Journal of Electronic Imaging 7(1), 145–154 (January 1998).

Rate constrained block matching algorithm for video coding*

Ulug Bayazit
Toshiba Advanced Television Technology Center
202 Carnegie Center, Suite 102
Princeton, New Jersey 08540
E-mail: bayazit@toshiba.com

William A. Pearlman
Rensselaer Polytechnic Institute
Electrical, Computer and Systems Engineering Department
Troy, New York 12180-3590

Abstract. *The rate constrained block matching algorithm (RCBMA) jointly minimizes displaced frame difference (DFD) variance and entropy, or conditional entropy of motion vectors for determining the motion vectors. It is intended for use in low rate video coding applications, where the contribution of the motion vector rate to the overall coding rate might be significant. The DFD variance versus motion vector rate performance of RCBMA employing size $K \times K$ blocks is shown to be superior to that of the conventional minimum distortion block matching algorithm (MDBMA) employing size $2K \times 2K$ blocks. Constraining of the entropy or conditional entropy of motion vectors in RCBMA results in smoother and more organized motion vector fields than those output by MDBMA. The motion vector rate of RCBMA can also be precisely controlled for each frame by adjusting a single parameter.* © *1998 SPIE and IS&T.*
[S1017-9909(98)02301-0]

1 Introduction

In motion estimation for video coding, the widely recognized minimum distortion block matching algorithm (MDBMA) minimizes the displaced frame difference (DFD) variance between two frames. Consider a temporally predicted target frame t and a reference frame r in a video sequence, as illustrated in Fig. 1. Let a block of size $K \times K$ with upper left corner P in frame t be denoted by $B_t(P)$, and the vectorized intensity values of the pixels in it be denoted by $X_t(P)$. The upper left corners of the blocks in frame t are at the vertices of a uniform grid, $V_{(m,n)} = (mKnK)^{\dagger}$, where † indicates a transpose operation. The index combination (m,n) assumes value in a finite set $(m,n) \in \mathcal{G}$. The motion model for MDBMA assumes uniform, translational motion of rigid objects rather than rotational motion, camera zooming, or occlusion effects, so that the same motion vector $\delta_{(m,n)}$ is assigned to all pixels within a particular block in frame t, $\delta(P) = \delta_{(m,n)}$ if $P \in B_t(V_{(m,n)})$. In exhaustive search MDBMA the motion vector $\delta_{(m,n)}$ minimizes the cost function $C^{MD}_{(m,n)}(\xi)$ among all the candidate motion vectors ξ in the search area $\mathcal{S}_0$,

$$C^{MD}_{(m,n)}(\xi) = d(X_t(V_{(m,n)}), X_r(V_{(m,n)} - \xi)) \tag{1}$$

$$\delta_{(m,n)} = \arg \min_{\xi \in \mathcal{S}_0} C^{MD}_{(m,n)}(\xi). \tag{2}$$

Without loss of generality, the search area is taken to be a square $\mathcal{S}_0$ of side length $2a$ concentric with the block $B_t(V_{m,n})$ and the distortion metric, $d(.,.)$, is induced by the Euclidean norm. In Eq. (2) each K^2 dimensional vector $X_t(V_{(m,n)})$, made up of the pixel intensity values in a size $K \times K$ block in frame t, is matched with the K^2 dimensional vector $X_r(V_{(m,n)} - \delta_{(m,n)})$, made up of the pixel intensity values in a size $K \times K$ block with upper left corner $V_{(m,n)} - \delta_{(m,n)}$ in the reference frame r. Figure 1 illustrates the relationship established by the candidate motion vector ξ between the two vectors of intensity values. Since the search area in the reference frame r is centered at the vector $\xi = 0$, the vectors in the set $\{X_r(V_{(m,n)} - \xi) : \xi \in \mathcal{S}_0\}$ are most likely highly correlated with $X_t(V_{(m,n)})$. MDBMA resembles an adaptive minimum distortion vector quantization scheme. However, in conventional vector quantization a single codebook is used to code all the vectors of intensity values (source vectors) in frame t, whereas in MDBMA a unique codebook is used to code each vector of intensity values in frame t. Let $\mathcal{S}_{+,0} = \{\xi \in \mathcal{S}_0 : p(\xi) > 0\}$, where $p(\cdot)$ is the probability mass function (pmf). MDBMA partitions the set of vectors of intensity values (source vectors) of a frame into $|\mathcal{S}_{+,0}|$ clusters, each of which is associated with a different motion vector (codevector index) with nonzero occurrence probability. Hence

*Awarded ''Best Student Paper'' at the IS&T/SPIE 1997 Visual Communications and Image Processing Conference in San Jose, California.

Paper VCP-02 received Sep. 1, 1997; accepted for publication Oct. 1, 1997.
1017-9909/98/$10.00 © 1998 SPIE and IS&T.

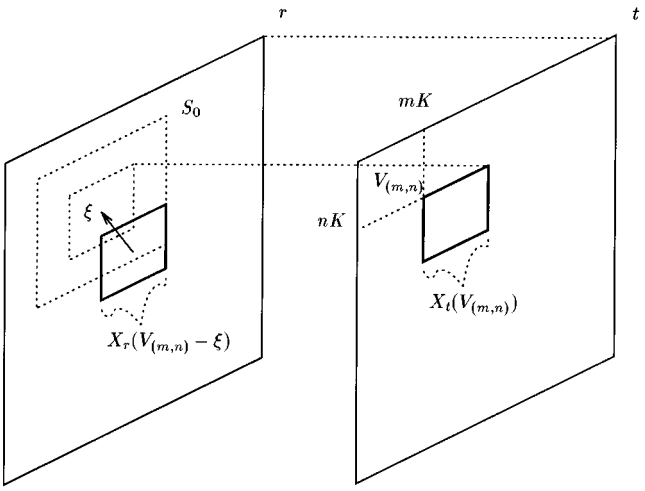


Fig. 1 Relationship between the motion vector and the vectors of intensity values (pixel blocks) in the target and reference frames.

the motion vector index for each block can be conveyed to the receiver at a rate of $\log_2|\mathscr{S}_{+,0}|$ bits per motion vector by a fixed-rate code.

A better alternative to fixed-rate coding of the motion vectors is the variable-rate entropy coding of the motion vectors, in which case the motion vectors $\{\xi:\xi\in\mathscr{S}_{+,0}\}$ are assigned variable length entropy codewords with a shorter expected length. One might assume that the motion present in each frame is concentrated only in certain directions and magnitudes without exhausting all possibilities (i.e., $|\mathscr{S}_{+,0}|\ll|\mathscr{S}_0|$). Under this assumption the entropy codeword lengths or probabilities can be transmitted on a frame-by-frame basis at a negligible overhead rate for frame adaptive entropy coding of motion vectors. Considerably low motion vector rates can be achieved by frame adaptive entropy coding of motion vectors estimated by MDBMA. However, MDBMA imposes no constraints on the entropy contribution of individual motion vectors, as they are determined by the minimum-distortion search. A candidate motion vector may be chosen over another candidate motion vector with significantly less contribution to entropy and slightly more contribution to distortion. As a result, the generated motion vector field is not smooth and contains numerous spurious motion vectors.

