Efficient Rate Control by Fast Adaptive Mode Selection

*Chul Ryu

※ This work was supported by a technical journal publishing program of Dongguk University

Abstract

A fast converging coding algorithm that adaptively selects the modes of macroblocks is introduced. For a given frame, the optimal modes are selected based on the decision curves that minimize the overall distortion at a given bit rate. The method proposed in this paper is different from the conventional ones in that it does not manipulate the quantizer to meet the target bit rate but it satisfies the target bit rate by finding optimal modes of macroblocks which result consistent visual quality. Lagrange multiplier of the unconstrained cost function is controlled to trigger decision curves to generate appropriate modes to meet bit rate and the curve is obtained by utilizing simulated annealing optimization technique. The algorithm is implemented within H.261 video codec and simulation results demonstrate superior visual quality.

Ⅰ. Introduction

It is well known that interframe coding can achieve better coding performance than intraframe coding at a given bandwidth but the resulting bit rate from interframe coding fluctuates in nature. In some applications, such as transmission over PSTN, constant bit rate is more desirable either for a simpler network configuration or for a fixed bandwidth. To achieve a constant bit rate transmission of video signals for such applications, a buffer is placed between the coder' s output and the channel to smooth out the bit rate fluctuation. Therefore rate control is a key importance to visual quality in digital video transmission over a constant bit rate channel.

Numerous methods have been introduced to improve visual quality at a given bit rate while still being compatible with the standard coder [6][10][14], where in the conventional strategies [2][3], the quantization parameter was manipulated based on the fullness of buffer. One published solution to this problem suggests the use of the previous bit counts as a prediction for the current macroblock or sub-group-of-block [4] to adjust the level of quantization. Other approach uses the estimates from the training sequences to predict the number of bits for the current macroblock under a stationary assumption [8]. In [15], rate-constrained product code is formalized to optimize the combination of quantization choices. In general, the more often the quantization is adjusted, the smoother the variation of the output bit rate and the smaller the required buffer size become. On the other hand, if the adjustment of quantization is based on a longer period of time, a larger buffer size may be required.

This paper proposes an efficient rate control algorithm by adaptively selecting the optimal modes of macroblocks in a block based video codec. Bit rate regulation is accomplished by manipulating Lagrangian multiplier of the unconstrained cost function, which triggers decision curves to generate modes to meet bit budget. Unique decision curves are utilized in each frame and are determined by iteratively comparing the output bit rate with a target bit rate which is based on the desirable bit rate, macroblock mode and frame complexity. This iterative procedure, which is independent of the channel buffer occupancy, guarantees optimal decision curves by utilizing the simulated annealing optimization technique [7]. Once optimal decision curves are determined, optimal modes of macroblocks are selected based on the decision curves that minimize the overall distortion at a given bit rate. Two different decision curves are optimizes in this paper, i.e., motion/no-motion compensation decision curve and intra/intercoding decision curve. Furthermore, bit rate

---

*Dept. of Information and Communications Eng.,
Dongguk University
Manuscript Received : November 8, 1999

is controlled within the macroblock layer in addition to the picture layer to aid a precise rate control. This paper is organized as follows. In Section 2, some fundamental problems existing on standard coders are described and adaptive mode selection algorithm is introduced. Rate control based on the adaptive mode selection algorithm is discussed in Section 3. The superior performance of the proposed algorithm is verified by simulation under various conditions and is shown in Section 4 and concluding remarks follow in Section 5.

Ⅱ. Problem Formulation

2.1. Background

In standard block based video codecs, the encoder has to make several decisions for each macroblock that which mode it will be coded with. These decisions can be: how to determine the best motion vectors to use, decide whether to code each macroblock as intra or predicted, and how to set the quantizer scale. Indeed an encoder has the difficult task of choosing between the different types of macroblocks. An exhaustive search method is to try coding a macroblock to the same degree of accuracy using each type, then choose the type that requires the least number of coding bits. Obviously exhaustive search is not suitable for a real coding scheme but a simpler method by making a series of decisions can be practical to implement and is computationally less expensive.

