A Rate-Constrained Block Matching Algorithm for Video Coding

Ulug Bayazit

Toshiba Advanced Television Technology Center

202 Carnegie Center, Suite 102

Princeton, NJ 08540

Tel: 609-951-8500

E-mail: bayazit@toshiba.com

William A. Pearlman

Electrical, Computer and Systems Engineering Department

Rensselaer Polytechnic Institute, Troy, New York 12180-3590

Tel: 518-276-6082

E-mail: pearlman@ecse.rpi.edu

January 19, 2000

Abstract

The Rate Constrained Block Matching Algorithm, (RC-BMA), introduced in this paper, 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 vs. motion vector rate performance of RC-BMA employing size KxK blocks is shown to be superior to that of the conventional Minimum Distortion Block Matching Algorithm (MD-BMA) employing size $2K$x$2K$ blocks. Constraining of the entropy or conditional entropy of motion vectors in RC-BMA results in smoother and more organized motion vector fields than those output by MD-BMA. The motion vector rate of RC-BMA can also be precisely controlled for each frame by adjusting a single parameter.

Subject terms : Motion estimation, block matching, vector quantization, rate constraint, entropy constraint, video coding

1   Introduction

In motion estimation for video coding the widely recognized Minimum Distortion Block Matching Algorithm (MD-BMA) minimizes the Displaced Frame Difference (DFD) variance between two frames. Consider two successive framesof a video sequence at times $t-1$ and t, as illustrated in Fig. 1. Let a block of size KxK with upper left corner P in a temporally predicted target 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, $V_{(m,n)} = (mK\ nK)^\dagger$, where $\dagger$ indicates a transpose operation. The index combination (m,n) assumes value in a finite set, $(m,n) \in \mathcal{G}$. The motion model for MD-BMA 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 MD-BMA the motion vector $\delta_{(m,n)}$

2

minimizes the cost function $C^{MD}_{(m,n)}(\xi)$ among all the candidate motion vectors ξ in the search area $\mathcal{S}_0$,

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

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

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 Eqn. (2) each (K^2) dimensional vector $X_t(V_{(m,n)})$, made up of the intensity values in a size KxK block in frame t, is matched with the (K^2) dimensional vector $X_r(V_{(m,n)} - \delta_{(m,n)})$, made up of the intensity values in a size KxK block with upper left corner $V_{(m,n)} - \delta_{(m,n)}$ in the reference frame $r = t - 1$. 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 $X_t(V_{(m,n)})$. MD-BMA 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 MD-BMA a unique codebook is used to code each vector of intensity values in frame t. Let $\mathcal{S}_{+,0} = \{\xi \in \mathcal{S}_0 : p(\xi) \neq 0\}$ where $p(.)$ is the probability mass function (pmf). MD-BMA partitions the set of vectors of intensity values (source vectors) of a frame into $\mid \mathcal{S}_{+,0} \mid$ clusters each of which is associated with a different motion vector (codevector index) with nonzero occurrence probability. Hence the motion vector index for each block can be conveyed to the receiver at a rate of $\log_2 \mid \mathcal{S}_{+,0} \mid$ 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 $\{\xi : \xi \in \mathcal{S}_{+,0}\}$ 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.

$\mid \mathcal{S}_{+,0} \mid \ll \mid \mathcal{S}_0 \mid$). 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 MD-BMA. However, MD-BMA 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 a significantly less contribution to entropy and a 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 BMA (RC-BMA)

Entropy coding of motion vectors can yield even better motion vector compression performance were the motion vectors generated by MD-BMA 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*[1] 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. 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 RC-BMA algorithm shares similarities with Entropy Constrained VQ, (ECVQ), ([1]) and Conditional Entropy Constrained VQ, (CECVQ), ([2]) algorithms.

---

[1]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.

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 which 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.

RC-BMA 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 only used for targeting a desired rate at which performance comparisons can be made with MD-BMA.

1.2   Related approaches for constraining the rate of motion vectors

Motion Vector Quantization, (MVQ), is a recently developed algorithm in [3, 4], and constrains the size of the index set, $\mid \mathcal{S}_{+,0} \mid$, by a clustering algorithm similar to LBG. The motion vector fields obtained by this technique are smoother than those obtained by MD-BMA. 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 MD-BMA when the motion vectors are entropy coded. Macro Motion Vector Quantization, (MMVQ), is also developed in [3], and 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 is used in [5] to determine the best spatial resolution of the motion vectors for region based very low rate video coding. A similar idea has been employed for variable block size motion estimation by variable-length quadtree structures in [6]. In both of these approaches the generated variable-length tree structures are rate-constrained, reminiscent of variable-

length tree-structured codebooks of [7, 8]. The rate, reported in [5, 6], 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.

