranging from 8% to 52% for I pictures and from 7% to 68% for P pictures.

II. Foveation Compression by Filtering

A. Image/Video Representation over Curvilinear Coordinates

Suppose there exists a coordinate transform $\mathbf{\Phi}(\mathbf{x}) = [\Phi_1(x_1, x_2), \Phi_2(x_1, x_2)]^t$ where the superscript t denotes transpose. If a one-to-one correspondence exists between $\mathbf{x} = (x_1, x_2)^t$ and $\mathbf{\Phi}(\mathbf{x})$, where Φ_1 and Φ_2 are continuous and uniquely invertible, then $\mathbf{\Phi}(\mathbf{x})$ are called "two-dimensional (2-D) curvilinear coordinates." Then, the *Jacobian* of the coordinate transformation is $J_{\mathbf{\Phi}}(\mathbf{x}) = J_{\mathbf{x}}^{-1}(\mathbf{\Phi})$.

We make the following notation for the stages of processing that occur between the camera and the human eye, as shown in Fig. 1, where $\mathbf{x} = (x_1, x_2)^t$ are the Cartesian coordinates, $\mathbf{\Phi}(\mathbf{x}) = [\Phi_1(x_1, x_2), \Phi_2(x_1, x_2)]^t$ are the curvilinear coordinates, $o(\mathbf{x})$ is the original image displayed on the monitor, $r(\mathbf{x})$ is the reconstructed (decompressed) image displayed on the monitor, $g(\mathbf{x})$ is the image formed on the human eye, $h(\mathbf{\Phi}(\mathbf{x}))$ is the image of $g(\mathbf{x})$ in the curvilinear coordinates, $v(\mathbf{x})$ is the foveated image of $o(\mathbf{x})$, and finally, $z(\mathbf{\Phi}(\mathbf{x}))$ is the image of $v(\mathbf{x})$ in the curvilinear coordinates. The relationships between the various images are given by $g(\mathbf{x}) = F_v^c(r(\mathbf{x}))$, $g(\mathbf{x}) = h(\mathbf{\Phi}(\mathbf{x}))$, $v(\mathbf{x}) = F_v^c(o(\mathbf{x}))$, $v(\mathbf{x}) = z(\mathbf{\Phi}(\mathbf{x}))$ where F_v^c denotes the process of *foveation filtering* in the continuous spatial domain, as given below in Definition 3. Fig. 2 shows an example of an original image against a foveated image over Cartesian coordinates and over curvilinear coordinates.

We now define terms associated with the frequency domain analysis. Let $\mathbf{\Omega} = (\Omega_1, \Omega_2)^t$ be continuous 2-D frequency. For $\mathbf{x}, \mathbf{\Omega} \in \mathcal{R}^2$, let $b(\mathbf{x})$ and $\mathcal{B}(\mathbf{\Omega})$ be a 2-D signal and its Fourier transform. When $\mathcal{B}(\mathbf{\Omega})$ is band-limited within a circle of radius Ω_o in the frequency domain, $\mathcal{B}(\mathbf{\Omega}) = 0$ for $|\mathbf{\Omega}| \geq \Omega_o$, then $b(\mathbf{x})$ is an Ω_o-band-limited signal, i.e., $b(\mathbf{x}) \in B^{\Omega_o}$. Then, we write $b(\mathbf{x}) \in B^{\Omega_o}$, where B^{Ω_o} is the space of Ω_o-band-limited signals. From the Whittaker-Shannon sampling theorem [25], the function $b(\mathbf{x})$ can be reconstructed from samples generated by a sampling matrix $\mathbf{V}$ that avoids aliasing.

Definition 1: Locally band-limited signals. For a given 2-D signal $z(\mathbf{x}) \in B^{\Omega_c}$, suppose that $z(\mathbf{\Phi}(\mathbf{x})) = v(\mathbf{x})$. Then, $v(\mathbf{x}) \in B^{\Omega_f(\mathbf{x})}$ is a *locally band-limited signal* with respect to the coordinate system $\mathbf{\Phi}(\mathbf{x})$ and radial frequency Ω_c, where $B^{\Omega_f(\mathbf{x})}$ is the space of locally band-limited signals.

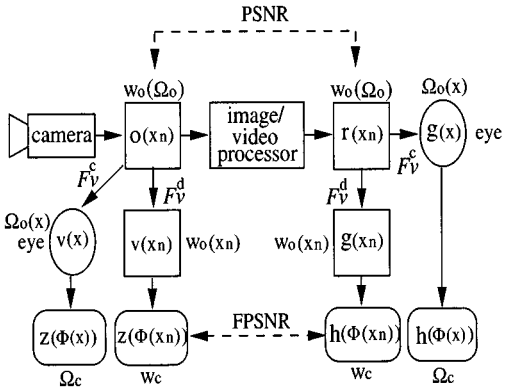


Fig. 1. Image representation.

Definition 2: Local bandwidth for one-dimensional signal. Suppose that a locally band-limited one-dimensional signal $v(x) \in B^{\Omega_f(x)}$ is mapped into $z(x) \in B^{\Omega_c}$ where $z(\Phi(x)) = v(x)$. The *local bandwidth* $\Omega(x)$ is defined in terms of the coordinate system $\Phi(x)$ according to

$$\Omega(x) = \Omega_c \frac{d\Phi(x)}{dx}. \tag{1}$$

In the 2-D case, we can use the Whittaker-Shannon sampling theorem to reconstruct the 2-D signal $z(\mathbf{x}) \in B^{\Omega_c}$ from its samples taken at the sampling points $\mathbf{x}_n$ defined in terms of sampling matrix $\mathbf{V}_c$. These sampling points correspond to the nonuniform samples of $v(\mathbf{\Phi}(\mathbf{x}))$ taken at $\mathbf{\Phi}(\mathbf{x}_n)$. At the point $\mathbf{x}_n$, the local sampling density $\det[\mathbf{V}_n]$ is given by the product of $\det[\mathbf{V}_c]J_{\mathbf{\Phi}}(\mathbf{x}_n)^{-1}$. Thus, for any given sampling matrix $\mathbf{V}_c$, the sampling density of $v(\mathbf{x})$ is $cJ_{\mathbf{\Phi}}(\mathbf{x})^{-1}$ where $c = \det(\mathbf{V}_c)$. Therefore, for the 2-D signal $v(\mathbf{x}) \in B^{\Omega_f(\mathbf{x})}$, the local bandwidth is proportional to the local sampling density

$$\Omega_f^2(\mathbf{x}) = cJ_{\mathbf{\Phi}}(\mathbf{x}). \tag{2}$$

Let Ω_o, $\Omega_f(\mathbf{x})$, and Ω_c be the global bandwidth of the original image, the foveated image, and the image corresponding to the foveated image in curvilinear coordinates, respectively. Then, $o(\mathbf{x})$, $r(\mathbf{x}) \in B^{\Omega_o}$, $g(\mathbf{x})$, $v(\mathbf{x}) \in B^{\Omega_f(\mathbf{x})}$, and $h(\mathbf{\Phi}(\mathbf{x}))$, $z(\mathbf{\Phi}(\mathbf{x})) \in B^{\Omega_c}$. For a given sampling matrix $\mathbf{V}$, the digital 2-D frequency is given by $\mathbf{w} = \mathbf{V}^t\mathbf{\Omega}$. When $\mathbf{x}_n = (x_{n_1}, x_{n_2})^t$ is a sampling point in Cartesian coordinates, the discrete images corresponding to the above functions are $o(\mathbf{x}_n)$, $v(\mathbf{x}_n)$, $z(\mathbf{\Phi}(\mathbf{x}_n))$, $r(\mathbf{x}_n)$, $g(\mathbf{x}_n)$, and $h(\mathbf{\Phi}(\mathbf{x}_n))$. Let w_o, $w_f(\mathbf{x}_n)$ and w_c denote the normalized bandwidth. Then, $o(\mathbf{x}_n)$, $r(\mathbf{x}_n) \in B^{w_o}$, $g(\mathbf{x}_n)$, $v(\mathbf{x}_n) \in B^{w_f(\mathbf{x})}$, and $h(\mathbf{\Phi}(\mathbf{x}_n))$, $z(\mathbf{\Phi}(\mathbf{x}_n)) \in B^{w_c}$.

The magnitudes of the global bandwidths Ω_o, $\Omega_f(\mathbf{x})$ and Ω_c depend on the picture size in the continuous domain. In Fig. 2(b), it can be observed that the local bandwidth $\Omega_f(\mathbf{x})$ varies with respect to the foveation point. In the frame memory, the image is sampled and stored. Since the maximum normalized frequency in the discrete domain is 0.5, $w_o = \pi$, and $w_f(\mathbf{x}) < \pi$. Fig. 2(d) shows the foveated image in curvilinear coordinates where $\Omega_o = \Omega_c$ and $w_o = w_c = \pi$. The region where the local bandwidth $w_f(\mathbf{x})$ is equal to π is not changed after the coordinate transformation, while the region where the local bandwidth $w_f(\mathbf{x}) < \pi$ is reduced in proportion to the magnitude of $w_f(\mathbf{x})$. Thus, area is unchanged near the center of the foveation point and contracts away from the foveation point toward the periphery. After the coordinate transformation, it is shown that the local bandwidth of the foveated image in Fig. 2(d) is globally uniform over curvilinear coordinates. Fig. 2(c) shows the original "News" image in curvilinear coordinates. In the image, we do not apply any low-pass filtering before the coordinate transformation. Thus, it can be seen that spatial frequency increases toward the periphery.

The image $v(\mathbf{x})$ is obtained from the orignal image $o(\mathbf{x})$ using *ideal foveation filtering*, which is also obtained from $\mathbf{\Phi}(\mathbf{x})$.

Definition 3: Ideal foveation filtering. Let $v(\mathbf{x}) \in B^{\Omega_f(\mathbf{x})}$ be the projection of $o(\mathbf{x})$ onto $B^{\Omega_f(\mathbf{x})}$. Let $\mathbf{\Phi}$ be the coordinate transformation: $z(\mathbf{\Phi}(\mathbf{x})) = v(\mathbf{x})$. Let $L_p^c(., \Omega_c)$ be an ideal low-pass filter with radial cutoff frequency Ω_c. Then, the *ideal foveation filtered image* $v(\mathbf{x})$ is $v(\mathbf{x}) = \tilde{v}(\mathbf{\Phi}(\mathbf{x}))$ where $\tilde{v}(\mathbf{x}) = L_p^c(o(\mathbf{\Phi}^{-1}(\mathbf{x})), \Omega_c)$.

For ideal foveation filtering, the band-limited signal $o(\mathbf{x}) \in B^{\Omega_o}$ is the corresponding locally band-limited signal $v(\mathbf{x}) \in B^{\Omega_f(\mathbf{x})}$. An example of the image $o(\Phi^{-1}(\mathbf{x}))$ is an expanded version of Fig. 2(c). The image $o(\Phi^{-1}(\mathbf{x}))$ is obtained from the original image $o(\mathbf{x})$ by the inverse curvilinear coordinate transformation. The region is expanded from the original image according to the local bandwidth. Thus, the region centered at the foveation point is expanded more than the periphery. Then, an ideal low-pass filter with cutoff frequency Ω_c is applied as $L_p^c(., \Omega_c)$. Since the region around the foveation point is expanded such that the local bandwidth is less than Ω_c, local spatial information is not lost by the low-pass filtering. However, spatial information over the peripheral region is removed inversely proportional to the area expansion. After taking the coordinate transformation $\tilde{v}(\mathbf{\Phi}(\mathbf{x}))$, the image size returns to the original image size in Fig. 2(b) and the image becomes the foveated image $v(\mathbf{x})$.

In practice, an ideal foveation filter can be approximated using a bank of low-pass filters. Let $S_o \subset \mathcal{R}^2$ be the original image region displayed on the monitor and A_o be the associated area. Each position vector $\mathbf{x}$ in the region S_o is given by $\mathbf{x} \in S_o$. Then, the foveation filtered image is $F_v^c(\mathbf{x}) = L_p^c(o(\mathbf{y}), \Omega_f(\mathbf{x}))$ where $\mathbf{y} \in S_o$.