1.1 *Rate Constrained Block Matching Algorithm*

Entropy coding of motion vectors can yield even better motion vector compression performance if the motion vectors generated by MDBMA were not so noisy and so discontinuous at the boundaries of moving objects. In this paper the discontinuity problem is addressed by partitioning the set of motion vectors of a frame into two classes. For the class of *predictable* motion vectors, which are highly correlated with their neighbors, the spatial prediction error vectors of motion vectors are entropy coded (or in a restricted sense, the motion vectors are conditional entropy coded). The class of *unpredictable* motion vectors are simply entropy coded. (Note that the names given to the classes may not be truly representative of all the motion vectors and are merely used to distinguish between the specific actions taken for the constituents of each class.) The motivation here is to exploit the local trends (in the form of correlation) in the motion vector field for the predictable motion vectors and global trends in the motion vector field for the unpredictable motion vectors. The minimized cost function for each block incorporates either an entropy or a conditional entropy constraint term. Imposing a constraint on entropy or conditional entropy helps reduce the disorderliness and noise in the motion vector field. In many respects the rate constrained block matching algorithm (RCBMA) shares similarities with entropy constrained vector quantization (ECVQ),[1] and conditional entropy constrained vector quantization (CECVQ)[2] algorithms.

Entropy or conditional entropy coding requires entropy or conditional entropy decoding tables to be constructed at the receiver. The approach adopted here is the frame adaptive transmission of three first order pmfs that are used to construct these tables at the receiver. Transmission of only first order pmfs is critical for keeping the overhead rate low. The approximations used to derive these functions will be explained in the following sections.

RCBMA allows the user to control the rate allocated to the motion vectors of each frame. Ideally the distribution of the overall rate to motion vectors and DFD compression must be optimized. However, this is a difficult problem, since the coding characteristics of DFD is dependent on the coding characteristics (or rate) of the motion vectors in a not-so-easily tractable manner. Therefore, in this work, rate control is employed on a frame by frame basis and is only used for targeting a desired rate at which performance comparisons can be made with MDBMA.

1.2 *Related Approaches for Constraining the Rate of Motion Vectors*

Motion vector quantization (MVQ)[3,4] constrains the size of the index set $|\mathscr{S}_{+,0}|$ by a clustering algorithm similar to the Linda, Buzo, Gray (LBG) algorithm. The motion vector fields obtained by this technique are smoother than those obtained by MDBMA. Yet, a size constraint on the motion vector set is equivalent to a fixed-rate constraint and does not ensure a distinct rate-distortion advantage over MDBMA when the motion vectors are entropy coded. Macro motion vector quantization (MMVQ)[3] extends the MVQ approach. The correlations between motion vectors of neighboring blocks are better exploited by constraining the size of the set of their joint occurences.

A variable-length tree-structured segmentation algorithm can be used to determine the best spatial resolution of the motion vectors for region based very low rate video coding. A similar idea has also been employed for variable block size motion estimation by variable length quadtree structures.[6] In both of these approaches the generated variable length tree structures are rate constrained, reminiscent of variable length tree structured vector quantization codebooks.[7,8] The rate includes the contribution due to the compression of DFD, and correspondingly, distortion is the variance of the quantization error of DFD. Although the variable length tree structures are rate constrained, the process employed to map a block to a node of the tree attempts to minimize only distortion.

Explicit rate constraints have previously been incorporated into the cost function of block matching.[9,10,11] In Stiller and Lappe,[9] the cost function minimized is heuristi-

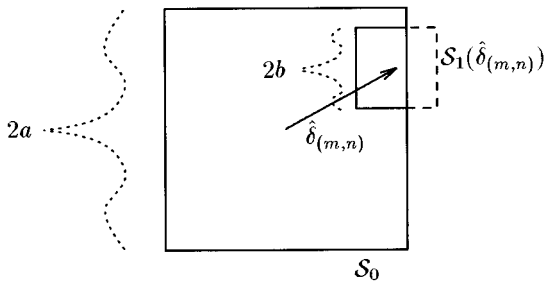


Fig. 2 Spatial relationship between the two search areas.

cally derived and is not optimal in the rate-distortion sense. The cost functions used in Chung, Kossentini, and Smith[10] and Hoang, Long, and Vitter[11] incorporate rate constraints similar to ours. In Chung, Kossentini, and Smith[10] only the most probable motion vectors are tested in a descending order to determine the best motion vector by comparing their cost function values against experimentally determined thresholds. The entropy constraints for estimating and the entropy codes for transmitting the motion vectors are not adapted to their occurence frequencies, which vary with the particular sequence or particular frame of the sequence coded, or with the coding rate. While this approach conforms to fixed entropy coding/decoding in the international standards such as MPEG-2 and H.263, we maintain that the adaptive transmission of first order pmfs at a low overhead rate is not only feasible, but also makes entropy constraining and coding more efficient and obviates this restriction.

2 Predictable Motion Vectors

By definition, a predictable motion vector is highly correlated with its neighbors and also with the prediction vector $\hat{\delta}_{(m,n)}$ for the motion vector. Therefore the prediction error vector for a predictable motion vector $\delta_{(m,n)}$ should lie in a small search area $\mathcal{S}_1(\underline{0})$ centered at the zero vector $\underline{0}$. Without loss of generality, the smaller search area for the prediction error vector is taken to be a square of size $2b \times 2b$ and centered at the zero vector such that $\mathcal{S}_1(\underline{0}) = [-b,b]\times[-b,b]$ with $b<a$. This definition is used consistently throughout this paper. The spatial relationship between $\mathcal{S}_0$ (the search area for a motion vector), and $\mathcal{S}_1(\hat{\delta}_{(m,n)})$ (the smaller search area for a predictable motion vector), is depicted in Fig. 2.

The overall cost functional minimized for predictable motion vectors between target and reference frames can then be written as $J_1 = D_1 + \mu R_1$. D_1 is the DFD variance of target frame blocks with predictable motion vectors. R_1 is the entropy of the spatial prediction error vectors of predictable motion vectors (or the conditional entropy of motion vectors) in $\mathcal{S}_1(\underline{0})$.