At each step of making decision, a fixed function or a fixed rule is used to speed up the decision making. For example, decisions of motion/no-motion compensation and intra/intercoding use preset functions (shown in Figure 1), whereas code/no-code decision is determined by the difference of error signals and quantization parameters are determined based on the buffer's contents. After a series of decisions, a mode of each macroblock is determined to be coded accordingly.

Most of the published solutions to rate control are concentrated on determining the levels of quantizer, which are adjusted based on the fullness of buffer. The proposed algorithm extends the coding decision options for rate control to the decision of motion/no-motion compensation as well as intra/intercoding. Instead of having fixed motion/no-motion compensation and intra/inter decision curves for the whole frame, it would be more natural to find adaptive decision curves based on the characteristics of each frame so that the PSNR at a given bit-rate is maximized. Therefore, the proposed approach provides better rate control than a simple quantizer feedback approach. The proposed approach simplifies the problem by seeking optimal decision curves for mode selection and decision curves are determined based on the rate-distortion theory [12][13].

2.2. Mode Selection

Let F_j and $\hat{F}_j$ be the jth input and reconstructed frames of the video codec, respectively, and they can be partitioned into groups of macroblocks,

$$F_j = (X_{o,j}, X_{1,j}, X_{2,j}, \dots, X_{L-1,j}) \qquad (1)$$

$$\hat{F}_j = (\hat{X}_{o,j}, \hat{X}_{1,j}, \hat{X}_{2,j}, \dots, \hat{X}_{L-1,j}). \qquad (2)$$

Macroblock $X_{i,j}$ in F_j can be coded using only one of N possible modes given by the set S,

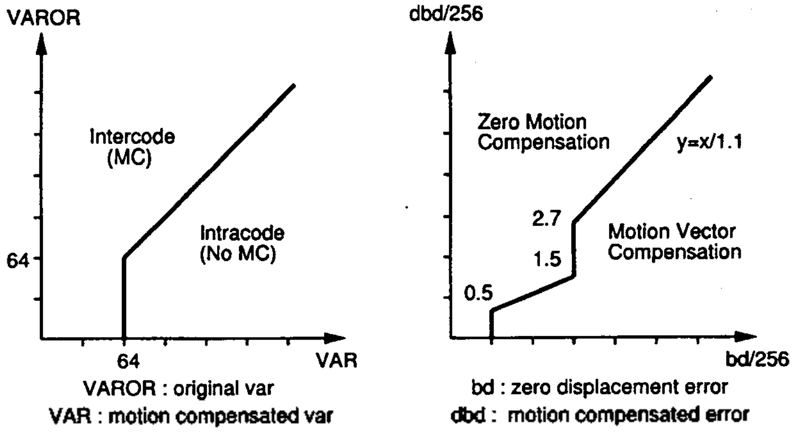


Figure 1. Courtesy of MPEG: Decision curves for motion/no-motion compensation and intra/intercoding for P-pictures.

$$S = (M_0, M_1, M_2, \ldots, M_{N-1}). \tag{3}$$

Let $M_k^i \in S$ where $k = 0, 1, \ldots, N-1$, be the mode selected to code the macroblock $X_{i,j}$ where the indices i and j represent the macroblock and the frame, respectively. Let $y_{i,j}$ be a coded macroblock with a selected mode M_k^i and $\hat{X}_{i,j}$ be the output of the video codec (Figure 2).

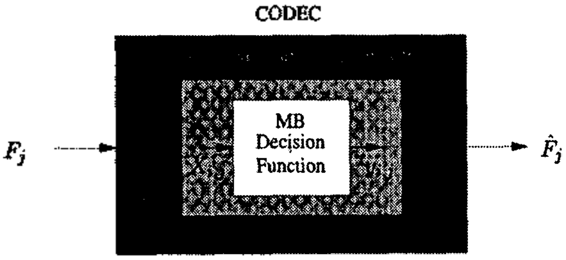


Figure 2. Simplified structure of the block based video encoder.