In the discrete domain, the discrete image $o(\mathbf{x}_n)$ is obtained from the original image $o(\mathbf{x})$ after sampling at $\mathbf{x}_n$. For example, let i_p be the number of sampling pixels in the horizontal line of the image and i_d be the length of the horizontal line image size. Suppose a pixel forms a square with the length of each side $\epsilon = i_d/i_p$. Then, set the sampling frequency $(1/\epsilon) = 1$. At the nth sampling point $\mathbf{x}_n = (x_{n_1}, x_{n_2})^t$, the unit area is $a_n^o = [x_{n_1} \pm (\epsilon/2)] \times [x_{n_2} \pm (\epsilon/2)] = 1$, and the area of the image is $A_o = a_n^o N = N$ where N is the number of pixels in the image. Therefore, the foveated image $v(\mathbf{x}_n)$ can be obtained by

$$v(\mathbf{x}_n) = F_v^d(o(\mathbf{x}_n)) = L_p^d(o(\mathbf{x}_m), w_f(\mathbf{x}_n)) \tag{3}$$

where $m \in \{1, 2, \ldots, N\}$ and $L_p^d(., w_f(\mathbf{x}_n))$ is an approximation to an ideal lowpass filter with bandwidth $w_f(\mathbf{x}_n)$.

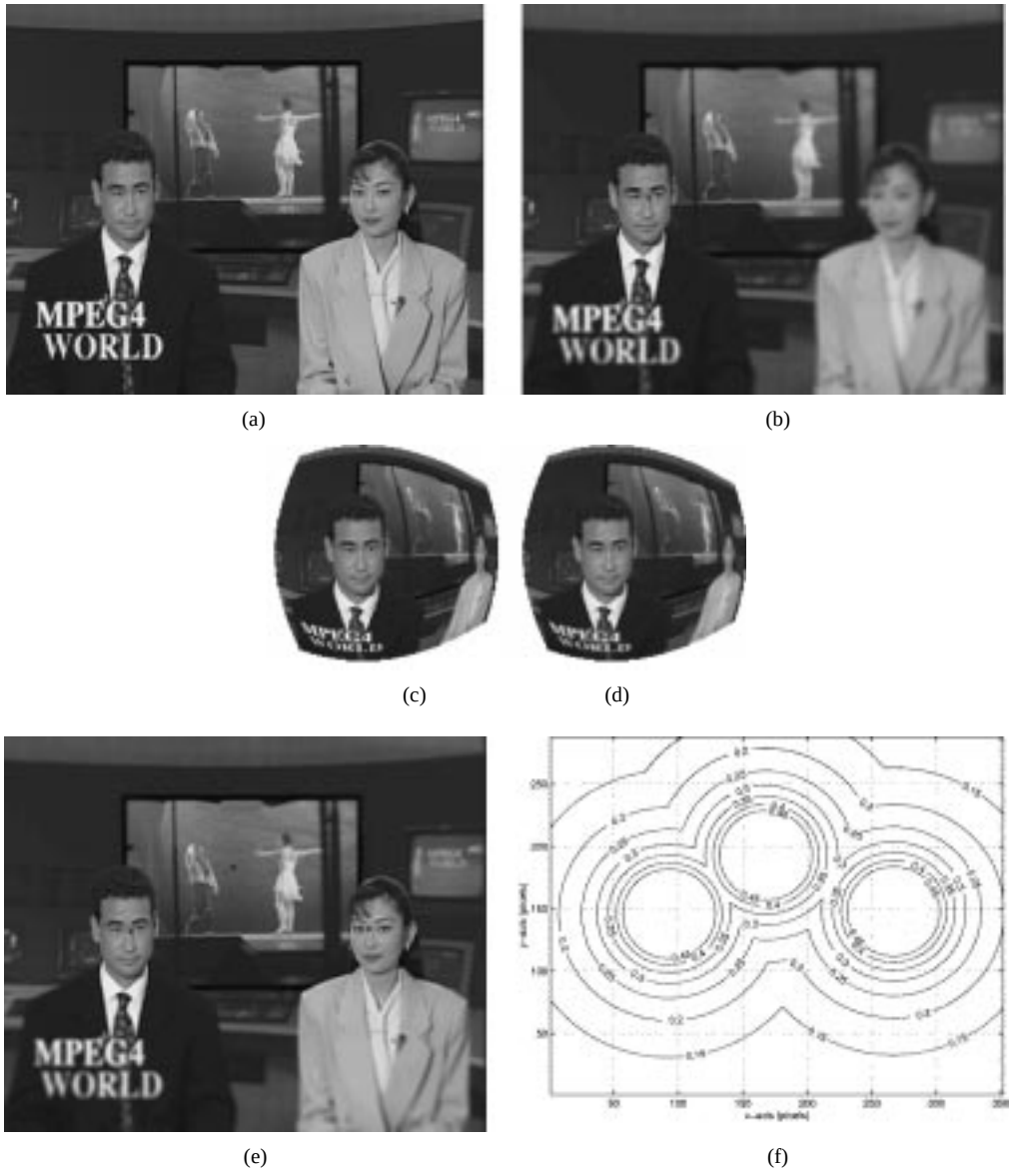


Fig. 2. Original and foveated "News" images. (a) Original "News" image in Cartesian coordinates: $o(\mathbf{x}) \in B^{\Omega_o}$, $o(\mathbf{x}_n) \in B^{\omega_o}$ where $\omega_o = \pi$. (b) Foveated "News" image in Cartesian coordinates: $v(\mathbf{x}) \in B^{\Omega_{f(\mathbf{x})}}$, $v(\mathbf{x}_n) \in B^{\omega_{f(\mathbf{x})}}$ where $\Omega_{f(\mathbf{x})} \leq \Omega_o$ and $\omega_{f(\mathbf{x})} \leq \pi$. (c) Original "News" image in curvilinear coordinates. (d) Foveated "News" image in curvilinear coordinates: $z(\mathbf{\Phi}(\mathbf{x})) \in B^{\Omega_c}$, $z(\mathbf{\Phi}(\mathbf{x}_n)) \in B^{\omega_c}$ where $\omega_c = \pi$. (e) Foveated image "News" with three foveation points. (f) Local bandwidth.

B. Nonuniform Sampling Theorem

Let a foveated image $g(\mathbf{x}) \in B^{\Omega_f(\mathbf{x})}$ be mapped into a image $h(\mathbf{\Phi}(\mathbf{x})) \in B^{\Omega_c}$ in the coordinate system $\mathbf{\Phi}(\mathbf{x})$. The foveated image $g(\mathbf{x})$ can be perfectly reconstructed by the uniform sampling matrix $\mathbf{V}_c$, which avoids any aliasing effects in the curvilinear coordinates $\mathbf{\Phi}(\mathbf{x})$ because of the following nonuniform sampling theorem. For example, Fig. 2(b) can be reconstructed by using the uniformly sampled points of Fig. 2(d) over the curvilinear coordinates. However, it is impossible to reconstruct the original image in Fig. 2(a) using the uniformly sampled points of Fig. 2(c) because of aliasing effects.

After mapping Cartesian coordinates into curvilinear coordinates, the Nyquist sampling frequency can be calculated by the uniform sampling theorem on curvilinear coordinates. Similarly, a locally band-limited signal can be reconstructed from a set of uniform sampling points on curvilinear coordinates whose sampling frequency is greater than the Nyquist frequency. This follows from the nonuniform sampling theorem [26], [27].

The 2-D uniform sampling theorem is described as follows: for $\mathbf{x}, \mathbf{\Omega} \in R^2$, $o(\mathbf{x}) \in B^{\Omega_o}$ is reconstructed from the sampling points $\mathbf{x}_n = \mathbf{V}_o\mathbf{n}, \mathbf{n} \in \mathcal{Z}^2$ where $\mathbf{V}_o$ is a sampling matrix which does not cause aliasing in the discrete frequency domain. Then

$$o(\mathbf{x}) = \frac{|\det \mathbf{V}_o|}{4\pi^2} \int_{B^{\Omega_o}} \sum_n o(\mathbf{x}_n) \exp\left[j\mathbf{\Omega}^t(\mathbf{x} - \mathbf{V}_o\mathbf{n})\right] d\mathbf{\Omega} \tag{4}$$

where $\mathbf{\Omega}^t$ is the transpose matrix of $\mathbf{\Omega}$. Let $\mathbf{w}$ denote discrete frequency for discrete signals. Then, the Fourier transform of $o(\mathbf{x}_n)$ becomes

$$O(\mathbf{w}) = \frac{1}{|\det \mathbf{V}_o|} \sum_k O(\mathbf{\Omega} - \mathbf{U}_o\mathbf{k}) \tag{5}$$

where $\mathbf{U}_o^t\mathbf{V}_o = 2\pi\mathbf{I}$, $\mathbf{k} \in \mathcal{Z}^2$ and $\mathbf{I}$ is the 2×2 identity matrix.

The 2-D nonuniform sampling theorem is developed on the curvilinear coordinate system. If the sampling frequency is greater than the Nyquist frequency along each axis, then aliasing will not occur on curvilinear coordinates either. Of course, the sampling points on the curvilinear coordinates generally correspond to nonperiodic points on Cartesian coordinates. From the nonuniform sampling points, the original function $g(\mathbf{x}) = h(\mathbf{\Phi}(\mathbf{x}))$ can be reconstructed by

$$g(\mathbf{x}) = \frac{|\det \mathbf{V}_c|}{4\pi^2} \int_{B^{\Omega_c}} \sum_n h(\mathbf{x}_n) \exp\left[j\mathbf{\Omega}^t(\mathbf{\Phi}(\mathbf{x}) - \mathbf{V}_c\mathbf{n})\right] d\mathbf{\Omega} \tag{6}$$

where $\mathbf{x}_n = \mathbf{\Phi}^{-1}(\mathbf{V}_c\mathbf{n})$ and $\mathbf{V}_c$ is the sampling matrix corresponding to Ω_c.

Since the bandwidth of the foveated image $\Omega_f(\mathbf{x})$ is less than Ω_o, the foveated image can also be reconstructed by using the sampling matrix $\mathbf{V}_o$. In the foveated image, sampling at the rate determined by $\mathbf{V}_o$ in Cartesian coordinates corresponds to a nonperiodic sampling rate which is always greater than $2\Omega_c$ for each coordinate. Therefore, $h(\mathbf{\Phi}(\mathbf{x}))$ can be always reconstructed from the sampling points using $\mathbf{V}_o$.

C. Target Bit Allocation in Curvilinear Coordinates

Given a curvilinear coordinate system, the locally band-limited (foveated) image is resampled into a new image which is globally band-limited. Suppose that the number of bits that is generated over some infinitesimal region of the curvilinear coordinate system is proportional to the area of the corresponding infinitesimal region in the rectangular Cartesian coordinate system (uniform domain). Then, the number of target bits required for the foveated image can be equally allocated into each unit region in the uniform domain, in proportion to the mapping ratio. Thus, the target bits are nonuniformly allocated according to the mapping ratio of the foveated image.

