RATE CONTROL FOR FOVEATED MPEG/H.263 VIDEO

Sanghoon Lee, Marios S. Pattichis and Alan C. Bovik

Laboratory for Image and Video Engineering
Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX 78712-1084

ABSTRACT

Given a set of target bits, video rate control algorithms that use Lagrange multipliers have been generally known as an optimal solution for maximizing the picture quality in uniform spatial domain. Even if the SNR(signal-to-noise) of a picture is maximized by the rate control scheme, the visual quality can be enhanced using a suitable algorithm for the human visual system. In this paper, we establish a new optimal rate control algorithm for maximizing the SNRC (signal-to-noise ration in curvilinear coordinates) using Lagrange multiplier. In addition, a target bit allocation technique for foveated video is introduced for simplified rate control over MEPG/H.263 video standards.

1. INTRODUCTION

When analyzing the anatomy of the human eye, the photoreceptor density monotonically decreases with increased distance from the fovea. If we maintain the high frequency components of an object corresponding to the foveation point, while removing the perceptually insignificant information, we can achieve a high compression ratio. An image in which the undetectable frequencies are removed by foveation filtering is called a "foveated image" which has a position-varying local bandwidth. Figure 1 shows an example of a foveated image where the fixation point is marked "x". Foveated video has great potential for very low bit rate coding applications such as wireless video phones and video conferencing systems[1].

Rate control is one of the most essential problems for maintaining high picture quality for a given target bit budget. Most optimal rate control algorithms attempt to maximize the SNR under the rate constraint using a Lagrange multiplier method. However, in terms of subjective quality assessment, the current rate control methods do not provide for the best resource allocation. In [2], SNRC is defined as an objective quality criterion suited for the human visual system. Using the SNRC, we can measure the subjective quality as an objective quantity, which provides for a more precise way to evaluate the foveated image/video quality. Thus, for obtaining the optimal visual quality, it is necessary to maximize the SNRC instead of the SNR for a given target bit rate.

The rate control method in MPEG-2 TM5 employed a virtual buffer and adaptive quantization[3]. In H.263 TMN5, the discrepancy between the target bits and the generated bits was measured to decide a quantization parameter for each macroblock[4]. In foveated video coding, the local bandwidth of each macroblock depends on the location of the foveation point. Therefore, in order to improve picture quality, the alloted bits to each macroblock must be decided according to the local bandwidth.

In this paper, we develop a new optimal rate control algorithm for maximizing the SNRC using a Lagrange multiplier technique in a curvilinear coordinate system, and introduce a bit allocation algorithm for the foveated image according to the local bandwidth.

2. TARGET BIT ALLOCATION IN CURVILINEAR COORDINATES

Given a curvilinear coordinate system, the locally band-limited image is resampled into a new image, and the resampled image is globally band-limited, for a particular fixed bandwidth value. Suppose that the number of generated bits in some infinitesimal region over the curvilinear coordinate system, is proportional to the corresponding infinitesimal area in the rectangular cartesian coordinate system (uniform domain). Then, the number of target bits for the foveated image can be equally allocated into each unit region in the obtained uniform domain in proportion to the mapping ratio, i.e., the target bits are non-uniformly allocated according to the mapping ratio for the foveated image. Let us be more precise.

Consider the coordinate mapping from (x_1, x_2) to

This work was supported by a grant from Motorola Inc.

0-8186-8821-1/98 $10.00 © 1998 IEEE

(Φ_1, Φ_2) given by $\mathbf{\Phi}(\mathbf{x}) = [\Phi_1(x_1, x_2)$, where $\Phi_2(x_1, x_2)]$ defines a one-to-one correspondence between $\mathbf{x}$ and $\mathbf{\Phi}(\mathbf{x})$ under the conditions: Φ_1 and Φ_2 are continuous, and have a single-valued inverse. Then, $\mathbf{\Phi}(\mathbf{x})$ is called the ”curvilinear coordinate”. Using the curvilinear coordinate transform, a one-to-one mapping is established between the $\mathbf{x}$-axis in cartesian coordinates and the $\mathbf{\Phi}(\mathbf{x})$-axis in curvilinear coordinates.