In general, $y_{i,j}$ can be represented depending on which mode it is assigned with,

$$y_{i,j} = \begin{cases} 0, & \text{if skipped} \\ X_{i,j}, & \text{if intra coded} \\ X_{i,j} - X_{i\Delta,j-1} & \text{if inter coded} \end{cases} . \tag{4}$$

$X_{i\Delta,j-1}$ is the ith macroblock of (j-1)th frame with the motion vector Δ where

$$\Delta = \begin{cases} 0, & \text{if no motion compensated} \\ \delta_{xy}, & \text{if motion conpenstated} \end{cases} . \tag{5}$$

Let' s assume that the decision curve for motion/no-motion compensation can be expressed as a polynomial of order P-1,

$$g(X) = \sum_{k=0}^{P-1} a_k X^k \tag{6}$$

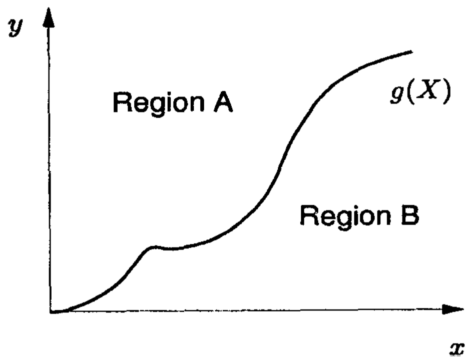


Figure 3. Sample decision curve dividing zero motion compensation and non-zero motion compensation.

EQ. 6 decides whether a macroblock should be coded using motion compensation or not. Figure 3 shows a sample decision curve which divides a region between motion compensation (region B) and no-motion compensation (region A).

Let the x-axis in Figure 3 be the sum of absolute differences between $X_{i,j}$ and $X_{i,j-1}$,

$$x(X) = | X_{i,j} - X_{i,j-1} | = \frac{1}{256} \sum_m \sum_n | X_{i,j}(m,n) - X_{i,j-1}(m,n) | . \tag{7}$$

And the variable y is defined as the sum of absolute differences between $X_{i,j}$ and $X_{i\Delta,j-1}$

$$y(X) = | X_{i,j} - X_{i\Delta,j-1} | = \frac{1}{256} \sum_m \sum_n | X_{i,j}(m,n) - X_{i\Delta,j-1}(m,n) | \tag{8}$$

where, the motion vector Δ is determined through the curve $g(X)$ as

$$\Delta = \begin{cases} 0, & \text{if } y(X) \geq g(X) \\ \delta_{xy}, & \text{if } y(X) < g(X) \end{cases} \tag{9}$$

Using the values of $x(X)$ and $y(X)$, a mode for motion and no-motion compensation will be decided by the coordination of $(x(X), y(X))$ with respect to $g(X)$.

The problem to solve, then, is that of finding the coefficients a_k in EQ. 6 which minimizes the global distortion D of the jth frame at a given bit rate R,

$$\min_{a_k} D_{F_j}(a_k) \tag{10}$$

$$\text{such that } R_{F_j}(a_k) \leq R. \tag{11}$$

Instead of solving the constrained optimization problem EQ. 10, let us consider the following unconstrained formulation. Let the Lagrangian cost function corresponding to the Lagrange multiplier $(\lambda \geq 0)$ for a_k of the jth frame be,

$$J_{F_j}(a_k) = D_{F_j}(a_k) + \lambda \cdot R_{F_j}(a_k). \tag{12}$$

For a *fixed* value of λ, the unconstrained problem specified below can be solved for the optimal a_k (a_k^*) that minimizes the Lagrangian cost function

$$J_{F_j}(a_k^*) = \min_{a_k} \{ D_{F_j}(a_k) + \lambda \cdot R_{F_j}(a_k) \} \tag{13}$$

while its constrained counter part of EQ. 10 becomes

$$D_{F_j}(a_k^*) = \min_{a_k} D_{F_j}(a_k) \text{ such that } R_{F_j}(a_k^*) \leq R. \tag{14}$$

As λ is searched through all positive real numbers from 0 (lowest distortion, highest rate) to ∞ (highest distortion, lowest rate), a set of solutions of (a_k^*) and constraints $R_{F_j}(a_k^*)$are created. The above approach identifies an optimal operating point on the convex functional of the rate-distortion curve for a fixed positive λ If the original constrained problem happened to have a budget constraint that meets one of the operating points on rate-distortion curve, then the unconstrained and constrained problems have identical solutions.