Let $S_o \subset \mathcal{R}^2$ be a spatial region of one frame of the original video sequence, and displayed on a monitor over the spatial $\mathbf{x}$ domain, and A_o be the associated area of this region. Denote $S_c \subset \mathcal{R}^2$ and A_c as the corresponding region and area of the image $z(\mathbf{\Phi}(x))$ over the $\mathbf{\Phi}(\mathbf{x})$ domain. Then

$$A_c = \int_{\mathbf{x}\in S_o} J_{\mathbf{\Phi}}(\mathbf{x}) d\mathbf{x}. \tag{7}$$

Now assume that the discrete function $v(\mathbf{x}_n)$ is obtained by sampling $v(\mathbf{x})$ at the sampling points $\mathbf{x}_n$. Suppose that each pixel is a square with side length ϵ. The unit area with respect to the nth sampling point is then $a_n^o = [x_{n_1} \pm (\epsilon/2)] \times [x_{n_2} \pm (\epsilon/2)]$, and the total area of the image is the sum of each unit area: $A_o = \sum_{n=1}^{N} a_n^o$ where N is the total number of pixels in a picture frame. Since ϵ is constant, a_n^o is independent of n, and $A_o = N a_n^o$. Hence $A_c = \sum_{n=1}^{N} a_n^c$ where $a_n^c = \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_n) = \int_{\mathbf{x}\in s_n^o} J_{\mathbf{\Phi}}(\mathbf{x}) d\mathbf{x}$ and $s_n^o \subset \mathcal{R}^2$ is the unit region. Fig. 2(d) shows the foveated image in the curvilinear coordinates where $\Omega_o = \Omega_c$ and $w_o = w_c = \pi$. The area in Fig. 2(d) becomes A_c which is unchanged near the center of the foveation point and decreases from the foveation point toward the periphery relative to the area A_o in Fig. 2(b). Since the nth pixel $\mathbf{x}_n$ also corresponds to the pth pixel of the mth macroblock, it can be denoted $\mathbf{x}_n = \mathbf{x}_{p,m}$. Let M and m_p be the number of macroblocks in a picture and the number of pixels in each macroblock, respectively. Then, A_c becomes $A_c = \sum_{m=1}^{M} \sum_{p=1}^{m_p} \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_{p,m})$. The corresponding area of the mth macroblock in curvilinear coordinates is $a_m^c = \sum_{p=1}^{m_p} \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_{p,m})$. If we allocate the target bits R_T into each macroblock according to the value of a_m^c, the number of bits allotted to the mth macroblock is $\hat{r}_m = R_T \times a_m^c / A_c$. In the foveated image, the local bandwidth depends on the nonuniform sampling density that corresponds to the uniform sampling density in curvilinear coordinates. Let $\mathbf{V}_n$ be a sampling matrix with the local bandwidth f_n at the nth point where $f_n = \Omega_f(\mathbf{x}_n)/2\pi$. Assuming that $\bar{J}_{\mathbf{\Phi}}(\mathbf{x}_n)$ is in proportion to the sampling density, then

$$\bar{J}_{\mathbf{\Phi}}(\mathbf{x}_n) = \frac{c_1}{|\det \mathbf{V}_n|} = c_2 f_n^2 \tag{8}$$

where c_1 and c_2 are constants.

The allocated rate $\hat{r}_m$ is obtained from the area ratio over the uniform spatial domain. However, in real image/video processing systems, the number of generated bits depends on the coding factor as well as the area ratio. Therefore, for a practical rate control implementation, it should be represented as a function of the local bandwidth such as $\tilde{a}_{p,m}^c(f_{p,m})$ not exactly as $f_{p,m}^2$. Then,

$$\hat{r}_m = R_T \times \sum_{p=1}^{m_p} \tilde{a}_{p,m}^c(f_{p,m}) \left[\sum_{m=1}^{M} \sum_{p=1}^{m_p} \tilde{a}_{p,m}^c(f_{p,m})\right]^{-1}. \tag{9}$$

D. Foveated Image/Video Quality Assessment

In [6] and [22], a quality assessment metric called foveal weighted signal-to-noise ratio (FWSNR) was defined. In the metric, the foveal weighting metric f_n^2 was utilized to take into account the spatial variation of visual resolution according to the direction of gaze in addition to the contrast sensitivity function (CSF). Since the foveal weighting metric f_n^2 effectively measures spatially-varying additive noise, the quality measurements made by the FWSNR and the FSNR can more accurately evaluate the localized visual quality when foveation is used [6], [22].

Under assumption (8), the foveal mean square error (FMSE) for discrete image frames is

$$\text{FMSE} = \frac{1}{\sum_{n=1}^{N} f_{p_n}^2} \sum_{n=1}^{N} [a(\mathbf{x}_n) - b(\mathbf{x}_n)]^2 f_{p_n}^2 \tag{10}$$

and the FPSNR is

$$\text{FPSNR} = 10 * \log_{10} \frac{\max[a(\mathbf{x}_n)]^2}{\text{FMSE}}, \quad 1 \le n \le N \tag{11}$$

where $b(\mathbf{x}_n)$ is a compressed version of an original image frame $a(\mathbf{x}_n)$, or where $b(\mathbf{x}_n)$ is a compressed version of a foveated image frame $a(\mathbf{x}_n)$. Here, we describe a new optimal rate control algorithm based on the above quality criteria.

E. Spatial Compression Gain

1) Visual Entropy: As a simple notation, we use x to represent $o(\mathbf{x})$ i.e., the value of the original image. Since x is an analog value, the cumulative distribution $\mathcal{D}(x)$ is obtained by $\mathcal{D}(x) = P(X \le x)$ where X is a continuous random variable

and P is the probability function. Let $\rho(x)$ be a probability density function of X. Then, the differential entropy $H(X)$ [28] becomes

$$H(X) = -\int_{x \in S_o} \rho(x) \log \rho(x) dx \tag{12}$$

where S_o is the image region in Cartesian coordinates. The entropy $H(X)$ is the minimum average number of bits needed to describe a random variable X for the image. If human fixation points are uniformly distributed over the image at a normal viewing distance, then the foveated image becomes the original image i.e., $v(\mathbf{x}) = o(\mathbf{x})$. If the fixation points are nonuniformly distributed, then the foveated image cannot be the original image i.e., $v(\mathbf{x}) \neq o(\mathbf{x})$. In the previous section, it was shown that $v(\mathbf{x}) \in B^{\Omega_f(\mathbf{x})}$ can be mapped onto a uniform resolution image $z(\mathbf{\Phi}) \in B^{\Omega_c}$ using curvilinear coordinates $\mathbf{\Phi} = (\Phi_1, \Phi_2)$. As a further simple notation, let ϕ be the value of the image $z(\mathbf{\Phi})$. Then, the cumulative distribution is given by $\mathcal{D}(\phi) = P(\Phi \leq \phi)$. Then, the visual entropy $H(\phi)$ is defined as follows.

Definition 4: $H(\Phi)$ (visual entropy) is the differential entropy of a foveated image over curvilinear coordinates, i.e., the minimum average number of bits required to describe a random variable Φ over curvilinear coordinates associated with the foveation points. Thus

$$H(\Phi) = -\int_{\phi \in S_c} \rho(\phi) \log \rho(\phi) d\phi \tag{13}$$

where S_c is the region over curvilinear coordinates.

If $v(\mathbf{x}) \in B^{\Omega_f(\mathbf{x})}$ is projected onto the space B^{Ω_o}, then the local bandwidth for both images $z(\mathbf{\Phi})$ and $o(\mathbf{x})$ becomes the same. Thus, we can assume that $\rho(\phi)$ in $z(\mathbf{\Phi})$ is equal to $\rho(x)$ in $o(\mathbf{x})$ and $H(X) = H(\Phi)$. The total entropy is obtained by $A_o H(X)$ for $o(\mathbf{x})$ and $A_c H(\Phi)$ for $z(\mathbf{\Phi})$ where A_o and A_c are the associated area. Then, the total saved entropy is a function of the area difference and expressed by $(A_o - A_c)H(X)$ where $A_o \geq A_c$. The saved entropy relative to the total original entropy becomes the mapping gain over curvilinear coordinates: $G^m = (A_o - A_c)/A_o$.

The differential entropy can be expressed by the discrete entropy [28]. Suppose that the discrete value is uniformly sampled and then quantized from $o(\mathbf{x})$ and $z(\mathbf{\Phi})$. Let N_o and N_c be the total number of sampling points over $o(\mathbf{x})$ and $z(\mathbf{\Phi})$. If we use the same QP for both images, then the mapping gain is expressed by $G^m = (N_o - N_c)/N_o$.

In real video, it is difficult to use G^m as the compression gain since the area ratio is not proportional to the ratio of generated bits, and the probability density functions ($\rho(x)$ and $\rho(\phi)$) depend on the image sequence. Further, video processing is a nonlinear operation so the generated bits cannot be obtained by any linear operation. The quantization errors measured over Cartesian coordinates map to different quantization errors over curvilinear coordinates. Thus, it is difficult to measure the saved entropy associated with the QPs. Finally, the quantization yields different distortions because each macroblock is mapped into a different area over curvilinear coordinates.

2) Compression Gain: The coding gain is obtained by removing high frequencies in a graded fashion away from fixation, while maintaining a high picture resolution over the foveated region. When high-frequency components occur in macroblocks around the periphery of fixation, foveation filtering effectively removes them without creating visual artifacts or distortions. Conversely, if high-frequency components exist within the fovea area, then more bits can be assigned, leading to potential improvements in picture quality. Therefore, the compression gain in a real system is obtained by two major factors. One is foveation filtering which yields the saved entropy. The other is nonuniform quantization which maximizes the FSNR subject to a rate constraint over the curvilinear coordinates. The compression gain afforded by foveation filtering is defined by the following.

Definition 5: G_c^f (compression gain due to foveation filtering). Let $o(\mathbf{x}_n)$ and $r(\mathbf{x}_n)$ be the discrete versions of the original image and the reconstructed (decompressed) image, and let $v(\mathbf{x}_n)$ and $g(\mathbf{x}_n)$ be the foveated versions of $o(\mathbf{x}_n)$ and $r(\mathbf{x}_n)$, respectively. Also, let R_o be the number of generated bits when $o(\mathbf{x}_n)$ is compressed to $r(\mathbf{x}_n)$ using QP q_o, and R_c be the number of generated bits when $v(\mathbf{x}_n)$ is compressed to $g(\mathbf{x}_n)$ using QP q_c. Then, the compression gain is

$$G_c^f = \frac{R_o - R_c}{R_o} \text{ subject to } q_o = q_c. \tag{14}$$

The goal of foveation in an image/video compression algorithm is ordinarily to create an image that appears the same as the original, provided that the fixation point of the eye coincides with the selected foveation point in the algorithm. The compression gain (14) is obtained using the fact that the visual quality for both $r(\mathbf{x}_n)$ and $g(\mathbf{x}_n)$ is assumed to be the same when $q_o = q_c$.

Since video standards utilize a macroblock (16×16 pixels) as a coding unit, it is unnatural to compress the foveated image over curvilinear coordinates. The distortion over curvilinear coordinates can be measured on a pixel-by-pixel basis. Using a Lagrange multiplier, an optimal bit allocation over curvilinear coordinates can be accomplished for single or multiple foveation points. The additional compression gain due to the nonuniform quantization is as follows.

Definition 6: G_c^q (compression gain due to nonuniform quantization). Let P_o and P_c be the obtained FPSNR using a constant QP $q_o = q_c$ for the images $o(\mathbf{x}_n)$ and $v(\mathbf{x}_n)$. Suppose that the FPSNR $\tilde{P}_o$, the rate $\tilde{R}_o$ for $o(\mathbf{x}_n)$ and the FPSNR $\tilde{P}_c$, the rate $\tilde{R}_c$ for $v(\mathbf{x}_n)$ are obtained using a nonuniform quantization. Then, the compression gain G_c^q for the image $o(\mathbf{x}_n)$ is defined as

$$G_c^q = \frac{R_o - \tilde{R}_o}{R_o} \text{ subject to } P_o = \tilde{P}_o \tag{15}$$

and the compression gain G_c^q for the image $v(\mathbf{x}_n)$ is defined as

$$G_c^q = \frac{R_c - \tilde{R}_c}{R_c} \text{ subject to } \tilde{P}_c = P_c. \tag{16}$$

Thus, the total compression gain G_c^t for foveated video becomes

$$G_c^t = G_c^f + \left(1 - G_c^f\right) G_c^q = \frac{R_o - \tilde{R}_c}{R_o}$$
$$\text{subject to } \tilde{P}_c = P_c \text{ and } q_o = q_c \tag{17}$$

and $G_c^t = G_c^q$ for regular video.

III. Optimal Rate Control for Foveated Video

A. Optimal Rate Control in Curvilinear Coordinates

Let $r_k(q_k)$, $d_k(q_k)$, and q_k be the rate, distortion, and QP of the kth macroblock. Let M be the number of macroblocks in a picture. The QPs for coding M macroblocks consist of a quantization state vector $\vec{Q} = (q_1, q_2, \ldots, q_M)$. Suppose that R_T target bits are assigned to the picture. Then the optimal rate control is to find the state vector $\vec{Q}$ which minimizes the overall distortion: $D(\vec{Q}) = \sum_{k=1}^{M} d_k(q_k)$ subject to the rate constraint $R(\vec{Q}) = \sum_{k=1}^{M} r_k(q_k) \le R_T$. By introducing a Lagrange multiplier $\lambda \ge 0$, the constrained problem can be defined and solved.