Motion vector information is conveyed to the receiver row by row, each row scanned from left to right. The neighboring motion vectors outside the nonsymmetric half plane (NSHP) support are not available to the receiver when the current motion vector is determined. The neighboring motion vectors to the left and to the top ($\delta_{(m-1,n)}$ and $\delta_{(m,n-1)}$ respectively) have the highest correlation with the motion vector of the current block (m,n). For a first order prediction, the prediction vector $\hat{\delta}_{(m,n)}$ for the current block is obtained as the MAP estimate of the current motion vector from the neighboring motion vectors to the left and to the top,

$$\hat{\delta}_{(m,n)} = \arg \max_{\xi \in \mathcal{S}_0} p(\xi | \delta_{(m-1,n)}, \delta_{(m,n-1)}). \tag{3}$$

The set of conditional probabilities $\{p(\xi|\delta_{(m-1,n)}, \delta_{(m,n-1)})\}$ for each possible pair $(\delta_{(m-1,n)}, \delta_{(m,n-1)})$ must be available at the receiver so that it can track the estimation process. This is usually not feasible with a moderately large $\mathcal{S}_0$ due to the order of the product space underlying the conditional pmf. The conditional pmf may be approximated by the product of horizontal and vertical marginals

$$\begin{aligned} p([\xi|\delta_{(m-1,n)}, \delta_{(m,n-1)}) &\simeq p_h(\xi|\delta_{(m-1,n)}) p_v(\xi|\delta_{(m,n-1)}) \\ &\simeq p_i(\xi|\delta_{(m-1,n)}) p_i(\xi|\delta_{(m,n-1)}), \end{aligned} \tag{4}$$

where the conditional pmf is further assumed to be isotropic in the second approximation. These approximations reduce the order of the product space by 1.

Once the prediction for the current motion vector is made in this manner, the conditional entropy constrained cost function for block (m,n) is written as

$$C^{CEC}_{(m,n)}(\xi) = \begin{cases} d(X_t(V_{(m,n)}), X_r(V_{(m,n)} - \xi)) - \mu \log_2 p(\xi|\hat{\delta}_{(m,n)}) \\ \qquad \text{for } \xi \in \mathcal{S}_1(\hat{\delta}_{(m,n)}) \cap \mathcal{S}_0, \\ \infty \qquad \text{for } \xi \in \mathcal{S}_0 \backslash \mathcal{S}_1(\hat{\delta}_{(m,n)}) \end{cases} \tag{5}$$

where $\mathcal{S}_1(P) = \{\tau + P : \tau \in \mathcal{S}_1(\underline{0})\}$.

This cost function incorporates the transmission cost of the predictable motion vector given by its conditional entropy codeword length $-\log_2[p(\xi|\hat{\delta}_{(m,n)})]$. Spatial prediction error vectors outside of the search area $\mathcal{S}_1(\underline{0})$ are automatically disregarded by setting the cost function to infinity. The blocks with such large spatial prediction error vectors are classified as unpredictable as is discussed in the following two sections. Also, when $\mu = 0$, it is worth noting that the above cost function reduces to that of MDBMA [Eq. (1)].

The conditional pmf $p(\xi|\zeta)$ for all $\zeta \in \mathcal{S}_0$ must also be available at the receiver for entropy coding/decoding. To keep the overhead rate low, the conditional pmf $p(\xi|\zeta)$ also governs spatial prediction by letting $p_i(\xi|\zeta) = p(\xi|\zeta)$. By Bayes's rule

$$p(\xi|\zeta) = \frac{p(\zeta|\xi)p(\xi)}{\Sigma_{\xi \in \mathcal{S}_0} p(\zeta|\xi)p(\xi)}. \tag{6}$$

The equality $p(\zeta|\xi) = p_n(\zeta - \xi)$ is valid for some first order pmf $p_n(\cdot)$ when the joint probability density function for ζ, ξ is Gaussian. Hence the conditional pmf can be approxi-

mated by the first order pmf $p(\zeta|\xi)\simeq p_n(\zeta-\xi)$, allowing us to work with spatial prediction error vectors of the form $\xi-\zeta$.

3 Unpredictable Motion Vectors

Classification of all the motion vectors as predictable leads to large prediction errors at the boundaries of moving objects or at places of nonuniform motion as a result of rotation or zooming of camera. The global information in the motion vector field may also be more important for a particular motion vector than the local information from its neighboring motion vectors. The cost functional minimized for the class of unpredictable motion vectors between target and reference frames can be written as $J_2=D_2+\mu R_2$. D_2 is the DFD variance of blocks with unpredictable motion vectors. R_2 is the rate of transmission of the unpredictable motion vectors in $\mathscr{S}_0$. The entropy constrained cost function for block (m,n) is written as

$$C^{EC}_{(m,n)}(\xi)=d(X_t(V_{(m,n)}),X_r(V_{(m,n)}-\xi))-\mu\ \log_2\ p(\xi). \tag{7}$$

This cost function incorporates the transmission cost of the unpredictable motion vector given by the entropy codeword length $-\log[p(\xi)]$.

4 Classification and Block Matching

If $p(\xi)>p(\xi|\hat{\delta}_{(m,n)})$ then $C^{EC}_{(m,n)}(\xi)<C^{CEC}_{(m,n)}(\xi)$ for $\xi\in\mathscr{S}_1(\hat{\delta}_{(m,n)})\cap\mathscr{S}_0$ follows from comparing Eq. (5) with Eq. (7). Hence the class bit of the candidate motion vector ξ for block (m,n) is set as

$$z_{(m,n)}(\xi)=\begin{cases}0 & \text{if } p(\xi)>p(\xi|\hat{\delta}_{(m,n)})\\ 1 & \text{otherwise}\end{cases}, \tag{8}$$

for $\xi\in\mathscr{S}_1(\hat{\delta}_{(m,n)})\cap\mathscr{S}_0$, and as

$$z_{(m,n)}(\xi)=0 \tag{9}$$

for $\xi\in\mathscr{S}_0\backslash\mathscr{S}_1(\hat{\delta}_{(m,n)})$. The overall cost function is defined as

$$C_{(m,n)}(\xi)=\begin{cases}C^{EC}_{(m,n)}(\xi) & \text{if } z_{(m,n)}(\xi)=0\\ C^{CEC}_{(m,n)}(\xi) & \text{otherwise}\end{cases}, \tag{10}$$

which is minimized by the motion vector $\delta_{(m,n)}$ as

$$\delta_{(m,n)}=\arg\min_{\xi\in\mathscr{S}_0} C_{(m,n)}(\xi). \tag{11}$$

The class bit map is the set of class bits for all blocks and is denoted as $\{z_{(m,n)}(\delta_{(m,n)})\}$.

4.1 *Modifications of pmfs for Entropy Coding and Decoding*

Once the set of bits $\{z_{(m,n)}(\xi):\xi\in\mathscr{S}_0\cap\mathscr{S}_1(\hat{\delta}_{(m,n)})\}$ is determined for a block with index (m,n), the estimates $\{p(\xi)\}$ and $\{p(\xi|\hat{\delta}_{(m,n)})\}$ are modified prior to entropy coding/decoding to prevent the overlap of nonzero probabilities of candidate motion vectors under different classes.

$$p'(\xi|\hat{\delta}_{(m,n)})=\begin{cases}\alpha_1^{-1}p(\xi|\hat{\delta}_{(m,n)}) & \text{if } z_{(m,n)}(\xi)=1\\ 0 & \text{otherwise}\end{cases}, \tag{12}$$

where $\alpha_1=1-\Sigma_{\{\tau:z_{(m,n)}(\tau)=0\}}p(\tau|\hat{\delta}_{(m,n)})$ and

$$p'(\xi)=\begin{cases}\alpha_2^{-1}p(\xi) & \text{if } z_{(m,n)}(\xi)=0\\ 0 & \text{otherwise}\end{cases}, \tag{13}$$

where $\alpha_2=1-\Sigma_{\{\tau:z_{(m,n)}(\tau)=1\}}p(\tau)$.