For a given 2-D original image $o(\mathbf{x})$ for $\mathbf{x} \in R^2$, let $O(\mathbf{\Omega})$ for $\mathbf{\Omega} \in R^2$ be the Fourier transform of $o(\mathbf{x})$. When $O(\mathbf{\Omega}) = 0$ for $\mathbf{\Omega} \geq \mathbf{\Omega}_o$ and $o(\mathbf{x}) \in L^2(R^2)$, the image $o(\mathbf{x})$ is defined as $\mathbf{\Omega}_o$-band-limited signal and denoted as $o(\mathbf{x}) \in B^{\mathbf{\Omega}_o}$ which is the space of 2-D finite energy signals. Suppose that a foveated image $f(\mathbf{x})$ with the local bandwidth $\mathbf{\Omega}_f(\mathbf{x}) \leq \mathbf{\Omega}_o$ is derived from the image $o(\mathbf{x})$. In such case, $B^{\mathbf{\Omega}_f(\mathbf{x})}$ becomes the space of locally band-limited signals and is denoted as $f(\mathbf{x}) \in B^{\mathbf{\Omega}_f(\mathbf{x})}$. Given $\mathbf{\Phi}(\mathbf{x})$, $f(\mathbf{x}) \in B^{\mathbf{\Omega}_f(\mathbf{x})}$ is mapped into $z(\mathbf{\Phi}(\mathbf{x})) \in B^{\mathbf{\Omega}_c}$ such as $f(\mathbf{x}) = z(\mathbf{\Phi}(\mathbf{x}))$.

Let S_o be the area of the image $f(\mathbf{x})$ displayed on the monitor over the spatial $\mathbf{x}$ domain. Then, the corresponding area S_c of the image $z(\mathbf{\Phi}(\mathbf{x}))$ over the $\mathbf{\Phi}(\mathbf{x})$ domain is

$$S_c = \int_{S_o} J_{\mathbf{\Phi}}(\mathbf{x}) d\mathbf{x}. \tag{1}$$

Assume the discrete function $f(\mathbf{x}_n)$ is obtained by sampling $f(\mathbf{x})$ with the sampling frequency $2\mathbf{\Omega}_o$. Since $f(\mathbf{x}) \in B^{\mathbf{\Omega}_f(\mathbf{x})}$ and $\mathbf{\Omega}_f(\mathbf{x}) \leq \mathbf{\Omega}_o$, $f(\mathbf{x})$ is always reconstructed by $f(\mathbf{x}_n)$ from the sampling theorem.

For the n^{th} point $\mathbf{x}_n = (x_{n_1}, x_{n_2})^t$ of the image $f(\mathbf{x})$, the unit area is $s_n^o = [x_{n_1} \pm \frac{\epsilon}{2}] \times [x_{n_2} \pm \frac{\epsilon}{2}]$ with respect to the point. Then, the total area of the image is the sum of each unit area : $S_o = \sum_n s_n^o$. Since the value ϵ is constant, s_n^o is independent on n and $S_o = N s_n^o$. Then, S_c in (1) can be represented by

$$S_c = \sum_n s_n^c \tag{2}$$

where

$$\begin{aligned} s_n^c &= \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_n) \\ &= \int_{\mathbf{x} \in s_n^o} J_{\mathbf{\Phi}}(\mathbf{x}) d\mathbf{x}. \end{aligned} \tag{3}$$

Since the n^{th} pixel $\mathbf{x}_n$ is also corresponding to the p^{th} pixel of the m^{th} macroblock, it can be denoted as $\mathbf{x}_n = \mathbf{x}_{p,m}$. The S_c in (2) becomes

$$S_c = \sum_m \sum_p \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_{p,m}). \tag{4}$$

The corresponding area of the m^{th} macroblock in curvilinear coordinates is

$$S_m^c = \sum_p \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_{p,m}). \tag{5}$$