For λ ranging from 0 to ∞, an optimal quantization state vector $\vec{Q}^*$ is obtained which minimizes the Lagrangian cost function $\mathcal{J}(\vec{Q}, \lambda) = D(\vec{Q}) + \lambda R(\vec{Q})$ while satisfying the rate constraint

$$\begin{aligned}\mathcal{J}(\vec{Q}^*, \lambda) &= \min\left[D(\vec{Q}) + \lambda R(\vec{Q})\right], \quad \forall \vec{Q}, \lambda \\ &= \sum_{k=1}^{M} j_k(q_k^*) = \sum_{k=1}^{M} \min[d_k(q_k) + \lambda r_k(q_k)], \\ q_1 &\le q_k \le q_M \end{aligned} \tag{18}$$

where $j_k(q_k)$ is the Lagrangian cost function for the kth macroblock and q_k^* is the optimal QP which minimizes $j_k(q_k)$ associated with the optimal Lagrange multiplier λ^*. Let Q be a set of allowable QPs. In MPEG/H.263 video coding, the set Q consists of positive integers from 1 to 31, and $\vec{Q}^*$ is the set of $q_k^* \in Q$.

In Cartesian coordinates, the distortion of the kth macroblock $d_k(q_k)$ is obtained by the mean square error (MSE) between the original image $o(\mathbf{x})$ and the reconstructed image $r(\mathbf{x})$ after coding with q_k. Suppose that $v(\mathbf{x}_n)$ and $g(\mathbf{x}_n)$ are foveated versions of image frames $o(\mathbf{x}_n)$ and $r(\mathbf{x}_n)$, respectively. In curvilinear coordinates, the normalized distortion $d_k(q_k)$ is given by

$$d_k(q_k) = \frac{1}{m_p} \sum_{p=1}^{m_p} [v(\mathbf{x}_{p,k}) - g(\mathbf{x}_{p,k})]^2 \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_{p,k}) \tag{19}$$

where $m_p = 384$ is the number of pixels in a macroblock and $\mathbf{x}_{p,k}$ is the pth pixel in the kth macroblock. Under assumption (8), (19) becomes $d_k(q_k) = (c_2/m_p) \sum_{p=1}^{m_p} [v(\mathbf{x}_{p,k}) - g(\mathbf{x}_{p,k})]^2 f_{p,k}^2$.

B. Exponential Expression for Rate–Distortion over Curvilinear Coordinates

The R–D function for a zero-mean, normally-distributed $N(0, \sigma^2)$ source with variance σ^2 is [28]

$$r(d) = \begin{cases} \frac{1}{2} \log \frac{\sigma^2}{d}, & 0 \le d \le \sigma^2 \\ 0, & d > \sigma^2. \end{cases} \tag{20}$$

In MPEG/H.263 video coding, an exponential expression is widely used for the R–D model

$$r(d) = \alpha \log_2 \frac{\sigma^2}{d} + \beta \tag{21}$$

where α and β are free variables. Applying this exponential form to the case of foveated video, the normalized variance σ_k^2 in curvilinear coordinates is obtained

$$\sigma_k^2 = \frac{1}{m_p} \sum_{b=1}^{m_b} \sum_{p=1}^{b_p} (o_b(p) - dc_b)^2 \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_{p,k}), \quad 1 \le b \le 6 \tag{22}$$

with

$$dc_b = \left[\sum_{p=1}^{b_p} \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_{p,k})\right]^{-1} \sum_{p=1}^{b_p} o_b(p) \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_{p,k}) \tag{23}$$

where

- $m_b = 6$ number of blocks in a macroblock consisting of four luminance blocks and two color blocks;
- $b_p = 64$ number of pixels in a block;
- $o_b(p)$ pth luminance pixel value of the bth block, which is a gray level value in I pictures and a differential value in P or B pictures.

C. Hierarchical Piece-Wise Rate–Distortion Model

To achieve optimal rate control, we characterize the R–D function in each macroblock for all $q \in Q$. Therefore, we encode each macroblock several times in order to estimate the R–D relation. To implement the optimal rate control algorithm efficiently, it is necessary to minimize the encoding time. In this paper, a hierarchical piece-wise (HPW) R–D model is introduced to enable efficient optimal rate control. If we assume that the Lagrange multiplier λ monotonically decreases in proportion to the rate, then the optimal value λ^* can be found along a local piece of the R–D curve constructed by all $q \in Q$. Via this HPW R–D model, we can construct the local R–D curve, including λ^*, hence reducing the computation load. Of course, this model can be applied for uniform video coding on Cartesian coordinates, as well as to foveated video coding on curvilinear coordinates.

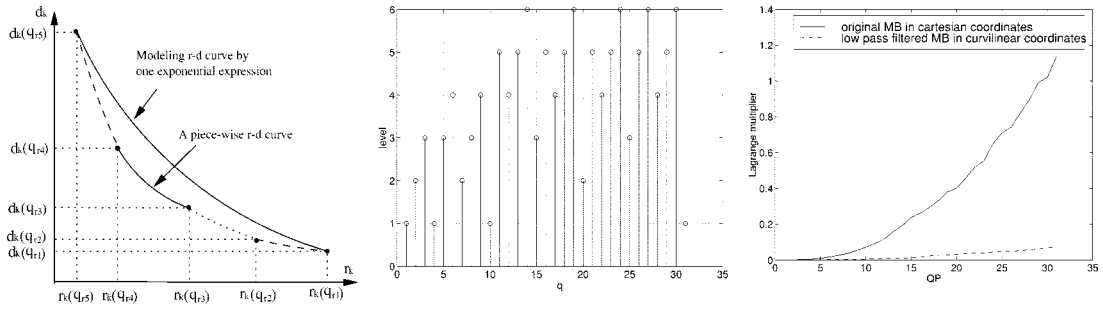
In order to obtain the R–D function, we employ the exponential model (21). One of the main advantages of this model is that computational redundancy can be reduced using only two variables (α, β), and the variance of macroblock is also used as one parameter in the model. Due to nonlinear effects in video coding, the global R–D curve for each macroblock is more heavily damped than the curve expressed by the model with two variables (α, β). To achieve more accurate modeling, the global R–D function must be obtained by several piece-wise local R–D functions, which are individually modeled by (21).

Given two reference quantization parameters (RQPs) q_{r_n} and $q_{r_{n+1}}$, a piece-wise R–D curve which represents the R–D function for $q_{r_n} \le q \le q_{r_{n+1}}$ is generated. For this piece-wise curve, α_k^n and β_k^n are, respectively, calculated according to

$$\alpha_k^n = \frac{r_k(q_{r_n}) - r_k(q_{r_{n+1}})}{\log_2 \left[\frac{d_k(q_{r_{n+1}})}{d_k(q_{r_n})}\right]} \tag{24}$$

$$\beta_k^n = r_k(q_{r_n}) - \alpha_k^n \log_2 \left[\frac{\sigma_k^2}{d_k(q_{r_n})}\right]. \tag{25}$$

Fig. 3(a) shows that a piece-wise R–D curve is constructed with the two RQPs (q_{r_4}, q_{r_5}), and that the curve expresses the real R–D function with increased accuracy. If the optimal value



(a) (b) (c)

Fig. 3. Optimal QP selection. (a) $r_k - d_k$ curve construction. (b) QP level. (c) Lagrange multiplier versus QP.

λ^* is not found on this R–D curve, then another local R–D curve is iteratively constructed incorporated into the current curve until λ^* is found.

D. Rate–Quantization Model

Suppose that we encode a macroblock with q_{r_n} and $q_{r_{n+1}}$, and obtain the rate and distortion values $r_k(q_{r_n})$, $r_k(q_{r_{n+1}})$, $d_k(q_{r_n})$, and $d_k(q_{r_{n+1}})$, and we construct the local R–D curve with α_k^n and β_k^n from (24) and (25). In order to estimate $r_k(q)$ for $q_{r_n} < q < q_{r_{n+1}}$, either the R–Q (rate–quantization) or the D–Q (distortion quantization) relation must be specified. The R–Q/D–Q relation can be more precisely estimated using the R–D model.

In MPEG/H.263 video, the rate is more predictable and is monotonic with the QP compared to the distortion. Here, the R–Q function is obtained by the slope of the R–D curve. Given two reference quantizers ($q_{r_n} < q_{r_{n+1}}$), define $\tilde{d}_k(q)$ for $q_{r_n} \le q < q_{r_{n+1}}$: $\tilde{d}_k(q) = d_k(q_{r_n}) + \mu_{r_n,r_{n+1}}(q - q_{r_n})$ where $\mu_{r_n,r_{n+1}} = [d_k(q_{r_{n+1}}) - d_k(q_{r_n})]/[q_{r_{n+1}} - q_{r_n}]$.

From (21), the slope of the R–D curve at each q becomes $\partial r_k(\tilde{d}_k(q))/\partial d_k = -\alpha_k^n/[(\log_2 e)(\tilde{d}_k(q'))]$. Therefore, the relation $r_k(q')$–q' for $q_{r_n} < q' < q_{r_{n+1}}$ is $r_k(q') = r_k(q'-1) - \gamma_{q'}(r_k(q_{r_n}) - r_k(q_{r_{n+1}}))$ where

$$\gamma_{q'} = \frac{\partial r_k(\tilde{d}_k(q'-1))}{\partial d_k}\left[\sum_{q=q_{r_n}}^{q_{r_{n+1}}-1}\frac{\partial r_k(\tilde{d}_k(q))}{\partial d_k}\right]^{-1}. \quad (26)$$

Using (21), $d_k(q') = \sigma_k^2 2^{[-r_k(q')/\alpha_k^n + \beta_k^n/\alpha_k^n]}$.

E. Hierarchical Piecewise R–D Model

The HPW R–D curve is constructed based on a reference level, which is needed to decide a RQP at each coding instant. Thus, each $q \in Q$ has a specified level. Let Q_{r^v} be the quantization set for the vth level and $q_{r_n^v} \in Q_{r^v}$ be the nth QP in the level. If Q consists of L levels, then $Q = Q_{r^1} \cup Q_{r^2} \cup Q_{r^3} \cdots \cup Q_{r^L}$. The local R–D curve is constructed based on a top–down method from level 1 to level L. Given the number of target bits $\hat{r}_k$, two RQPs $q_{r_n^1}$ and $q_{r_{n+1}^1} \in Q_{r^1}$ are initially selected for $r_k(q_{r_n^1}) \le \hat{r}_k \le r_k(q_{r_{n+1}^1})$. The next time, $q_{r_n^v}$ $(v > 1)$ is used to search for λ^*, according to the hierarchical structure of Q.

The number of QPs at each level and the maximum level L are important factors for achieving a fast convergence. In particular, the QP $q_{r_n^1} \in Q_{r^1}$ is chosen to characterize the global R–D curve with a low resolution. Therefore, the method for constructing Q_{r^1} is another factor to be considered for reducing the convergence time. In MPEG/H.263 video coding, the rate rapidly increases for small QPs. In order to characterize such abrupt rate changes for small QPs, the set Q_{r^1} must be well organized. The set Q_{r^v}, $v > 1$ is constructed from the set Q_{r^1} by a top–down approach:$q_{r_m^v} = (q_{r_n^{v-1}} + q_{r_{n-1}^{v-1}})/2$.

Therefore, for two consecutive $q_{r_n^1}$ and $q_{r_{n-1}^1}$ $(q_{r_n^1} > q_{r_{n-1}^1})$, the maximum level L is decided by

$$L = H\left[\log_2\left(\max\left[q_{r_n^1} - q_{r_{n-1}^1}\right]\right)\right] + 1, \quad q_{r_n^v} \in Q_{r^v} \quad (27)$$

where the function H rounds the input value to the next largest integer. For example, when $Q_{r_1} = \{1, 3, 10, 20\}$ and $1 \le q \le 31$, the level of each QP is obtained, as shown in Fig. 3(b).

Now let $C(r_1, r_2)$ be a piecewise R–D curve between q_{r_1} and q_{r_2} as constructed in (21), (24), and (25). For example, when $r_k(q_{r_{n+1}^1}) < \hat{r}_k \le r_k(q_{r_n^1})$, then $C(r_n^1, r_{n+1}^1)$ is obtained. Then, a QP $\hat{q}_k$ which minimizes $|\hat{r}_k - r_k(q)|$, $q_{r_n^1} \le q \le q_{r_{n+1}^1}$ is selected. If there exists $q_{r_m^2} \in Q_{r^2}$ and $q_{r_n^1} < q_{r_m^2} < q_{r_{n+1}^1}$, the next RQP becomes $q_{r_m^2}$. After coding with $q_{r_m^2}$, $C(r_n^1, r_{n+1}^1)$ is replaced by $C(r_n^1, r_m^2)$ and $C(r_m^2, r_{n+1}^1)$. In a similar way, an RQP in the higher level set is selected and the HPW R–D curve is constructed until a prescribed level is attained.