5 RCBMA Motion Estimation Algorithm

The RCBMA algorithm iteratively estimates the motion vectors $\{\delta_{(m,n)}:(m,n)\in\mathscr{G}\}$ and the sets of probabilities $\{p(\xi):\xi\in\mathscr{S}_0\}$, $\{p_n(\gamma):\gamma\in\mathscr{S}_1(\underline{0})\}$, $p_{CEC}=1-p_{EC}=Pr\{z_{(m,n)}(\delta_{(m,n)})=1\}$. The probabilities are estimated from the observed frequencies of motion vectors or their prediction error vectors, and are, in turn, used to form the rate constraint terms in the cost functions and spatial predictor at the next iteration to yield a new set of motion vectors.

The first part of each iteration consists of three stages. For block (m,n), the first stage is the prediction of $\hat{\delta}_{(m,n)}$. Several special circumstances are handled in different ways. For instance, if the two neighboring motion vectors conflict with each other (i.e., $p(\xi|\delta_{(m-1,n)})p(\xi|\delta_{(m,n-1)})=0$), then the spatial prediction vector for the current motion vector is their mean instead of the MAP estimate given by Eq. (3). In the second stage, the class bits $\{z_{(m,n)}(\xi):\xi\in\mathscr{S}_0\}$ and the overall cost function $\{C_{(m,n)}(\xi):\xi\in\mathscr{S}_0\}$ are evaluated in accordance with Eqs. (8) and (9) and Eqs. (5), (7), and (10). Then the minimum of $C_{(m,n)}(\xi)$ over all $\xi\in\mathscr{S}_0$ is determined by Eq. (11) to yield $\delta_{(m,n)}$. Rate R, distortion D, and total cost J for frame t are updated by the contributions of block (m,n) before the next block is processed. After the motion vectors for all blocks are determined in this manner, R and J are further corrected by ΔR_{ov}, the overhead rate for the transmission of p_{CEC}, $\{p(\xi):\xi\in\mathscr{S}_0\}$, $\{p_n(\gamma):\gamma\in\mathscr{S}_1(\underline{0})\}$. The computation of ΔR_{ov} will be explained in Section 7.

The second part of each iteration is the estimation of the probabilities from the observed frequencies. Let $N_\xi=|\{(m,n):\delta_{(m,n)}=\xi\}|$, $N^{CEC}_\gamma=|\{(m,n):z_{(m,n)}(\delta_{(m,n)})=1,\ \delta_{(m,n)}=\gamma+\hat{\delta}_{(m,n)}\}|$, and $N=|\mathscr{G}|=\Sigma_{\xi\in\mathscr{S}_0}N_\xi$, $N^{CEC}=\Sigma_{\gamma\in\mathscr{S}_1(\underline{0})}N^{CEC}_\gamma$, where $|\cdot|$ denotes cardinality. $\{p(\xi):\xi\in\mathscr{S}_0\}$, $\{p_n(\gamma):\gamma\in\mathscr{S}_1(\underline{0})\}$ and p_{CEC} are determined from frequencies as

$$p(\xi)=\frac{N_\xi}{N},\quad p_n(\gamma)=\frac{N^{CEC}_\gamma}{N^{CEC}},\quad p_{CEC}=\frac{N^{CEC}}{N}. \tag{14}$$

The total cost J, total distortion D, and total rate R can be expressed as

$$J = J_1 + J_2 + \mu \Delta R_{\text{ov}} = \frac{1}{K^2 N} \sum_{m,n} C_{(m,n)}(\delta_{(m,n)}) + \mu \Delta R_{\text{ov}}, \tag{15}$$

$$D = \frac{1}{K^2 N} \sum_{m,n} d(X_t(V_{(m,n)}), X_r(V_{(m,n)} - \delta_{(m,n)})), \tag{16}$$

$$\begin{aligned} R = -\frac{1}{K^2 N} \Bigg(& \sum_{\{m,n: z_{(m,n)}(\delta_{(m,n)})=1\}} \log_2 \\ & \times [p_{CEC} p(\delta_{(m,n)} | \hat{\delta}_{(m,n)})] \\ & + \sum_{\{m,n: z_{(m,n)}(\delta_{(m,n)})=0\}} \log_2 [(1 - p_{CEC}) p \delta_{(m,n)})] \Bigg) \\ & + \Delta R_{\text{ov}}, \end{aligned} \tag{17}$$

where $C_{(m,n)}(\delta_{(m,n)})$ in Eq. (15) is defined by Eqs. (10), (7), and (5).

For a given μ, J decreases for the first few iterations and either converges to or oscillates around a final value for the rest of the iterations. There is no guarantee that J will monotonically decrease with the iteration number. Therefore RCBMA is terminated after a predetermined number of iterations. Let $*$ indicate the best iteration with the smallest total cost J_*. The set of motion vectors $\{\delta^*_{(m,n)} : (m,n) \in \mathcal{G}\}$, class bit map $\{z^*_{(m,n)}(\delta^*_{(m,n)}) : (m,n) \in \mathcal{G}\}$ and the set of probabilities $\{p^*_n(\gamma) : \gamma \in \mathcal{S}_1(\underline{0})\}$, $\{p^*(\xi) : \xi \in \mathcal{S}_0\}$, p^*_{CEC} are transmitted.

The computational complexity of the algorithm can be kept low by storing $\{d(X_t(V_{(m,n)}), X_r(V_{(m,n)} - \xi)) : \xi \in \mathcal{S}_0, (m,n) \in \mathcal{G}\}$. During each iteration, the distance values can be read off from a table for the evaluation of the cost functions.

6 Rate Control Mechanism

The motion vector rate for a particular frame can be controlled to fall within a target rate interval, $(R_{t1}, R_{t2}]$, by varying the constraint parameter μ. Increasing μ usually results in a decrease in the motion vector rate (R_*) and vice versa. The way μ is varied is governed by the rate control mechanism, which is described next.

The mechanism is started with a given $\mu = \mu^1$. After each run $j-1$ of RCBMA, the constraint parameter μ^j for the current run is set equal to $\kappa \mu^{j-1}$ if the output rate of RCBMA from the previous run, R^{j-1}_*, is above the target interval $(R_{t1}, R_{t2}]$, and is set equal to μ^{j-1}/κ if R^{j-1}_* is below the target interval. κ is a constant and satisfies $\kappa > 1$. If R^{j-1}_* falls inside the target interval, the rate control mechanism is terminated after a final run of RCBMA. If R^{j-1}_* and R^{j-2}_* are on opposite sides of the target interval, μ^j is set equal to the geometric mean of μ^{j-1} and μ^{j-2}. In this case κ is reduced in magnitude. If $\kappa < 1 + \epsilon$, where ϵ is a small constant, change from μ^{j-1} to μ^j is negligible and the mechanism is terminated.