If we allocate the target bits R_T into each macroblock according to the value S_m^c, the alloted bits for the m^{th} macroblock is

$$\hat{r}_m = R_T \times \frac{S_m^c}{S_c}. \tag{6}$$

In the foveated image, the local frequency is continuously changed according to the position. Let f_n be this local frequency at the n^{th} point. Then, a sampling matrix $\mathbf{V}_n$ can be found, which avoids aliasing for the band-limited image to $2\pi f_n$. Assume that the value $\bar{J}_{\mathbf{\Phi}}(\mathbf{x}_n)$ is in proportion to the sampling density, then

$$\begin{aligned} \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_n) &= \frac{c_1}{|\det \mathbf{V}_n|} \\ &= c_2 f_n^2 \end{aligned} \tag{7}$$

where c_1 and c_2 are constants. In such case, the allocated bits $\hat{r}_m$ is denoted :

$$\hat{r}_m = R_T \times \frac{\sum_{p=1}^{P} f_{p,m}^2}{\sum_{m=1}^{M} \sum_{p=1}^{P} f_{p,m}^2} \tag{8}$$

The allocated rate $\hat{r}_m$ is just obtained by considering the area ratio on 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 the rate control implementation, it is necessary to estimate the effective area $\tilde{s}_n^c(f_n)$ according to the local bandwidth. Then,

$$\hat{r}_m = R_T \times \frac{\tilde{s}_{p,m}^c(f_{p,m})}{\sum_{m=1}^{M} \sum_{p=1}^{P} \tilde{s}_{p,m}^c(f_{p,m})} \tag{9}$$

In MPEG TM5 [3], the target bits R_T(bits/frame) are temporally allocated into each picture within a GOP (group of picture) based on the channel rate R_c(bits/sec.), the GOP structure, and the PCT(picture coding type). For better performance, the target bits are spatially allocated into each macroblock using a virtual buffer and adaptive quantization. The encoding rate for each PCT is determined from its corresponding virtual buffer fullness d. The buffer fullness d is initialized to a constant d^o for every sequence layer. The initial value d^o is also updated by the buffer fullness of the previous GOP at every GOP layer. The virtual buffer size is decided according to the channel rate, which is one of the most important parameters in rate control. As the size decreases, the variation of QPs(quantization parameters) is more sensitive to the magnitudes d^c so that the magnitude of consecutive QPs fluctuates highly. In the opposite case, there is less variation so that the buffer overflow or underflow may continue for a long time. Generally, as the QP value is more frequently updated, the number of generated bits is precisely adjusted to the number of target bits.

In order to decide a QP for macroblock, the rate controller monitors the capacity of the virtual buffer. The k^{th} virtual buffer fullness d_k is calculated by

$$d_k = d' + \sum_{j=1}^{k} r_j - \frac{R_T}{M}k \tag{10}$$

where d' is the virtual buffer fullness of the previous picture and r_j is the number of generated bits in the j^{th} macroblock.

Similar to TM5, TMN5 in the H.263 video coding also uses the buffer fullness for determining QP. The discrepancy b_k between the generated bits and the target bits in the k^{th} macroblock is

$$b_k = \sum_{j=1}^{k} r_j - \frac{R_T}{M}k \tag{11}$$

which is a reference value for the decision of the k^{th} QP. Let G denote the frame rate of the original video sequence coming from the camera. In the H.263 video coding, the frame rate can be changed into a target frame rate F which is the actual encoded frame number per second. Suppose that R_T is the same for every consecutive frame after the first I frame. Then, for a given R_c(bits/sec), R_T is equal to R_c/F.

For foveated video coding, the target bits are determined from the local bandwidth. If the number of target bits for the j^{th} macroblock is decided by the effective area as (9), then the virtual buffer fullness (10) and the amount of the discrepancy (11) are updated by:

$$d_k = d' + \sum_{j=1}^{k} r_j - \sum_{j=1}^{k} \hat{r}_j \tag{12}$$

and

$$b_k = \sum_{j=1}^{k} r_j - \sum_{j=1}^{k} \hat{r}_j, \tag{13}$$

respectively.

