Optimizing INTRA/INTER Coding Mode Decisions

Yen-Kuang Chen*, Anthony Vetro+, Huifang Sun+, and S. Y. Kung*
Princeton University* and Mitsubishi Electric ITA+

Abstract

In this paper, we present a new approach for INTRA/INTER mode decisions. In the current MPEG verification model (VM), the coding mode for each macroblock is selected by comparing the energy of predictive residuals. Specifically speaking, the INTRA/INTER decision is determined by comparing the variance of the macroblock pixels against the predictive residuals. However, the coding mode selected by the VM criteria does not result in the optimal coding performance. We observe that in deciding which of the various coding modes is best, one should not only consider spending bits for coding the prediction residuals but also consider spending bits for coding motion vectors. The simulation results show that our method provides better SNR while using the same amount of bits.

1 Introduction

The MPEG video coding standards specify a general coding methodology and syntax for generating compliant MPEG bitstreams. However, many opportunities still exist to improve the coding efficiency. This flexibility in the encoder design has encouraged a great deal of research in the areas of image pre-processing, motion estimation, coding mode decisions, and rate control. In most cases, the objective is to minimize subjective distortion for a prescribed bit rate, while observing constraints on operating delay and computational complexity.

In [3], Chen and Willson formulate the motion estimation problem as a shortest path finding problem, which minimizes the number of bits for texture and for motion assuming all the blocks are coder in the INTER mode. They use Viterbi-type dynamic programming to determine the optimal motion vectors. In [2], Chen, Villasenor, and Park present an alternative motion estimation algorithm that considers rate-distortion trade-offs in a low complexity framework. However, both techniques, while achieving good bit rates, are computational too complex for practical video coding.

In [6], Sun, Kwok, Chien, and Ju present a new algorithm for determining the optimal MPEG coding strategy in terms of the selection of macroblock coding modes and quantizer scales. They observe that the processes of coding mode decision and rate control are intimately related to each other. Hence, the two processes should be determined jointly in order to achieve optimal coding performance. They formulate the constrained optimization problem and present solutions based upon rate-distortion characteristics, for all the macroblocks that compose the picture being coded. The determination of the optimal solution is complicated by the MPEG differential encoding of motion vectors and DC coefficients, which introduce dependencies that carry over from macroblock to macroblock for a duration equal to the slice length. As an approximation, a near-optimum greedy algorithm is proposed.

In [4], we observe that piecewise continuous motion field reduces the bit rate for differentially encoded motion vectors. We propose a rate-optimized motion estimation based on a “true” motion tracker. Since the algorithm provides true motion vectors for encoding, the blocking artifacts are decreased and, hence, the pictures look better subjectively.

In this paper, we would like to extend our proposed neighborhood relaxation method to be used with coding mode decisions. To code a particular macroblock in a P- or B- video-object-plane (VOP), many choices as specified by the emerging MPEG-4 standard [1] exist. For one, the encoder must choose between INTRA and INTER coding. Then, if INTER coding is chosen, there are a variety of prediction modes to choose from. In this paper, we offer a new approach on the coding mode decision between the INTRA mode and the 16×16 INTER mode. In this approach, we incorporate the concepts of true motion tracking and take into account the number of bits used to code the motion.

The remainder of the paper is organized as follows. In Section 2, some preliminaries on the current INTRA/INTER decision process and motion vector encoding is given. In Section 3, the proposed approach is introduced. The results of our simulations which compare the conventional mode decision with our new mode decision is discussed in Section 4. Finally, in Section 5, some conclusion remarks are given.

2 INTRA/INTER Decision

In the INTRA mode, the texture of the macroblock is coded via a DCT transformation, quantization, and variable length coding (VLC). No motion compensation is required in this mode.

