VERY LOW BIT RATE FOVEATED VIDEO CODING FOR H.263

Sanghoon Lee 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 USA
{slee, bovik}@ece.utexas.edu

ABSTRACT

Recently, foveated video has been introduced as an important emerging method for very low bit rate multimedia applications [1][2]. In this paper, we develop several rate control algorithms, and measure the performance of foveated video. We utilize H.263 video, and compare the performance with regular video based on the SNRC (signal-to-noise ratio in curvilinear coordinates). In order to maximize compression, we use a maximum quantization parameter (QP = 31) for the regular video, and code a foveated video sequence at the equivalent bit rate. In simulation, we improve the PSNRC to 3.64 (1.62)dB under 30 (14) Kbits/sec for P pictures in CIF "News" ("Akiyo") standard video sequence.

1. INTRODUCTION

In MPEG/H.263 video standards, we can maximize compression ratio when using a maximum QP (quantization parameter) for an input video sequence. If we remove unessential spatial high frequency components from the video sequence, the spatial redundancy decreases. In addition, motion compensation errors are also reduced. Because of such spatial/temporal redundancy reduction, the coding efficiency will be improved, and the compressed bit rate reduced.

Then, consider how to remove such high frequency components while maintaining high visual qualities. In the human visual system, the spatial resolution depends on the distribution of photoreceptors on the retina[3]. The region where photoreceptors are densest is called as the "fovea", and located along the visual axis. A point at which a human observer is focusing in an image along the visual axis is the foveation point. When undetectable visual frequencies from the image (with reference to the foveation point) are removed, the image becomes the "foveated image". For interactive multimedia applications, this foveation point can be dynamically selected using a mouse, touch screen, or an eye tracker[1][2].

Using a rate control algorithm, we can non-uniformly allocate a bit budget into each macroblock. In [1][4], the SNRC is defined as an objective quality criterion suited for the human visual system, and an optimal rate control algorithm is developed by using a Lagrange multiplier in curvilinear coordinates.

This work was supported by a grant from Motorola Inc.

In this paper, we develop a sub-optimal rate control which can adopt to the normal/modified quantization mode in H.263. For real-time implementation, a piecewise *R-D* (rate-distortion)/*R-Q* (rate-quantization) model is described. Based on these models, a fast algorithm for searching an optimal Lagrange multiplier λ^* is presented. We demonstrate the out-performance of foveated video relative to normal video.

2. *R-D* MODEL FOR FOVEATED VIDEO

2.1. *R-D model in curvilinear coordinates*

Assuming a mapping ratio to curvilinear coordinates at each point is proportional to the sampling density, then the MSEC(mean square error in curvilinear coordinates) for discrete images is

$$\text{MSEC} = \frac{1}{\sum_{n=1}^{N} f_n^2} \sum_{n=1}^{N} [a(\mathbf{x}_n) - b(\mathbf{x}_n)]^2 f_n^2, \tag{1}$$

and the PSNRC is

$$\text{PSNRC} = 10 * log_{10} \frac{max[a(\mathbf{x}_n)]^2}{\text{MSEC}} \tag{2}$$

where f_n is the local frequency at the n^{th} point, $a(\mathbf{x}_n)$ is the original image or the foveated image, and $b(\mathbf{x}_n)$ is the coded image of $a(\mathbf{x}_n)$ [1][5].

Let the subscripts p and k denote the p^{th} pixel of the k^{th} macroblock. In curvilinear coordinates, the normalized distortion $d_k(q_k)$ for the k^{th} macroblock is obtained by

$$d_k(q_k) = \frac{1}{m_p} \sum_{p=1}^{m_p} [a(\mathbf{x}_{p,k}) - b(\mathbf{x}_{p,k})]^2 f_{p,k}^2 \tag{3}$$

where m_p is the number of pixels in a macroblock.

