A Theory for the Optimal Bit Allocation between Displacement Vector Field and Displaced Frame Difference

*Guido M. Schuster and †Aggelos K. Katsaggelos

* U.S. Robotics, Advanced Technologies Research Center, Network System Division, Mount Prospect, IL 60056, (847) 222-2486, gschuste@usr.com

† Northwestern University, Department of Electrical and Computer Engineering, McCormick School of Engineering and Applied Science, Evanston, IL 60208, (847) 491-7164, aggk@ece.nwu.edu

Abstract

In this paper, we address the fundamental problem of optimally splitting a video sequence into two sources of information, the displaced frame difference (DFD) and the displacement vector field (DVF).

We first consider the case of a lossless motion compensated video coder (MCVC) and derive a general Dynamic Programming (DP) formulation which results in an optimal tradeoff between the DVF and the DFD. We then consider the more important case of a lossy MCVC and present an algorithm which solves the tradeoff between the rate and the distortion. This algorithm is based on the Lagrange multiplier method and the DP approach introduced for the lossless MCVC. We then present an H.263-based MCVC which uses the proposed optimal bit allocation and compare its results to H.263. As expected, the proposed coder is superior in the rate-distortion sense. In addition to this, it offers many advantages for a rate control scheme.

The presented theory can be applied to build new optimal coders and to analyze the heuristics employed in existing coders. In fact whenever one changes an existing coder, the proposed theory can be used to evaluate how the change affects its performance.

List of Figures

List of Tables

1 Introduction

Video compression attracted considerable attention over the last decade [1, 2, 3, 4]. Several standards for video coding such as MPEG-1 [5], MPEG-2 [6], H.261 [7] and most recently H.263 [8] have been established. There is a large redundancy in any video sequence which has to be exploited by every efficient video coding scheme. This redundancy is divided into temporal and spatial. The temporal redundancy is usually reduced by motion compensated prediction of the current frame from a previously reconstructed frame, whereas the spatial redundancy left in the prediction error is commonly reduced by a transform coder or a vector quantizer. Video coders which use the concept of motion compensated prediction are henceforth called motion compensated video coders (MCVC). All existing video standards belong to this class of video coders. In an MCVC, the original video sequence is represented by the displacement vector field (DVF) and the displaced frame difference (DFD).

A fundamental problem of MCVC is the bit allocation between the DFD and the DVF. In this paper we present a general theory which uses operational rate-distortion curves to solve this problem for a finite set of admissible quantizers and motion vectors.

There have been previous attempts to solve the optimal tradeoff between DVF and DFD. In the standard coders, such as MPEG-1, MPEG-2, H.261 and H.263, the bit allocation among DFD and DVF is not explicitly defined and there is great freedom of how the motion vectors and the quantizers are selected. Block matching is the most common approach for finding the DVF. The resulting DFD is encoded with quantizers selected by a rate control algorithm. In such a scheme, the tradeoff between DVF and DFD is implicit, without taking into account the resulting rate and distortion.

In [9], the authors assume a stochastic model for the distribution of the DFD and proceed to calculate the entropy of a given block based on some observed statistics. This entropy is then used to decide if a block should be split into four smaller blocks with their own motion vectors, or if the block should be kept as a basic unit.

In [10], the problem of rate-constrained motion estimation is considered and the optimal bit allocation condition for a strictly convex and everywhere differentiable multivariate rate distortion function is derived. It is applied to the problem of optimal bit allocation between the DVF and the DFD and a rate-constrained, region based motion estimator is introduced. In this paper, we do not assume knowledge of a convex and everywhere differentiable multivariate rate distortion function, but instead we deal with a set of finite quantizers and motion vectors. Therefore the operational rate-distortion functions are not differentiable everywhere and not convex.

In [11], a variable block size motion estimator is presented. It is implied that the motion vectors are

encoded by pulse code modulation (PCM) and hence the resulting optimization procedure is quite simple and in fact equivalent to the one presented in [12]. In contrast to this work, we allow for more sophisticated encoding schemes of the DVF, such as, the popular differential PCM (DPCM). This leads to a more complex optimization problem for which we derive the optimal solution.

In [13], the optimal bit allocation problem for lossless video coders is studied. The authors use a stochastic model which has been derived in [9] to find a formula for the entropy of the DFD as a function of the DVF accuracy. A similar formula can also be found in [14, 15]. As mentioned above, the stochastic model derived in [9] cannot be applied if the DFD is encoded by a sophisticated encoding scheme.

The main contribution of this paper is that the theory we present is very general and allows for the optimization of a wide range of schemes. Furthermore, the presented theory allows for the rate and the distortion of a given region to depend on quantizers and motion vectors of other regions. This enables the efficient encoding of the DVF using a DPCM based scheme and the use of distortion measures which also include region boundary effects which are very important for human observers.

Since we work with the operational rate distortion functions, we do not assume any stochastic models or convexity properties. Stochastic models are commonly used to estimate the entropy of the prediction error, but our experiments have shown that when a sophisticated encoding scheme is used, such as DCT and run-length encoding, these estimates can be quite inaccurate. The convexity and differentiability properties usually invoked for the rate-distortion function imply a continuous function though in every real video coder the set of motion vectors and quantizers is finite and hence the operational rate distortion curves are neither differentiable everywhere nor necessarily convex. It is commonly assumed or implied that the motion vectors are encoded by a simple PCM scheme which is inefficient, since neighboring motion vectors are highly correlated and a DPCM scheme is more appropriate. The problem of DPCM is the dependency it introduces in the optimization procedure which cannot be solved by any of the previously proposed approaches. As we will show, these dependencies can be handled efficiently in the proposed framework. The commonly used distortion measures are intra region based. In other words, these measures do not capture the boundary effects, such as blocking artifacts. It is well known that these artifacts are highly visible to a human observer. Again, the proposed framework does allow for such measures.

The paper is organized as follows: In section 2, we define the problem under consideration. In section 3 we derive the optimal solution for a lossless MCVC. This solution is then extended in section 4 to include lossy MCVC. In section 5 we develop a lossy video coder based on the presented theory and in section 6 we discuss some implementation issues which reduce the computational complexity of the presented coder. The experimental results of this coder are presented in section 7 and the paper is summarized in section 8.

2 Notation and assumptions

In this section we introduce the necessary notation and state the assumptions which will be used in the rest of the paper. Our study of the optimal bit allocation between the DVF and DFD is restricted to the frame level. In other words we do not attempt to optimally allocate the bits among the different frames of a video sequence. The reader interested in that problem is referred to [16]. For the rest of this paper we assume that a rate control algorithm has given us the maximum number of bits available (R_{max}) or the maximum acceptable distortion (D_{max}) for a given frame.

Let $f_k(\vec{r})$ denote the current frame, $\vec{d_k}(\vec{r})$ the DVF and $\tilde{f}_{k-1}(\vec{r})$ the previously reconstructed frame, which will be used to predict the current frame. Note that $\tilde{f}_{k-1}(\vec{r})$ does not need to be a frame from the past, but as in MPEG, this could be a future frame when backwards prediction is used. The predicted frame $\hat{f}_k(\vec{r})$ is defined as,

$$\hat{f}_k(\vec{r}) = \tilde{f}_{k-1}(\vec{r} - \vec{d_k}(\vec{r})), \tag{1}$$

and the DFD ($f_k^{DFD}(\vec{r})$) is defined by,

$$f_k^{DFD}(\vec{r}) = f_k(\vec{r}) - \hat{f}_k(\vec{r}). \tag{2}$$

In a lossy MCVC, the DFD is quantized and is denoted by,

$$f_k^{QDFD}(\vec{r}) = Q[f_k^{DFD}(\vec{r})], \tag{3}$$

where $Q[\cdot]$ is the quantization operator. Finally, the reconstructed frame $\tilde{f}_k(\vec{r})$, which is the frame displayed at the decoder is defined for a lossy MCVC as follows,

$$\tilde{f}_k(\vec{r}) = \hat{f}_k(\vec{r}) + f_k^{QDFD}(\vec{r}), \tag{4}$$

whereas for a lossless MCVC, by definition, $\tilde{f}_k(\vec{r}) = f_k(\vec{r})$.

We assume that the current frame is segmented into N regions, $o_1, \ldots, o_N$, and that this segmentation and the associated scanning path are known to both the encoder and the decoder. We then number the regions such that the scanning path visits them in ascending order. Every region o_i has a motion vector $m_i \in M_i$, and a quantizer $q_i \in Q_i$ associated with it, where M_i is the set of all admissible motion vectors for region o_i and Q_i is the set of all admissible quantizers for region o_i. As in every practical video coding scheme, we assume that the sets M_i and Q_i are finite. Let us define a decision vector $v_i = [m_i, q_i] \in V_i$, for every region o_i which contains the motion vector and the quantizer for that region. $V_i = M_i \times Q_i$ is the admissible decision vector set for region o_i.

We assume that the frame distortion D_k is a function of $\tilde{f}_k$ and f_k. As we can see from the above definitions, $\tilde{f}_k$ is a function of f_k, $\tilde{f}_{k-1}$, $\vec{d_k}$ and $Q[]$. Therefore $D_k(v_1, \ldots, v_N)$ will be considered a function of

$\tilde{f}_{k-1}$, f_k and the decision vectors $v_1, \ldots, v_N$. Equivalently, we also assume that the frame rate $R_k(v_1, \ldots, v_N)$ is a function of $\tilde{f}_{k-1}$, f_k and all the decision vectors $v_1, \ldots, v_N$. Note that the term "frame rate" represents the number of bits used to encode a certain frame and not the number of frames per second.