3. OPTIMAL RATE CONTROL IN CURVILINEAR COORDINATES

Denote $r_k(q_k)$, $d_k(q_k)$ and q_k be the rate, distortion and QP of the k^{th} macroblock. Let M be the number of macroblocks in a picture. The QPs for coding M macroblocks consist of a quantization state vector $\vec{Q} = (q_1, q_2, ..., q_M)$. Suppose that the number of target bits R_T are assigned into the picture, then the optimal rate control is finding the state vector $\vec{Q}$ which minimizes the overall distortion:

$$D(\vec{Q}) = \sum_{k=1}^{M} d_k(q_k) \tag{14}$$

subject to the rate constraint

$$R(\vec{Q}) = \sum_{k=1}^{M} r_k(q_k) \leq R_T. \tag{15}$$

By introducing a Lagrange multiplier $\lambda \geq 0$, the constrained problem can be solved.

By sweeping λ from 0 to ∞, an optimal quantization state vector $\vec{Q}^*$ which minimizes the Lagrangian cost function $J(\vec{Q}, \lambda) = D(\vec{Q}) + \lambda R(\vec{Q})$ is obtained while satisfying the constraint(15).

$$\begin{aligned} J(\vec{Q}^*, \lambda) &= min[D(\vec{Q}) + \lambda R(\vec{Q})] \\ &= \sum_{k=1}^{M} min[d_k(q_k) + \lambda r_k(q_k)] \\ &= \sum_{k=1}^{M} j_k(q_k^*) \end{aligned} \tag{16}$$

where $j_k(q_k)$ is the Lagrangian cost function for the k^{th} 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 allowed QPs. In the 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 k^{th} macroblock $d_k(q_k)$ is obtained by the mean square error between the original image $o(\mathbf{x})$ and the reconstructed image $r(\mathbf{x})$ after coding with q_k.

$$d_k(q_k) = \frac{1}{m_p} \sum_{p=1}^{m_p} [o(\mathbf{x}_{p,k}) - r(\mathbf{x}_{p,k})]^2 \tag{17}$$

where m_p=384 is the number of pixels in a macroblock, $o(\mathbf{x}_{p,k})$ is the p^{th} pixel in the k^{th} original macroblock, and $r(\mathbf{x}_{p,k})$ is the p^{th} pixel in the k^{th} reconstructed macroblock.

Suppose that $f(\mathbf{x}_n)$ and $g(\mathbf{x}_n)$ are the foveated images of $o(\mathbf{x}_n)$ and $r(\mathbf{x}_n)$ respectively. In curvilinear coordinates, the normalized distortion $d_k(q_k)$ is obtained by

$$d_k(q_k) = \frac{1}{m_p}\sum_{p=1}^{m_p}\left[f(\mathbf{x}_{p,k}) - g(\mathbf{x}_{p,k})\right]^2 \bar{J}_{\mathbf{\Phi}}(\mathbf{x}_{p,k}). \quad (18)$$

Under the assumption (7), the equation (18) becomes

$$d_k(q_k) = \frac{c_2}{m_p}\sum_{p=1}^{m_p}\left[f(\mathbf{x}_{p,k}) - g(\mathbf{x}_{p,k})\right]^2 f_{p,k}^2 \quad (19)$$

4. SIMULATION RESULTS

For the foveated video quality assessment, we employ the PSNRC (peak signal-to-noise ratio in curvilinear coordinates) defined in [2].

$$\text{PSNRC} = 10 * log_{10} \frac{max[a(\mathbf{x}_n)]^2}{\frac{1}{S_d}\sum_{n=1}^{N}\left[a(\mathbf{x}_n) - b(\mathbf{x}_n)\right]^2 f_{p_n}^2} \quad (20)$$

where

$$S_d = \sum_{n=1}^{N} f_{p_n}^2. \quad (21)$$

where $a(\mathbf{x}_n)$ is the original image, $b(\mathbf{x}_n)$ is the coded image of $a(\mathbf{x}_n)$ or $a(\mathbf{x}_n)$ is the foveated image, $b(\mathbf{x}_n)$ is the coded image of the foveated image $a(\mathbf{x}_n)$.