On the contrary, the INTER mode uses motion compensation. First, the block matching motion estimation is used, for every block in the current frame (called current block), to find the best matched block within a range in the previous frame (called predicted block). The displacement of the predicted block relative to the current block is called a motion vector (MV). A motion compensated difference block (called residual block) is formed by subtracting the pixel values of the predicted block from that of the current block point by point. Texture coding is then performed on the difference block. The coded motion vector and the coded texture information of the difference block are transmitted to the decoder. Using this information, the decoder can then reconstruct an approximated current block by adding the quantized difference block to the predicted block according to the motion vector.

Because most of the blocks in the current frame are similar to a predicted block in the existing frames, the residue of the block subtracted by a similar predicted block is usually very small. In this case, a small amount of bits can encode the residual block via a similar DCT transformation. Therefore, most of the time, coding the motion vector as well as coding the residue costs fewer bits than coding the texture of the block.

Nevertheless, in some cases (e.g., scene change, object occlusion), coding the difference block and the motion vector may require more bits than coding the original texture of the current block. Consequently, there must be a way to decide when to choose INTRA mode and when to choose INTER mode.

A sound solution of choosing INTRA/INTER mode for a block B_i should compare $\mathit{bits}(\mathit{VLC}(\mathit{Q}(\mathit{DCT}(B_i))))$ with $\mathit{bits}(\mathit{VLC}(\mathit{Q}(\mathit{DCT}(\mathit{MC}(B_i))))) + \mathit{bits}(\mathit{motion\ vector})$ when the distortion is the same. If the first term is smaller, then the INTRA mode is preferred; otherwise, the INTER mode is preferred.

Original Criterion in Current MPEG-4 Verification Model

It is computational intensive to calculate the entire bit counts and distortion. For simplicity, in current MPEG-4 Verification Model (VM), the following parameters are calculated to make the INTRA/INTER decision:

$$\mathit{SAD} = \min_{MV_x, MV_y} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |p(i + MV_x, j + MV_y) - c(i,j)| \quad (1)$$

$$\mathit{mean} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} c(i,j)$$

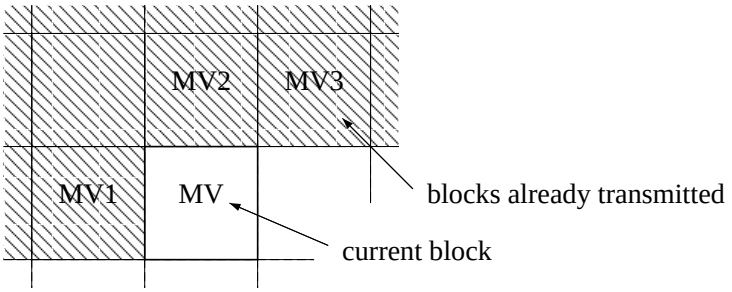

Figure 1: Motion vector prediction for differential coding of motion vectors used in MPEG-4.

$$VAR = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |c(i,j) - mean| \tag{2}$$

where $N = 16$ is the size of the block. The INTRA mode is chosen if

$$VAR < SAD - 2N_B \tag{3}$$

where $N_B = N^2$ is the number of pixels of the macroblock. In order to favor the INTER mode when there is no significant difference in the energy of predictive residuals and the energy of original texture, *SAD* is reduced by $2N_B$.

Nevertheless, it is too simple to cover all the cases. For example, it doesn't take in account the number of bits for encoding motion vectors. Because it takes no bits on coding motion vector in the INTRA, it can take more bits on texture coding.

Motion Vector Encoding

Because motion vectors in a neighborhood usually show good similarities, motion vectors are encoded differentially for coding efficiency. Similar to 1-dimensional DPCM, a difference is calculated between the current motion vector and the prediction of current motion vector. However, unlike 1-dimensional DPCM, the prediction of current motion vector is not equal to the previous transmitted motion vector. Current MPEG-4 use a spatial neighborhood of three motion vectors already transmitted to predict the current motion vector as shown below:

$$\begin{aligned} P_x &= \text{Median}(MV1_x, MV2_x, MV3_x) \\ P_y &= \text{Median}(MV1_y, MV2_y, MV3_y) \end{aligned}$$