The next assumption expresses the idea that the frame rate and frame distortion can be decomposed into a sum of region rates $r_i(v_{i-a}, \ldots, v_{i+b})$ and region distortions $d_i(v_{i-a}, \ldots, v_{i+b})$, which only depend on a local neighborhood. We assume that there exist integers $a \geq 0$, and $b \geq 0$ and a family of functions d_i and r_i such that the following holds,

$$D_k(v_1, \ldots, v_N) = \sum_{i=1}^{N} d_i(v_{i-a}, \ldots, v_{i+b}), \tag{5}$$

$$R_k(v_1, \ldots, v_N) = \sum_{i=1}^{N} r_i(v_{i-a}, \ldots, v_{i+b}), \tag{6}$$

where the decision vectors v_j not belonging to any region ($j \notin [1, \ldots, N]$), represent the boundary parameters and can be set to any desired value. The above assumption is very important since the efficiency of the optimization procedure introduced later will directly depend on $a + b$ which defines the size of the neighborhood.

It is noted here that assumptions (5) and (6) are quite general and valid for every existing video coding standard. One contribution of this paper is the formulation of the optimal bit allocation problem for a MCVC. It is important to realize that assumptions (5) and (6) are essential for the development which follows. Most commonly used distortion measures, such as the mean squared error (MSE) or the peak signal to noise ratio (PSNR), satisfy assumption (6). For example, in MPEG-1 and MPEG-2, the rate for a given block depends not only on the quantizer and motion vector of that block, but also on the motion vector of the previous block, which is used as a predictor for the current motion vector. Usually the block distortion measured by the popular MSE measure depends only on the motion vector and the quantizer of the current block. A noteworthy exception is H.263 when the "Advanced Prediction Mode" is used. Then, overlapped block motion compensation is employed and the MSE of a given 8×8 block depends now on four spatial neighbors.

3 Lossless MCVC

In this section we study the case of a lossless MCVC. Since the reconstructed frame is identical to the original frame, the frame distortion will be zero and the goal is to minimize the number of bits required for the DVF and the DFD. This can be stated as follows,

$$\min_{[v_1, \ldots, v_N] \in V_1 \times \ldots \times V_N} R_k(v_1, \ldots, v_N). \tag{7}$$

Since this is a lossless MCVC, the DFD is not quantized, but encoded losslessly. With a slight abuse of notation let q_i represent different lossless encoding schemes for region o_i (i.e., DPCM with different predictor order, etc.), instead of different quantizers. Since we will refer to this algorithm later on, we will still call the q_i's quantizers in the following derivation.

Since we deal with a finite number of admissible motion vectors and quantizers, the above optimization problem can clearly be solved by an exhaustive search. The time complexity for such an exhaustive search is $O(|V_i|^N)$, where $|V_i|$ denotes the cardinality of V_i, and we assume that all the V_i have the same cardinality. We will show that the proposed algorithm reduces this complexity significantly. Note that when we use the term time complexity, we refer to the number of comparisons necessary to find the optimal solution. This does not include the time spent to evaluate the operational rate distortion functions, since this strongly depends on the implementation of a given MCVC.

As we stated in assumption (6), the frame rate R_k is the sum of rates which only depend on local neighborhoods. We will now employ this assumption to derive a Dynamic Programming (DP) [17] solution to problem (7). We will use generic terms, ($g_i(v_{i-a}, \ldots, v_{i+b})$ and $G_k(v_1, \ldots, v_N)$), in this derivation since they will be defined differently for the lossless and the lossy cases. For the presentation in this section, $g_i(\cdot)$ represents the region rate, that is,

$$g_i(v_{i-a}, \ldots, v_{i+b}) = r_i(v_{i-a}, \ldots, v_{i+b}), \tag{8}$$

and $G_k(\cdot)$ represents the frame rate, that is,

$$G_k(v_1, \ldots, v_N) = \sum_{i=1}^{N} g_i(v_{i-a}, \ldots, v_{i+b}). \tag{9}$$

Let $g_l^*(\cdot)$ be the minimum of $G_k(\cdot)$ up to and including region l, that is,

$$g_l^*(v_{l-a+1}, \ldots, v_{l+b}) = \min_{[v_1, \ldots, v_{l-a}] \in V_1 \times \ldots \times V_{l-a}} \sum_{i=1}^{l} g_i(v_{i-a}, \ldots, v_{i+b}). \tag{10}$$

From Eq. (10) it follows that,

$$g_{l+1}^*(v_{l+1-a+1}, \ldots, v_{l+1+b}) \tag{11}$$

$$= \min_{[v_1, \ldots, v_{l+1-a}] \in V_1 \times \ldots \times V_{l+1-a}} \sum_{i=1}^{l+1} g_i(v_{i-a}, \ldots, v_{i+b}) \tag{12}$$

$$= \min_{v_{l+1-a} \in V_{l+1-a}} \left[\min_{[v_1, \ldots, v_{l-a}] \in V_1 \times \ldots \times V_{l-a}} \left(\sum_{i=1}^{l} g_i(v_{i-a}, \ldots, v_{i+b}) + g_{l+1}(v_{l+1-a}, \ldots, v_{l+1+b}) \right) \right] \tag{13}$$

Since $g_{l+1}(v_{l+1-a}, \ldots, v_{l+1+b})$ does not depend on $v_1, \ldots, v_{l-a}$, it can be moved outside the inner minimization. Then the resulting inner minimization is equal to $g_l^*(v_{l-a+1}, \ldots, v_{l+b})$ in Eq. (10) and the following

DP recursion formula results,

$$g^*_{l+1}(v_{l+1-a+1},\ldots,v_{l+1+b}) = \min_{v_{l+1-a}\in V_{l+1-a}} \left[g^*_l(v_{l+1-a},\ldots,v_{l+b}) + g_{l+1}(v_{l+1-a},\ldots,v_{l+1+b})\right]. \quad (14)$$

Forward DP (also called the Viterbi algorithm [18]) can now be used to solve problem (7). First we need to initialize the recursion which is achieved in the following way,

$$g^*_a(v_1,\ldots,v_{a+b}) = \sum_{i=1}^{a} g_i(v_{i-a},\ldots,v_{i+b}), \qquad \forall [v_1,\ldots,v_{a+b}] \in V_1 \times \ldots \times V_{a+b}. \quad (15)$$

Next, the recursion is started, hence the DP recursion formula (14) is applied for $l = a$ up to and including $l = N-1$, that is,

$$\begin{aligned} g^*_{l+1}(v_{l+1-a+1},\ldots,v_{l+1+b}) \quad = \quad & \min_{v_{l+1-a}\in V_{l+1-a}} \left[g^*_l(v_{l+1-a},\ldots,v_{l+b}) + g_{l+1}(v_{l+1-a},\ldots,v_{l+1+b})\right], \\ & \forall [v_{l+1-a+1},\ldots,v_{l+1+b}] \in V_{l+1-a+1} \times \ldots \times V_{l+1+b}. \end{aligned} \quad (16)$$

Then the final solution is found by observing that,

$$\min_{[v_1,\ldots,v_N]\in V_1\times\ldots\times V_N} G_k(v_1,\ldots,v_N) = \min_{[v_{N-a+1},\ldots,v_N]\in V_{N-a+1}\times\ldots\times V_N} g^*_N(v_{N-a+1},\ldots,v_{N+b}). \quad (17)$$

As we have seen the time complexity for the exhaustive search is exponential. The time complexity for the DP approach depends directly on the size of the neighborhood and is $O(N * |V_i|^{a+b+1})$, where we again assume that all the V_i have the same cardinality.

3.1 Example

We now consider a simple example to illustrate the above points. Assume that a lossless MCVC has an admissible motion vector set $M_i = \{ma, mb\}$, and two different lossless encoding schemes, i.e., $Q_i = \{qa, qb\}$. Let the entire frame be split into $N = 4$ regions and the motion vectors be encoded using a first order DPCM along the scanning path. Using these definitions, the goal is to minimize the frame rate $R_k(v_1, v_2, v_3, v_4)$ with respect to the motion vector and quantizer choices. First we have to identify the size of the neighborhood involved in this problem. Since a first order DPCM along the scanning path is used for the encoding of the motion vectors, only the previous motion vector is required for determining the rate of the current region. Therefore $a = 1$ and $b = 0$.