7 Computation of Overhead Rate ΔR_{ov} for Adaptive Transmission of Probabilities

A fixed-rate code is used to adaptively transmit the significant probabilities. First $\overline{s} = \max_{\{\xi: p^*(\xi)>0\}} |\xi|$ is transmitted with full precision. Next a significance map for $\{p^*(\xi) : |\xi| \leq \overline{s}\}$ is transmitted. Specifically 1 is sent if $p^*(\xi) > 0$, and 0 is sent if $p^*(\xi) = 0$ for $\xi \in \{\xi : |\xi| \leq \overline{s}\}$. Finally $\{p^*(\xi) : p^*(\xi) > 0\}$ are coded with high precision (12 bits per ξ) and transmitted. The same method is also used to transmit $\{p^*_n(\gamma)\}$ and p^*_{CEC}.

As it may be desirable for rate control, ΔR_{ov} increases as the overall rate R_* increases and decreases as R_* decreases. This is due to the fact that large μ forces the first order pmfs, $p^*(\xi)$ and $p^*_n(\gamma)$, to be concentrated at or near $\xi = 0$ and $\gamma = 0$, respectively.

Note that $\{p^*(\xi) : \xi \in \mathcal{S}_0\}$ are also transmitted in the same fashion for MDBMA. The increase in overhead rate for RCBMA over that of MDBMA is due to the additional transmission of $\{p^*_n(\gamma) : \gamma \in \mathcal{S}_1(\underline{0})\}$, which is usually small.

8 Simulations

In this section we provide performance comparisons between the RCBMA and MDBMA. All simulations are performed with an exhaustive search of the search areas at half-pixel accuracy. Search area $\mathcal{S}_0 = [-7,7] \times [-7,7]$ is used for both algorithms to allow a fair comparison. RCBMA is only tested on sequences with motion low enough to be sufficiently represented with vectors in $\mathcal{S}_0$. More challenging sequences such as ''Flower Garden,'' ''Table Tennis,'' and ''Football'' have not been coded, since an exhaustive search of a sufficiently large search area was too time consuming and/or these sequences had large areas of occluded regions or objects.

8.1 *Operational Distortion Rate Characteristics for Selected Frame Pairs*

In this section the operational distortion-rate (DFD variance versus motion vector rate) characteristics obtained by the application of the RCBMA algorithm on selected pairs of original frames from several image sequences are analyzed. The operational distortion-rate characteristics for two special cases of the RCBMA algorithm are also reported. In the first special case, all motion vectors are classified in the predictable class (by letting $z_{(m,n)}(\xi) = 1$, $\forall (m,n)$, $\forall \xi$) and are conditional entropy coded and constrained using Eq. (5). In the second special case, all motion vectors are classified in the unpredictable class and are unconditional entropy coded and constrained using Eq. (7). The classification decision making is bypassed for the special cases.

The characteristics obtained for the frame pairs Trevor001-002, Salesman000-002, and Claire000-002, by the application of the RCBMA algorithm and the two special cases, are shown in Figure 3. These characteristics have been traced using the rate control mechanism initialized with $\mu^1 = 10$, $\kappa = 1.25$, and $R_{t2} = 0$. The block size was 8×8 and the search area for the prediction error vector was $\mathcal{S}_1(\underline{0}) = [-2,2] \times [-2,2]$. Two other distortion-rate

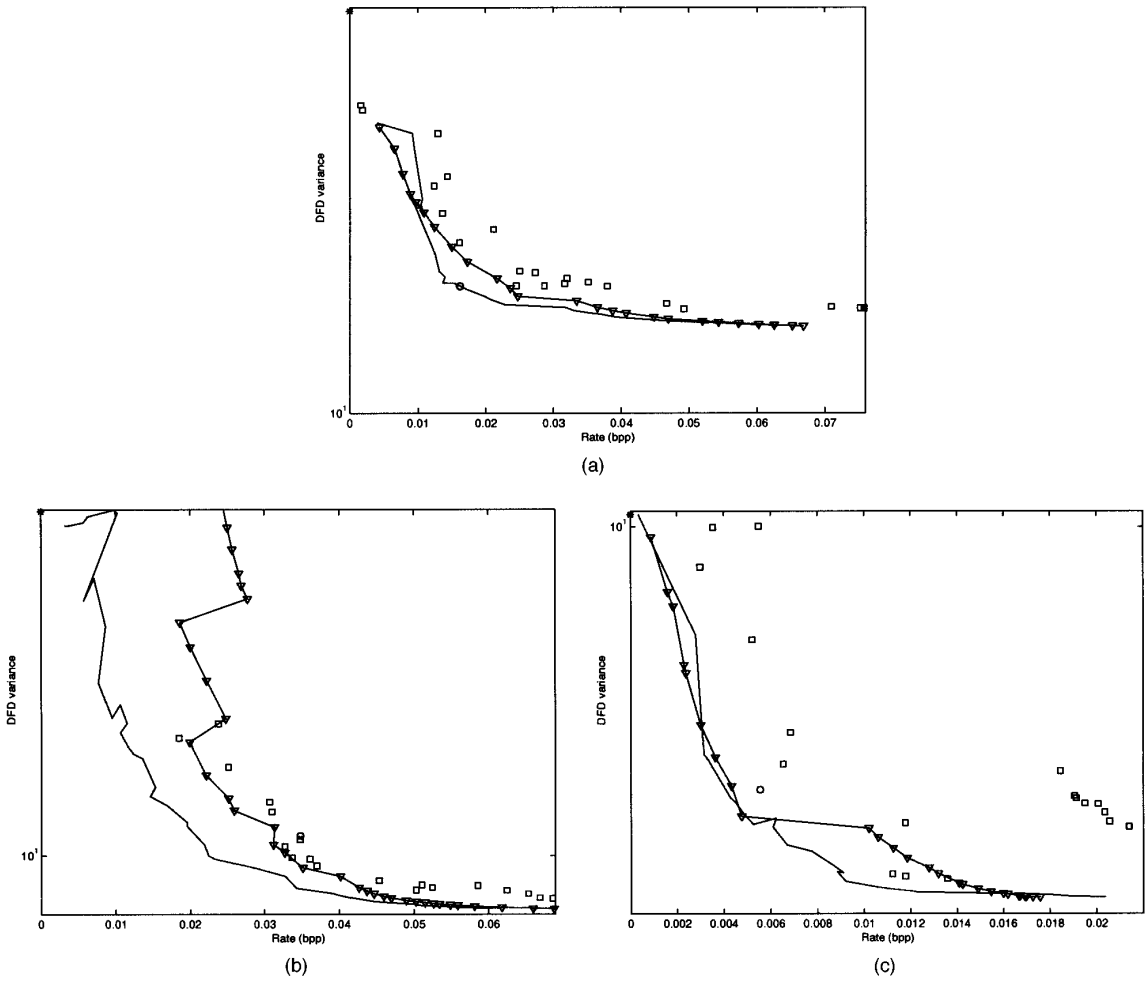


Fig. 3 Variation of (motion vector estimation) distortion with (motion vector) rate for various rate constraint scenarios. □ represents the first case; ▽ is the second case; — represents RCBMA; ∗ is the frame difference replenishment; and ○ represents MDBMA. (a) Trevor001–002; (b) Claire000–002; and (c) Salesman000–002.

points, corresponding to frame-difference replenishment (zero rate) and entropy coded MDBMA with size 16×16 blocks, are also shown in all three plots.