In [9, 10, 11] explicit rate-constraints have been incorporated into the cost function of block matching. In [9], the cost function minimized is heuristically derived and is not optimal in the rate-distortion sense. The cost functions used in [10, 11] incorporate rate-constraints similar to ours. In [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. In both [10] and [11], 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 and H.263, we maintain that the adaptive transmission of first order pmf's 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 $2b$x$2b$ and centered at the zero vector such that $\mathcal{S}_1(\underline{0}) = [-b, b]x[-b, b]$ 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 Figure 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 NSHP (Non-symmetric Half Plane) 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.

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

The set of conditional probabilities $\{p(\xi \mid \delta_{(m-1,n)}, \delta_{(m,n-1)})\}$ 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

$$\begin{aligned} p(\xi \mid \delta_{(m-1,n)}, \delta_{(m,n-1)}) &\simeq p_h(\xi \mid \delta_{(m-1,n)}) p_v(\xi \mid \delta_{(m,n-1)}) \\ &\simeq p_*(\xi \mid \delta_{(m-1,n)}) p_*(\xi \mid \delta_{(m,n-1)}) \end{aligned} \tag{4}$$

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

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

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

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

The above cost function incorporates the transmission cost of the predictable motion vector given by its conditional entropy codeword length $-\log_2[p(\xi \mid \hat{\delta}_{(m,n)})]$. 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 shall be discussed in the following two sections. Also, when $\mu = 0$, it is worth noting that the above cost function reduces to that of MD-BMA (Eqn. (1)).

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

$$
p(\xi \mid \zeta) = \frac{p(\zeta \mid \xi)p(\xi)}{\sum_{\xi \in \mathcal{S}_0} p(\zeta \mid \xi)p(\xi)} \tag{6}
$$

The equality $p(\zeta \mid \xi) \simeq p_n(\zeta - \xi)$ is valid for some first order pmf $p_n(.)$ when the joint probability density function for ζ, ξ is Gaussian. Hence the conditional pmf can be approximated by the first order pmf $p_n(\zeta - \xi)$ 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 carry more importance 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 $\mathcal{S}_0$. The entropy constrained cost function for block (m, n) is written as

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

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

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

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

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

for $\xi \in \mathcal{S}_0 \setminus \mathcal{S}_1(\hat{\delta}_{(m,n)})$. 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

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

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

4.1 Modifications of pmf's for entropy coding/decoding

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

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

where $\alpha_1 = 1 - \sum_{\{\tau : z_{(m,n)}(\tau) = 0\}} p(\tau \mid \hat{\delta}_{(m,n)})$ 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 $\alpha_2 = 1 - \sum_{\{\tau : z_{(m,n)}(\tau) = 1\}} p(\tau)$.

5 RC-BMA motion estimation algorithm

The RC-BMA algorithm iteratively estimates the motion vectors $\{\delta_{(m,n)} : (m,n) \in \mathcal{G}\}$ and the sets of probabilities $\{p_n(\gamma) : \gamma \in \mathcal{S}_1(\underline{0})\}$, $\{p(\xi) : \xi \in \mathcal{S}_0\}$ and $p_{CEC} = 1 - p_{EC} = Pr\{\mathbf{z} = 1\}$. 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. $(p(\xi \mid \delta_{(m-1,n)})p(\xi \mid \delta_{(m,n-1)}) = 0)$), then the spatial prediction vector for the current motion vector is their mean instead of the MAP estimate given by Eqn. (3). In the second stage the class bits $\{z_{(m,n)}(\xi) : \xi \in \mathcal{S}_0\}$ and the overall cost function $\{C_{(m,n)}(\xi) : \xi \in \mathcal{S}_0\}$ are evaluated in accordance with Eqn. (8), (9) and (10). Then the minimum of $C_{(m,n)}(\xi)$ over all $\xi \in \mathcal{S}_0$ is determined 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}, $\{p(\xi) : \xi \in \mathcal{S}_0\}$, $\{p_n(\gamma) : \gamma \in \mathcal{S}_1(\underline{0})\}$, 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 $N_\xi =\mid \{(m,n) : \delta_{(m,n)} = \xi\} \mid$, $N_\gamma^{CEC} =\mid \{(m,n) : z_{(m,n)}(\delta_{(m,n)}) = 1, \delta_{(m,n)} = \gamma + \hat{\delta}_{(m,n)}\} \mid$, and $N =\mid \mathcal{G} \mid= \sum_{\xi \in \mathcal{S}_0} N_\xi$, $N^{CEC} = \sum_{\gamma \in \mathcal{S}_1(\underline{0})} N_\gamma^{CEC}$, where $\mid . \mid$ denotes cardinality. $\{p(\xi) : \xi \in \mathcal{S}_0\}$, $\{p_n(\gamma) : \gamma \in \mathcal{S}_1(\underline{0})\}$ and p_{CEC} are determined from frequencies as