We can now use Eq. (15) to initialize the forward DP, i.e.,

$$g^*_1(v_1) = g_1(v_0, v_1), \qquad \forall v_1 \in V_1, \quad (18)$$

where v_0 can be set to any value, say $v_0 = [ma, qa]$. Next, the recursion is started, hence the DP recursion formula (14) is applied. First for $l = 1$,

$$g^*_2(v_2) = \min_{v_1\in V_1} \left[g^*_1(v_1) + g_2(v_1, v_2)\right], \qquad \forall v_2 \in V_2, \quad (19)$$

then for $l = 2$,

$$g_3^*(v_3) = \min_{v_2 \in V_2} \left[g_2^*(v_2) + g_3(v_2, v_3) \right], \qquad \forall v_3 \in V_3, \tag{20}$$

and finally,

$$g_4^*(v_4) = \min_{v_3 \in V_3} \left[g_3^*(v_3) + g_4(v_3, v_4) \right], \qquad \forall v_4 \in V_4. \tag{21}$$

The final solution can now be found using Eq. (17),

$$\min_{[v_1, v_2, v_3, v_4] \in V_1 \times V_2 \times V_3 \times V_4} G_k(v_1, v_2, v_3, v_4) = \min_{v_4 \in V_4} g_4^*(v_4). \tag{22}$$

A good tool to visualize DP is a trellis. In Fig. 1 the trellis corresponding to the above example is displayed. The upper trellis in Fig. 1 shows the entire trellis for this example whereas the lower trellis shows a specific minimization. The different quantizer and motion vector configurations are indicated on the left and the direction of the scanning path is from left to right starting at region o_1 and ending at region o_4. Each node in the trellis represents a certain decision vector choice for a given region. In the lower trellis, it is shown how $g_3^*([ma, qb])$ is calculated; it can be interpreted in this example as the smallest rate needed to encode region o_1 up to and including region o_3, where region o_3 uses the motion vector ma and the quantizer qb.

In this simple, first order ($a + b = 1$) example, we can assign the bit rate needed to encode the DFD of a given region to the associated node, which is the result of using a particular motion vector with a particular quantizer. The transitional bit rate between the nodes occurs because of the DPCM encoding of the DVF and this dependency is the reason why DP is used to solve this example. For higher order dependencies, ($a + b > 1$), drawing a trellis and indicating the associated costs for the DFD and DVF encoding is not as clear and hence it is important to understand the algebraic derivation of the DP recursion formula.

Note that for an exhaustive search $4^4 = 256$ comparisons are necessary, whereas the DP solution requires only $3 * 4^2 + 4 = 52$ comparisons.

4 Lossy MCVC

So far we have considered lossless MCVC. In this section we study the more interesting case of lossy MCVC. Clearly for a lossy MCVC it does not make sense to minimize the frame rate R_k with no additional constraints, since this would lead to a very high frame distortion D_k.

The most common approach to solve the tradeoff between the frame rate and the frame distortion is to minimize the frame distortion D_k subject to a given maximum frame rate R_{max}. Clearly since the total number of regions N is known, minimizing the total distortion is equivalent to minimizing the average distortion. This problem can be formulated in the following way,

$$\min_{[v_1, \ldots, v_N] \in V_1 \times \ldots \times V_N} D_k(v_1, \ldots, v_N), \qquad \text{subject to:} \qquad R_k(v_1, \ldots, v_N) \leq R_{max}. \tag{23}$$

This constrained discrete optimization problem is very hard to solve in general. In fact the approach we propose will not necessarily find the optimal solution but only the solutions which belong to the convex hull of the rate-distortion curve. On the other hand, as we show in Sec. 7, the solutions on the rate-distortion curve tend to be quite dense and hence the convex hull approximation is very good.

We solve this problem using the concept of Lagrangian relaxation [19, 20], which is a well known tool in operations research. It is mainly used to relax some constraints which destroy the integrality property of an integer program. The relaxed integer program can then be solved by linear programming which leads to an efficient method for certain problems. In this application we will use Lagrangian relaxation to relax the constraint so that the relaxed problem can be solved by DP. This is the same strategy employed in [21, 22] for the problem of optimal mode selection for H.263.

First we introduce the Lagrangian cost function which is of the following form,

$$J_\lambda(v_1, \ldots, v_N) = D_k(v_1, \ldots, v_N) + \lambda * R_k(v_1, \ldots, v_N), \tag{24}$$

where $\lambda \geq 0$ is called the Lagrangian multiplier. It has been shown in [23, 19, 20] that if there is a λ^* such that,

$$[v_1^*, \ldots, v_N^*] = \arg \min_{[v_1, \ldots, v_N] \in V_1 \times \ldots \times V_N} J_{\lambda^*}(v_1, \ldots, v_N), \tag{25}$$

leads to $R_k(v_1^*, \ldots, v_N^*) = R_{max}$, then $v_1^*, \ldots, v_N^*$ is also an optimal solution to (23). It is well known that when λ sweeps from zero to infinity, the solution to problem (25) traces out the convex hull of the rate distortion curve, which is a non-increasing function. Hence bisection [24] could be used to find λ^*. A faster converging algorithm which uses some knowledge about the convexity of the curve is employed in [25] and we present an even faster algorithm in [26].

Therefore the problem at hand is to find the optimal solution to problem (25). We next show how the original DP approach can be modified to find the global minimum of problem (25). For a given λ, let the $g_i(v_{i-a}, \ldots, v_{i+b})$ functions be defined as follows,

$$g_i(v_{i-a}, \ldots, v_{i+b}) = d_i(v_{i-a}, \ldots, v_{i+b}) + \lambda * r_i(v_{i-a}, \ldots, v_{i+b}), \tag{26}$$

which implies that, $G_k(v_1, \ldots, v_N) = J_\lambda(v_1, \ldots, v_N)$. Hence the DP algorithm presented in section (3) leads to the optimal solution of problem (25).

Note that the dual problem, which can be stated as follows,

$$\min_{[v_1, \ldots, v_N] \in V_1 \times \ldots \times V_N} R_k(v_1, \ldots, v_N), \quad \text{subject to:} \quad D_k(v_1, \ldots, v_N) \leq D_{max}, \tag{27}$$

can be solved with exactly the same technique using the following relabeling of function names, $R_k(v_1, \ldots, v_N) \leftarrow D_k(v_1, \ldots, v_N)$ and $D_k(v_1, \ldots, v_N) \leftarrow R_k(v_1, \ldots, v_N)$.

5 A video compression scheme with optimal bit allocation between DVF and DFD

In this section we present an example of how the proposed theory can be applied to existing coders and how the understanding of the theory can foster new coders well suited for the proposed optimization algorithm.

A coder can be considered a model for the video source it is intended to compress. Clearly, one wants to match the model as closely as possible with the source, since this results in a good compression performance. On the other hand, since a particular coder is required to compress a wide variety of sequences, many parameters need to be found so that the coder can be adapted to a particular sequence. If the modeling of the sequence is to detailed, i.e., the coder is to complex, then finding the optimal parameters can be nearly impossible. Hence a good model is powerful enough to capture the essence of a source and simple enough so that its optimal parameters can be found efficiently.

We will apply the presented theory to the optimal allocation of the bits between the DFD and the DVF for a video coder which is largely based on the ITU standard for very low bit rate video coding H.263 [8].

The presented theory can be used to optimize H.263 with all its options activated. For simplicity, we base our coder on H.263 without the recently added options, which are the "Unrestricted Motion Vector Mode", the "Syntax-based Arithmetic Coding mode", the "Advanced Prediction mode" and the "PB-frames mode". In fact the proposed coder is almost identical to test model 4 (TMN4) [27] of H.263 with some noteworthy exceptions which we will point out later on. The changes we incorporated are mainly to reduce the complexity of the optimization procedure and are meant to lead to a video coder which is ideally suited for the presented theory of optimal bit allocation among DFD and DVF. Note that the proposed coder is closely related to the one we presented in [28].

Because of its popularity, we use the peak signal to noise ratio (PSNR) as the distortion measure. It is defined in the following way,

$$D_k = \frac{255^2}{\mathrm{MSE}(\tilde{f}_k, f_k)}, \tag{28}$$

where $\mathrm{MSE}(\tilde{f}_k, f_k)$ is the mean squared error between the luminance channel of the reconstructed and the original frames. In the case of TMN4, the PSNR frame distortion can be written as,

$$D_k(v_1, \ldots, v_N) = \sum_{i=1}^{N} d_i(v_i), \tag{29}$$

since the region distortion $d_i(v_i)$ only depends on the selected motion vector and the selected quantizer for that region. We use "quarter common intermediate format" (QCIF) sequences, which have dimensions 176×144 pixels. Since TMN4 breaks the frame into 11×9 macro blocks of size 16×16, the regions $o_1, \ldots, o_N$ are defined to be these macro blocks. Clearly, the total number of regions N equals 99 for this

implementation. Note that the TMN4 scanning path is a raster-scan, in other words, the upper left block is encoded first and the lower right block is encoded last.

The frame rate for TMN4 can be written as follows,

$$R_k(v_1, \ldots, v_N) = \sum_{i=1}^{N} r_i(v_{i-11}, \ldots, v_i). \tag{30}$$

The reason for $a = 11$ can be found in the way TMN4 encodes the current motion vector. TMN4 uses the vector median of three neighboring motion vectors as the prediction for the current motion vector. The three motion vectors employed for the prediction belong to the macro block to the left of the current macro block, to the macro block above the current macro block and to the macro block to the right and above the current macro block. The macro block directly above the current macro block has been visited by the raster scan 11 macro blocks earlier than the current macro block. Therefore, the neighborhood has to contain the information about the encoding decisions made for the last 11 macro blocks, because this is the knowledge needed to make the future of the optimization process independent from its past.

Since the computational complexity of the proposed optimal bit allocation algorithm is exponential in a, we would like to keep a as small as possible. The smallest a, which is still useful when $b = 0$ as in this case, is $a = 1$. In other words, we will employ a single predictor DPCM for the DVF encoding. This is the same way all the other standards encode the DVF. There are other dependencies in TMN4, which we will discuss later, which all can be captured by an a of one. Hence the total frame rate can now be expressed by,

$$R_k(v_1, \ldots, v_N) = \sum_{i=1}^{N} r_i(v_{i-1}, v_i), \tag{31}$$

where

$$r_i(v_{i-1}, v_i) = r_i^{QDFD}(v_i) + r_i^{DVF}(v_{i-1}, v_i), \tag{32}$$

$r_i^{QDFD}(v_i)$ are the bits needed to encode the DFD of block o_i using the quantizer q_i and the motion vector m_i, and $r_i^{DVF}(v_{i-1}, v_i)$ the bits needed to encode the motion vector difference $(m_i - m_{i-1})$.