In MPEG/H.263 video, an exponential expression is widely used for *R-D* model :

$$r_k(d_k) = \alpha_k \log_2 \frac{\sigma_k^2}{d_k} + \beta_k \tag{4}$$

where σ_k^2 is the variance, r_k is the rate, and α_k and β_k are variables for the k^{th} macroblock. If we apply this exponential form to foveated video, the normalized variance σ_k^2 in curvilinear coordinates is

$$\sigma_k^2 = \frac{1}{m_p} \sum_{b=1}^{m_b} \sum_{p=1}^{b_p} (a_b(p) - dc_b)^2 f_{p,k}^2, \quad 1 \le b \le 6, \tag{5}$$

while

$$dc_b = \frac{1}{\sum_{p=1}^{b_p} f_{p,k}^2} \sum_{p=1}^{b_p} a_b(p) f_{p,k}^2 \tag{6}$$

where m_b=6 is the number of blocks in a macroblock which consists of four luminance blocks and two color blocks, b_p=64 is the number of pixels in a block, and $a_b(p)$ represents the p^{th} pixel value of the b^{th} block. The pixel $a_b(p)$ is an original pixel value in I pictures and is a differential value between an original pixel and a motion compensated pixel after temporal prediction in P or B pictures.

2.2. *Hierarchical piece-wise R-D model*

Due to non-linear coding effects in video coding, the global *R-D* curve for each macroblock is heavily damped. For more accurate modeling, the global *R-D* function must be obtained by several piece-wise local *R-D* functions which are individually modeled by (4).

For given two RQPs(reference quantization parameters) 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 calculated by

$$\alpha_k^n = \frac{r_k(q_{r_n}) - r_k(q_{r_{n+1}})}{\log_2(\frac{d_k(q_{r_{n+1}})}{d_k(q_{r_n})})} \tag{7}$$

$$\beta_k^n = r_k(q_{r_n}) - \alpha_k^n \log_2(\frac{\sigma_k^2}{d_k(q_{r_n})}) \tag{8}$$

In this paper, a HPW (hierarchical piece-wise) *R-D* algorithm is introduced for real-time optimal rate control. The HPW *R-D* curve is constructed based on a level which is a reference value needed to decide a RQP at each coding time. Thus, each $q \in Q$ has the specified level. Let Q_{r^v} be the quantization set for the v^{th} level and $q_{r_n^v} \in Q_{r^v}$ be the n^{th} QP in the level. If Q consists of L levels, then

$$Q = Q_{r^1} \cup Q_{r^2} \cup Q_{r^3} ... \cup Q_{r^L}. \tag{9}$$

We construct local *R-D* curves while searching an optimal Lagrange multiplier λ^* based on a top-down method from the level 1 to the level L.

3. FOVEATED VIDEO CODING IN H.263

3.1. *Convergence into an optimal Lagrange multiplier*

An optimal Lagrange multiplier λ^* is found by an iterative method while sweeping λ on *R-D* curves. From the slope of *R-D* curve, the corresponding λ_k is obtained by

$$\lambda_k = -\frac{\partial d_k}{\partial r_k}. \tag{10}$$

For the i^{th} iteration, denote $\vec{Q}^i$ and $\vec{\lambda}^i$ as the state vectors whose components are q_k^i and λ_k^i for $1 \le k \le M$ respectively. Then,

$$R(\vec{\lambda}^i) = \sum_k r_k(\lambda_k^i), \tag{11}$$

and the range of λ^* can be found by the following Lemma 1.

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 $\lambda_m^i = min[\lambda_k^i]$, $1 \le k \le M$. Suppose that $R(\vec{\lambda}^i) = R_T$, then there exists λ^* in $\lambda_m^i \le \lambda^* \le \lambda_M^i$.

Since the value of each $q \in Q$ is discrete, each λ_k^i can be approximately obtained by $\lambda_k^i(q_k^i) = d_k^i(q_k^i)/r_k^i(q_k^i)$. From the monotonic property of *R-D* function in Lemma 1, λ is converged into an optimal λ^* within a limited encoding time instead of sweeping λ from 0 to ∞. In order to reduce the encoding time while maintaining a rate constraint, an iterative convergence algorithm is developed.