The curve for the two class RCBMA algorithm lies below the ones for the special one class cases, showing the importance of classification of the motion vectors as predictable or unpredictable and employing both conditional (for predictable motion vectors) and unconditional (for unpredictable motion vectors) entropy coding and constraints. It can be seen that as rate steadily decreases for the second special case, distortion gracefully increases. However, the plot for Trevor001-002 indicates that the performance of MDBMA with size 16×16 blocks may still turn out to be better, and MDBMA might be more advantageous to use due to its simplicity. For example at the same rate as MDBMA, the second special case yields an improvement over MDBMA of 0.7 dB for Claire000-002 and 0.47 dB for Salesman000-002, and is inferior to MDBMA by 0.38 dB for Trevor001-002. Exploiting only the global information in the motion vector field may not be sufficient.

If the increase in complexity is not an issue for the application, performance can be improved for the second case by exploiting the memory between some size 8×8 blocks with the two class RCBMA algorithm. For example, at the same rate as MDBMA, RCBMA yields the same DFD variance for Trevor001-002, while the PSNR gains for Claire000-002 and Salesman000-002 are 1.39 dB and 0.47 dB, respectively.

On the other hand, the first special case employing only conditional entropy coding and constraint leads to unacceptably poor performance, and the rate and distortion are not tractable by the adjustment of μ. Even the convex hull of the distortion-rate pairs for the first special case lies above the other two characteristics for the three selected frame pairs.

8.2 *Video Coding Simulations*

In this section results are presented and summarized for the motion estimation/compensation and subsequent compression of the DFD frames of several video sequences. Important parameters about the simulations are summarized in Table 1. Let the motion vector rate output by MDBMA be R_T. The rate control mechanism has been operated with

Table 1 Video coding simulations and parameters.

Sim. No.	Sequence Name	Frame Dim.	Frame Freq.	Block size MDBMA	Block size RCBMA	$\mathcal{S}_1(0)$ size for RCBMA	SPIHT Rate (bpp.)
1	Claire	352×288	15Hz	16×16	8×8	9×9	0.025
2	Missa	352×288	15Hz	16×16	8×8	9×9	0.025
3	Salesman	352×288	15Hz	16×16	8×8	9×9	0.040
4	Caltrain	512×400	30Hz	16×16	8×8	9×9	0.040
5	Caltrain	512×400	30Hz	8× 8	4×4	5×5	0.040
6	Susie	352×240	30Hz	16×16	8×8	9×9	0.040
7	Susie	352×240	30Hz	8× 8	4×4	5×5	0.040
8	Trevor	256×256	30Hz	16×16	8×8	9×9	0.040

$R_{t1}=0.95*R_T$, and $R_{t2}=1.05*R_T$ for Simulations 3 and 6, and with $R_{t1}=0.9*R_T$ and $R_{t2}=1.0*R_T$ for the other simulations. Since only the memory between adjacent blocks is exploited by spatial prediction, RCBMA block dimensions are half of those of MDBMA to allow a fair comparison.

In motion compensated predictive video coding DFD frames have to be compressed and coded with reasonable efficiency so that the reconstruction quality does not degrade throughout the sequence. Ideally the technique used must take full advantage of the roughness of the DFD spectrum. In this work, set partitioning in hierarchical trees (SPIHT) coding method,[12] which efficiently allocates bits to the subbands of a low-pass spectrum and exploits the dependencies between the subbands, has been used to code DFD.

The output bit stream of the SPIHT coder, the unpredictable motion vectors, the prediction error vectors of predictable motion vectors, and the class bits are all adaptive arithmetic coded. The details of arithmetic coding of DFD compressed with SPIHT can be found in Said and Pearlman.[12] The spatial prediction vector for the motion vector and/or the class bit information yields the pmfs $p'(\xi|\hat{\delta}_{(m,n)})$, $p'(\xi)$ used for arithmetic coding/decoding of each motion vector, or its spatial prediction error vector. In this operation, the total number of bits, output by the arithmetic coder for the motion vectors of a frame, approximates the sum of the ideal codeword lengths for the predictable and unpredictable motion vectors and class bits in that frame. (The ideal codeword length of ξ is $-\log_2 p'(\xi|\hat{\delta}_{(m,n)})$ for a predictable and $-\log_2 p'(\xi)$ for an unpredictable motion vector. The ideal codeword length of class bit z is $-\log_2(Pr\{z_{(m,n)}(\delta_{(m,n)})=z\})$.) The fixed-rate coded probability estimates are also transmitted with the method outlined in Section 7.

Table 2 summarizes the average values of the PSNR and rate curves before and after the coding of DFD for each of the simulations in Table 1. Curves for two of the simulations are also plotted in Figure 4. RCBMA with size 8 ×8 blocks has a better temporal estimation performance

Table 2 Average PSNR and rate (before and after SPIHT coding of DFD) for the simulations in Table 1.

Sim. No.	Motion Est. Method	Avg. Motion Vec. Rate (bpp)	Avg. Motion Est. PSNR	Avg. Total Rate. (bpp)	Avg. Frame Reconst. PSNR
1	MDBMA	0.0445	37.2641	0.0695	38.5900
	RCBMA	0.0406	38.5674	0.0656	39.6163
2	MDBMA	0.0563	36.8680	0.0813	37.8779
	RCBMA	0.0506	37.8844	0.0756	38.6014
3	MDBMA	0.0204	33.6742	0.0604	34.9613
	RCBMA	0.0185	34.2458	0.0585	35.3424
4	MDBMA	0.0308	29.0304	0.0708	30.0852
	RCBMA	0.0282	29.5367	0.0682	30.2163
5	MDBMA	0.1310	30.7097	0.1710	31.4461
	RCBMA	0.1242	32.1570	0.1642	32.6794
6	MDBMA	0.0429	33.3571	0.0829	34.6414
	RCBMA	0.0394	33.8494	0.0794	34.7800
7	MDBMA	0.1505	34.7814	0.1905	35.8006
	RCBMA	0.1348	35.6849	0.1748	36.4429
8	MDBMA	0.0361	32.3899	0.0761	33.3252
	RCBMA	0.0308	32.7133	0.0708	33.5227