In the next two paragraphs we define what we exactly mean by q_i and m_i. Let $e_i \in E_i$ be the encoding mode of macro block o_i, where $E_i = \{\text{Intra, Inter, Skip, Prediction}\}$. The encoding mode can be set differently for each macro block. When the Intra mode is selected, then the macro block is encoded using a "JPEG"-like scheme and its associated motion vector is set to zero. In the Inter mode, the motion vector is used to create the predicted block and then the difference between the original and the predicted block is encoded using a similar scheme as in the Intra mode. The Skip mode means that the current block is replaced by the block at the same location in the previous reconstructed frame and its motion vector is considered to be zero. TMN4 has all the above modes but the next mode, the Prediction mode has been

introduced by us. This mode is identical to the Inter mode with the exception that no difference signal is sent. Hence in the case where the prediction is good enough, one wants to use the Prediction mode.

Let $QP_i \in Z_i$ be the DCT domain quantizers for block o_i, where Z_i is the set of all admissible DCT domain quantizers for block o_i. In TMN4, 31 different DCT domain quantizers are admissible. Note the distinction made between quantizers and DCT domain quantizers. The reason for this is that the modes can be considered quantizers too. Therefore we can define the new set of quantizers for block o_i as $q_i = [e_i, QP_i] \in Q_i$ where $Q_i = E_i \times Z_i$. As defined before, $m_i \in M_i$ is the motion vector for block o_i, where M_i is the set of all admissible motion vectors for block o_i.

It is well known that the DC values of the luminance (Y) and the chrominance channels (Cb, Cr) of neighboring blocks are highly correlated and therefore an encoding scheme should take advantage of this. One way of exploiting this fact is by encoding the DC values of consecutive Intra blocks by a single predictor DPCM. Therefore an additional dependency has been introduced and the $r_i(v_{i-1}, v_i)$ from Eq. (32) is now equal to

$$r_i(v_{i-1}, v_i) = r_i^{QDFD}(v_i) + r_i^{DVF}(v_{i-1}, v_i) + r_i^{DC}(v_{i-1}, v_i), \tag{33}$$

where $r_i^{DC}(v_{i-1}, v_i)$ is zero whenever the blocks o_{i-1} and o_i are not both Intra coded, and equal to the number of bits needed to encode the difference between the DC coefficients of the two Intra coded blocks, otherwise.

For an Intra frame, the DPCM of the DC values is very important since many bits can be saved by this technique. Hence by using this DPCM the proposed coder will also efficiently encode an Intra frame, such as the first frame of a sequence.

In TMN4 a quantizer is selected by transmitting a quantizer step size QP. QP is encoded using a modified delta modulation with a range of ± 2. Hence the quantizer step size of block o_i, QP_i, is equal to $QP_{i-1} + \delta_i$, where $\delta_i \in [-2, -1, 0, 1, 2]$. At the beginning of the frame, QP_1 of the first block is coded using PCM. Clearly, this delta modulation introduces another dependency which can be captured by modifying $r_i(v_{i-1}, v_i)$ from Eq. (33),

$$r_i(v_{i-1}, v_i) = r_i^{QDFD}(v_i) + r_i^{DVF}(v_{i-1}, v_i) + r_i^{DC}(v_{i-1}, v_i) + r_i^{QP}(v_{i-1}, v_i), \tag{34}$$

where $r_i^{QP}(v_{i-1}, v_i)$ corresponds to the bits needed to encode $\delta_i \in [-2, -1, 0, 1, 2]$. $r_i^{QP}(v_{i-1}, v_i)$ is set to infinity for a QP_i which is out of reach. This will force the optimal path to select only accessible quantizers.

In most MCVCs the blocks are processed along a simple raster scan. Clearly DPCM is most effective when the data is highly correlated. In the presented approach, the motion vector, the DC values (for Intra blocks) and the quantizer step size of the previous block are used as predictors for the values of the current block. Therefore, the higher the correlation between the blocks along the scanning path is, the better the

performance of the coding scheme.

Hilbert curves have been used in image and video processing as scanning paths on the pixel level in the luminance domain for lossless coding [29] and lossy coding [30]. They have also been used as a scanning path for the coefficients in the transform domain [31]. In all these cases, the fact that a scanned image according to a Hilbert curve creates an one-dimensional representation of the image, which is more correlated than a raster scan, is exploited.

Since in the presented approach the correlation of the blocks along the scanning path is of foremost importance, a Hilbert curve should be employed for the scanning of the blocks. Since the proposed video coder works with QCIF video sequences, a modified version of a Hilbert scan is used (see Fig. 2), since a perfect Hilbert scan requires an image format of $2^n \times 2^n$, where n is an integer. Note that a Hilbert scan is also ideal for higher order predictors since the previous blocks along the scanning path are closer to the current block than in a raster scan.

6 Implementation Issues

In this section we discuss how the computational complexity can be further reduced by restricting the set of admissible decision vectors and using a fast evaluation of the operational rate-distortion functions.

From a theoretical point of view, every possible motion vector of block o_i should be included in the set of admissible motion vectors M_i. This means that in the case of the TMN4 implementation, where the search window is ± 15 pixels and the accuracy of the motion estimation is 1/2 pixel, $|M_i| = 63 * 63 = 3969$, which is quite large. Most of those motion vectors are not likely candidates for the optimal path since they do not correspond well to the real motion in the scene and therefore they lead to a high distortion and a high rate.

To make the optimization process faster, such prior knowledge should be used. Even though this is complicated in general, it can be achieved easily in the presented framework by constraining the set M_i of admissible motion vectors of block o_i. We propose the following strategy to constrain the set M_i. An initial motion vector is first found by using block matching with integer accuracy and the sum of absolute error matching criterion. Then the set M_i is defined as the set which contains this motion vector plus the K neighboring motion vectors at half pixel locations. This leads to $|M_i| = K + 1$ which can be much smaller than the original size of 3969. Our experiments have shown (see Section 7.1.1) that a $K = 8$ results in a performance loss which is negligible compared to the achieved reduction in computational complexity.

A similar situation arises for the quantizer selection. In TMN4, the quantizer parameter QP_i can take on values between 1 and 31. Since a nearly constant distortion is usually targeted, a reduced admissible quantizer set, which is centered around the quantizer step size which leads to the desired distortion, can

be used without any noticeable loss of performance. The set employed in the presented experiments is $Z_i = \{8, 9, 10, 11, 12\}$.

Since for TMN4 (and for most other coders) the Skip and Intra modes imply a zero motion vector, this knowledge can be used to further reduce the set of admissible decision vectors V_i. For the Inter and Prediction encoding mode, $|M_i|$ motion vectors are considered, but for the Skip and Intra encoding modes only the zero motion vector is needed. This leads to the final admissible decision vector set $V_i = (\{\text{Inter,Prediction}\} \times Z_i \times M_i) \cup (\{\text{Skip,Intra}\} \times Z_i \times \{\vec{0}\})$. The cardinality of this set equals $2 * |Z_i| * (|M_i| + 1)$ (recall that the time complexity of DP is $O(N * |V_i|^{a+b+1})$).

Note that we restricted the sets M_i and Z_i to reflect our prior knowledge about the solution. One of the great advantages of DP, besides finding a global optimum, is that it can incorporate difficult constraint sets such as the one we formulated above.

For every member v_i of the set V_i a new rate and distortion pair needs to be calculated. The number of required DCTs per block is equal to $|M_i| + 1$, since the DFD of every admissible motion vector needs to be transformed and for the Intra encoding mode, the DCT of the original block has to be calculated. In general it takes as many inverse DCTs as it takes DCTs, since for the calculation of the distortion, the reconstructed block must be available. By selecting the mean squared error, however, or a block weighted MSE distortion measure, these inverse DCTs are not necessary. Recall that the block MSE between an original block f of dimensions $K \times I$, and a reconstructed block $\tilde{f}$ of the same dimensions is defined as follows,

$$\text{MSE} = \frac{1}{K \cdot I} \cdot \sum_{k=1}^{K} \sum_{i=1}^{I} (f[i,k] - \tilde{f}[i,k])^2. \tag{35}$$

Since the two dimensional DCT used (DCT-II, [32]), is a linear and distance preserving transformation, the following holds true,

$$\frac{1}{K \cdot I} \cdot \sum_{k=1}^{K} \sum_{i=1}^{I} (f[i,k] - \tilde{f}[i,k])^2 = \frac{1}{K \cdot I} \cdot \sum_{k=1}^{K} \sum_{i=1}^{I} (F[i,k] - \tilde{F}[i,k])^2, \tag{36}$$

where $F = \text{DCT}(f)$ and $\tilde{F} = \text{DCT}(\tilde{f})$. This means that the squared sum of the error in the original domain, is equal to the squared sum of the error in the DCT domain. Therefore, the mean squared error can be computed in the DCT domain and no inverse DCT operation is required.