F. Convergence to an Optimal Lagrange Multiplier λ^*

An optimal Lagrange multiplier λ^* is iteratively searched for by sweeping λ along the r_k–d_k curves. From the slope of r_k–d_k, the corresponding λ_k is obtained by $\lambda_k = (-\partial d_k/\partial r_k)$. At the ith iteration, denote $\vec{Q}^i$ and $\vec{\lambda}^i$ as the state vectors whose components are q_k^i and λ_k^i, respectively, for $1 \le k \le M$. Then, $R(\vec{\lambda}^i) = \sum_k r_k(\lambda_k^i)$, and the range of λ^* can be found by the following lemma.

Lemma 1: Assume that the monotonic property is satisfied: if $\lambda_1 < \lambda_2 < \lambda_3$, then $r(\lambda_3) < r(\lambda_2) < r(\lambda_1)$ and $d(\lambda_1) < d(\lambda_2) < d(\lambda_3)$. Let $\lambda_M^i = \max[\lambda_k^i]$ and $\lambda_m^i = \min[\lambda_k^i]$, $1 \le k \le M$, and suppose that $R(\vec{\lambda}^i) = R_T$. Then there exists λ^* in $\lambda_m^i \le \lambda^* \le \lambda_M^i$.

Proof: Since $\lambda_M^i \le \lambda_k^i \le \lambda_m^i$, then $r_k(\lambda_m^i) \le r_k(\lambda_k^i) \le r_k(\lambda_M^i)$ and

$$\sum_k r_k\left(\lambda_M^i\right) \le \sum_k r_k\left(\lambda_k^i\right) = R_T \le \sum_k r_k\left(\lambda_m^i\right). \quad (28)$$

The biased Lagrangian cost is $W(\lambda) = J(\lambda) - \lambda R_T = \sum_k [d_k(\lambda) + \lambda r_k(\lambda)] - \lambda R_T$. The derivative of W with respect to λ is $(\partial W/\partial \lambda) =$

$\sum_k[(\partial d_k(\lambda)/\partial\lambda) + \lambda(\partial r_k(\lambda)/\partial\lambda)] + \sum_k r_k(\lambda) - R_T$. Since $\lambda = -(\partial d_k(\lambda)/\partial r_k(\lambda))$, $(\partial W/\partial\lambda) = \sum_k r_k(\lambda) - R_T$. At the optimal Lagrange multiplier λ^*, $\partial W/\partial\lambda = 0$ which yields $\sum_k r_k(\lambda^*) = R_T$. From (28), $\sum_k r_k(\lambda_M^i) \le \sum_k r_k(\lambda^*) \le \sum_k r_k(\lambda_m^i)$ and by the monotonic property, $\lambda_m^i \le \lambda^* \le \lambda_M^i$. Thus, there exists an optimal value λ^* which meets the constraint $\sum_k r_k(\lambda^*) = R_T$ on the bound of $\lambda^* \le \lambda_M^i$. ■

In order to find the optimal Lagrange multiplier λ^*, a reference Lagrange multiplier $\bar{\lambda}^i$ is selected in the range $\lambda_m^i < \bar{\lambda}^i < \lambda_M^i$. Then, the rate difference $\Delta r_k^i = r_k(\lambda_k^i) - r_k(\bar{\lambda}^i)$ is

$$\Delta r_k^i = \int_{\bar{\lambda}^i}^{\lambda_k^i} \frac{\partial r_k^i}{\partial\lambda} d\lambda = c_k^i \left(\lambda_k^i - \bar{\lambda}^i\right) \tag{29}$$

where c_k^i is a constant. For a given Lagrange multiplier vector $\vec{\lambda}^i$ for all macroblocks, suppose that $R(\vec{\lambda}^i) = R_T$ where $R(\vec{\lambda}^i)$ is the sum of rates for all macroblocks at λ_k^i. Then, $\delta R^i = \sum_k \Delta r_k^i = R_T - \sum_k r_k^i(\bar{\lambda}^i)$. From (29), $\bar{\lambda}^i = (\delta R^i + \sum_k c_k^i \lambda_k^i)/\sum_k c_k^i$. The optimal Lagrange multiplier λ^* is obtained by sweeping $\bar{\lambda}^i$ from λ_m^i to λ_M^i. When $\bar{\lambda}^i$ approaches to λ^*, δR^i converges to zero: $\lim_{\bar{\lambda}_i \to \lambda^*} \delta R^i = 0$.

Given $\vec{\lambda}^i$, assume that $\sum_k r_k(\lambda_M^i) < R_T < \sum_k r_k(\lambda_m^i)$. If $\lambda_m^i < \bar{\lambda}^i < \lambda_M^i$, the rate bound of λ^* with respect to $\bar{\lambda}^i$ is $r_k(\bar{\lambda}^i) < r_k(\lambda^*) < r_k(\lambda_m^i)$ for $\lambda^* < \bar{\lambda}^i$, and $r_k(\lambda_M^i) < r_k(\lambda^*) < r_k(\bar{\lambda}^i)$ for $\bar{\lambda}^i < \lambda^*$. Therefore, $\bar{\lambda}^i < \lambda^* < \lambda_M^i$ for $R_T < R(\bar{\lambda}^i)$, and $\lambda_m^i < \lambda^* < \bar{\lambda}^i$ for $R(\bar{\lambda}^i) < R_T$.

When $R(\lambda_M^i) < R(\bar{\lambda}^i) < R(\lambda_m^i)$ for $\lambda_m^i < \bar{\lambda}^i < \lambda_M^i$, $\lim_{i\to\infty} \bar{\lambda}^i = \lambda^*$ can be found by the following method. Let $\delta\lambda^i$ be the search range for λ^* at the ith iteration which is given by $\delta\lambda^i = \lambda_M^i - \lambda_m^i$. When $R(\bar{\lambda}^i) < R_T < R(\lambda_m^i)$, set $\lambda_m^{i+1} = \lambda_m^i$, $\lambda_M^{i+1} = \bar{\lambda}^i$, and $\bar{\lambda}^i = (\lambda_m^i + \lambda_M^i)/2$. Then, $\delta\lambda^{i+1} = \lambda_M^{i+1} - \lambda_m^{i+1} = (\lambda_M^i - \lambda_m^i)/2$. Conversely, when $R(\lambda_M^i) < R_T < R(\bar{\lambda}^i)$, we set $\lambda_M^{i+1} = \lambda_M^i$, $\lambda_m^{i+1} = \bar{\lambda}^i$, and $\bar{\lambda}^i = (\lambda_m^i + \lambda_M^i)/2$. Generally, $\delta\lambda^{i+1} = \delta\lambda^i/2 = (\lambda_M^1 - \lambda_m^1)/2^i$. Therefore, $\lim_{i\to\infty} \delta\lambda^i = 0$ and $\bar{\lambda}^\infty = \lambda^*$.

Since the value of each $q \in Q$ is discrete, each λ_k^i is not equal to λ^*, and we obtain $\lambda_k^i(q_k^i) = d_k^i(q_k^i)/r_k^i(q_k^i)$ approximately. In very low bit rate video coding, λ^* tends to approach λ_M to satisfy a rate constraint. Since λ_k is maximum when QP is 31, λ_k is thresholded to λ_k (31) which can be less than λ^*. In such a case, it is difficult to find an r–d relation around λ^*, and to expect a good performance improvement using an optimal rate control. However, in foveated video coding, the minimum bound of the bit rate is lower than that of normal video coding. Thus, the optimal value λ^* exists near a median value between λ_M and λ_m, so that the performance improvement should be larger than in normal (nonfoveated) video coding.

G. Iterative Convergence Algorithm

From the monotonic property of the R–D function proved in Lemma 1, estimates of λ converge to optimal λ^* within a limited encoding time instead of sweeping λ from 0 to ∞. In order to reduce the encoding time while maintaining the rate constraint, an iterative convergence algorithm is employed. The desired optimal constant slope value λ^* is not known prior to coding, and it is dependent on the desired target budget. The procedure for finding the value λ^* is encapsulated in the following steps.

Step 1) Allocate target bits into each macroblock. Let $\hat{r}_k^1$ be the number of assigned bits for the kth macroblock for the first iteration.

Step 2) Search two RQPs $q_{r_n^1}$ and $q_{r_{n+1}^1} \in Q_{r^1}$ which satisfy $r_k(q_{r_n^1}) \le \hat{r}_k^1 \le r_k(q_{r_{n+1}^1})$, and calculate $d_k(q_{r_n^1})$ and $d_k(q_{r_{n+1}^1})$.

Step 3) Construct a piece-wise R–D curve for the level 1 using (24) and (25).

Step 4) Repeat Step 1–Step 3 to a specified level and construct R–D curves using the HPW algorithm.

Step 5) Find $\hat{q}_k^i = \arg\min[q]$ for minimizing $|\hat{r}_k^i - r_k(q)|$.

Step 6) Calculate a Lagrange multiplier $\hat{\lambda}_k^i = d_k(\hat{q}_k^i)/r_k(\hat{q}_k^i)$.

Step 7) Based on $\hat{\lambda}_k^i$, obtain a Lagrange multiplier $\bar{\lambda}^i$ for all macroblocks in a picture.

Step 8) Find $\hat{q}_k^i = \arg\min[q]$ for minimizing $|\lambda(\hat{q}_k^i) - \bar{\lambda}^i|$.

This iterative procedure (Step 1–Step 8) is continued until the rate constraint is satisfied. After convergence, $\bar{\lambda}^i = \lambda^*$ and $\hat{q}_k^i = q_k^*$ which is the optimal QP of the kth macroblock. The optimal QP vector $\vec{Q}^*$ consists of q_k^* for $1 \le k \le M$.

H. Suboptimal Rate Control in H.263 Video Coding

To achieve optimal rate control on curvilinear coordinates, the FMSE in (10) must be minimized. The normalized distortion in (19) is proportional to the Jacobian of the coordinate transformation. Since the magnitude of the Jacobian is less than or equal to one, the distortion along curvilinear coordinates is always less than, or equal to the distortion on Cartesian coordinates. The distortion along curvilinear coordinates is generally reduced in proportional to the transform ratio.

Fig. 3(c) shows an example of the variation of the Lagrange multiplier magnitude in an original macroblock, and the corresponding low-pass filtered macroblock with respect to the QP. For a given QP, the magnitude of the Lagrange multiplier in curvilinear coordinates is much less than that in Cartesian coordinates due to the abovementioned factors. The Lagrange multiplier in the low-pass filtered macroblock also slowly varies according to the QP as compared to the original macroblock. Therefore, the ratio of the distortion rate relative to the bit rate is much higher in the original macroblock. In other words, the lowpass filtered macroblock is less sensitive to changes in the QP than is the original macroblock. In foveated video coding, since the high-frequency components in the background are removed, a large QP may be used without degrading performance.

Next, we develop an approach for suboptimal rate control. First, we find an optimal QP set from the proposed iterative procedure, under the assumption that the QP difference is zero. Here, we denote q_k^* as the optimal QP of the kth macroblock, and r_k^* as the obtained rate using the model.

Let U be a macroblock index set. The set consists of macroblock indices whose average local bandwidth is less than a threshold f^{th}. Then, determine a QP $\bar{q}$ which is the minimum QP that satisfies the following rate constraint for macroblocks in U : $\sum_{k\in U} r_k(q) \le \sum_{k\in U} r_k^*(q_k^*)$ where r_k is generated bits in real video coding, and $f^{\text{th}} = 0.25$ in the simulation.