2.3. Fast Solution by Fixing λ

In order to solve EQ. 13, a_k as well as λ which minimize the unconstrained cost function are needed to be determined. It is found empirically that the value of λ does not fluctuate severely between frames when the scene transition is consistent which indicates that the variation of coded bit rate is small between frames. Therefore, instead of finding λ for each frame, computation complexity can be reduced by repeating the same λ for the subsequent frames having consistent scene change. Since λ can be interpreted as a quality index, it is crucial to find the optimal λ for the starting frame.

A biased Lagrangian cost function is introduced in [9] to seek for optimal λ Let x be the set of all permissible combinations of a_k and $J(\lambda) = D(x) + \lambda R(x)$ be the Lagrangian sub-cost function associated with x for quality criterion λ Then, the biased Lagrangian cost function W becomes,

$$W(\lambda) = \left(\min_x [D(x) + \lambda R(x)] - \lambda R\right). \tag{15}$$

Using the above biased cost function, we can find λ^* for the starting frame and the subsequent frames having consistent scene change. By expanding [9], the following result is obtained : $\lambda^*$is the optimal solution to

$$W_{F_j}(\lambda^*) = \max_{\lambda \geq 0}\left(\min_{a_k}\left\{D_{F_j}(a_k) + \lambda R_{F_j}(a_k)\right\} - \lambda R\right) \tag{16}$$

EQ. 16 becomes a new unconstrained cost function to be solved by finding the coefficients a^* with λ^* Therefore, the biased Lagrangian cost function becomes,

$$\begin{aligned} W_{F_j}(\lambda^*) &= \max_{\lambda \geq 0}\left(\min_{a_k}\left\{D_{F_j}(a_k) + \lambda R_{F_j}(a_k)\right\} - \lambda R\right) \\ &= \max_{\lambda \geq 0}\left(\min_{a_k}\left\{\frac{1}{L}\sum_{i=0}^{L-1} | X_{i,j} - (X_{i|\Delta,j-1} + \hat{X}_{i,j}) |^2 + \lambda R_{F_j}(a_k)\right\} - \lambda R\right) \end{aligned} \tag{18}$$

where,

$$\hat{X}_{i,j} = D^{-1} \cdot Q^{-1} \cdot Q \cdot D(X_{i,j} - X_{i|\Delta,j-1})$$

$$F_j = j\text{th input frame}(= \sum_i X_{i,j}). \tag{18}$$

$$\hat{F}_j = j\text{th reconstructed frame}(= \sum_i \hat{X}_{i,j})$$

D and D^{-1} denote DCT and IDCT, and Q and Q^{-1} denote quantization and inverse quantization, respectively.

For the case of intra- and interframe coding decision function, EQ. 7 and EQ. 8 are replaced with

$$x(X) = \mathbf{var}(X_{i,j}, X_{i,j-1}) = \frac{1}{256}\sum_m \sum_n (X_{i,j}(m,n) - X_{i,j-1}(m,n))^2 \tag{19}$$

$$y(X) = \mathbf{var}(X_{i,j}, X_{i|\Delta,j-1}) = \frac{1}{256}\sum_m \sum_n (X_{i,j}(m,n) - X_{i|\Delta,j-1}(m,n))^2 \tag{20}$$

This is then the unconstrained optimization formulation to our problem and a stochastic annealing optimization algorithm is employed to find the coefficients a_k in EQ. 6 which minimizes the distortion in EQ. 17.

Ⅲ. Adaptive Mode Selecting Rate Control

Buffer control is a necessary step to encounter when it comes to control of averaged bit rate to the desired values. Usually, the state of the buffer is fed back to the coder who selects the level of quantizers so that the buffer overflow or underflow can be avoided [2][3]. However, the number of coded bits generated from different modes of frame (I, P, and B) vary widely, thus the conventional rate control schemes that exploit a mapping between buffer occupancy and quantizers do not yield good coding results. In this Section, a simple buffer control strategy based on the AMS algorithm is described.

