A VERY LOW BIT-RATE VIDEO CODEC WITH OPTIMAL TRADE-OFF AMONG DVF, DFD AND SEGMENTATION

*Guido M. Schuster and †Aggelos K. Katsaggelos
Northwestern University, Department of Electrical and Computer Engineering
2145 Sheridan Road, Evanston, Illinois 60208-3118, USA
E-mail: *gschuster@nwu.edu, †aggk@eecs.nwu.edu

ABSTRACT

In this paper we present a theory for the optimal bit allocation among quad-tree (QT) segmentation, displacement vector field (DVF) and displaced frame difference (DFD). The theory is applicable to variable block size motion compensated video coders (VBSMCVC), where the variable block sizes are encoded using the QT structure, the DVF is encoded by first order differential pulse code modulation (DPCM), the DFD is encoded by a block based scheme and an additive distortion measure is employed. We consider the case of a lossless VBSMCVC first, for which we develop the optimal bit allocation algorithm using Dynamic Programming (DP). We then consider a lossy VBSMCVC, for which we use Lagrangian relaxation and show how an iterative scheme, which employees the DP-based solution, can be used to find the optimal solution. We finally present a VBSMCVC, which is based on the proposed theory, which employees a DCT-based DFD encoding scheme. We compare the proposed coder with H.263. The results show that it outperforms H.263 by about 25% in terms of bit rate for the same quality reconstructed image.

1 INTRODUCTION

In video compression, the temporal redundancy is usually reduced by motion compensated prediction of the current frame from a previously reconstructed frame. The spatial redundancy left in the prediction error is commonly reduced by a transform coder or a vector quantizer. One of the drawbacks of most video coders is the fixed block size used for motion estimation, since a real DVF is inherently inhomogeneous. It is well known that the block size controls the tradeoff between the reliability of the DVF estimate, which increases with the size of the block, and the ability to reduce the energy of the DFD, which decreases with the size of the block. Therefore, a variable block size should be used which adapts to the contents of the scene. The nesting of blocks of different sizes in the image plane requires some underlying structure. For this purpose, most VBSMCVC use a QT to segment the image plane [1, 2]. A fundamental problem of a VBSMCVC is the bit allocation between the QT, DFD and DVF. In this paper we present an algorithm which finds the optimal QT decomposition, the optimal DVF and the optimal quantizers, when the DVF is encoded by first order DPCM.

2 NOTATION AND ASSUMPTIONS

Our study of the optimal bit allocation between the QT, DVF and DFD is restricted to the frame level. We assume that a rate control algorithm has set the maximum number of bits available (R_{max}) or the maximum acceptable distortion (D_{max}) for a given frame.

We assume that the current frame is segmented by a QT. The QT data structure decomposes a $2^N \times 2^N$ image (or block of an image) down to blocks of size $2^{n_0} \times 2^{n_0}$. This decomposition results in an $(N-n_0+1)$-level hierarchy ($0 \leq n_0 \leq N$), where all blocks at level n ($n_0 \leq n \leq N$) have size $2^n \times 2^n$. This structure corresponds to an inverted tree, where each $2^n \times 2^n$ block (called a *tree node*) can either be a *leaf*, i.e., it is not further subdivided, or can branch into four $2^{n-1} \times 2^{n-1}$ blocks, each a *child* node. The tree can be represented by a series of bits that indicate termination by a leaf with a "0" and branching into child nodes with an "1". Let $b_{l,i}$ be block i at level l, and the children of this block are therefore $b_{l-1,4*i+j}$, where $j \in [0,1,2,3]$. The QT is denoted by $\mathcal{T}$. Each leaf of $\mathcal{T}$ represents a block which will be used for motion estimation and DFD encoding. For future convenience, let the leafs be numbered from one to the total number of leafs in the QT ($N_{\mathcal{T}}$), from left-to-right.

Let $q_{l,i} \in Q_{l,i}$ be the quantizer for block $b_{l,i}$, where $Q_{l,i}$ is the set of all admissible quantizers for block $b_{l,i}$. Let $m_{l,i} \in M_{l,i}$ be the motion vector for block $b_{l,i}$, where $M_{l,i}$ is the set of all admissible motion vectors for block $b_{l,i}$. Let $s_{l,i} = [l, i, q_{l,i}, m_{l,i}] \in S_{i,l} = \{l\} \times \{i\} \times Q_{l,i} \times M_{l,i}$ be the local state for block $b_{l,i}$, where $S_{l,i}$ is the set of all admissible state values for block $b_{l,i}$. Let $x = [l, i, q, m] \in X = \bigcup_{l=N}^{n_0} \bigcup_{i=0}^{4^{N-l}-1} S_{l,i}$ be the global state and X the set of all admissible state values. Finally let $x_0, \ldots, x_{N_{\mathcal{T}}-1}$ be a global state sequence, which represents the left-to-right ordered leaves of a valid QT $\mathcal{T}$.

