Rate Control of H.263 for Low Bit Rate Visual Communication

Hwangjun Song and C.-C. Jay Kuo
Integrated Media Systems Center
Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, CA 90089-2564

Abstract

H.263 is an emerging video compression standard for low bit rate visual communication. A new rate control for H.263 is studied in this research. The requirements for rate control algorithm depend on the applications. For more flexible rate control, the frame rate is considered as a control variable. A frame rate control algorithm, which can minimize the temporal quality degradation caused by variable frame rate and estimate the supportable frame rate fast, is proposed. Also, an efficient global bit allocation algorithm of H.263, which can guarantee the reasonable sub-optimality and reduce the computational complexity, is proposed. Finally, experimental results are provided to demonstrate the better performance of proposed scheme.

1. INTRODUCTION

Visual communication has been growing very fast due to developments of wired and wireless communication techniques and video compression techniques. Although wireless networks such as PCS, CDPD, mobile IP and wireless ATM are growing tremendously fast, the available channel bandwidth is very limited, generally. Therefore, video compression technique, which is efficient at low bit rate, is inevitable. Up to now, International standards such as MPEG-1, MPEG-2, H.261 were proposed to accommodate different applications. MPEG-4, MPEG-7 and H.263 are under development. H.263 was proposed as a component of H.324, which was proposed for video-conferencing for PSTN and mobile radio. Hence, the target bit rate of H.263 is very low. The rate control algorithm plays a key role to satisfy the channel conditions or storage medium characteristics. Depending on channel bandwidth and medium characteristics, different rate control algorithms are required. It is one of the reasons that rate control algorithm can not be defined by video compression standards. Generally speaking, existing rate control algorithms are over-constraint for visual communication. The controllable variables are only quantization parameters. For high bandwidth and relatively reliable and fast storage medium, existing rate control algorithms work well. But, for narrow bandwidth and time-varying communication channels, the situation can be changed. In this paper, the proposed rate control treats frame rate as a control variable. It is obvious that the variable frame rate can degrade the temporal human perception quality. Therefore, a frame rate control algorithm that can minimize the temporal human perception quality is proposed. In addition, GOP of H.263 is too long to be used as a basic control unit because of higher computational complexity and time-delay. It is one of the reasons that there are little researches about H.263 global bit allocation. In this work, we propose efficient H.263 global bit allocation which can guarantee the reasonable sub-optimality in performance and reduce the computational complexity, significantly.

2. PROPOSED H.263 RATE CONTROL ALGORITHM

2.1. Frame Rate Control Algorithm

Two problems must be considered: how to change the frame rate while preserving motion smoothness and when the frame rate should be changed. We assume that the video camera generates 30 frames per second. For the first problem, we have to determine the encoded frame positions to minimize the degradation in motion smoothness. The sub-GOPs of our choice consist of 12 frames, which is the least common number of 2,3,4 and 6. The relation between the frame rate and encoded frame positions are shown in Tables 1 and 2.

1058-6393/98 $10.00 © 1998 IEEE

Frame rate	Encoded frame No.
1/1	1,2,3,4,5,6,7,8,9,10,11,12
1/2	2,4,6,8,10,12
1/3	3,6,9,12
1/4	4,8,12
1/6	6,12
1/12	6

Table 1. Even-encoded frame number.

Frame rate	Encoded frame No.
1/1	1,2,3,4,5,6,7,8,9,10,11,12
1/2	1,3,5,7,9,11
1/3	1,4,7,10
1/4	1,5,9
1/6	1,7

Table 2. Odd-encoded frame number.

Human eyes are sensitive to abrupt (temporal) interval change between adjacent frames, where unsmooth motion is perceived more obviously. By imposing the frame rate change pattern, we can minimize the interval change between adjacent frames. For example, if the 1/12 frame rate in Table 1 is chosen, then the 1/6 frame rate in Table 2 must be chosen for the next sub-GOP. As a result, the proposed algorithm can control the frame rate from about 3.7 fps to 30 fps.