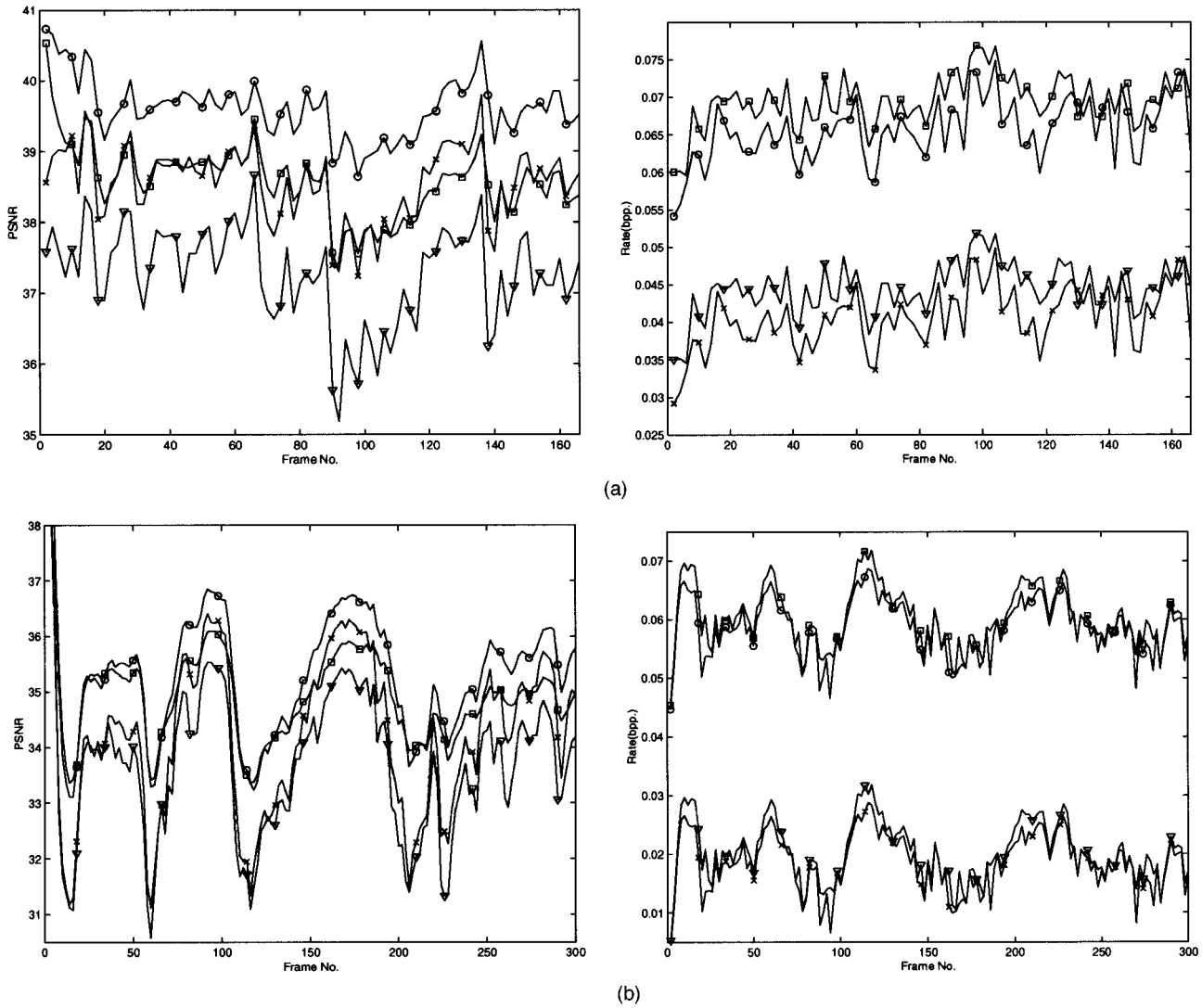
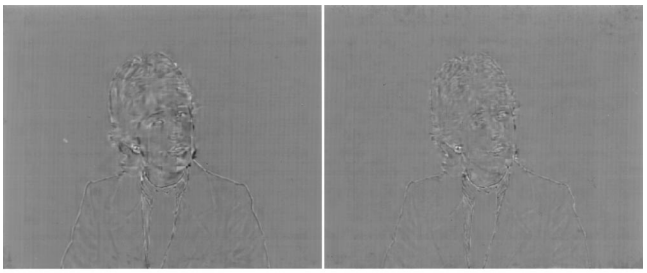


Fig. 4 Comparison of MDBMA (16×16 blocks) with RCBMA (8×8 blocks) and variation of PSNR and rate (before and after SPIHT coding) with frame number. ▽ represents MDBMA motion estimation PSNR and motion vector rate; × is RCBMA motion estimation PSNR and motion vector rate; □ represents MDBMA+SPIHT reconstructed frame PSNR and total rate; and ○ is RCBMA+SPIHT reconstructed frame PSNR and total rate. (a) Simulation 1 (''Claire'') and (b) Simulation 3 (''Salesman'').

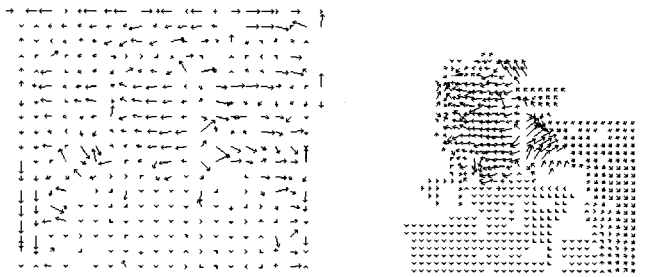
than MDBMA with size 16×16 blocks. This is largely due to the fact that both local and global information about the motion vector field are exploited. For the six simulations employing RCBMA with size 8×8 blocks, average motion estimation gains in the range of 0.32 to 1.30 dB over MDBMA have been obtained with a lower average motion vector rate than that for MDBMA. The average gains in some cases are even higher if one ignores the first few frames of each sequence when computing the averages. Simulations on sequences with more uniform motion and less occlusion such as ''Missa'' and ''Claire'' have yielded the larger gains. A comparison of the result for Simulations 5 and 7 with those for Simulations 4 and 6 indicates the appreciable increase in motion estimation advantage of RCBMA compared to MDBMA when smaller size blocks are used. However, the use of smaller size blocks may not be justifiable for MDBMA or for RCBMA if the gain in motion estimation PSNR is offset by a large increase in motion vector rate.

The quantitative performance advantage of RCBMA is also accompanied by the improvement in visual video signal quality. For example, for Simulation 1 with MDBMA reported for ''Claire,'' large blockiness and distortion on the chin and cheek areas of the woman's face was observed, which became very distracting and unpleasant between frames 90 to 100. There was also some flickering at the boundary between the arms, shoulder, and the stationary background. For Simulation 1 with RCBMA reported for ''Claire,'' only slight flickering at the chin boundary and even less flickering at the boundary between the arms, shoulder, and the stationary background was observed. For Simulation 6 reported for ''Susie,'' both motion estimation methods resulted in blockiness at the boundary of the face with the background. The size 16×16 MDBMA blocks



(a) (b)

Fig. 5 DFD frames between Claire090-092 in Simulation 1: (a) MDBMA and (b) RCBMA.



(a) (b)

Fig. 7 Motion vector fields between Claire090-092 in Simulation 1: (a) MDBMA and (b) RCBMA.

could actually be distinguished. Blockiness was less distracting for RCBMA, since the size of RCBMA blocks are a quarter of the size of the MDBMA blocks and the reconstruction PSNR was higher. For Simulation 7 reported for ''Susie,'' smooth reconstruction with very small visible granular distortion on the face was achieved with RCBMA. MDBMA yielded better visual results in Simulation 7 than in Simulation 6 due to the small size blocks, but distortion was still largely visible on the face of ''Susie.'' This became quite distracting between frames 40 to 60. For Simulation 8 reported for ''Trevor,'' both algorithms yielded large distortion in the form of blur, and the stripes of the shirt were not distinguishable in both cases. Background near the human figure boundary was more blurry, and blockiness along the left arm was more conspicuous for MDBMA.