We assume that the frame distortion of the reconstructed frame ($D_k(x_0, \ldots, x_{N_\mathcal{T}-1})$) is the sum of the individual block distortions $d(x_j)$, that is,

$$D_k(x_0, \ldots, x_{N_\mathcal{T}-1}) = \sum_{j=0}^{N_\mathcal{T}-1} d(x_j). \quad (1)$$

Most common distortion measures, such as the mean squared error (MSE), a weighted MSE or the peak signal to noise ratio ($PSNR = 10 * log_{10}(255^2/MSE)$) fall into this class. We assume that the DVF is encoded by first order DPCM. In other words, the difference between consecutive motion vectors is encoded using entropy coding. In [3], we have developed a theory for the optimal bit allocation between DFD and DVF for motion compensated video coders with fixed segmentation, which also covers higher order DPCM schemes. Based on the first order DPCM assumption, the frame rate $R_k(x_0, \ldots, x_{N_\mathcal{T}-1})$ can be expressed as follows,

$$R_k(x_0, \ldots, x_{N_\mathcal{T}-1}) = \sum_{j=0}^{N_\mathcal{T}-1} r(x_{j-1}, x_j), \quad (2)$$

where $r(x_{j-1}, x_j)$ is the block rate which depends on the encoding of the current and previous blocks, since the motion vector difference between these two blocks is entropy coded.

Note that the DPCM encoding is along the scanning path, which is different for each QT decomposition. Clearly for the DPCM to be effective, consecutive blocks should be spatial neighbors. We propose that level n_0 is scanned using a Hilbert curve [4] of order $(N - n_0)$ (see Fig. 3, where $(N - n_0) = 3$). Then, for any given QT decomposition, the overall scanning path is derived by merging the blocks of level n_0 according to the Hilbert scan. This results in an overall scanning path where consecutive blocks are always neighboring blocks (again, see Fig. 3). Hence the DPCM of the motion vectors is very efficient.

3 LOSSLESS VBSMCVC

Since the reconstructed frame is identical to the original frame, the frame distortion is equal to zero and the goal is to minimize the required frame rate for QT, DVF and DFD. This can be stated as follows,

$$\min_{x_0, \ldots, x_{N_\mathcal{T}-1}} R_k(x_0, \ldots, x_{N_\mathcal{T}-1}). \quad (3)$$

Since this is a lossless VBSMCVC, the DFD is not quantized, but encoded losslessly. We show next that the dependency between the previous and the current block leads to an optimization problem which can be solved by forward Dynamic Programming (DP) also called the Viterbi algorithm.

3.1 Multilevel trellis

To be able to employ the Viterbi algorithm, a DP recursion formula needs to be established. A graphical equivalent of the DP recursion formula is a trellis where the admissible nodes and the permissible transitions are explicitly indicated. Consider Fig. 1 which represents the multilevel trellis for a 32×32 image block ($N = 5$), with a QT segmentation developed down to level 3 ($n_0 = 3$, block size 8×8). The QT structure is indicated by the white boxes with the rounded corners. These white boxes are not part of the trellis used for the Viterbi algorithm but indicate the set of admissible state values $S_{l,i}$ for the individual blocks $b_{l,i}$. The black circles inside the white boxes are the nodes of the trellis (i.e., the state values $s_{l,i}$). Note that for simplicity, only two trellis nodes per QT node are indicated, but in general, a QT node can contain any number of trellis nodes. The auxiliary nodes, start and termination (S and T) are used to initialize the DPCM of the motion vectors and to select the path with the smallest cost. Each of the trellis nodes represents a different way of encoding the block it is associated with. Since the state of a block is defined to contain its motion vector and its quantizer, each of the nodes contains the rate (and distortion, for the lossy VBSMCVC) occurring by predicting the associated block with the given motion vector and encoding the resulting DFD with the given quantizer. As can be seen in Fig. 1, not every trellis node can be reached from every other trellis node. By restricting the permissible transitions, we are able to force the optimal path to select only valid QT decompositions. Such valid decompositions are based on the fact that at level l, block $b_{l,i}$ can replace four blocks at level $l-1$, namely $b_{l-1,4*i+0}, b_{l-1,4*i+1}, b_{l-1,4*i+2}$ and $b_{l-1,4*i+3}$. As is explained later in this section, the QT encoding cost can be distributed recursively over the QT so that each path picks up the right amount of QT segmentation overhead. Assume that no QT segmentation is used and the block size is fixed at 8×8. In this case, only the lowest level in the trellis in Fig. 1 is used. The transition costs between the trellis nodes would be the rate required to encode the motion vector differences between consecutive blocks along the scanning path. Assume now that the next higher level, level 4, of the QT is included. Clearly the transition cost between the trellis nodes of level 3 stay the same. In addition, there are now transition costs between the trellis nodes of level 4 and also transition cost between trellis nodes of level 3 to trellis nodes of level 4 and vice versa, since each cluster of four blocks at level 3 can be replaced by a single block at level 4. The fact that a path can only leave and enter a certain QT level at particular nodes results in paths which all correspond to valid QT decompositions. Note that every QT node in a path is considered a leaf of the QT which is associated with this path. In this example, a tree of depth 3 has been used to illustrate