3.1. Contents of buffer

Let R_f and R_c be the frame rate (fps) and channel rate (bps). Let B be the size of buffer which is k msec of the channel rate, kR_c Let the input rate to the buffer be expressed as

$$R_{in} = R_1 + R_2 + R_3 + \ldots + R_n \tag{21}$$

$$= R_f \cdot r_1 + R_f \cdot r_2 + R_f + r_3 + \ldots + R_f \cdot r_n \tag{22}$$

assuming that only n frames and the unit of $R_i, i = 1, 2, \ldots, n,$ is in bps.

Using the above notations, we can formulate the contents of buffer at each frame as follows,

$$\sum_{i=0}^{n-1} R_f r_i - nR_c \tag{23}$$

$$\sum_{i=0}^{n-1} R_f r_i - (n+1)R_c. \tag{24}$$

EQ. 23 and EQ. 24 represent the fullness of buffer just after the nth frame is coded and just before the $(n+1)$th frame is coded, respectively.

In order to avoid the buffer being overflowed or underflowed, the buffer' s level, B_L must be bounded all the time

$$0 \le B_L < B. \tag{25}$$

A more generalized form can be expressed as follows

$$\alpha B \le B_L < \beta B \tag{26}$$

where, $\alpha + \beta = 1$, $\alpha \ge 0$, $\beta \ge 0$, i.e., $\alpha = 0$, $\beta = 1$ yield EQ. 25

The buffer can avoid its overflow or underflow if EQ. 23 and EQ. 24 satisfy EQ. 26,

$$\alpha B \le \sum_{i=0}^{n-1} R_f r_i - (n+1)R_c \tag{27}$$

$$\sum_{i=0}^{n-1} R_f r_i - nR_c < \beta B. \tag{28}$$

Since the B can be represented in terms of the channel rate, the above EQs can be further expanded as,

$$\alpha k R_c \le \sum_{i=0}^{n-1} \left(\frac{R_f}{R_c}\right) r_i - (n+1) \tag{29}$$

$$\sum_{i=0}^{n-1} \left(\frac{R_f}{R_c}\right) r_i - n < \beta k. \tag{30}$$

By varying α and β , we can preset the bounds to alert when the buffer reaches underflow or overflow.

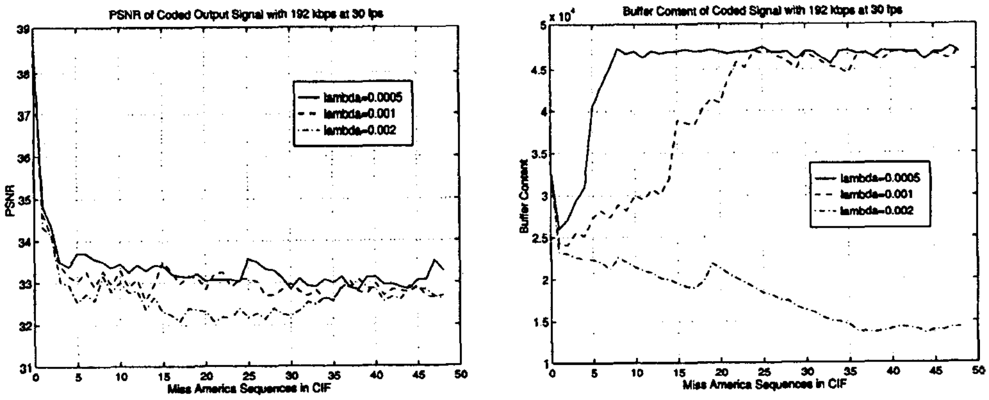
3.2. Updating Lagrangian Multiplier

In rate control using adaptive mode selection, the Lagrangian multiplier is adjusted depending on the state of buffer. When the buffer underflows, dummy bits are filled so that the buffer maintains the minimum of αB bits. If the content of buffer reaches its upper bound, βB , the overflow state is triggered and the decision curve is now more concentrated to generate macroblocks with block skipping mode by increasing the Lagrangian multiplier λ in EQ. 17. By increasing λ the algorithm now gives more favor to bit rate constraint than distortion constraint. Once the optimal coefficients of the decision curve are found, each macroblock is coded according to the mode assigned with.

Figure 4 shows that small increase in λ reduces buffer fullness and the resulting lower PSNR, whereas decrease in λ increases buffer fullness. Therefore, we could control the target bit rate by adjusting λ depending on the state of buffer while sacrificing visual quality.