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

11

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

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

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

$$R = -\frac{1}{K^2 N} \Big(\sum_{\{m,n:z_{(m,n)}(\delta_{(m,n)})=1\}} \log_2 [p_{CEC} p(\delta_{(m,n)} \mid \hat{\delta}_{(m,n)})]$$

$$+ \sum_{\{m,n:z_{(m,n)}(\delta_{(m,n)})=0\}} \log_2 [(1 - p_{CEC}) p(\delta_{(m,n)})] \Big) + \Delta R_{ov} \tag{17}$$

where $C_{(m,n)}(\delta_{(m,n)})$ in Eqn. (15) is defined by Eqns. (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 RC-BMA algorithm 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 frame t 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 below.

The mechanism is started with a given $\mu = \mu^1$. After each run $j - 1$ of RC-BMA the constraint

12

parameter μ^j for the current run is set equal to $\kappa\mu^{j-1}$ if the output rate of RC-BMA 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 RC-BMA. 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 $\bar{s} = \max_{\{\xi : p^*(\xi) > 0\}} \mid \xi \mid$ is transmitted with full precision. Next a significance map for $\{p^*(\xi) :\mid \xi \mid \leq \bar{s}\}$ is transmitted. Specifically 1 is sent if $p^*(\xi) > 0$ and 0 is sent if $p^*(\xi) = 0$ for $\xi \in \{\xi :\mid \xi \mid \leq \bar{s}\}$. 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 MD-BMA. The increase in overhead rate for RC-BMA over that of MD-BMA 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 RC-BMA and MD-BMA. All simulations are performed with exhaustive search of the search areas at half-pixel accuracy. Search region $\mathcal{S}_0 = [-7, 7] \mathrm{x} [-7, 7]$ is used for both algorithms to allow a fair comparison. RC-BMA 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, Football have not been coded since 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 vs. Motion Vector Rate) characteristics obtained by the application of the RC-BMA algorithm on selected pairs of original frames from several image sequences are analyzed. The operational distortion-rate characteristics for two special cases of the RC-BMA algorithm are also reported. In Special Case I, all motion vectors are classified in the predictable class (by letting $z_{(m,n)} = 1$, $\forall (m, n)$) and are conditional entropy coded and constrained using Eqn. (5). In Special Case II, all motion vectors are classified in the unpredictable class and are unconditional entropy coded and constrained using Eqn. (7). The classification decision making is bypassed for the special cases.

The characteristics obtained for the frame pairs Trevor 001-002, Salesman 000-002 and Claire 000-002, by the application of the RC-BMA 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 8x8 and the search area for the prediction error vector was $\mathcal{S}_1(\underline{0}) = [-2, 2] \mathrm{x} [-2, 2]$. Two other distortion-rate points, corresponding to zero rate frame-difference variance and entropy coded MD-BMA with size 16x16 blocks, are also shown in all three plots.

The curve for the two-class RC-BMA algorithm lies below the ones for the special one-class cases

14

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 Special Case II distortion gracefully increases. However, the plot for Trevor 000-001 indicates that the performance of MD-BMA with size 16x16 blocks may still turn out to be better and MD-BMA might be more advantageous to use due to its simplicity. For example at the same rate as MD-BMA, Special Case II yields an improvement over MD-BMA of 0.7dB for Claire000-002 and 0.47dB for Salesman000-002 and is inferior to MD-BMA by 0.38dB for Trevor000-001. 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 over Special Case II by exploiting the memory between some of size 8x8 blocks with the two-class RC-BMA algorithm. For example, at the same rate as MD-BMA, RC-BMA yields the same DFD variance for Trevor000-001, while the PSNR gains for Claire000-002 and Salesman000-002 are 1.39dB and 0.47dB respectively.

On the other hand, Special Case I 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 Special Case I 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 MD-BMA be R_T. The rate control mechanism has been operated with $R_{t1} = 0.95 * R_T$ and $R_{t2} = 1.05 * R_T$ for Simulations No. 3

15

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 are exploited by spatial prediction, RC-BMA block dimensions are half of those of MD-BMA 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 class bits are all adaptive arithmetic coded. The details of arithmetic coding of DFD compressed with SPIHT can be found in [12]. The spatial prediction vector for the motion vector and/or the class bit information yield(s) the pmfs $p'(\xi \mid \hat{\delta}_{(m,n)})$, $p'(\xi)$ used for arithmetic coding/decoding of each motion vector, or its spatial prediction error vector. In this operation, the length of the bit stream, output by the arithmetic coder for a frame, approximates the sum of the ideal codeword lengths for the motion vectors and motion prediction error vectors in that frame [2] for an unpredictable motion vector. 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 simulations are also plotted in Figure 4. RC-BMA with size 8x8 blocks has a better temporal estimation performance than MD-BMA with size 16x16 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 RC-BMA with size 8x8 blocks, average motion estimation