Using the above admissible decision vector reduction and the fast distortion calculation, we observed that the current implementation of the proposed coder requires about three times the amount of time to encode a video sequence than the available TMN4 implementation, on which the proposed coder implementation is based. Since the motion vector search is the most time consuming part of TMN4, it is is important to note that the TMN4 implementation uses an exhaustive search to first find the best motion vector with pixel accuracy. Then the best motion vector with half pixel accuracy is selected from a set containing the best

motion vector with pixel accuracy and its 8 half pixel neighbors. Even though the increase of complexity by a factor of three is significant, it is well known that the speed of desktop computers doubles roughly every 18 months. On the other hand, for applications for which the encoding can be done off-line, such as a the encoding of a video clip for a multimedia encyclopedia, the encoding speed is clearly not as critical.

7 Experiments

In this section, the results of the proposed coder which is based on our general theory for optimal bit allocation between DFD and DVF, are compared to TMN4. As we pointed out in Sec. 4 the Lagrangian relaxation can only find solutions which belong to the convex hull of the rate-distortion curve. If these solutions are dense enough, the convex hull approximation can for practical purposes be considered the optimal solution. Based on our experiments we have found that the convex hull solution is usually within $\pm 2.5\%$ of the desired solution, which is certainly sufficient for practical systems.

Note that the presented coder, like TMN4, writes a bit stream which is uniquely decodable by our decoder. Hence the listed bit rates are the effective number of bits used and not an estimate of the entropy. Since TMN4 was selected to implement the proposed optimal bit allocation scheme, most parts of the proposed coder and TMN4 are identical. The deviations of the proposed coder from the TMN4 implementation consist of the use of the following:

- First order DPCM encoding of the DVF along a modified Hilbert scan.
- DPCM encoding of the DC values (Y, Cb and Cr) for consecutive intra blocks.
- Optimal bit allocation between DFD and DVF.

The first order DPCM encoding of the DVF is selected to keep the computational complexity of the optimization procedure reasonable. TMN4 uses a more sophisticated DVF DPCM encoding which involves three previous motion vectors. The Hilbert scanning path is used to maximize the correlation along the scanning path, which improves the efficiency of the first order DPCM. The DPCM encoding of the DC values for consecutive blocks enables the proposed coder to encode Intra frames, such as the first frame, in an efficient and optimal way without having to treat these frames differently.

The first two changes to TMN4 listed above have been incorporated so that the resulting coder presents a good framework for the employed optimal decision strategy. Note that these changes alone do not lead to a better coder than TMN4. Therefore, all the improved results presented in the following are achieved by the proposed optimal bit allocation between DFD and DVF.

In order to compare TMN4 and the proposed coder, TMN4 was used to encode every 4^{th} frame of the first

200 frames of the QCIF color sequence "Mother and Daughter" with a fixed quantizer step size $QP = 10$. The first frame was Intra coded using the same quantizer step size. Since the "Mother and Daughter" sequence is considered to be recorded at 30 frames/second, this leads to an encoded rate of 7.5 frames/second. The resulting frame rate and frame distortion were used for the comparison between TMN4 and the proposed coder. Remember that the term "frame rate" is used for the number of bits required to encode a certain frame and not for the number of encoded frames per second. The employed distortion measure is the peak signal to noise ratio (PSNR) which is defined in Eq. (28).

7.1 Matched distortion

The goal of this experiment is to compare the proposed coder with TMN4 in the case where their frame distortions are matched. This can be achieved by setting D_{max}, the maximum frame distortion from Eq. (27) equal to the frame distortion of TMN4. Clearly D_{max} changes from frame to frame, following the distortion profile generated by the TMN4 run. The proposed coder will lead to the smallest number of bits needed to encode a given frame for the given maximum distortion D_{max}.

The resulting rate and distortion are displayed in Figs. 3 and 4. The average rates and distortions for the first frame, the sequence without the first frame and the entire sequence are listed in Table 1. We list these three entities separately since in a very low bit rate video coding scheme, the contribution of the first frame can be quite large. Furthermore, since the first frame is completely intra coded, the optimal DPCM of the intra DC values, discussed in the previous section, results in a very efficient intra coding scheme, as can be seen from the first column in Table 1. The distortion of the proposed coder follows the TMN4 distortion extremely closely. Clearly the proposed coder is superior to TMN4 when their frame distortions are matched, based on the resulting difference in bit rates.

Fig. 5 shows the reconstructed 12^{th} frame of the sequence which is used to predict the 16^{th} frame. The optimal encoding mode selection, the optimal quantizer selection and the optimal motion vector field are displayed in Figs. 6, 7 and 8 for the 16^{th} frame of the "Mother and Daughter" sequence. Note in Fig. 6 how the new object (the hand) and the uncovered areas (left of the hand) are Intra coded and the stationary background is replaced by the blocks from the previously decoded frame (Skip mode). Also note the smoothness of the motion vector field in Fig. 8, which can be encoded very efficiently by DPCM.

7.1.1 The influence of the constrained search space on the rate for the matched distortion case

As mentioned in section 6, the optimization process can be accelerated by using prior knowledge about the admissible motion vectors and quantizers. In this section we experimentally compare various solutions with differently constrained search spaces. Table 2 is discussed which is a collection of encoding results for the

"Mother and Daughter" sequence which all have the same distortion profile as TMN4, resulting in an average distortion of 33.0 dB PSNR. In Table 2 the following information is shown: in column one, the available encoding modes; in column two, the admissible motion vectors; in column three, the admissible quantizer step sizes; in column four, the cardinality of V_i; and in column five, the resulting bit rate. The coders are listed in decreasing order of search space constraints. In other words, the top coder has the most constrained search space and the bottom coder has the least constrained search space. As expected the bit rate drops as the search space get less constrained but also the computational complexity rises as the square of the cardinality of V_i. For all the results presented in this paper, the coder with a cardinality of $|V_i| = 100$ is used, since for this cardinality, the tradeoff between speed and performance is very good.

It is interesting to notice that the inclusion of the 8 half-pel neighbors in the motion vector search achieved by far the biggest gain in the bit rate, and that additional inclusion of more motion vectors did not improve the result significantly. Hence the TMN4 generated motion vectors are very close to the optimal motion vectors, but the small error made ($\pm$ 0.5 pels) can lead to a loss in performance of about 10%. It is interesting to note that this observation is quite general since we observed similar effects for other sequences with varying degree of motion activity, such as "Miss America", "Carphone" and "Foreman".

7.2 Matched rate

In this experiment, the proposed coder is compared to TMN4 in the case where their frame rates are matched. This can be achieved by setting R_{max}, the maximum frame rate from Eq. (23) equal to the frame rate obtained by TMN4. Clearly R_{max} changes from frame to frame, following the rate profile of TMN4 and the proposed coder will minimize the resulting frame distortion for the given frame rate.

The resulting rate and distortion are displayed in Figs. 9 and 10. The average rate and distortion for the entire sequence are shown in Table 1. Again, note that the rate of the proposed coder follows the TMN4 rate very closely. Besides being able to outperform TMN4 for matched rates, this experiments also shows the enormous potential of this approach with respect to rate control since the optimal coder can follow an arbitrary bit assignment per frame and produce the smallest possible distortion for the given bit budget.

7.3 Constant distortion

So far TMN4 and the proposed coder have been compared in terms dictated by the TMN4 run. This gives TMN4 an advantage, since it sets the rate distortion profile, which is then followed by the proposed coder.

An interesting application of the proposed coder is for channels which can accept a variable bit rate, such as an ATM network. For such applications, one would like to keep the distortion constant which can be achieved by setting D_{max} equal to the desired frame distortion. The proposed coder allows therefore for

quality scalability.

For the experiment discussed next, D_{max} was selected to be equal to the minimum frame PSNR of the TMN4 run, since this is the best quality TMN4 can guarantee over the entire sequence. For this experiment the set of admissible quantizer step sizes Z_i has been changed to $\{9, 10, 11, 12, 13\}$, since on the average, a coarser quantization will be needed to achieve the new target distortion.

The resulting rate and distortion are displayed in Figs. 11 and 12. The average rate and distortion for the entire sequence are shown in Table 1. Clearly the goal of constant distortion (quality) has been achieved and the resulting average rate is much lower than the TMN4 rate, even though visually these two encoded sequences cannot be distinguished. Some observers even prefer the constant quality sequence over the TMN4 sequence. One possible explanation for this fact is based on the globally optimal selection of the DFD and the DVF. Recall that a Hilbert scan is used for the DPCM encoding of the DVF and a smooth DVF along this path leads to a low bit rate. Hence the optimal solution enforces a global smoothness constraint on the DVF which in turn leads to predicted frames which are more visually pleasing than the ones produced by block matching.

8 Summary and conclusions

We have presented a general theory for the optimal bit allocation between displacement vector field (DVF) and displaced frame difference (DFD). The theory can be applied to all region based motion compensated video coders (MCVC), which includes all current video standards.

We first considered a lossless MCVC and derived the optimal bit allocation algorithm which is based on dynamic programming (DP). We then addressed the problem of lossy MCVC and we showed that Lagrangian relaxation and DP can find the convex hull approximation to the optimal solution.

We then presented a video coder which is largely based on H.263, and uses this optimal bit allocation between the DVF and the DFD. We pointed out the changes we incorporated to reduce the computational complexity and presented results which clearly show the superiority of this coder.