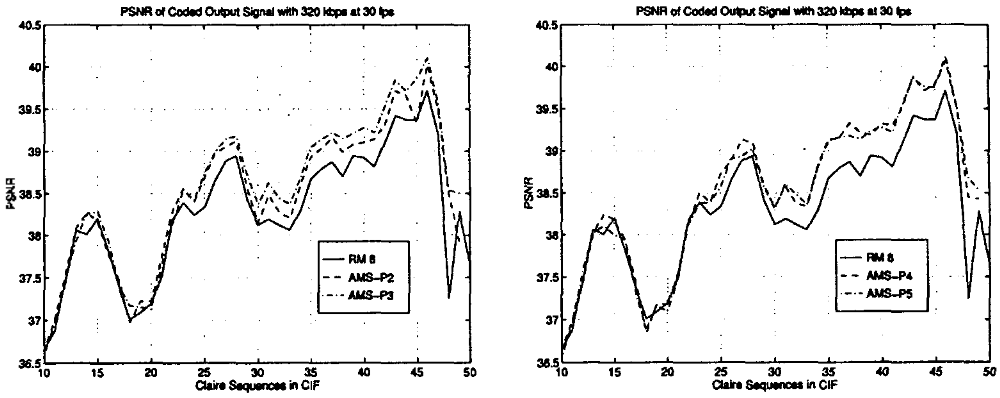
Ⅳ. Simulation Results

Simulations were conducted using *Claire and Miss America* sequences in the CIF format (352 pixels X 240 lines) for different degrees of the polynomial P from 2 to 9 in EQ. 6. Throughout the simulations, the frame rate is held constant at 30 fps and the average bit rate is varied from 160 kbps to 320 kbps. As a part of the encoding process, the modes are selected using the procedure described above: first finding the optimal intracoding curve then intercoding curve. The coding results are then compared with coded sequences generated by the video codec test model RM 8.



a) b)

Figure 4. Change of λ in intra/interframe coding decision curve.



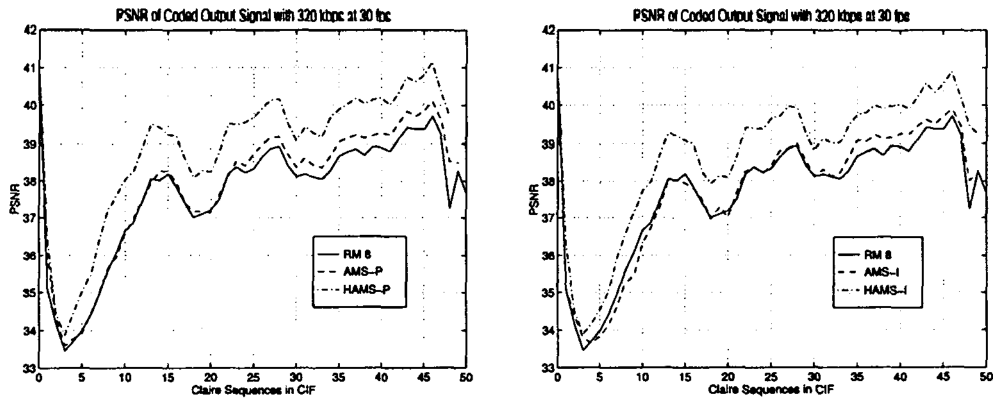
a) b)

Figure 5. PSNR and frame rate of coded Claire sequences using hierarchical search based adaptive mode selection.

During the experiments, the order of polynomial was found to be optimal when $P=3$ for motion/no-motion compensation decision curve Figure 5. The good suboptimal coefficients were not found when the order below 3 was used, and the PSNR was increased a very small amount while the computation became large when the order above 3 was used. The optimal order of polynomial for intra/inter decision curve was found to be $P=2$.