In real video coding, the number of generated bits non-linearly depends on the local bandwidth so that the bit allocation (8) must be modified according to the coding algorithm. In this simulation, we allocate the number of bits for each macroblock dependent on the mean of local bandwidth. For the performance comparison, two methods are employed.

TM5 BA : TM5 bit allocation.
FV BA : FV(foveated) bit allocation where $\hat{r}_m$ is decided by the equation (9).

Figure 2 shows the performance comparison for the above two methods where we use the foveated images "Flower Garden" with 704×480 frame size and 4:2:0 chrominance format. The effective area $\tilde{s}_n^c(f_n)$ is set by f_n instead of f_n^2. The PSNRC in the rate control using FV BA is always larger than the PSNRC in the TM5 rate control for I and P pictures. The FV BA method is better for the subjective quality in terms of the human visibility with a small objective quality difference.

To demonstrate foveated video compared to regular video, optimal rate control using a Lagrange multiplier λ is accomplished for minimizing MSE(mean square error) or MSEC(mean square error in curvilinear coordinates) over H.263 video.

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

Figure 3-6 show the coded I pictures for a given compression ratio(bpp = 0.3748) in the above four methods. In Method 4, the PSNRC is the largest and the subjective quality is better compared to the other methods.

5. CONCLUSIONS

Most traditional rate control algorithms are focused on maximizing the SNR of reconstructed picture using Lagrange multiplier λ. In this paper, we established the optimal rate control algorithm for maximizing the visual quality while maximizing the SNRC using Lagrange multiplier λ in curvilinear coordinates. Moreover, we introduced the nonuniform target bit allocation for MPEG TM5/H.263 TMN5, which gives higher visual quality compared to uniform bit allocation.

6. REFERENCES

[1] W. S. Geisler and J. S. Perry. A real-time foveated multiresolution system for low-bandwidth video communincation. In *SPIE Proceedings*, volume 3299, 1998.

[2] S. Lee, M. S. Pattichis, and A. C. Bovik. Foveated image/video quality assessment in curvilinear coordinates. In *Int'l. Workshop on Very Low Bitrate Video Coding*, Urbana, IL, October 8 - 9 1998.

[3] Mpeg-2 video test model 5. Technical report, ISO/IEC JTC1/SC29/WG11 and ITU-TS SG15 EG for ATM Video Coding, April 1993.

[4] Video codec test model tmn5. Technical report, ITU-T/SC15, Jan. 1995.

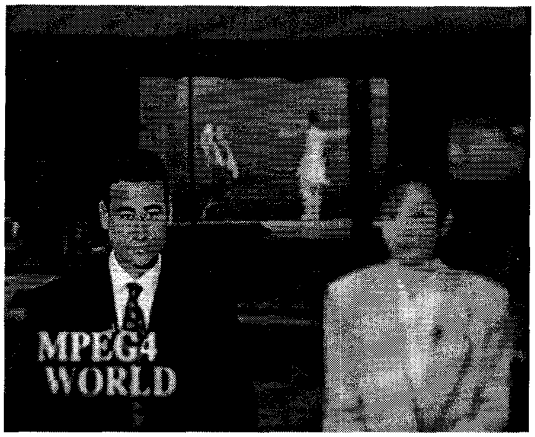

Figure 1: Foveated image "News"