We showed that the presented theory can be used to optimize existing video coders. We also showed that it can foster new coders which are designed in such a way that the optimization procedure can be achieved in real time. We pointed out in the matched rate experiments that a video coder employing the optimal bit allocation scheme for the frame encoding is ideal for a rate control algorithm.

One of the main applications of this theory could be the evaluation of changes to an existing coding scheme. Conceptually every encoding scheme can be broken down into methods and decisions. The problem in evaluating a certain method is that a decision rule needs to be formulated. During the design process,

one can use the presented theory to evaluate the method first (such as a new encoding mode, like the Prediction mode we used), employing the optimal decisions found by the proposed algorithm. If the results are satisfactory, one can use the statistical data of the optimal algorithm to formulate a fast heuristic and evaluate that heuristic versus the optimal decisions. Hence the proposed theory can be used to separately evaluate the method and the decision rule which should speed up the development of new video coders.

References

[1] R. Forchheimer and T. Kronander, "Image coding-from waveforms to animation," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, vol. 37, pp. 2008–2023, Dec. 1989.

[2] H. G. Musmann, P. Pirsch, and H. Grallert, "Advances in picture coding," *Proceedings of the IEEE*, vol. 73, pp. 523–548, Apr. 1985.

[3] A. K. Jain, "Image data compression: A review," *Proceedings of the IEEE*, vol. 69, pp. 349–389, Mar. 1981.

[4] A. N. Netravali and J. O. Limb, "Picture coding: A review," *Proceedings of the IEEE*, vol. 68, pp. 366–406, Mar. 1980.

[5] "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbits/s." International Standard ISO/IEC IS-11172, Oct. 1992.

[6] "Generic coding of moving pictures and associated audio." International Standard ISO/IEC IS-13818, Nov. 1994.

[7] ITU-T Recommendations H.261, *Video codec for audiovisual services at $p \times 64$ kbits,*.

[8] Expert's group on very low bit rate video telephony, ITU-Telecommunication standardization sector, *Draft Recommendation H.263.*

[9] F. Moscheni, F. Dufaux, and H. Nicolas, "Entropy criterion for optimal bit allocation between motion and prediction error information," in *Proceedings of the Conference on Visual Communications and Image Processing*, vol. 2094, pp. 235–242, SPIE, 1993.

[10] B. Girod, "Rate-constrained motion estimation," in *Proceedings of the Conference on Visual Communications and Image Processing*, vol. 2308, pp. 1026–1034, SPIE, 1994.

[11] J. Lee, "Optimal quadtree for variable block size motion estimation," in *Proceedings of the International Conference on Image Processing*, vol. 3, pp. 480–483, Oct. 1995.

[12] G. J. Sullivan and R. L. Baker, "Efficient quadtree coding of images and video," *IEEE Transactions on Image Processing*, vol. 3, pp. 327–331, May 1994.

[13] J. Ribas-Corbera and D. L. Neuhoff, "Optimal bit allocations for lossless video coders: motion vectors vs. difference frames," in *Proceedings of the International Conference on Image Processing*, vol. 2, pp. 180–183, 1995.

[14] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE Journal on selected Areas in Communications,*, vol. SAC-5, pp. 1140–1154, Aug. 1987.

[15] B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Transactions on Communications*, vol. 41, pp. 604–612, Apr. 1993.

[16] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Transactions on Image Processing*, vol. 3, pp. 533–545, Sept. 1994.

[17] D. P. Bertsekas, *Dynamic programming: Deterministic and stochastic models.* Prentice-Hall, 1987.

[18] G. D. Forney, "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[19] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, pp. 399–417, 1963.

[20] M. L. Fisher, "The Lagrangian relaxation method for solving integer programming problems," *Management Science*, vol. 27, pp. 1–18, Jan. 1981.

[21] G. M. Schuster and A. K. Katsaggelos, "Fast and efficient mode and quantizer selection in the rate distortion sense for H.263," in *Proceedings of the Conference on Visual Communications and Image Processing*, pp. 784–795, SPIE, Mar. 1996.

[22] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, "Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard," *Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 182–190, Apr. 1996.

[23] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1445–1453, Sept. 1988.

[24] C. F. Gerald and P. O. Wheatley, *Applied numerical analysis.* Addison Wesley, fourth ed., 1990.

[25] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Transactions on Image Processing*, vol. 2, pp. 160–175, Apr. 1993.

[26] G. M. Schuster and A. K. Katsaggelos, "A video compression scheme with optimal bit allocation among segmentation, motion and residual error," *IEEE Transactions on Image Processing*, 1997. to appear.

[27] Expert's Group on Very Low Bitrate Visual Telephony, *Video Codec Test Model, TMN4 Rev1.* ITU Telecommunication Standardization Sector, Oct. 1994.

[28] G. M. Schuster and A. K. Katsaggelos, "A video compression scheme with optimal bit allocation between displacement vector field and displaced frame difference," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, May 1996.

[29] J. A. Provine and R. M. Rangayyan, "Lossless compression of peanoscanned images," *Journal of Electronic Imaging*, vol. 3, pp. 176–181, Apr. 1994.

[30] B. Moghaddam, K. J. Hintz, and C. V. Steward, "Space-filling curves for image compression," in *Automatic Object Recognition*, vol. 1471, pp. 414–421, SPIE, 1991.

[31] T. Ebrahimi, F. Dufaux, I. Moccagatta, T. G. Campbell, and M. Kunt, "A digital video codec for medium bitrate transmission," in *Proceedings of the Conference on Visual Communications and Image Processing*, vol. 1605, pp. 2–15, SPIE, 1991.

[32] K. R. Rao and P. Yip, *Discrete cosine transform: algorithms, advantages, applications.* Boston: Academic Press, 1990.

[33] G. M. Schuster and A. K. Katsaggelos, *Rate-distortion based video compression, Optimal video frame compression and object boundary encoding.* Kluwer academic publishers, 1997.

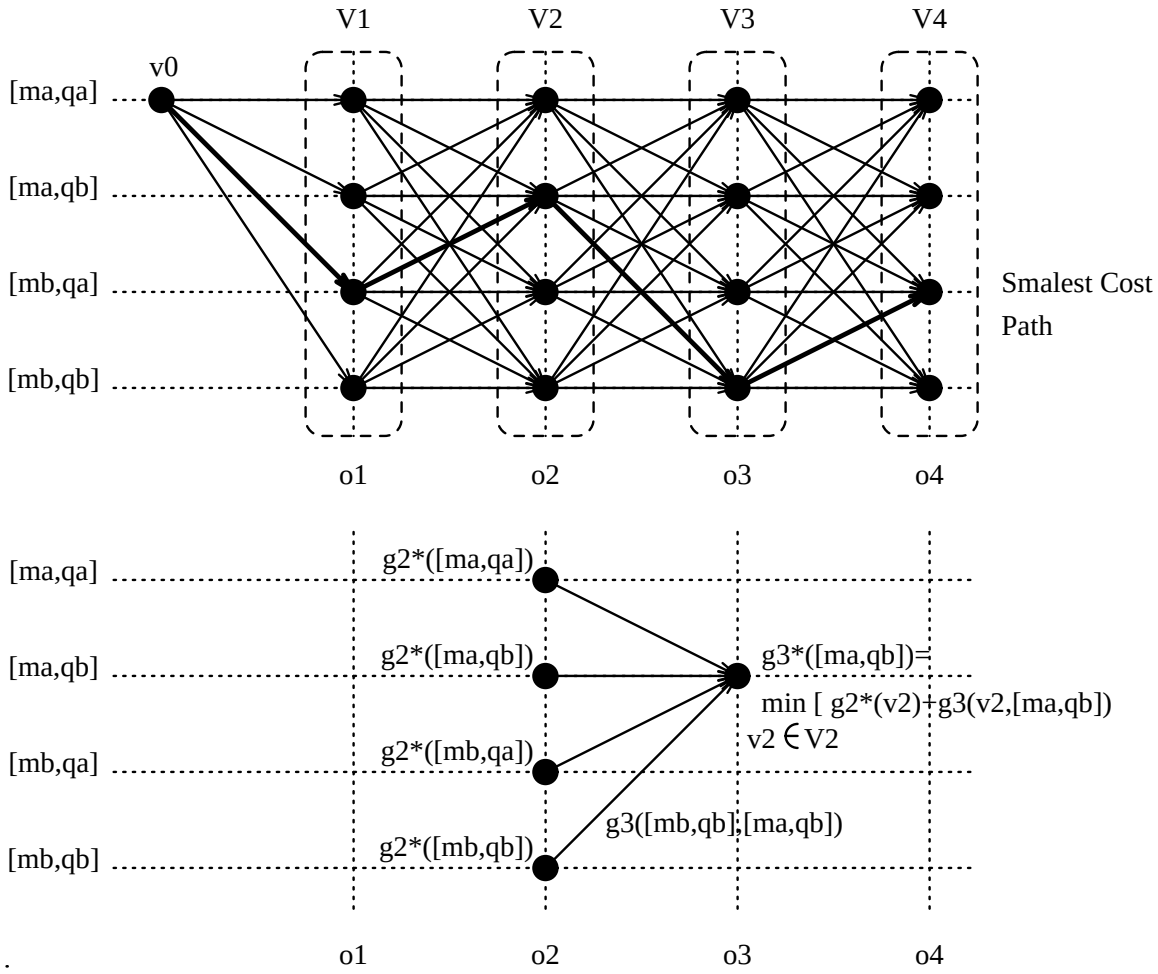

Figure 1: The trellis of the lossless MCVC example

	First frame only		Sequence without first frame		Entire Sequence	
	Bits	PSNR [dB]	Rate [kbits/s]	PSNR [dB]	Rate [kbits/s]	PSNR [dB]
TMN4	18297	33.7	21.0	33.0	23.4	33.0
Matched Distortion	17558	33.7	17.7	33.0	20.0	33.0
Matched Rate	18479	34.1	21.0	33.5	23.4	33.5
Constant Distortion	14514	32.4	15.4	32.4	17.3	32.4