Finally, we find the QPs for macroblocks which are not in U. Assume that the (increasing) rate of change of the Lagrange

cost function j_k, with respect to the change in the QP is proportional to the average local frequency of each macroblock. Since the value of j_k is minimum at q_k^* for λ^*, we must maintain the value q_k^* without change to obtain the minimum Lagrange cost function. Thus, we determine the QPs for the remaining macroblocks in descending order of average local bandwidth. In the normal quantization mode, the dynamic range of the current QP is limited by the QP values of the previous macroblocks. Let k and i be the index of the current macroblock and the adjacent previous macroblock, respectively. Since the coding order can be backward or forward according to the local bandwidth, we must consider the adjacent previous macroblock index i for both directions. Then, the allowed range of q_k is

$$\max[1, q_i - 2\times|k-i|] \le q_k \le \min[31, q_i + 2\times|k-i|]. \quad (30)$$

In the modified quantization mode, we can choose any QP value for the current macroblock. Then, we select q_k which minimizes $j_k(q)$ for a given set of available quantizers while satisfying the rate constraint

$$r_k(q) \le \sum_{p\in V} r_p^*\left(q_p^*\right) + r_k^*\left(q_k^*\right) - \sum_{p\in V} r_p(q_p) \quad (31)$$

where V is a previously coded macroblock index set. To calculate $j_k(q)$, we use the optimal Lagrange multiplier λ^* obtained by Step 8.

I. Real-Time Rate Control in H.263 Video

Section III-H focuses on developing standard compatible rate coding schemes using *normal mode* and *modified mode*. When using the rate–distortion model, it is possible to reduce the number of encodings to obtain an optimal Lagrange multiplier λ^*. Nevertheless, it is necessary to develop very simple coding algorithms for real-time embedded systems by using one-time encoding. This paper focuses on how well visual quality can be improved by optimal rate control based on the quality criterion FPSNR rather than PSNR, and on developing standard compatible optimal-rate control algorithms.

By exploiting the nonuniform resolution property of the retina, it is possible to implement a very simple rate control algorithm. Defining the local bandwidth to be circularly symmetric with respect to a foveation point, with maximum at the foveation point, it is possible to construct a set of circularly symmetric QPs. In other words, a minimum QP is decided at the foveation point, then the rest of the QPs are determined using the minimum QP and the average value of the local bandwidth. This method can be also generalized for the generation of foveated videos that have multiple foveation points. Furthermore, it is possible to reduce the coding redundancy due to *DQUANT* information as long as both the encoder and the decoder use a protocol that constructs such a circularly symmetric QP set. In other words, the encoder only sends single/multi-foveation points and the associated minimum QP values. The decoder then constructs the set of QPs according to the protocol. In this way it is possible to reduce operational redundancies at low bit rates and also afford high visual quality.

Let Q_m be the minimum QP value and q_k be the QP of the kth macroblock. Then, q_k is given by

$$q_k = \frac{Q_m}{[2\bar{f}_k]^n} \quad (32)$$

where $\bar{f}_k$ is the average local bandwidth of the kth macroblock and $n \in \mathbf{N}$. If $\bar{f}_k$ is the maximum discrete frequency 0.5, then $q_k = Q_m$. Then, the decoder can reconstruct the QP q_k using (32).

IV. Simulation Results

A. Optimal Rate Control Based on H.263 Video

In our simulations with H.263 video coding, a reference frame rate of 30, two skip frames, and a target frame rate of 10 are used. To measure P picture quality, we use the previous original image as a reference image for current P pictures. To demonstrate the efficacy of foveated video compared to normal video, optimal rate control using Lagrange multiplier λ is implemented for minimizing the MSE or the FMSE in (10). The codeword due to the QP difference is not counted at the following four methods.

- Method 1: Optimal rate control for minimizing the MSE between the original image $o(\mathbf{x}_n)$ and the reconstructed image $r(\mathbf{x}_n)$.
- Method 2: Optimal rate control for minimizing the FMSE between $o(\mathbf{x}_n)$ and $r(\mathbf{x}_n)$.
- Method 3: Optimal rate control for minimizing the MSE between the foveated image $v(\mathbf{x}_n)$ of the original image $o(\mathbf{x}_n)$ and the reconstructed image $g(\mathbf{x}_n)$ of $v(\mathbf{x}_n)$.
- Method 4: Optimal rate control for minimizing the FMSE between $v(\mathbf{x}_n)$ and $g(\mathbf{x}_n)$.

Fig. 4 shows the coded images for a given compression ratio ($\mathrm{bpp} = 0.36$, i.e., 36.13 kb) in the above four methods. Because Method 1 minimizes the MSE in Cartesian coordinates, the PSNR in Method 1 is higher than that in Method 2. However, the FPSNR in Method 1 is lower than that in Method 2, i.e., when we focus our attention on the fixation point from an appropriate distance, the quality of the reconstructed image using Method 2 is better. In Method 4, the FPSNR is the largest compared to the other methods, and provides the best subjective quality for foveated video, at least in this case. Fig. 4(d) shows that the reconstructed image around the foveation point is similar in quality to the original image.

B. Performance Measurement of the HPW Model

To measure the performance of the HPW model, we simulate the following two cases.

- *Ideal RC:* we obtain $r_k(q)$ and $d_k(q)$ for all $q \in Q$ and decide the optimal QP set $\vec{Q}^*$ for a given rate constraint.
- *Approximate RC:* we code each macroblock with the QPs of levels 1 and 2, where the QP number of those levels is 7, as shown in Fig. 3(b). Based on the piece-wise R–D model, the values of $r_k(q)$ and $d_k(q)$ for levels higher than 2 are estimated. Then, the set $\vec{Q}^*$ is obtained in the same way as the first method.

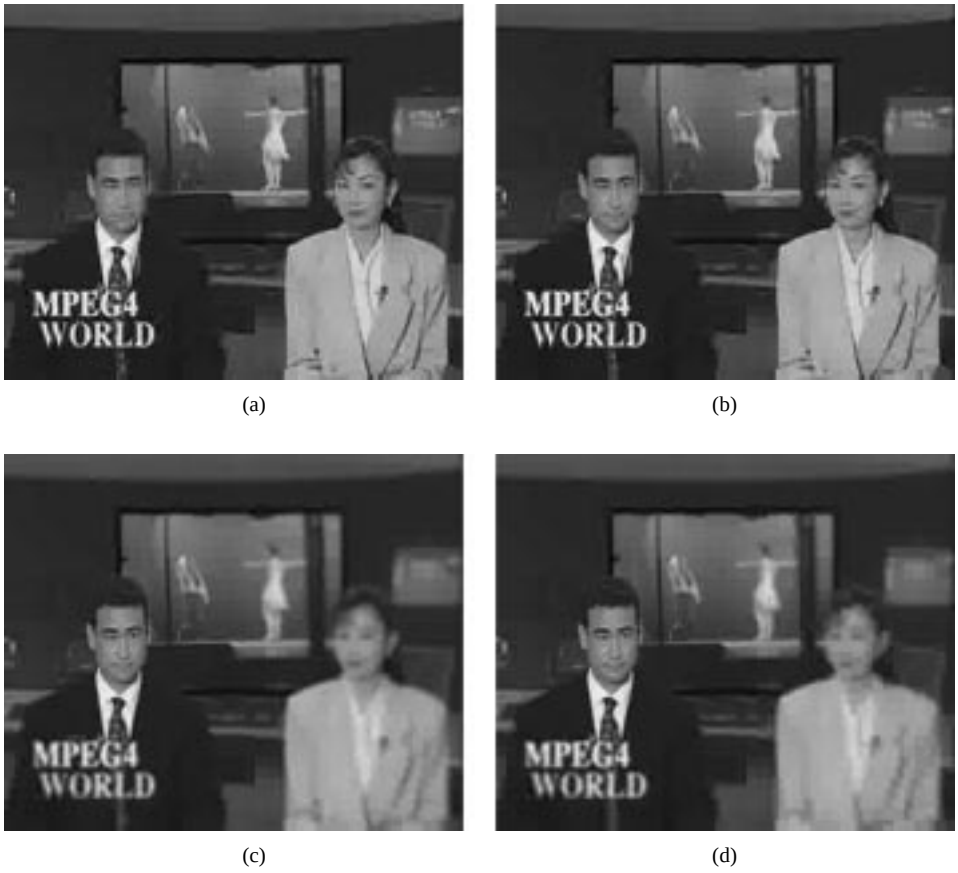


Fig. 4. Optimal rate control for an I picture in "News" CIF image (bpp $= 0.36$, i.e., 36.13 kb/s). (a) Method 1: PSNR $= 30.10$, FPSNR $= 29.54$. (b) Method 2: PSNR $= 30.05$, FPSNR $= 30.45$. (c) Method 3: PSNR $= 34.94$, FPSNR $= 33.53$. (d) Method 4: PSNR $= 33.92$, FPSNR $= 35.01$.

The piece-wise R–D model demonstrates exact $r(q)$–$d(q)$ estimation. The FPSNR and the PSNR using the approximate rate control are nearly identical to those in using the ideal rate control.

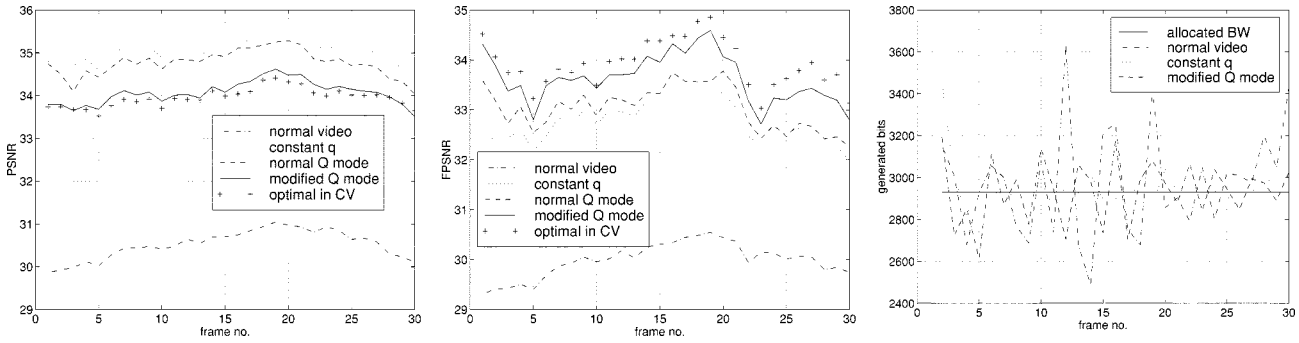
For optimal rate control, the iteration number for finding the optimal state vector $\vec{Q}^*$ is measured. Here, it is assumed that the monotonic properties of rate and distortion relative to λ are satisfied. When we approximately code the macroblock from level 2, the coding number is always less than five.

C. Performance Measurement for Suboptimal Rate Control

For H.263 video coding, the number of bits allocated to consecutive QP differencing must be considered. When the number of generated bits due to the QP difference is ignored, then the reconstructed picture quality using the optimal rate control becomes an upper bound on the coding performance. To compare the performance relative to the optimal rate control, the following coding methods are employed.

- *Normal video:* In order to maximize the compression ratio, the original image sequence is coded by $\mathrm{QP} = 31$.
- *Constant q:* For the foveated image sequence, constant QP is decided for each picture. The transmission rate is set to the average rate in *normal video.*
- *Normal Q mode:* The foveated image sequence is coded by using the suboptimal rate control algorithm in curvilinear coordinates with the normal quantization mode.
- *Modified Q mode:* The foveated image sequence is coded by using the suboptimal rate control algorithm in curvilinear coordinates with the modified quantization mode.
- *Optimal in CV:* Optimal rate control for minimizing the FMSE for the foveated image sequence. In this method, we do not consider the number of generated bits due to the QP difference so that the performance of this method can be used as the upper bound of the above four methods. On the other hand, the QP difference is taken into account in the above four methods.

When the QP is set to 31 for the 30 frames of the "News" CIF image sequence (with two skip frames), the number of generated bits is 35.4 kb for the I picture and the coding rate is 29.3 kb/s for the following P pictures. These rates are used as the target rates for the foveated video. The reconstructed picture quality and the average values are shown in Fig. 5 and Table I. The ratio PSNR/FPSNR for normal video coding is 3 dB less than that of the foveated video coding. A trade-off between the PSNR and the FPSNR is shown among the rate control methods. The PSNR of the optimal rate control method is the lowest, even if the number of bits due to the QP difference is not counted, but the FPSNR of this method is an upper bound in curvilinear coordinates. Because of the flexibility in changing the QP value, the FPSNR for rate control using the modified quantization mode is improved to 0.6 dB compared to the normal quantization mode. The average FPSNR for suboptimal rate control using the modified quantization mode is 0.3



(a) (b) (c)

Fig. 5. Reconstructed picture quality/transmission rate of P-pictures according to the rate control methods in H.263, where the allocation bandwidth is 29.3 kb/s for P-pictures and 354 kb for the first I picture. (a) PSNR versus frame no. (b) FPSNR versus frame no. (c) Generated bits versus frame no.