3.2. *Iterative method for searching an optimal QP vector*

The value λ^* is not known prior to coding, and dependent on a desired target budget. The whole procedure for finding λ^* is described by :

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

Step 2: Search two reference quantization parameters $q_{r_n^1}$ 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 (7) and (8).

Step 4: Repeat *Step* 1 - *Step* 3 to a specified level and construct *R-D* curves using the HPW algorithm.

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

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 $\hat{q}_k^i$ = argmin[q] for minimizing $|\lambda(\hat{q}_k^i) - \bar{\lambda}^i|$. This iterative procedure 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 k^{th} macroblock. The optimal QP vector $\vec{Q}^*$ consists of q_k^* for $1 \le k \le M$.

In regular video coding at very low bit rate, λ^* tends to approach λ_M. Since λ_k is maximum when QP is 31, λ_k is thresholded to $\lambda_k(31)$ which can be less than λ^*. In such case, it is difficult to find a *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 bit rate is lower than that of normal video coding. Thus, λ^* exists near a median value between λ_M and λ_m so that the performance improvement should be larger than normal video coding.

3.3. *Sub-optimal rate control in H.263 video coding*

In foveated video, there are two major factors to reduce coding distortion compared to normal video. One is an undetectable frequency elimination, and the other is a distortion measurement in curvilinear coordinates. After removing high frequency components in background, we can save bits, and reallocate them into a focused region. For optimal rate control in curvilinear coordinates, the MSEC in (1) must be minimized. The normalized distortion in (3) is proportional to the *Jacobian* of a coordinate transformation. Since the magnitude of the *Jacobian* is equal to or less than 1, a distortion in curvilinear coordinates is always equal to, or less than the corresponding distortion in cartesian coordinates. Therefore, the distortion in curvilinear coordinates is reduced in proportional to the transform ratio.

Thus, the magnitude of distortion in curvilinear coordinates is much less than that in cartesian coordinates due to the above two major factors. Therefore, the distortion variation with respect to QP change is less sensitive in curvilinear coordinates. Specially, for background, we can choose a QP without much degrading the performance.

Here, we develop the following sub-optimal rate control. First, an optimal vector $\vec{Q}^*$ from the proposed iterative procedure where we do not count the generated bits due to QP change. Then, denote r_k^* as the rate of the k^{th} macroblock with respect to q_k^*.

Next, let U be a macroblock index set which consists of macroblock indexes whose average local bandwidth is less than a threshold value f^{th}. Then, determine a QP $\bar{q} = \text{argmin}[q]$ satisfying the following rate constraint for macroblocks in U :

$$\sum_{k\in U} r_k(q) \le \sum_{k\in U} r_k^*(q_k^*) \tag{12}$$

where r_k is generated bits in real video coding and $f^{th} = 0.25$ in the simulation.

Finally, we find QPs for remaining macroblocks whose set is denoted as V. For the k^{th} macroblock where $k \in V$, we consider the Lagrange cost function $j_k = d_k(q_k) + \lambda^* r_k(q_k)$, and the range of QP change according to the quantization mode. We choose q_k with satisfying both conditions : minimizing j_k and being included in the range. In the normal quantization mode, the dynamic range of current QP is limited by the QP value of previous macroblock. In the modified quantization mode, we can choose any QP value for current macroblock.

4. SIMULATION RESULTS

For the H.263 video coding, the reference frame rate 30, two skip frames, and the target frame rate 10 are used. For the performance comparison, the following coding methods are used.

normal video : In order to maximize the compression ratio, an original video sequence is coded by using QP = 31.

constant q : For a foveated image sequence, a constant QP is decided for each picture. The transmission rate is set to the equivalent rate generated in the *normal video* method.

normal Q mode : A foveated image sequence is coded by using the sub-optimal rate control algorithm for the normal quantization mode in curvilinear coordinates.

modified Q mode : A foveated image sequence is coded by using the sub-optimal rate control algorithm for the modified quantization mode in curvilinear coordinates.

optimal in CV : An optimal rate control for minimizing the MSEC for a foveated image sequence. Here, we do not count generated bits due to QP change. The reconstructed picture quality using this method becomes an upper bound on the coding performance.