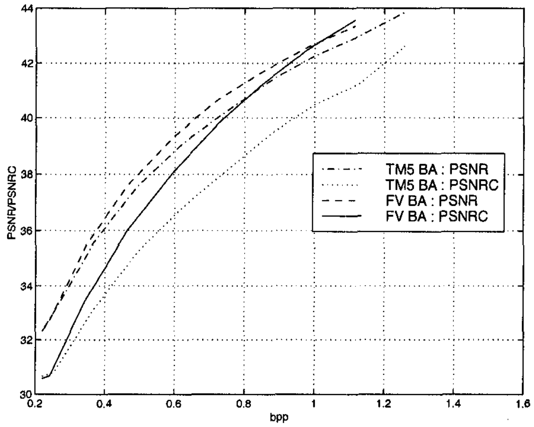

Figure 2: Performance comparison in I-pictures : TM5 v.s. FV bit allocation on MPEG-2 video

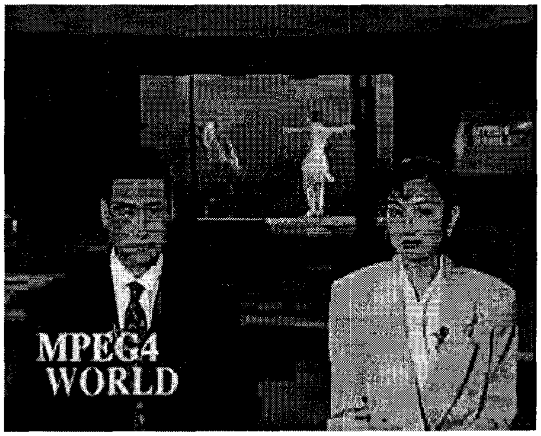

Figure 3: Method 1 : PSNR = 30.22, PSNRC = 29.50

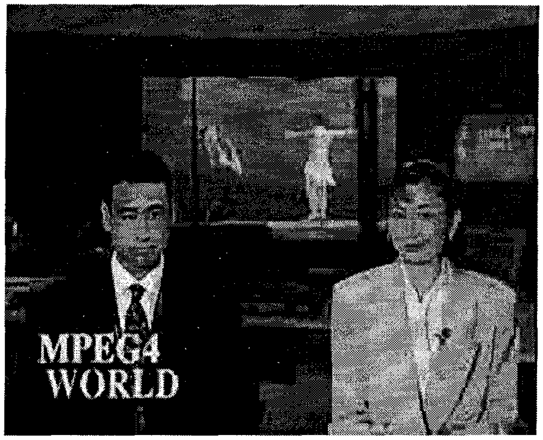

Figure 4: Method 2 : PSNR = 30.03, PSNRC = 30.55

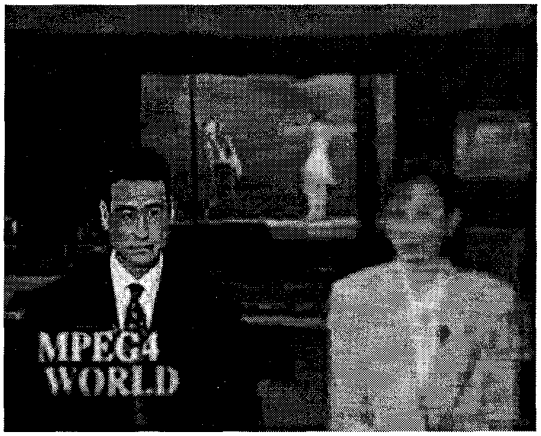

Figure 5: Method 3 : PSNRC = 33.83

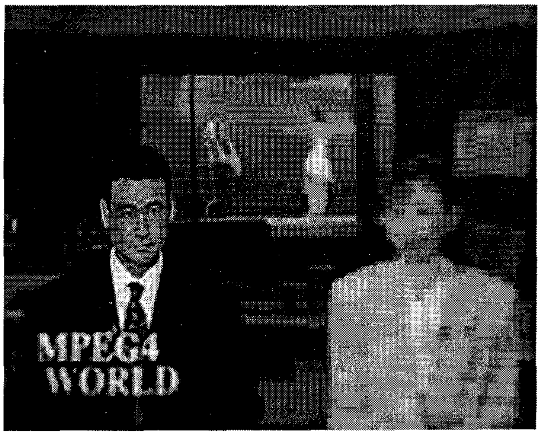

Figure 6: Method 4 : PSNRC = 35.65