To estimate the supportable frame rate, the rate-distortion (R-D) curve can be used. Many models for the R-D curve such as the MPEG test model, the statistical model, the exponential model, and the approximating spline model were proposed to speed up the rate control algorithm. The R-D curve can provide a good estimate of the frame rate if the model is accurate. However, a larger computational complexity is needed for more accurate R-D models. Furthermore, R-D models for every frame are required to predict the exact frame rate for the following sub-GOP. This is again computationally intensive. Here, we consider a simplified approaches based on the histogram of difference. It is able to detect the motion smoothness with a low computational cost.

Histogram-based methods [5] such as DOH (difference of histograms), BH (block histogram difference) and HOD (histogram of difference image) can be used to detect the motion in video. Since required bit rates for predictive frames depend on the motion (i.e. higher bit rates for faster motion), histogram based methods can be used to estimate the supportable frame rate. BH and DOH are better in detecting global changes rather than local motion while HOD is more sensitive to local motion. Besides histogram-based methods, other methods [5] such as BV (block variance difference) and MCE (motion compensation error) were also proposed. MCE is nearly ideal but with the drawback that it is too expensive computationally.

In this work, HOD is adopted. We define a distortion measure between two frames f_n and f_m as follows:

$$D_h(f_n, f_m) = \frac{\sum_{i>|TH_1|} hod(i)}{N_{pixel}}, \tag{1}$$

where *hod(i)* is the histogram of the difference image, TH_1 is threshold value for detecting the closeness of the position to zero, and N_{pixel} is the number of pixels. If HOD values in a sub-GOP are calculated, the estimated HOD value $\hat{D}_e$ for the following sub-GOP can be calculated by

$$\hat{D}_e = D_h + \omega_h \times a_h, \tag{2}$$

where D_h is the HOD value between the two last encoded frames in current sub-GOP, ω_h is a weighting factor and a_h is the slope of approximating line which minimizes the mean square error of HODs in current sub-GOP.

The proposed frame rate control algorithm can be summarized as:

- If $\hat{D}_e < TH_2$, increase the frame rate by one step.
- If $TH_2 < \hat{D}_e < TH_3$, keep the current frame rate.
- If $\hat{D}_e > TH_3$, decrease the frame rate by one step.

Values TH_2 and TH_3 in above are thresholds.

2.2. Global Bit Allocation Algorithm

Once the frame rate is determined, efficient global bit allocation is required for these chosen frames. The bit budget and image quality must be considered simultaneously. In this work, the dependent frame coding scheme without buffer constraint is used. We can formulate the following mathematical problem for the optimal global bit allocation of a GOP under the assumption that the structure of GOP is IPPP ... PP.

Problem Formulation: (Dependent frame coding without the buffer constraint)

Determine $\vec{q} = (q_1, q_2, \ldots, q_{N_P})$ so that the following constrained optimization problem is solved.

$$minimize \quad D(\vec{q}) + \omega_q \times E(\vec{q}), \tag{3}$$

$$subject\ to \quad \sum_{i=1}^{N_P} B_{i,used}(q_1, q_2, \cdots, q_i) \le B \tag{4}$$

$$D(\vec{q}) = \frac{1}{N_P} \cdot \sum_{i=1}^{N_P} d_i(q_1, q_2, \cdot\cdot, q_i) \qquad (5)$$

$$E(\vec{q}) = \frac{1}{N_P} \sum_{i=1}^{N_P} (d_i(q_1, q_2, \cdot\cdot, q_i) - d_{i-1}(q_1, q_2, \cdot\cdot, q_{i-1}))^2, \qquad (6)$$

where N_P is the encoded P-frame number of a GOP, B is total bit budget for a GOP except for the I-frame, $B_{i,used}(q_1, q_2, \ldots, q_i)$ is the used bit budget for the ith frame and ω_q is the weighting factor for abrupt quality change and flickering.

The bit allocation algorithm proposed by Ramchandran *et al.* [6] determined a set of optimal quantization parameters for all encoded frames by the Viterbi algorithm. It mainly serves as a benchmarking due to the tremendous computational complexity required and the resulting time-delay. It is impractical to get the optimal global bit allocation for a GOP by using the Lagrange multiplier method because of the huge length of GOP. Even though approximating R-D models can be used to reduce the computational complexity significantly, the optimal global bit allocation for a GOP is still not efficient enough.