Figure 5 shows the DFD frames and Figure 6 shows the final reconstructed frames obtained with MDBMA and with RCBMA for Claire092. DFD frame obtained with RCBMA has noticeably less energy content. Figure 7 displays the motion vector fields output by MDBMA and RCBMA for Claire092. Although RCBMA employs small size (8×8) blocks, the motion vectors for these blocks are much more organized than those for the large size (16×16) blocks of MDBMA. As a result, the motion vector fields of RCBMA have fewer spurious motion vectors than those of MDBMA. Nevertheless, RCBMA does not completely prevent some of the stationary blocks with little detail from getting assigned nonzero motion vectors.

(a) (b)

Fig. 6 Final reconstructed Claire092 in Simulation 1: (a) MDBMA and (b) RCBMA.

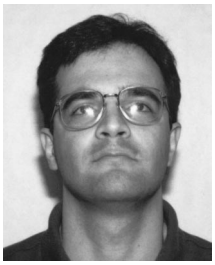
9 Conclusion

This paper has extended the minimum distortion motion vector estimation technique of MDBMA by incorporating rate constraint terms into the cost function of estimation. In RCBMA, the imposed rate constraint for a motion vector is either conditional or unconditional depending on its predictability from its neighbors. The algorithm alternatingly and iteratively estimates the probabilities (rate constraint terms) and the motion vectors, and transmits the estimated probabilities as overhead for frame adaptive entropy coding/decoding. It allows the motion vector rate to be gracefully traded off for DFD variance and either to be controlled and set at a desired level. Simulations on various sequences have shown significant visual improvement in video quality as well as rate-distortion performance with RCBMA employing size $K\times K$ blocks over MDBMA employing size $2K\times 2K$ blocks. Motion vector fields output by RCBMA are also smoother and more organized.

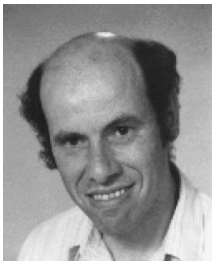
References

1. P. A. Chou, T. L. Lookabough, and R. M. Gray, ''Entropy constrained vector quantization,'' *IEEE Trans. Info. Theory* **37**(1), 31–42 (1989).
2. P. A. Chou and T. Lookabaugh, ''Conditional entropy constrained vector quantization,'' in *IEEE Intl. Conf. Acoust., Speech, Sign. Proc. (ICASSP)*, Vol. 1, 197–200 (1990).
3. Y. Y. Lee, ''Motion vector compression for digital video,'' PhD thesis, Rensselaer Polytech. Inst., Troy, NY (1994).
4. Y. Y. Lee and J. W. Woods, ''Motion vector quantization for video coding,'' *IEEE Trans. Image Processing* **4**(3), 378–382 (1995).
5. B. Girod, ''Rate-constrained motion compensation,'' *Proc. SPIE* **2308**, 1026–1034 (1994).
6. J. Lee, ''Optimal quadtree for variable block size motion estimation,'' in *Proc. IEEE Intl. Conf. on Image Proc.* **3**, 480–483 (1995).
7. M. Balakrishan, W. A. Pearlman, and L. Lu, ''Variable-rate tree structured vector quantizers,'' *IEEE Trans. Inform. Theory* **41**(7), 917–930 (1995).
8. E. A. Riskin and R. M. Gray, ''A greedy tree growing algorithm for the design of variable rate vector quantizers,'' *IEEE Trans. Acoust., Speech Sign. Proc.* **73**(11), 1551–1558 (1991).
9. C. Stiller and D. Lappe, ''Gain/cost controlled displacement estimation for image sequence coding,'' in *IEEE Intl. Conf. Acoust., Speech, Sign. Proc.*, Vol. 4, 2729–2732 (1991).
10. W. C. Chung, F. Kossentini, and M. J. T. Smith, ''Rate-distortion constrained statistical motion estimation for video coding,'' in *IEEE Intl. Conf. on Image Proc.* **3**, 184–187 (1995).
11. D. Hoang, P.M. Long, and J. S. Vitter, ''Efficient cost measures for motion compensation at low bit rates,'' in *Proc. of Data Compression Conference*, IEEE Computer Society Press, pp. 102–111 (1996).
12. A. Said and W. A. Pearlman, ''A new fast and efficient image codec based on set partitioning in hierarchical trees,'' *IEEE Trans. on Cir-*

cuits and Systems for Video Technology **6**(3), 243–250 (1996).
13. U. Bayazit and W. A. Pearlman, ''Rate-constrained block matching algorithm,'' in *Proc. SPIE* **3024**, 1110–1121 (1997).



Ulug Bayazit received his BS degree in electrical engineering in 1991 from Bosporus University, Turkey, his MS degree in electrical engineering in 1993, and MS and PhD degrees in math and electrical engineering, respectively, in 1996 from Rensselaer Polytechnic Institute, Troy, New York. His doctoral studies were supported by a scholarship from Bosporus University and research and teaching assistantships from Rensselaer Polytechnic Institute. Currently he is with the Advanced Television Technology Center of Toshiba America Consumer Products, Incorporated. His current research interests include data compression, video coding and video signal processing.



William A. Pearlman received the SB and SM degrees in electrical engineering from the Massachusetts Institute of Technology in 1963, and the PhD degree in electrical engineering from Stanford University in 1974. Between 1963 and 1974 he was employed as an engineer and consultant by Lockheed Missiles and Space Company in Sunnyvale, California, and as an engineer by GTE-Sylvania in Mountain View, California. He left industry in 1974 to become an assistant professor in the Department of Electrical and Computer Engineering at the University of Wisconsin-Madison. In 1979 he joined the electrical, computer, and systems engineering department at Rensselaer Polytechnic Institute, where he is currently a professor. During the 1985 to 1986 academic year, and again for the spring semester in 1993, he was on sabbatical leave as a visiting professor and Lady Davis Scholar in the Department of Electrical Engineering at the Technion-Israel Institute of Technology, Haifa. He also held a Visiting Professor Chair at Delft University of Technology in The Netherlands in January and February 1993, and was an IBM Visiting Scientist in the IBM-Rio Scientific Center in Rio de Janeiro, Brazil, in 1988. His research interests are in digital image, video, and audio compression, information theory, and signal and image processing. He has authored or coauthored more than 100 technical publication in these fields. He has served on the steering committee of SPIE's Visual Communications and Image Processing Conference since its inception in 1986, and in 1989 was the conference chairman. He has served on technical program committees of several IEEE conferences and on NSF program review panels. He has been Associate Editor of Coding for the IEEE *Transactions on Image Processing*. He is a Fellow of SPIE and a Senior Member of the IEEE.