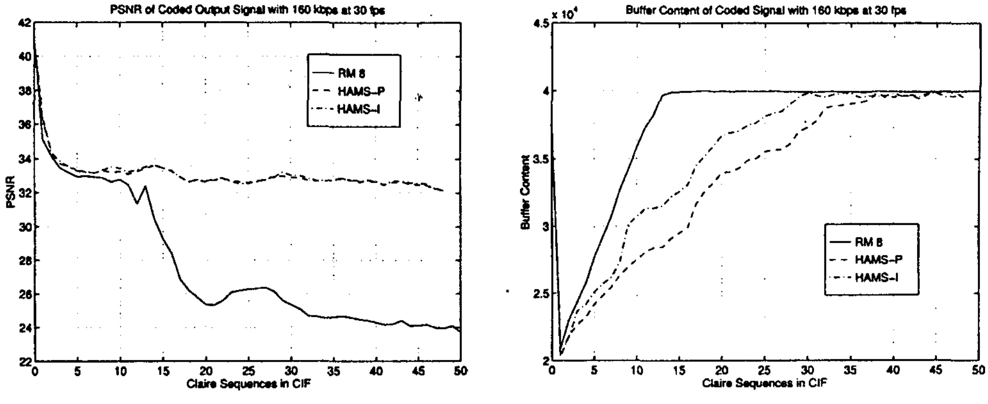
Hierarchical search motion estimation [11] is adopted in order to reduce the burden of complexity in finding motion vectors and to increase coding performance. The modes are affected from the motion vectors found by hierarchical search motion estimation. The modes found using hierarchical search by adaptive mode selection (HAMS) would be different from the modes selected based on adaptive mode selection. This is because different motion vectors result different absolute difference or variance between the current macroblock and the macroblock corresponding with that motion from the previous frame. It is shown in Figure 6. that HAMS approach performs better than the AMS approach.

Two different models of finding optimal decision curve using hierarchical search based adaptive mode selection (HAMS) are compared with RM 8, i.e. using optimal motion/no-motion compensation (HAMS-P) and inter/intracoding curve (HAMS-I). Proposed models demonstrate consistent visual quality within and between frames as shown in Figure 6. To investigate the effect of buffer' s overflow and underflow, a relatively low bit rate was applied while other parameters remain the same. In the normal coding procedure, bit rate should be greater



a) b)

Figure 6. PSNR and frame rate of coded Claire sequences using adaptive mode selection versus hierarchical search based adaptive mode selection.



a) b)

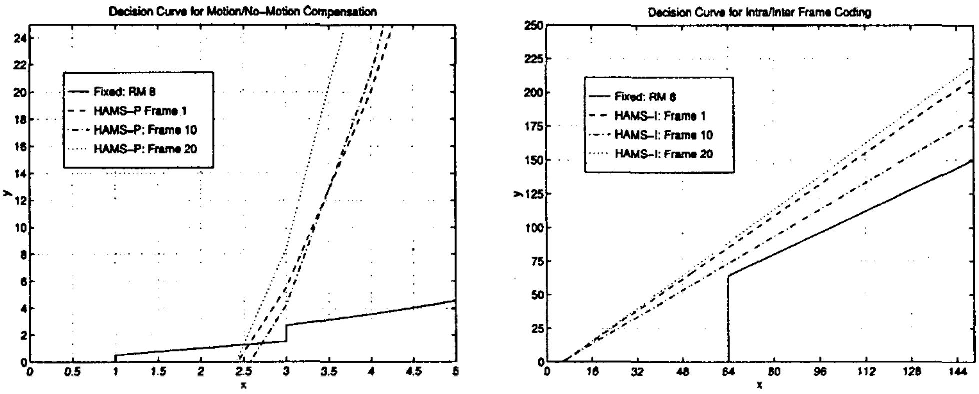
Figure 7. PSNR and buffer contents of coded Claire sequences using selective mode.

than or equal to 320 kbps in order to avoid overflow when CIF-type frames with the frame rate at 30 fps are coded. Figure 7 illustrates the corresponding PSNR and buffer contents when frames are coded with 160 kbps at 30 fps.

The buffer begins to fill significantly after frame 13 (Figure 7-b) which forces RM 8 to repeat the previous frame resulting degraded visual quality (Figure 7-a). With the proposed scheme, however, the buffer doesn' t reach overflow state until frame 35 as macroblock modes are selected based on the optimal decision curves in all models and show superior visual quality. Figure 8 illustrates some motion/no-motion compensation and intra/intercoding decision curve compared with fixed ones. It can be seen from Figure 8-a) that if the error magnitude is low (about 2.5 in HAMS case compared to 1.0 in RM 8) then no motion compensation is used. Intra/intercoding decision curves using HAMS favor intraframe coding as shown in Figure 8-b) and explains why the optimal order of the polynomial of intra/intercoding decision curve is 2.