In this work, we simplify the globally optimal problem to reduce the required computational complexity, and consider a reasonable sub-optimal solution. In our approach, we divide one GOP into several sub GOPs. Dependent models are used within all sub-GOPs while dependency among sub-GOPs is also considered. Then, we are lead to the following simplified problem.

Simplified Sub-optimal Problem:

Determine $\vec{q_m}$, $m = 1, 2, \ldots, M$ so that the following constrained optimization problem is solved.

$$minimize \quad \sum_{m=1}^{M} (D_m(\vec{q_m}) + \omega_q \times E_m(\vec{q_m})), \qquad (7)$$

$$subject \ \ to \quad \sum_{m=1}^{M} B_{m,used}(\vec{q_m}) \leq B_{subgop} \cdot M, \qquad (8)$$

$$\sum_{m=1}^{M} N_m = N_P, \qquad B_{subgop} = \frac{N_{sub}}{N_{gop}} \cdot B, \qquad (9)$$

$$D_m(\vec{q_m}) = \frac{1}{N_m} \cdot \sum_{i=1}^{N_m} d_i(q_1, \cdot\cdot, q_i), \qquad (10)$$

$$E_m(\vec{q_m}) = \frac{1}{N_m} \cdot \sum_{i=1}^{N_m} (d_i(q_1, \cdot\cdot, q_i) - d_{i-1}(q_1, \cdot\cdot, q_{i-1}))^2 \qquad (11)$$

where $\vec{q_m} = (q_{m,1}, q_{m,2}, \ldots, q_{m,N_m})$, N_m is the encoded frame number of the m_{th} sub-GOP, $B_{m,used}(\vec{q_m})$ is the used bit budget for the m_{th} sub-GOP, M is the number of sub-GOPs in a GOP, N_{sub} is the total frame number of a sub-GOP and N_{gop} is the total frame number of a GOP.

Note that in above $E_m(\vec{q_m})$ is inserted to reduce the flickering effect and to avoid the abrupt quality change. By adjusting the weighting factor ω_q, the abrupt quality change and the flickering effects are controllable. Now, the Lagrange multiplier method can be used to solve the optimization problem with constraints. The dependency of sub-GOPs is taken into account based on the following facts.

1. The penalty function for the mth sub-GOP, which is derived from the bit budget constraint in Equation (8), can be written as

   $$P_m(\vec{q_m}) = (\sum_{j=1}^{m} B_{j,used}(\vec{q_j}) - B_{subgop} \cdot m)^2. \qquad (12)$$

   The sum of bit rates used by 1st, 2nd, ... and m*th* sub-GOPs is used as the penalty function so that a GOP satisfies the constraint of the total bit budget will be more smooth.

2. The dependency among sub-GOPs is considered by controlling the weighting factors $\{\lambda_m, m = 1, 2, \ldots, M\}$ of the Lagrange multiplier method in Equation (13).

3. d_0 in Equation (11) is the error of the last encoded frame in the previous sub-GOP.

By combining the Equations (7) and (12), we can define a new cost function for the mth sub-GOP as:

$$\Phi_m(\vec{q_m}, \lambda_m) = J_m(\vec{q_m}) + \lambda_m \times P_m(\vec{q_m}), \qquad (13)$$

$$\text{where } J_m(\vec{q_m}) = D_m(\vec{q_m}) + \omega_q \times E_m(\vec{q_m}).$$

The conventional Lagrange multiplier method uses a constant for the weighting factor of the constraint. That is,

$$\lambda_i = \lambda_j, \quad for \quad i \neq j.$$

It can cause the quality degradation. To compensate the dependency among sub-GOPs properly, the weighting factors for sub-GOPs are set to satisfy the following equation:

$$\lambda_i \leq \lambda_j, \quad if \ \ i \leq j. \qquad (14)$$

It is similar to the monotonicity property proposed by Ramchandran [6]. In this paper, the gradient search techniques are used to find the optimal solution of

Equation 13. Since the $\Phi_m(\vec{q_m}, \lambda_m)$ is a convex function [7] generally, the globally optimal solution in each sub-GOP can be found by gradient search techniques. By inserting a factor for buffer constraint into the cost function, this algorithm can be extended to rate control problem with the buffer constraint [7, 6].