Table 1: Average rate distortion comparison for the "Mother and Daughter" sequence between TMN4 and the proposed coder for different modes of operation

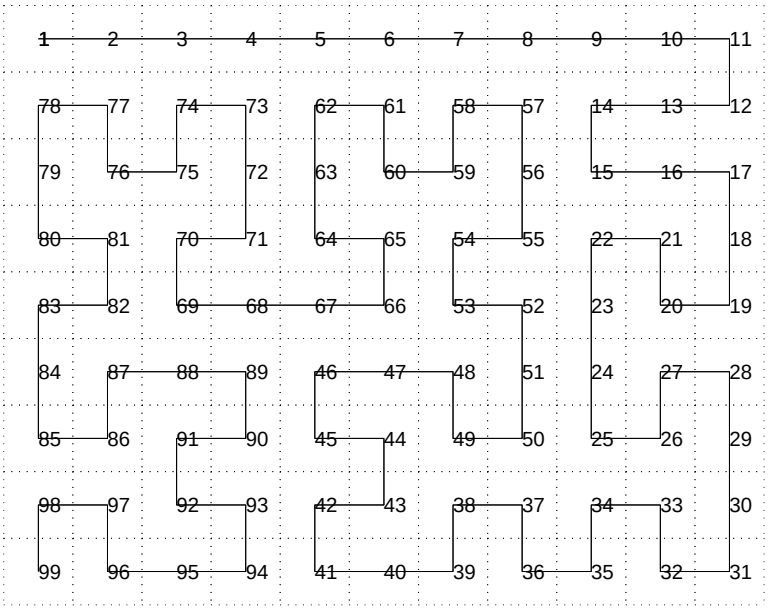

Figure 2: Modified Hilbert scanning curve for TMN4

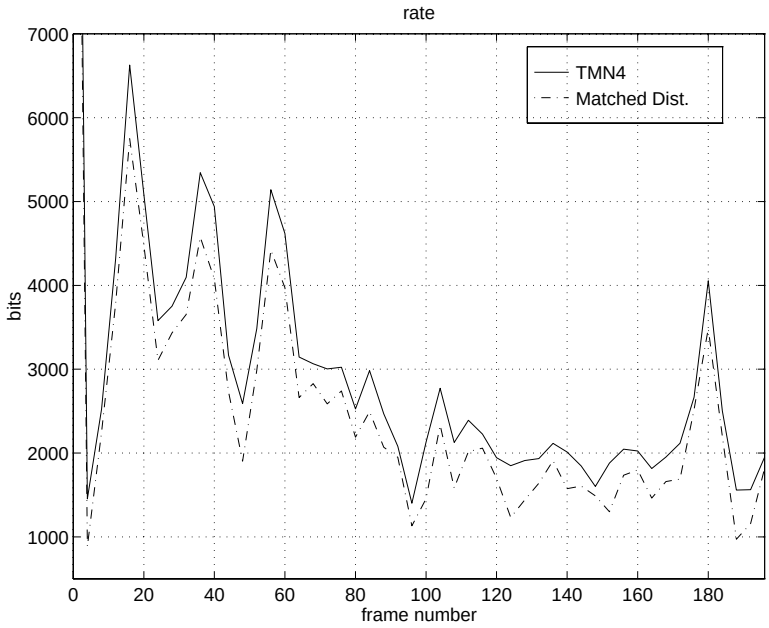

Figure 3: Rate comparison between TMN4 and the proposed coder, where the TMN4 distortion is the target distortion of the proposed coder.

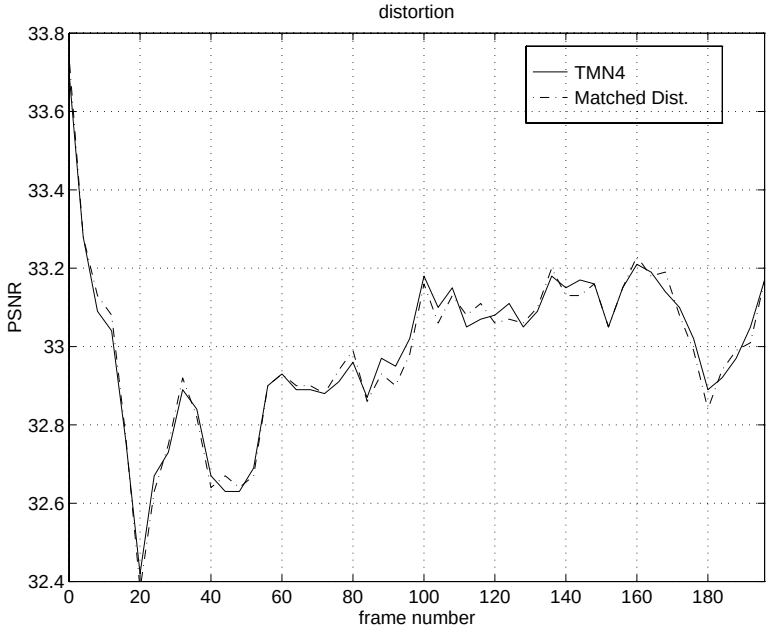

Figure 4: Distortion comparison between TMN4 and the proposed coder, where the TMN4 distortion is the target distortion of the proposed coder.

Figure 5: The 12^{th} reconstructed frame of the "Mother and Daughter" sequence. This frame is used to predict the 16^{th} frame.

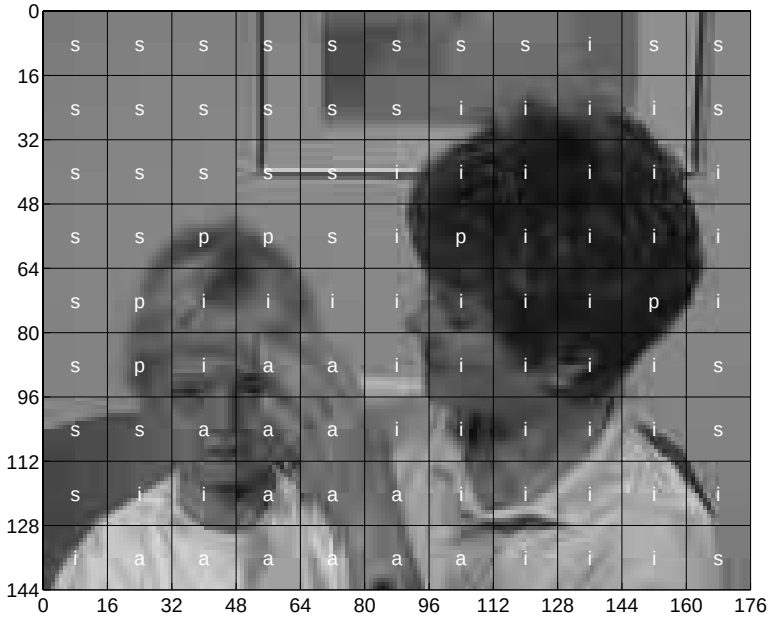

Figure 6: The optimal mode selection for the 16^{th} frame of the "Mother and Daughter" sequence. (i) Inter mode, (s) Skip mode, (p) Prediction mode and (a) Intra mode.

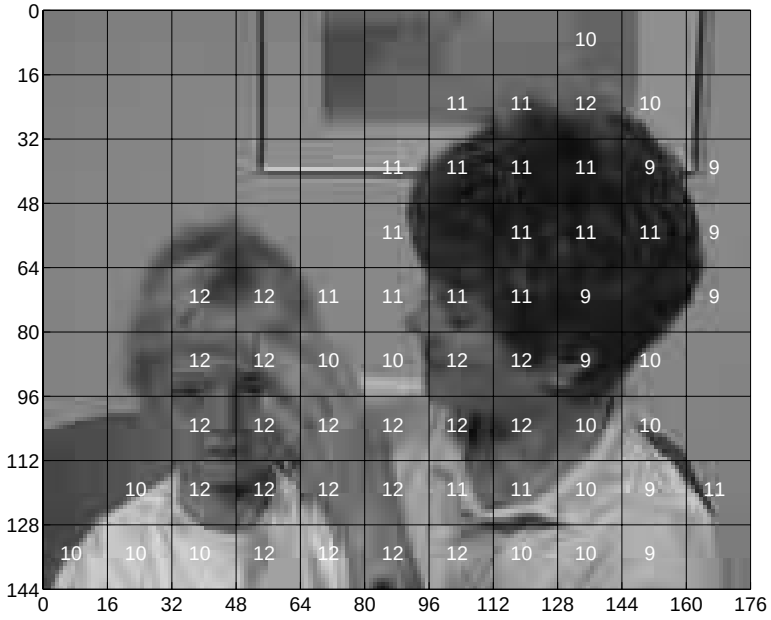

Figure 7: The optimal quantizer selection for the 16^{th} frame of the "Mother and Daughter" sequence. The numbers stand for the quantizer step size QP used for that block.