TABLE I
AVERAGE VALUE OF PSNR/FPSNR FOR H.263 VIDEO AT 29.3 kb/s FOR P-PICTURES AND 354 kb FOR THE FIRST I PICTURE

	normal video	*constant q*	*normal Q mode*	*modified Q mode*	*optimal in CV*
PSNR	30.52	35.04	34.76	34.06	33.95
FPSNR	29.99	32.74	33.03	33.63	33.92

dB less than the upper bound of the FPSNR. The number of bits generated for P pictures using the various rate control methods are shown in Fig. 5(c).

D. Measuring Performance over Wireless Networks

In order to compare the performance of the various rate control methods using standard video compression algorithms (MPEG-4 baseline and H.263++), the following simulation environments were employed. We set the target transmission rate to 51.1 kb/s and the target frame rate to 10 frames/s for the CIF *News, Mobile,* and *Akiyo* image sequences. In addition, we used the QCIF *carphone* (176×144) image sequence with a reference frame rate of 30 frames/s, a target frame rate of 15 frames/s, and a target transmission rate of 64 kb/s. The decoder includes the error resilience features supported by MPEG-4/H.263++, e.g., independent segment decoding, data partitioning, reversible variable length codes (RVLCs), and reference picture selection [29]. Since the foveated video bitstreams maintain 100% compatibility with the bitstream syntax of the standard videos (MPEG and H.263), we measured coding performance by varying the rate control algorithm for each encoding/decoding technique.

In order to measure the coding performance over fading channels, real fading statistics collected in the downtown area of Austin, Texas, at 1.9 GHz were used in our simulations [30]. For channel coding, the rate compatible punctured convolutional (RCPC) codes in H.223 Annex C were implemented. The coding rate was adaptively changed according to the feedback channel SNR and the punctured pattern in the H.223 standard. To compare the performance of the suboptimal rate control algorithms, the following coding methods were employed.

- Method 1: constant q + regular video sequence
- Method 2: modified Q mode + regular video sequence
- Method 3: constant q + foveated video sequence
- Method 4: modified Q mode + foveated video sequence

Fig. 6(a) shows an original *carphone* image and Fig. 6(b) is the foveated version of the original image. Table II shows the reconstructed video quality measured in PSNR and FPSNR and the number of skip frames (per 30 frames) at the average channel SNR of 10 dB.

For a given video sequence, it is possible to increase the *FPSNR* by around 0.2 dB and reduce the number of skip frames by 0.3–1.68 frames by using *modified Q mode* compared to *constant q*. In the CIF image sequences, the magnitude of the PSNR in Method 2 and Method 4 was reduced by 0.57–0.89 dB relative to Method 1 and Method 3 while increasing the temporal resolution. Since the CIF image sequences *Akiyo* and *News* do not contain much high-frequency information in the background regions, increasing the FPSNR and the temporal resolution results in a decrease in the PSNR. In the *QCIF* image sequence, the value of the PSNR also increased by 0.21–0.34 dB by using Method 2 and Method 4. Motion compensation errors are effectively reduced in the background. The bits were allocated on regions about the foveated face, preventing degradation in quality from temporal error propagation.

To measure performance at low bit rates, the following simulation was done. The QCIF image sequences (*carphone, claire, coastguard, foreman, salesman*) were compressed by using the above four methods at a target transmission rate of 30 kb/s over error-free channels where each sequence consists of 300 frames. Table III shows the reconstructed video quality and the number of skip frames (per 30 frames). It is observed that the FPSNR is improved and the number of skip frames is reduced by Metho 4.

E. Quality and Compression Gain Measurements

In order to compare the performance of the PSNR versus the FPSNR, we compress both the regular and the foveated image sequences and measure the quality as a function of the QPs. In

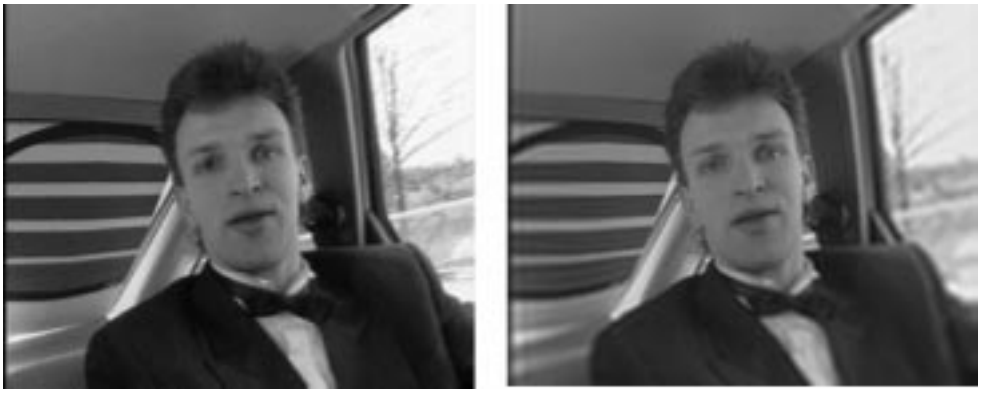


Fig. 6. Original image versus foveated image. (a) Original *carphone* image. (b) Foveated *carphone* image.

TABLE II
AVERAGE VALUE OF PSNR (dB)/FPSNR (dB)/THE NUMBER OF SKIP FRAMES (PER 30 FRAMES) FOR H.263++ VIDEO AT THE AVERAGE CHANNEL SNR 10 dB

	CIF images			QCIF images		
	PSNR	FPSNR	# of skip frames	PSNR	FPSNR	# of skip frames
Method 1	33.64	32.52	7.23	31.95	31.56	2.01
Method 2	33.07	32.74	5.55	32.29	31.75	1.74
Method 3	36.40	34.27	5.89	33.85	32.96	1.98
Method 4	35.51	34.54	5.02	34.06	33.08	1.68

TABLE III
AVERAGE VALUE OF PSNR/FPSNR (dB) AND THE NUMBER OF SKIP FRAMES FOR THE QCIF IMAGE SEQUENCES (300 FRAMES) AT 30 kb/s: PS (PSNR), FS (FPSNR), NS (THE NUMBER OF SKIP FRAMES)

	carphone		*claire*		*coastguard*		*foreman*		*salesman*	
	PS/FS	NS	PS/FS	NS	PS/FS	NS	PS/FS	NS	PS/FS	NS
Method 1	30.6/30.3	2.1	37.2/36.8	1.6	26.7/26.3	3.6	28.9/28.4	2.4	32.4/32.0	1.5
Method 2	30.2/30.8	2.1	36.9/37.2	1.7	26.4/26.9	3.6	28.2/29.2	2.4	32.1/32.4	1.5
Method 3	32.5/31.9	1.8	38.3/38.1	1.5	28.3/27.9	3.3	30.4/30.1	2.1	34.4/34.1	1.2
Method 4	31.9/33.1	1.8	38.0/38.5	1.5	28.0/28.8	3.3	29.5/31.2	2.1	34.0/34.5	1.2

Fig. 7(a) and (b), the picture quality is measured using the PSNR and the FPSNR for the "News" image sequences. The relative variation of the PSNR and the FPSNR relative to the QPs is similar in both I and P frames. In the regular sequence, the PSNR is approximately equal to the FPSNR. In the foveated sequence, the PSNR is greater than the FPSNR because the high-frequency reduction in the background is more heavily weighted as compared to the FPSNR. The FPSNR in the regular and foveated sequences are similar. In the foveated sequence with three foveation points, the motion activity is high relative to the background. The FPSNR is slightly less than that of the regular sequence. In the CIF "Mobile" sequence, there are a lot of motion compensated errors over the background. Therefore, the FPSNR of the foveated sequence is relatively larger than that of the regular sequence in Fig. 7(c) and (d). The FPSNR drift is taken into account to obtain the compression gain in (15) and (16).

Foveation is an effective way to improve visual quality at low bit rates. Here, the compression gain is measured at $q \geq 16$. In Fig. 8, the compression gains G_c^f and G_c^q are measured by (14)–(16) according to q and the FPSNR. When q is decreasing, higher frequency components are coded in the regular sequence. Thus, the compression gain increases at smaller q and higher rates. Fig. 8(a) shows the compression gain G_c^f for the "News" image sequence. Higher compression gains are obtained in the I frames than the P frames. In addition, foveated image sequences with a single foveation point demonstrate higher compression gains than sequences with multiple foveation points. In the "Mobile" image sequence, a higher compression gain is found in the P frames in Fig. 8(b). Generally, the compression gain depends on the number of motion compensation errors and on the complexity of the image sequence. The compression gain in the P frames is 23% for the single foveation "News" sequence, 5% for the multiple foveated "News" sequence, and 65% for the "Mobile" sequence.

The compression G_c^q gains due to nonuniform quantization are demonstrated in Fig. 8(c) and (d) for the regular sequence

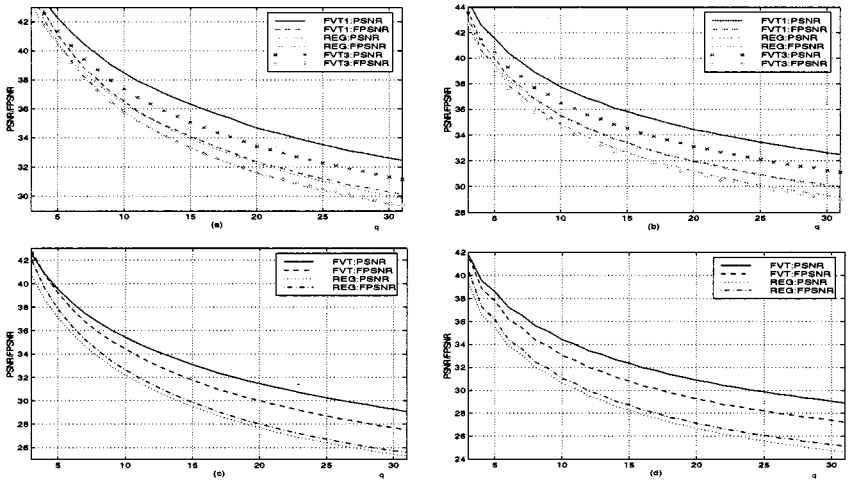


Fig. 7. Comparison of PSNR versus FPSNR against foveation and QPs where FVT1(FVT 3) is the foveated image with 1(3) foveation point(s) and REG is the regular image. (a) "News" I frame, (b) "News" P frame, (c) "Mobile" I frame, (d) "Mobile" P frame.

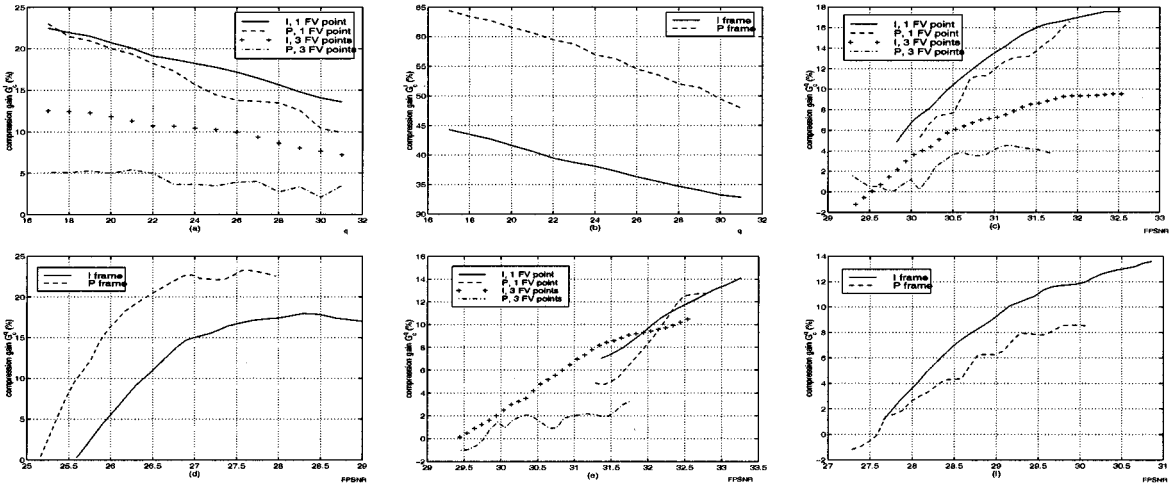


Fig. 8. Compression gain G_s^f due to foveation filtering (a), (b): (a) G_s^f for "News," (b) G_s^f for "Mobile." Compression gain G_s^q due to nonuniform quantization (c)–(f): (c) G_s^q for regular "News," (d) G_s^q for regular "Mobile," (e) G_s^q for foveated "News," (f) G_s^q for foveated "Mobile."

and Fig. 8(e) and (f) for the foveated sequence. The value of G_c^q increases as the FPSNR increases. For the regular sequences, the compression gain is in the range 5–17% for the "News" sequence with a single foveation point, 0–10% for the "News" sequence with the multiple foveation points, and 0–23% for the "Mobile" sequence. For the foveated sequences, the gain is obtained in the range of 5–14% for the "News" sequence with a single foveation point, 0–10% for the "News" sequence with multiple foveation points, and 0–14% for the "Mobile" sequence. The gain due to the quantization is less than the gain G_c^f due to the foveation filtering. The obtained compression gains due to nonuniform quantization are similar for the regular and the foveated video or the I and P frames.