3. Experimental Results

The Salesman image sequences of the CIF format (352×288) is used as the test video in our experiment. Let the image sequence be generated at a rate of 30 fps, and the given channel bandwidth is 32 kbps. The GOP consists of 97 frames, including an I-frame and 96 P-frames. Let the I-frame be encoded with 35.30 kbits with 28.577 dB. The remaining 96 P-frames are divided into 8 sub-GOPs, each of which consists of 12-frames.

3.1 Frame Rate Control

To estimate the supportable frame rate of the following sub-GOP, HOD values are calculated in current sub-GOP. Based on these information, the frame rate for next sub-GOP is predicted. First, the difference images between two encoded adjacent frames are calculated. TH_1 of Equation 1 is set to 32 in this experiment. By Equation 2, $\hat{D}_e$ is calculated. The frame rate change is determined by $\hat{D}_e$. The results are shown in Table 3. The negative slope in Table 3 means that the motion is becoming slow while positive slope means that the motion is becoming fast. TH_2 and TH_3 are set to 0.01 and 0.06, respectively. Now, for 2_{nd} and 6_{th} sub-GOPs, frame rate is changed.

Sub	Slope	last HOD	m	Rate
1	0.006290	0.061553	0.0804	3
2	-0.006145	0.026407	0.0080	2
3	0.000492	0.018742	0.0202	3
4	0.000300	0.021100	0.0220	3
5	0.004858	0.549830	0.0696	2
6	-0.006279	0.180120	-0.0008	3
7	0.000631	0.008967	0.0109	3

Table 3. Calculation of $\hat{D}_e$ and encoded frame numbers of sub-GOPs ($\omega = 3$).

3.2 Global Bit Allocation

The optimization of quantization parameters is achieved on the basis of frame rate determined by frame rate control algorithm. The new cost function $\Phi_m(\vec{q_m}, \lambda_m)$ in Equation 13 is minimized by adjusting the quantization parameters for encoded frames in sub-GOP. λ_m is related to dependency among the sub-GOPs. Gradient search algorithm is used for optimal solution of Equation 13. In this experiment, Table 4 is used for λ_m. The λ_m in Table 4 satisfy the Equation 14.

No. of sub-GOP(m)	λ_m
1	0.0001
2	0.0002
3	0.0002
4	0.001
5	0.001
6	0.001
7	0.01
8	0.1

Table 4. Weighting factors for quantization parameter optimization.

By proposed frame rate control and global bit allocation algorithms, we can get the following results. Proposed algorithm is compared with Telenor tmn version 2.0 [2]. The PSNR plot of encoded frames is shown in Figure 1 (a), and statistical data of PSNR is shown in Table 5,

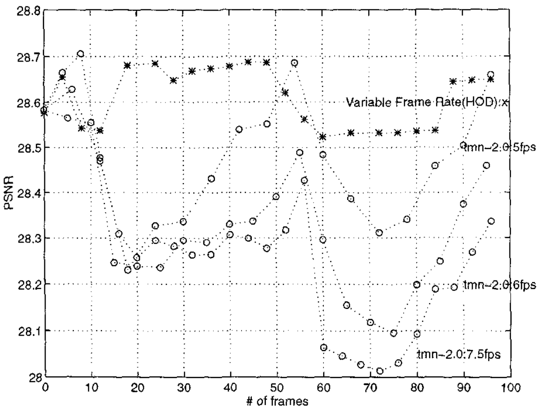

Figure 1. Performance comparison : PSNR plot of Telenor tmn 2.0 and proposed algorithm.