V. Conclusion

In this paper we introduced a new scheme of rate control algorithm in video coding by adaptively selecting the optimal operation mode of block-based video coding framework. The algorithm is based on the Lagrangian method for optimal bit allocation at a given bit rate. Simulation results demonstrate higher PSNR and better visual quality compared with standard quantizer feedback based bit rate control approach. A drawback of this method is the requirement of rather high complexity due to simulated annealing optimization algorithm. However, it could be useful in certain applications where complexity may not be a critical issue, such as storage of video on



a) b)

Figure 8. PSNR of change in order of motion/no-motion decision curve. The order of the polynomial was varied from 2 to 5.

CD-ROM. More work can be done to improve the algorithm and reduce its complexity, based on the same basic idea. Experimental results, provided in this paper, demonstrate that the algorithm is effective and can be very useful in many applications.

References

1. CCITT SDM XV-R 37-E, *Video codec for audiovisual services at px64 kbps*, CCITT Recommendation H.261, Aug. 1990.
2. ISO/IEC 11172-2, *Information technology-Generic coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s -Part 2: Video*, International Standard, Nov, 1993.
3. CCITT SC XV, *Description of reference model 8 (RM 8)*, Document 525, June 1989.
4. C. T. Chen and A. Wong, A self-governing rate buffer control strategy for pseudoconstant bit rate video coding, *IEEE Trans. on Image Processing*, vol. 2, No, 1, pp. 50-59, Jan. 1993
5. W. Ding and B. Liu, Rate control of MPEG video coding and recording by rate-quantization modeling, *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 6, No. 1, pp. 12-20, Feb. 1996.
6. C. A. Gonzales and E. Viscito, Motion video adaptive quantization in the transform domain, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, no. 4, pp. 374-378, Dec. 1991
7. L. Ingber and B. Rosen, Genetic algorithms and very fast simulated reannealing: a comparison, *Mathematical and Computer Modelling*, vol. 16, no. 11, pp. 87-100, 1992.
8. A. Puri and R. Aravind, Motion-compensated video coding with adaptive perceptual quantization, *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 1, No. 4, pp.351-361, Dec. 1991.
9. K. Ramchandran and M. Vetterli, Best wavelet packet bases in a rate-distortion sense, *IEEE Trans. Image Processing*, vol. 2, no. 2, pp. 160-175, Apr. 1993.
10. K. Ramchandran, A. Ortega, and M. Vetterli, Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders, *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 533-545, Sep. 1994.
11. C. Ryu and S. P. Kim, Bit-plane hierarchical search motion estimation, *IEEE 5th Signal and Image Processing Conference*, Nov. 11 - Nov. 14, 1996.
12. Y. Shoham and A. Gersho, Efficient bit allocation for an arbitrary set of quantizers, *IEEE Trans. of Acoust. Speech, Signal Processing*, vol. 36, pp. 1445-1453, Sept. 1988.
13. Claude E. Shannon, Coding Theorems for a discrete source with a fidelity criterion, *IRE Nat. Conv. Rec.*, pt. 4, pp. 142-163, March 1959.
14. E. Viscito and C. Gonzales, A video compression algorithm with adaptive bit allocation and quantization, *SPIE Visual Commun. Image Processing*, vol 1605-1, pp. 58-72, Nov. 1991.
15. T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell and S. Mitra, Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard, *IEEE. Trans. on Circuits Syst. Video Technol.*, vol. 6, No. 2, pp. 182-190, April 1996.

▲ Chul Ryu

Chul Ryu received the M.S and Ph.D degrees of electrical engineering in 1991 and 1997 respectively, both from the Polytechnic University, New York. From 1998 to 1999, he served as a research engineer of advanced wireless technologies lab at LG Information and Communications. He recently joined the Department of information and communications engineering at the Dongguk University, Korea. His research interests are visual communications, encryption and data compression.