The total gain G_c^t is obtained by (17) in Table IV. In the regular video, there is no gain obtained by the quantization at very low bit rates. Hence, the compression gain G_c^t is around zero at low rates. As the rates are increased, it is possible to manipulate the total bit budget over the curvilinear coordinates to obtain larger compression gains. Since the total gain for foveated video includes the gain due to foveation filtering and nonuniform quantization, the compression gains can be improved in the range of 5%–40% relative to the normal video. In the Mobile sequence, motion compensation errors are effectively reduced in the background. Thus, the compression gain is improved 68% for P frames. On the other hands, in the "News" sequence with three foveation points, the compression gain is slightly increased due to the low-/high-frequency reduction in the P frames (2%–7%).

TABLE IV
TOTAL COMPRESSION GAIN: NEWS 1 HAS ONE FOVEATION POINT, NEWS 3 HAS THREE FOVEATION POINTS

	Regular video						Foveated video					
	News 1		News 3		Mobile		News 1		News 3		Mobile	
coding type	I	P	I	P	I	P	I	P	I	P	I	P
MIN(G_c^t %)	5	5	0	0	0	0	20	14	8	2	33	48
MAX(G_c^t %)	20	17	10	5	18	23	56	33	21	7	52	68

V. Conclusions

The potential benefit that can accrue from using foveated image/video coding is a possible dramatic improvement in visual quality for a given number of target bits. Given a fixation point(s) and a set of viewing parameters, the local bandwidth of the image can be from models of the human visual system [6]. In order to evaluate foveated image/video quality, the FSNR has been utilized to measure the reconstructed picture quality. The development of optimal rate control algorithms makes it possible to improve the picture quality for a given target bit rate.

Most traditional rate control algorithms are focused on maximizing the SNR of the reconstructed picture using a Lagrange multiplier method. In this paper, we established the optimal rate control algorithm for maximizing the visual quality while maximizing the FSNR using a Lagrange multiplier method along curvilinear coordinates. Moreover, we introduced several efficient rate control algorithms: the nonuniform target bit allocation for MPEG TM5, the exponential expression for the rate–distortion model in curvilinear coordinates, the HPW R–D (rate–distortion) model for efficient rate control, and so on.

For H.263 video coding, a suboptimal rate control algorithm was developed. In this algorithm, we classify macroblocks into two groups according to the average local frequency. For the group whose frequency is less than a specified threshold, we used a constant QP. For the other group, we decided a QP for each macroblock while considering the quantization mode, the Lagrange cost function, and the average local frequency.

From the simulation results, we demonstrated that the optimal rate control for minimizing the FMSE is supplies excellent visual fidelity. In addition, using the HPW R–D model, the reconstructed picture quality using the suboptimal rate control algorithm is close to that obtained using optimal rate control. For efficient implementation, an iterative algorithm was introduced to reduce the search time required to find the optimal Lagrange multiplier λ^*.

In order to demonstrate the performance of foveated video coding at very low bit rates, we compared the reconstructed picture quality of the normal video with that of foveated video. For foveated video, we set the rate to be equivalent to the number of bits generated by normal coding. Using foveated video, we demonstrated a significant increase in the FPSNR compared to normal video. Overall, for certain very low bit rate coding applications, where foveation is a reasonable technology, foveated compression and rate control shows great potential in terms of compression performance, speed, and visual quality.

References

[1] P. G. J. Barten, "Evaluation of subjective image quality with the square-root integral method," *J. Opt. Soc. Amer.*, vol. 7, pp. 2024–2031, Oct. 1990.

[2] 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.*, vol. 8, pp. 1775–1787, Nov. 1991.

[3] G. Beach, C. J. Cohen, J. Braun, and G. Moody, "Eye tracker system for use with head mounted displays," in *Proc. IEEE ICSMC*, vol. 5, 1998, pp. 4348–4352.

[4] W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communincation," *Proc. SPIE*, vol. 3299, 1998.

[5] *Operate Your PC with Your Eye Eyetech Digital Systems, Inc.*, http://www.eyetechds.com/.

[6] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," IEEE Trans. Multimedia, submitted for publication.

[7] I. Sutherland, "The ultimate display," in *Proc. IFIP Congr.*, 1965, pp. 506–508.

[8] H. Ohzu and K. Habara, "Behind the scenes of virtual reality: vision and motion," *Proc. IEEE*, vol. 84, pp. 782–798, May 1996.

[9] F. Kishino, T. Miyasato, and N. Terashima, "Virtual space teleconferencing 'communication with realistic sensation," in *Proc. RO-MAN*, July 1995, pp. 205–210.

[10] J. M. Rosen, H. Soltanian, R. J. Redett, and D. R. Laub, "Evolution of virtual reality from planning to performing surgery," *IEEE Eng. Med. Biol. Mag.*, pp. 16–22, Mar./Apr. 1996.

[11] S. Daly, K. Matthews, and J. Ribas-Corbera, "Visual eccentricity models in face-based video compression," *Proc. SPIE*, vol. 3644, Jan. 1999.

[12] P. L. Silsbee, A. C. Bovik, and D. Chen, "Visual pattern image sequence coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 291–301, Aug. 1993.

[13] T. H. Reeves and J. A. Robinson, "Adaptive foveation of MPEG video," in *Proc. 4th ACM Int. Multimedia Conf.*, Boston, MA, 1996, pp. 231–241.

[14] S. Lee and A. C. Bovik, "Very low bit rate foveated video coding for H .263," in *Proc. IEEE ICASSP*, Phoenix, AZ, Mar. 1999, pp. VI3113–VI3116.

[15] N.-D. Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Processing*, vol. 9, pp. 636–650, Apr. 2000.

[16] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 525–536, July 1974.

[17] S. Lee, A. C. Bovik, and B. L. Evans, "Efficient implementation of foveation filtering," in *Proc. Texas Instruments DSP Educator's Conf.*, Houston, TX, Aug. 1999.

[18] S. Lee and A. C. Bovik, "Motion estimation and compensation for foveated video," in *Proc. IEEE ICIP*, Kobe, Japan, Oct 1999.

[19] S. Lee, A. C. Bovik, and Y. Y. Kim, "Low delay foveated visual communications over wireless channels," in *Proc. IEEE ICIP*, Kobe, Japan, Oct. 1999.

[20] S. Lee, C. Podilchuk, V. Krishnan, and A. C. Bovik, "Unequal error protection for foveation-based error resilience over mobile networks," in *Proc. IEEE ICIP*, Vancouver, BC, Canada, Sept. 2000.

[21] S. Lee and A. C. Bovik. (1999) Foveated video demonstration. [Online]. Available: http://pineapple.ece.utexas.edu /class/Video/demo.html

[22] ——, "Foveated video image analysis and compression gain measurements," in *IEEE Southwest Symp. Image Analysis and Interpretation*, Austin, TX, Apr. 2000, pp. 63–67.

[23] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated image/video quality assessment in curvilinear coordinates," in *Proc. VLBV*, Oct. 1998, pp. 189–192.

[24] IL——, "Rate control for foveated MPEG/H.263 video," in *Proc. IEEE ICIP*, vol. 2, Chicago, Oct. 1998, pp. 365–369.

[25] A. J. Jerri, "The shannon sampling theorem-its various extentions and applications: A tutorial review," *Proc. IEEE*, vol. 65, pp. 1565–1596, Nov. 1977.

[26] J. J. Clark, M. R. Palmer, and P. D. Lawrence, "A transformation method for the reconstruction of functions from nonuniformly spaced samples," *IEEE Trans. Acoust., Speech, Signal, Processing*, vol. ASSP-33, pp. 1151–1165, Oct. 1985.

[27] Y. Zeevi and E. Shlomot, "Nonuniform sampling and antialiasing in image representation," *IEEE Trans. Signal Processing*, vol. 41, pp. 1223–1236, Mar. 1993.

[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Philadelphia, PA: Willey Interscience, 1991.

[29] *Description of error resilient core experiments*, ISO/IEC JTC1/SC29/WG11 N1646 MPEG97, 1997.

[30] H. Ling. (1997) Wireless channel modeling. [Online]. Available: http://ling0.ece.utexas.edu/comm/ comms.html

Sanghoon Lee was born in Korea in 1966. He received the B.S. degree in electrical engineering from Yonsei University, Korea, in 1989, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST) in 1991, and the Ph.D. degree in electrical engineering from the University of Texas, Austin, in 2000.

From 1991 to 1996, he was with Korea Telecom, where he was involved in the software development of the MPEG standard, channel coding, network protocols, and VLSI implementations for MPEG2 chip sets. In the summer of 1999, he was with Bell Laboratories, Lucent Technologies, working on wireless multimedia communications. Since February 2000, he has been working at Bell Labs., Lucent Technologies, Murray Hill, NJ, developing real-time embedded software and communication protocols for 3G W-CDMA networks. He is interested in mobile internet, real-time embedded software, and wireless multimedia communications.

Marios S. Pattichis received the B.S. degree in mathematics and computer sciences in 1991, and the M.S. and Ph.D. degrees in electrical and computer engineering in 1993 and 1998, respectively, all from the University of Texas, Austin.

His research areas are focused in the general area of digital image and video processing and communication. After his graduation, he was a Post-doctoral Fellow at the University of Texas (summer 1998) and a Visiting Assistant Professor at Washington State University, Pullman, (September 1998–August 1999). He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, where he is the Director of the Image and Video Processing and Communication Laboratory (ivPCL).

Alan Conrad Bovik (S'80–M'81–SM'89–F'96) received the B.S. degree in computer engineering in 1980, and the M.S. and Ph.D. degrees in electrical and computer engineering in 1982 and 1984, respectively, all from the University of Illinois, Urbana-Champaign.

He is currently the Robert Parker Centennial Endowed Professor in the Department of Electrical and Computer Engineering at the University of Texas at Austin, where he is the Associate Director of the Center for Vision and Image Sciences. During the Spring of 1992, he held a visiting position in the Division of Applied Sciences, Harvard University, Cambridge, MA. His current research interests include digital video, image processing, computer vision, wavelets, three-dimensional microscopy, and computational aspects of biological visual perception. He has published over 300 technical articles in these areas and holds two U.S. patents. He is also the editor/author of the *Handbook of Image and Video Processing*, (New York: Academic, 2000).

Dr. Bovik was named Distinguished Lecturer of the IEEE Signal Processing Society in 2000, received the IEEE Signal Processing Society Meritorious Service Award in 1998, the IEEE Third Millennium Medal in 2000, the University of Texas Engineering Foundation Halliburton Award and is a two-time Honorable Mention winner of the international Pattern Recognition Society Award for Outstanding Contribution (1988 and 1993). He is a Fellow of the IEEE and has been involved in numerous professional society activities. He is Editor-in-Chief of IEEE TRANSACTIONS ON IMAGE PROCESSING, and is on the editorial board of PROCEEDINGS OF THE IEEE. He also serves on the editorial boards of several other technical journals. He was the Founding General Chairman of the *First IEEE International Conference on Image Processing*, held in Austin, TX in November 1994. He is a registered Professional Engineer in the State of Texas and is a frequent consultant to industry and academic institutions.