- Quality change
  To reduce the abrupt change in quality, ω_q in Equation 7 is set to 2. As shown in Table 5, proposed rate control algorithm increases the average PSNR by about 0.3 - 0.4 dB while it reduces the

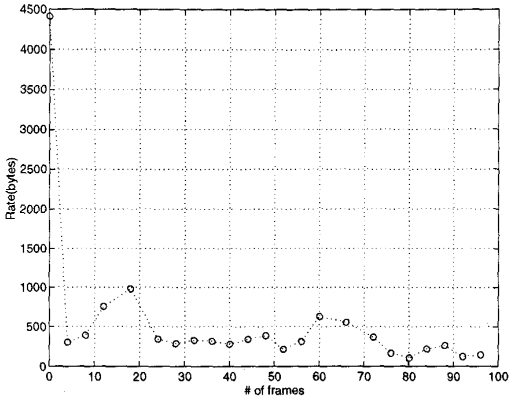

Figure 2. Rate generated by proposed algorithm.

Method	Average	S.D.	Frames
tmn 2.0(1/4)	28.2796	0.1870	24
tmn 2.0(1/5)	28.3251	0.1450	19
tmn 2.0(1/6)	28.4663	0.1322	16
Proposed	28.6410	0.0813	21

Table 5. Comparison of average and standard deviation of PSNR.

standard deviation of PSNR by about 40 - 65 %. Although the standard deviation of PSNR is not an exact measure of flickering effect, it is fair to say that the flickering effect can be reduced by the smaller standard deviation of PSNR in the case of slow moving head & shoulder video. In this experiment, we saw a significant flickering effect in the neck-tie area coded by Telenor tmn 2.0. Proposed rate control algorithm reduces the flickering of this part, greatly.

- Motion smoothness
  As said earlier, intra-filter method is used for frame interpolation. A subjective measure is used for motion smoothness. With proposed frame rate control algorithm, the motion smoothness was not degraded. That is, we can not distinguish the difference between Telenor tmn 2.0(1/4) and proposed rate control algorithm(21 frames(HOD)), subjectively. Proposed algorithm can improve the spatial quality of each frame, reduce the quality change and provide more smooth motion than Telenor tmn 2.0(1/6) at the same bit rates.

4. Conclusion

Two new rate control schemes were proposed for H.263 in this work. The first scheme treats the frame rate as a control variable so that we can exploit the difference of degradation due to smooth and rapid motion patterns in video. Second, we considered a global bit allocation algorithm for H.263+. Up to now, there has been little research about global bit allocation due to the large amount of computational complexity required. With this algorithm, we can determine a good trade-off between the computational complexity and the rate-distortion performance. As a result, a good rate-distortion performance is achieved at the expense of a low computational complexity of rate control.

References

[1] *DRAFT ITU-T Recommendation H.263: Video coding for low bitrate communication*, July, 1997.

[2] Telenor Research,"TMN(H.263) encoder/decoder, version 2.0," TMN(H.263) codec, June, 1996.

[3] *ISO/IEC 13818(MPEG-2):Generic coding of moving pictures and associated audio information*, Nov. 1994.

[4] C. T. Chen and A. Wong, "A self-governing rate-buffer control strategy for pseudo constant bit rate video coding," *IEEE Trans. Image Proc.*, Vol. 2, pp. 50–59, Jan. 1993.

[5] Jungwoo Lee and Bradly W. Dickinson, "Temporally adaptive motion interpolation exploiting temporal masking in visual perception," *IEEE Trans. on Image Proc.*, Vol. 3, No. 5, pp. 513–526,Sep., 1994.

[6] K. Ramchandran, A. Ortega and M. Vetterli, " Bit allocation for dependent quantization with application to multiresolution and MPEG video coder," *IEEE Trans. on Image Proc.*, Vol. 3, No. 5, pp. 533–545,Sep., 1994.

[7] L. J. Lin, A. Ortega, and C. C. J. Kuo, "Rate control using spline interpolated R-D characteristics," in *Proc. of SPIE Visual Communication and Image Processing '96*, pp. 111–122, Mar. 1996.

[8] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell and S. K. Mitra, "Rate-distortion optimized mode for very low bit rate video coding and emerging H.263 standard," *IEEE Trans. on CAS for video technology*, Vol. 6, No. 2, April 1996.