where $MV1$, $MV2$, and $MV3$ are the three motion vectors already transmitted as shown in Figure 1. Thereafter, the vector differences MVD_x and MVD_y are

$$\begin{aligned} MVD_x &= MV_x - P_x \\ MVD_y &= MV_y - P_y \end{aligned}$$

Then, the variable length coding is applied to the difference between the actual motion vector and the predicted motion vector. Figure 2 shows the bit requirement for different vector difference. The smaller the difference, the less the bits required. The zero differential motion vector saves a lot of bits in encoding. That is, we can use more bits to code the residual component. In order to favor the zero difference vector when there is no significant difference in the sum of absolute difference (SAD), $SAD(P_x, P_y)$ is reduced by $(N_B/2 + 1)$. After that, the displacement vector resulting in the lowest SAD is chosen as the motion vector.

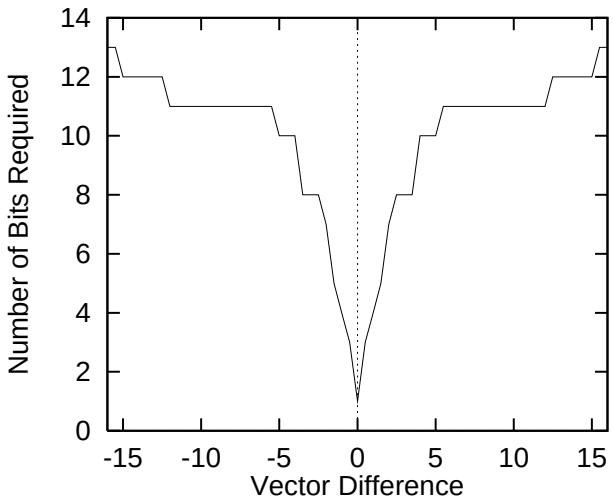

Figure 2: Variable length coding in motion vector difference used in MPEG-4.

The decision of coding modes gets more complicated when the motion vectors are coded differentially by using a spatial neighborhood of motion vectors already transmitted. The *bits*$(VLC(Q(DCT(MC(B_i)))))+$ *bits*$($*motion vector*$)$ not only depends on the texture of the block but also depends on the motion vectors which are already transmitted. Moreover, the next optimal motion vector also depends on the current decision.

3 Proposed Approach

Mathematically speaking, the INTRA/INTER mode decision for a single block should consider

$$\min\left\{\textit{bits}(\textit{Texture}(B_i), Q_I), \min_{\vec{v}}\{\textit{bits}(\textit{Residue}(B_i,\vec{v}), Q_i) + \textit{bits}(\Delta\vec{v})\}\right\} \tag{4}$$

where Q_I stands for the INTRA quantization level, Q_i stands for the INTER quantization level, *bits*$($*Texture*$(B_i), Q_I)$ stands for the number of bits for texture coding, *bits*$($*Residue*$(B_i,\vec{v}), Q_i)$ stands for the number of bits for residual coding, and *bits*$(\Delta\vec{v})$ stands for the number of bits used for differentially coding the motion vectors.

At the start of each slice, prediction motion vectors are reset to zero. As each macroblock is encoded in raster scan order, a macroblock motion vector is encoded differentially with respect to previous motion vectors. *In the event that the previous macroblock was coded as INTRA or skipped, the previous motion vector is assumed to be zero.* Therefore, an INTRA block is more preferable when the surrounding blocks are static or moving slowly while an INTER block is more preferable when the surrounding blocks are moving fast.

Assume the blocks in the neighborhood are coded in INTER mode.

We could add a look-ahead penalty term (non-causal) for INTRA mode as the following:

$$\min\left\{\textit{bits}(\textit{Texture}(B_i), Q_I) + \textit{bits}(\Delta\vec{v}_0), \min_{\vec{v}}\{\textit{bits}(\textit{Residue}(B_i,\vec{v}), Q_i) + \textit{bits}(\Delta\vec{v})\}\right\} \tag{5}$$

where *bits*$(\Delta\vec{v}_0)$ counts the bits for the motion vector difference from $\vec{v}$ to zero motion.