When the QP is set to be 31 for 30 frames of the "News" ("Akiyo") CIF video sequence (with two skip frames), the number of generated bits is 35.1 (27.8) Kbits for the I picture and the coding rate is 29.3 (13.1) Kbits/sec for the following P pictures. The PSNRC of the normal video coding is 3 (1.42) dB less than that of the foveated video coding. Because of the flexibility for changing the QP value, the PSNRC in the rate control using *modified Q mode* is improved to 0.6 (0.2) dB compared to *normal Q mode*. The average PSNRC of *modified Q mode* is the closest to the upper bound of the PSNRC. In Figures 1 - 6, it is shown that the reconstructed picture qualities of the foveated video is much higher than those of the regular video at the very low bit rate.

5. CONCLUSIONS

The potential benefit that can accrue from using foveated video coding is to improve the visual quality for a given number of target bits. Using H.263 video standard, we measured the performance of the foveated coding according to rate control method. For the normal video, we set QP to 31 and coded the "News" ("Akiyo") CIF video sequence. For the foveated video, we controlled the rate which is equivalent to the generated bits of the normal coding. In the foveated video, we can improve the PSNRC to 3.64 (1.62) dB compared to the normal video and show a great potential for low bit rate video coding.

6. REFERENCES

[1] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment and compression gain," IEEE Trans. Image Processing, submitted.

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

[3] B. A. Wandell, Foundations of Vision. Sunderland, MA: Sinauer Associates, Inc, 1994.

[4] S. Lee, M. S. Pattichis, and A. C. Bovik, "Rate control for foveated MPEG/H.263 video," in Proc. IEEE Int'l. Conf. Image Proc., (Chicago), Oct. 1998.

[5] 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), Oct. 1998.

Figure 1: Regular I picture : PSNRC = 29.31, bpp = 0.346

Figure 2: Foveated I picture : PSNRC = 34.31, bpp = 0.346

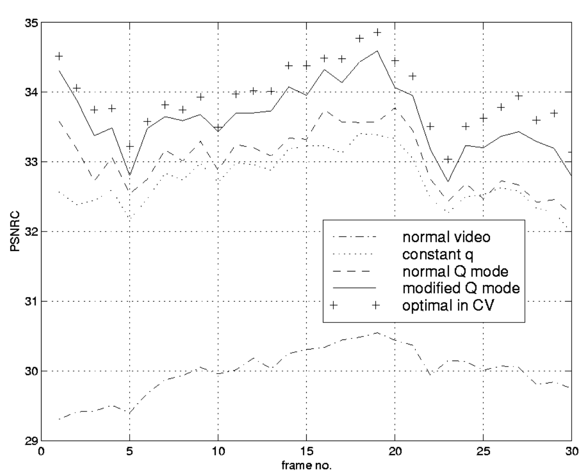

Figure 3: PSNRC for the CIF "News" sequence

Figure 4: Regular I picture : PSNRC = 29.98, bpp = 0.274

Figure 5: Foveated I picture : PSNRC = 33.08, bpp = 0.274

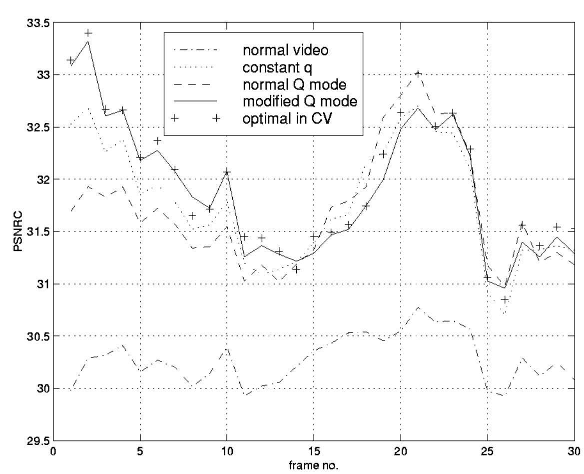

Figure 6: PSNRC for the CIF "Akiyo" sequence