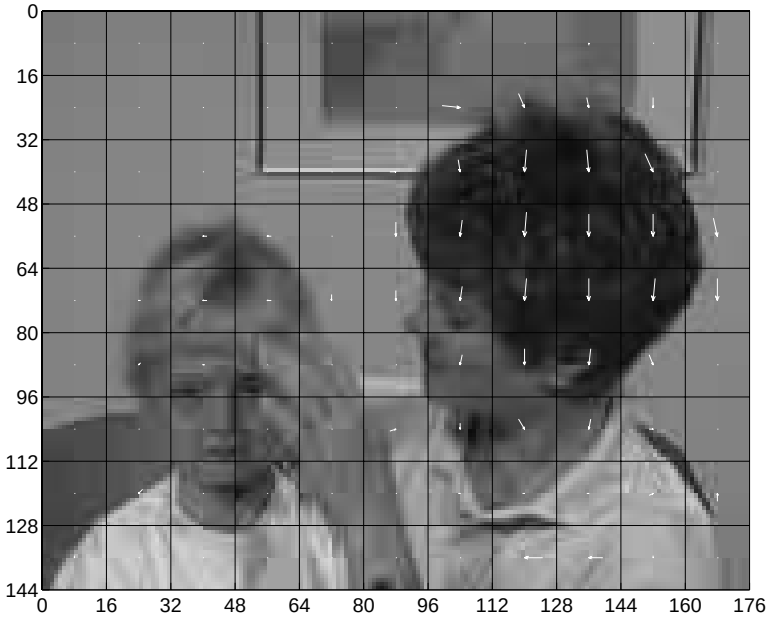

Figure 8: The optimal motion vector field for the 16^{th} frame of the "Mother and Daughter" sequence.

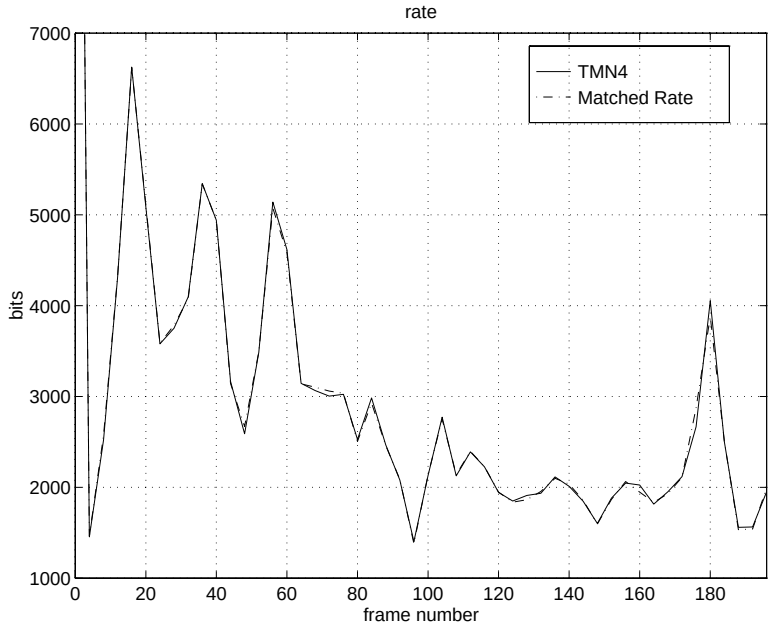

Figure 9: Rate comparison between TMN4 and the proposed coder, where the TMN4 rate is the target rate of the proposed coder.

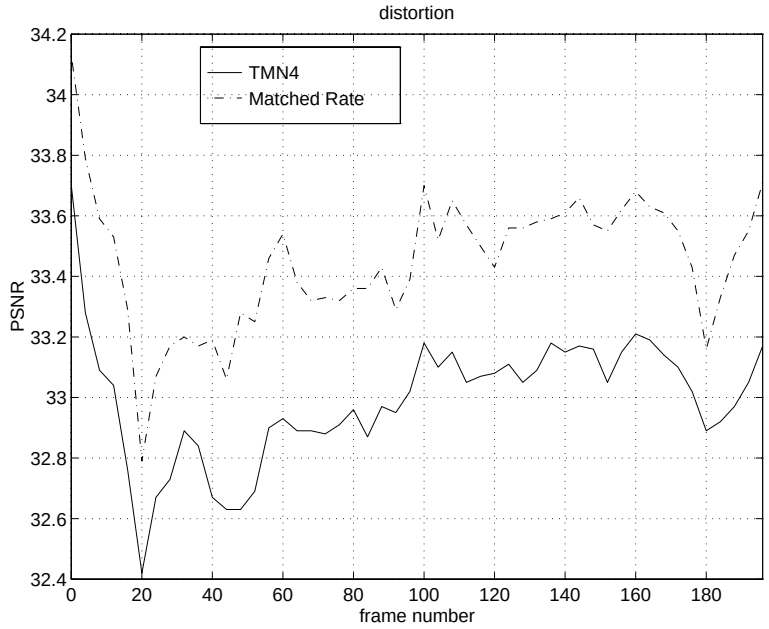

Figure 10: Distortion comparison between TMN4 and the proposed coder, where the TMN4 rate is the target rate of the proposed coder.

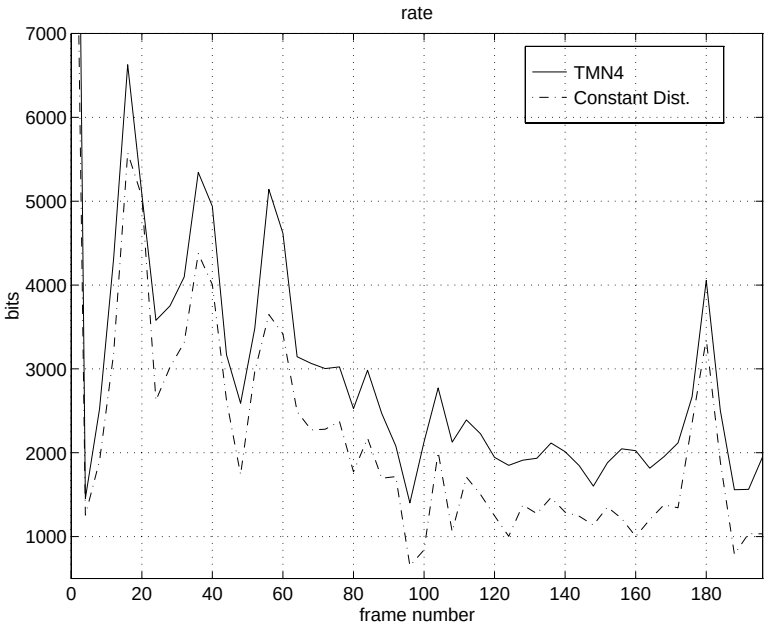

Figure 11: Rate comparison between TMN4 and the proposed coder, where the distortion of the proposed coder is fixed.

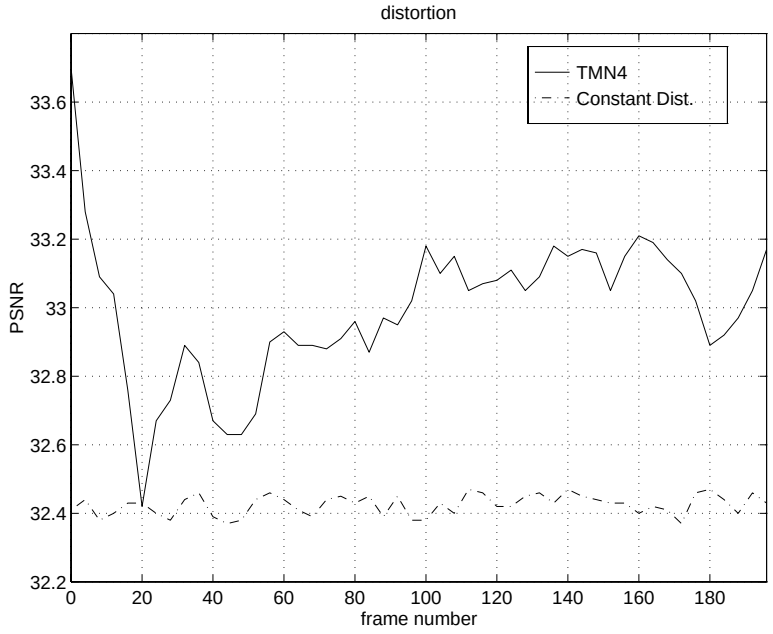

Figure 12: Distortion comparison between TMN4 and the proposed coder, where the distortion of the proposed coder is fixed.

	encoding modes	motion vectors	quantizer step sizes	$\|V_i\|$	average rate kbits/s
TMN4	Intra, Inter Skip	TMN4	10	NA	23.36
matched Distortion	Intra, Inter Skip, Pred.	TMN4	10	4	22.89
matched Distortion	Intra, Inter Skip, Pred.	TMN4 + 8 half-pel neighbors	10	20	20.56
matched Distortion	Intra, Inter Skip, Pred.	TMN4 + 8 half-pel neighbors	8,9,10,11,12	100	20.03
matched Distortion	Intra, Inter Skip, Pred.	TMN4 + 80 half-pel neighbors	8,9,10,11,12	820	19.80

Table 2: Average rate comparison for the "Mother and Daughter" sequence between TMN4 and the distortion matched proposed coder with differently constrained search spaces