The *bits*$($*Texture*$(B_i), Q_I)$ grows when the variance of the texture of B_i grows and the Q_I drops. The *bits*$($*Residue*$_i(\vec{v}), Q_i)$ and *bits*$(\Delta\vec{v})$ grow when *SAD*$(B_i,\vec{v})$ and $||\Delta\vec{v}||$ grow, respectively. Because it is diffi-

cult to mathematically express the bit costs for different *Textures*, *SADs*, and $\Delta\vec{v}$, the above equation is first simplified into the following approximation

$$\min\left\{\frac{\alpha_I}{Q_I}(\mathit{VAR}(B_i)) + \beta||\Delta\vec{v}_0)||, \min_{\vec{v}}\{\frac{\alpha_i}{Q_i}\mathit{SAD}(B_i,\vec{v}) + \beta||\Delta\vec{v}||\}\right\} \tag{6}$$

where $\mathit{VAR}(B_i)$ stands for the variance of the texture of B_i as shown in Eq. (2).

Assume that B_j is a neighbor of B_i, $\vec{v_j^*}$ is the true motion vector for the neighborhood, and that $SAD(B_j,\vec{v})$ increases as $\vec{v}$ deviates from $\vec{v_j^*}$according to

$$\mathit{SAD}(B_j,\vec{v}) \approx \mathit{SAD}(B_j,\vec{v_j^*}) + \gamma||\vec{v} - \vec{v_j^*}|| \tag{7}$$

or

$$||\Delta\vec{v}|| = ||\vec{v} - \vec{v_j^*}|| \approx \gamma^{-1}(\mathit{SAD}(B_j,\vec{v}) - \mathit{SAD}(B_j,\vec{v_j^*})) \tag{8}$$

Substituting Eq. (8) into Eq. (6), we have

$$\min\left\{\frac{\alpha_I Q_i}{\alpha_i Q_I}(\mathit{VAR}(B_i)) + \mu\sum_{B_j\in\mathcal{N}(B_i)}\left(\mathit{SAD}(B_j,\vec{0}) - \mathit{SAD}(B_j,\vec{v_j^*})\right),\right.$$
$$\left.\min_{\vec{v}}\{\mathit{SAD}(B_i,\vec{v}) + \mu\sum_{B_j\in\mathcal{N}(B_i)}\left(\mathit{SAD}(B_j,\vec{v}) - \mathit{SAD}(B_j,\vec{v_j^*})\right)\}\right\} \tag{9}$$

where $\mathcal{N}(B_i)$ means the neighboring blocks of B_i.

Our Proposed Criterion

Dropping the $\mathit{SAD}(B_j,\vec{v_j^*})$ (which can be considered as constant), we have

$$\min\left\{\frac{\alpha_I Q_i}{\alpha_i Q_I}(\mathit{VAR}(B_i)) + \mu\sum_{B_j\in\mathcal{N}(B_i)}\mathit{SAD}(B_j,\vec{0}), \min_{\vec{v}}\{\mathit{SAD}(B_i,\vec{v}) + \mu\sum_{B_j\in\mathcal{N}(B_i)}\mathit{SAD}(B_j,\vec{v})\}\right\} \tag{10}$$

Fast Motion in the Neighborhood

Eq. (10) provides a more flexible decision than Eq. (3). To compare Eq. (10) with Eq. (3), first assume $\alpha_i Q_I = \alpha_I Q_i$. If the true motion $\vec{v^*}$ around this neighborhood is far way from zero motion, then $\mathit{SAD}(B_j,\vec{v^*}) \ll \mathit{SAD}(B_j,\vec{0})$. Eq. (10) will be

$$\min\left\{\mathit{VAR}(B_i), \mathit{SAD}(B_i,\vec{v^*}) - \mu\sum_{B_j\in\mathcal{N}(B_i)}\left(\mathit{SAD}(B_j,\vec{0}) - \mathit{SAD}(B_j,\vec{v^*})\right)\right\} \tag{11}$$

which is very similar to Eq. (3).