---

[2]Ideal codeword length of ξ is $-\log_2 p'(\xi \mid \hat{\delta}_{(m,n)})$ for a predictable and $-\log_2 p'(\xi)$

gains in the range of $0.32 - 1.30$ dB over MD-BMA have been obtained with a lower average motion vector rate than that for MD-BMA. 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 curves for the Simulations No. 5 and 7 with those for the Simulations No. 4 and 6 shows the appreciable increase in motion estimation advantage of RC-BMA over MD-BMA when smaller size blocks are used. However, the use of smaller size blocks may not be justifiable for MD-BMA or for RC-BMA if the gain in motion estimation PSNR is offset by a large increase in motion vector rate.

The quantitative performance advantage of RC-BMA is also accompanied by the improvement in visual video signal quality. For example for Simulation No. 1 with MD-BMA 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 $90 - 100$. There was also some flickering at the boundary between the arms, shoulder and the stationary background. For RC-BMA, 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 No. 6 reported for Susie, both motion estimation methods resulted in blockiness at the boundary of the face with the background. The size 16x16 MD-BMA blocks could actually be distinguished. Blockiness was less distracting for RC-BMA since the size of RC-BMA blocks are a quarter of the size of the MD-BMA blocks and reconstruction PSNR was higher. For Simulation No. 7 reported for Susie smooth reconstruction with very small visible granular distortion on the face was achieved with RC-BMA. MD-BMA yielded better visual results in Simulation No. 7 than No. 6 due to the small size blocks, but distortion was still largely visible on the face of Susie. This became quite distracting around frames $40 - 60$. For Simulation No. 8 reported for Trevor both algorithms yielded large distortion in the form of blur, and 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 MD-BMA.

Figure 5 shows the DFD frames and Figure 6 shows the final reconstructed frames obtained with MD-BMA and with RC-BMA for Claire092. DFD frame obtained with RC-BMA has noticeably less energy content. Figure 7 displays the motion vector fields estimated with MD-BMA and RC-BMA. It is seen that although RC-BMA employs smaller size blocks the motion vectors for these blocks are much more organized than those for the large size blocks of MD-BMA. As a result, the motion vector fields have fewer spurious motion vectors for RC-BMA than for MD-BMA. Nevertheless, RC-BMA does not completely prevent some of the stationary blocks with little detail from getting assigned nonzero motion vectors.

9 Conclusion

This paper has extended the minimum distortion motion vector estimation technique of MD-BMA by incorporating rate constraint terms into the cost function of estimation. In RC-BMA, 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 RC-BMA employing size KxK blocks over MD-BMA employing size $2K$x$2K$ blocks. Motion vector fields output by RC-BMA are also smoother and more organized.

References

[1] P. A. Chou, T. L. Lookabough, and R. M. Gray, "Entropy constrained vector quantization", *IEEE Trans. on Info. Theory* 37(1),31–42 (1989).

[2] P. A. Chou and T. Lookabaugh, "Conditional entropy constrained vector quantization", in *Proc. of 1990 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 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. on Image Processing* 4(3),378–382, (1995).

[5] B. Girod. "Rate-constrained motion compensation", in *Visual Comm. and Image Proc.*, Proc. SPIE 2308 (1994).

[6] J. Lee, "Optimal quadtree for variable block size motion estimation". in *Proc. 1995 IEEE Int. Conf. on Image Proc.*,Vol. III, 480–483 (1995).

[7] M. Balakrishan, W. A. Pearlman, and L. Lu, "Variable-rate tree structured vector quantizers", *IEEE Trans. on 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. on Acoust., Speech and Signal Proc.* 73(11), 1551–1558 (1991).

[9] C. Stiller and D. Lappe, "Gain/cost controlled displacement estimation for image sequence coding", in *Proc. of 1991 IEEE Int. Conf. on Acoust., Speech, and Signal Proc. (ICASSP*, 2729–2732 (1991).

[10] W. C. Chung, F. Kossentini, and M. J. T. Smith, "Rate-distortion constrained statistical motion estimation for video coding", in *Proc. of 1995 IEEE Int. Conf. on Image Proc.*, Vol. III, 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, 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 Circuits and Systems for Video Technology* 6(3),243–250 (1996).

[13] U. Bayazit and W. A. Pearlman, "Rate-constrained block matching algorithm", in *Visual Comm. and Image Proc.*, Proc. SPIE 3024, 1110–1121 (1997).

LIST OF TABLES

LIST OF FIGURES