Slow Motion in the Neighborhood

On the other hand, if the true motion $\vec{v^*}$ around this neighborhood is static, then $\mathit{SAD}(B_j,\vec{v^*}) \approx \mathit{SAD}(B_j,\vec{0})$. Eq. (10) will be

$$\min\left\{\frac{\alpha_I Q_i}{\alpha_i Q_I}\mathit{VAR}(B_i), \mathit{SAD}(B_i,\vec{v^*})\right\} \tag{12}$$

which provides a different mode decision from Eq. (3).

INTRA Blocks in the Neighborhood

If $VAR(B_j) \ll SAD(B_j, \vec{v_j^*})$, then it is likely that block B_j is coded in the INTRA mode. There should be less penalty to code the current block B_i in the INTRA mode. Therefore, we change Eq. (10) to the following:

$$\min \left\{ \frac{\alpha_I Q_i}{\alpha_i Q_I}(VAR(B_i)) + \mu \sum_{B_j \in \mathcal{N}(B_i)} \min\{SAD(B_j, \vec{0}), VAR(B_j)\}, \right. \\ \left. \min_{\vec{v}}\{SAD(B_i, \vec{v}) + \mu \sum_{B_j \in \mathcal{N}(B_i)} SAD(B_j, \vec{v})\} \right\} \quad (13)$$

4 Simulation

Because Eq. (13) assumes the neighboring blocks will move in the same translational motion, it may not adequately model rotational motion. Since the neighboring blocks may not have uniform motion vectors, a further relaxation on the neighboring motion vectors is introduced [5]:

$$\min \left\{ \frac{\alpha_I Q_i}{\alpha_i Q_I}(VAR(B_i)) + \mu \sum_{B_j \in \mathcal{N}(B_i)} \min\{SAD(B_j, \vec{\delta}), VAR(B_j)\}, \right. \\ \left. \min_{\vec{v}}\{SAD(B_i, \vec{v}) + \mu \sum_{B_j \in \mathcal{N}(B_i)} SAD(B_j, \vec{v} + \vec{\delta})\} \right\} \quad (14)$$

where a **small** $\vec{\delta}$ is incorporated to allow local variations of motion vectors among neighboring blocks due to the non-translational motions.

In the frame-level rate control of the current VM, the quantization parameter (QP) is applied for the inter-frame (P) regardless of its mode, i.e., $Q_I = Q_i$ for simplicity. In our simulation, we further assume that $\alpha_i Q_I = \alpha_I Q_i$. The criterion will be

$$\min \left\{ VAR(B_i) + \mu \sum_{B_j \in \mathcal{N}(B_i)} \min\{SAD(B_j, \vec{\delta}), VAR(B_j)\}, \right. \\ \left. \min_{\vec{v}}\{SAD(B_i, \vec{v}) + \mu \sum_{B_j \in \mathcal{N}(B_i)} SAD(B_j, \vec{v} + \vec{\delta})\} \right\} \quad (15)$$

Figure 3 shows the first simulation result when we incorporated the above algorithm into the MPEG-4 VM7 provided by MoMuSys. The motion fields around the dancer are fast. Therefore, an INTER block saves 10 bits in the encoding the motion vectors compared with the INTRA block. That is, it supports our argument that an INTER block is more preferable when (1) there is no significant difference between the *VAR* and the *SAD*, and (2) the neighborhood is moving fast.

As shown in Figure 3(d), the rate-distortion curve of encoding the 8th frame using our method is slightly better than the rate-distortion curve using the original method. It is hard to distinguish since there are only 10 bits saving for the whole frame. However, the collective saving over a sequence of 300 frames could be significant. Table 1 shows the simulation results on some standard low bit-rate test conditions[1]. In the table, the significant difference in SNR (greater than 0.05 dB) are highlighted. On the average, our method indeed provides better SNR while using the same amount of bits.

[1]Note that the PSRN takes account of the skipped frames, which will be inserted at the decoder so that decoder can display the video at the original frame rate. For instance, we encode 100 frames at 10 fps for a sequence length of 300 frames with original frame rate of 30 fps. We include the full 300 frames in the calculation of PSNR. (That is, 100 encoded frames and the 200 skipped frames.)

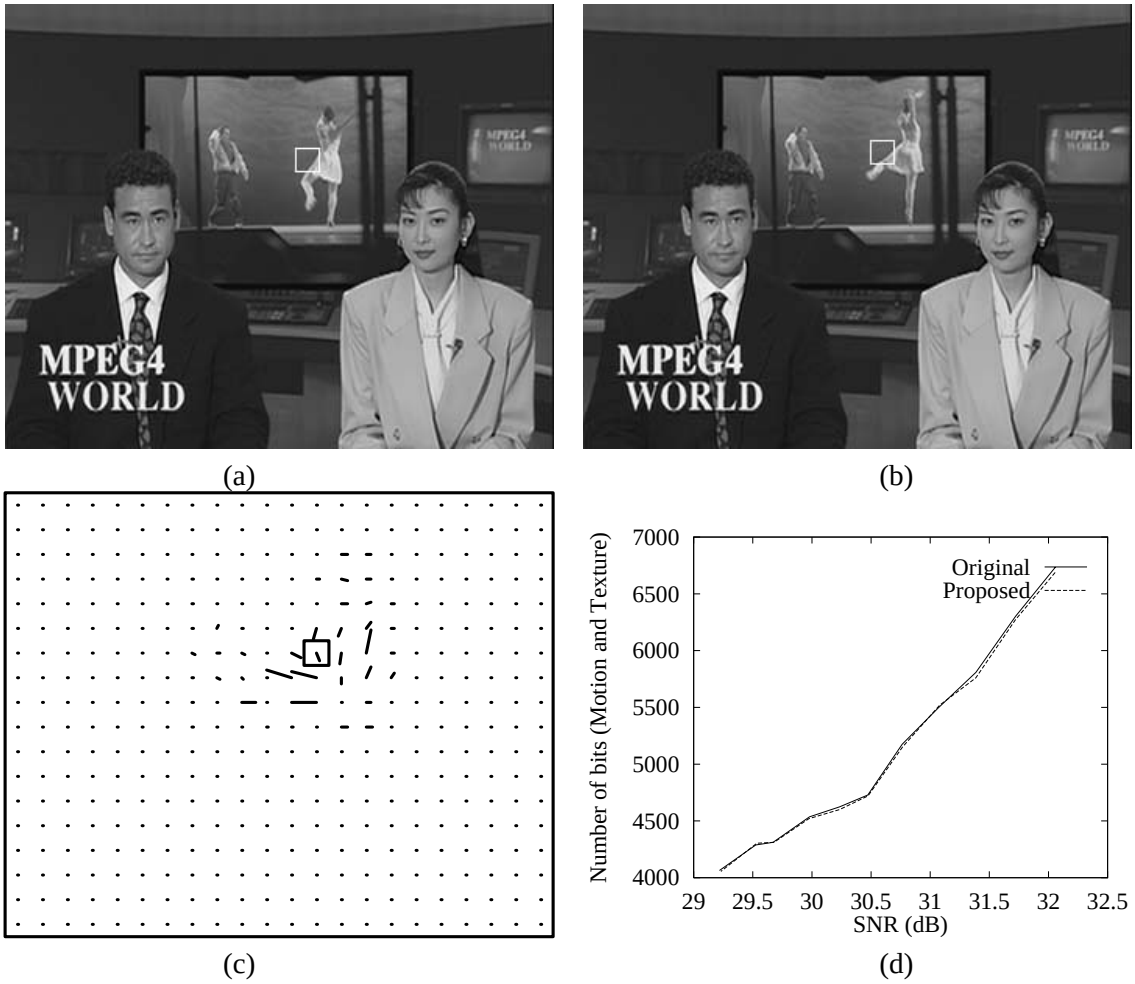

Figure 3: (a)(b) shows the 4th frame and the 8th frame of the "news" sequence. (c) shows the motion vectors found by our true motion estimation which is based on a neighborhood relaxation method [4]. In the original mode decision, the marked block on the dancer's leg is coded with INTRA. However, the number of bits for motion vectors is reduced when the INTER mode is used. (d) shows the rate-distortion curves of encoding the 8th frame.

5 Conclusions

In this paper, a new criterion for INTRA/INTER mode selection has been proposed. This method is unique in that it incorporates the true motion tracking algorithm [4] to better estimate the costs involved in choosing one mode over another. The simulation results verify that the proposed criterion improves the quality of the decoded image sequences. However, since we have only considered one of the many possible decisions which the encoder is faced with, large margins of improvement are not expected. This work will serve as a foundation though so that other mode decisions can be analyzed within the context of true motion.

References

[1] "MPEG-4 video verification model V7.0 ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG97/N1642," Apr. 1997. Bristol, UK.

[2] F. Chen, J. D. Villasenor, and D. S. Park, "A Low-Complexity Rate-Distortion Model for Motion Estimation in H.263," in *Proc. of ICIP'96*, vol. II, pp. 517–520, Sept. 1996.

Sequence ID	Target Bit Rate (kps)	Frame Rate (Hz)	Format	PSNR (dB)		Actual Bit Rate	
				VM7	Ours	VM7	Ours
Container	10	7.5	QCIF	28.71	**28.79**	10.0	10.0
Hall monitor	10	7.5	QCIF	28.51	28.50	10.0	10.0
Mother&Daughter	10	7.5	QCIF	30.88	**31.07**	10.1	10.0
Container	24	10	QCIF	31.75	31.76	24.1	24.1
Hall monitor	24	10	QCIF	31.99	**32.04**	24.0	24.0
Mother&Daughter	24	10	QCIF	33.85	33.86	24.0	24.1
Coastguard	48	10	QCIF	26.51	26.49	48.0	48.0
Foreman	48	10	QCIF	27.04	27.04	48.1	48.0
News	48	7.5	CIF	27.87	27.88	48.1	48.0
Coastguard	112	15	CIF	25.03	**25.16**	111.3	111.3
Foreman	112	15	CIF	25.93	**25.99**	111.3	111.3
News	112	15	CIF	31.77	31.78	111.2	111.3

Table 1: Side-by-side comparison of the coding efficiency by the original mode decision and our proposed mode decision criterion with rate-optimized motion estimation algorithm. We use the following options: (1) motion vector search range: -16.0 to +15.5, (2) advanced prediction mode and unrestricted motion mode, (3) no shape coding (rectangular) and no sprite coding, (4) VM5 rate control, (5) H.263 quantization, (6) ac/dc prediction, and (7) separate motion & texture coding. On the average, our method provides better SNR while using the same amount of bits.

[3] M. C. Chen and A. N. Willson, Jr., "Rate-Distortion Optimal Motion Estimation Algorithm for Video Coding," in *Proc. of ICASSP'96*, vol. IV, pp. 2098–2111, May 1996.

[4] Y.-K. Chen and S. Y. Kung, "Rate Optimization by True Motion Estimation," in *Proc. of IEEE Workshop on Multimedia Signal Processing*, pp. 187–194, June 1997.

[5] Y.-K. Chen, Y.-T. Lin, and S. Y. Kung, "A Feature Tracking Algorithm Using Neighborhood Relaxation with Multi-Candidate Pre-Screening," in *Proc. of ICIP'96*, vol. II, pp. 513–516, Sept. 1996.

[6] H. Sun, W. Kwok, M. Chien, and C.-H. J. Ju, "MPEG Coding Performance Improvement by Jointly Optimizing Coding Mode Decisions and Rate Control," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 7, no. 3, pp. 449–458, June 1997.