how the multilevel trellis is built. For QTs of greater depth, a recursive rule has been derived [5] which leads to the proper connections between the QT levels. In the presented multilevel trellis, the nodes of the respective blocks hold the information about the rate (and in case of a lossy VBSMCVC, the distortion) occurring when the associated block is encoded using the quantizer and motion vector of the node. The rate needed to encode the inhomogeneous motion field is incorporated into the transition cost between the nodes, but so far, the rate needed to encode the QT decomposition has not been addressed. Since the Viterbi algorithm will be used to find the optimal QT decomposition, each node needs to contain a term which reflects the number of bits needed to split the QT at its level. Clearly, trellis nodes which belong to blocks of smaller size have a higher QT segmentation cost than nodes which belong to bigger blocks. When the path includes only the top QT level N, then the QT is not split at all, and only one bit is needed to encode this. Therefore the segmentation cost $A_{N,0}$ equals one. For the general case, if a path splits a given block $b_{l,i}$ then a segmentation cost of $A_{l,i}+4$ bits has to be added to its overall cost function, since by splitting block $b_{l,i}$, 4 bits will be needed to encode whether the four child nodes of block $b_{l,i}$ are split or not. Since the path only visits trellis nodes and not QT nodes, this cost has to be distributed to the trellis nodes of the child nodes of block $b_{l,i}$. How the cost is split among the child nodes is arbitrary since every path which visits a sub-tree rooted by one child node, also has to visit the other three sub-trees rooted by the other child nodes. Therefore the path will pick up the segmentation cost, no matter how it has been distributed among the child nodes. Since the splitting of a node at level n_0+1 leads to four child nodes at level n_0, which can not be split further, no segmentation cost needs to be distributed among its child nodes. In other words, since it is known that the n_0 level blocks cannot be split, no information needs to be transmitted for this event. In Fig. 2 the recursive distribution of the segmentation cost is indicated, where the cost is passed along the left-most child. Having established the multilevel trellis, the Viterbi algorithm can be used to find the optimal state sequence $x_0^*,\ldots,x_{N_\mathcal{T}-1}^*$ which will minimize the frame rate in problem (3). In Fig. 2, a QT of depth 4 is displayed and the optimal state sequence is indicated which leads to the segmentation shown in Fig. 3. Note that consecutive blocks in the overall scanning path are neighboring blocks and the segmentation cost along the optimal path adds up to 13 bits, which is the number of bits needed to encode this QT decomposition. The bit stream for this QT decomposition is "1010000011001".

4 LOSSY VBSMCVC

Clearly for a lossy VBSMCVC 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}. This problem can be formulated in the following way,

$$\min_{x_0,\ldots,x_{N_\mathcal{T}-1}} D_k(x_0,\ldots,x_{N_\mathcal{T}-1}) \quad \text{subject to:} \quad R_k(x_0,\ldots,x_{N_\mathcal{T}-1}) \le R_{max}. \quad (4)$$

This constrained discrete optimization problem is generally very hard to solve. 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 operational rate distortion curve.

We solve this problem using the Lagrangian multiplier method [6]. First we introduce the Lagrangian cost function which is of the following form,

$$J_\lambda(x_0,\ldots,x_{N_\mathcal{T}-1}) = D_k(x_0,\ldots,x_{N_\mathcal{T}-1}) + \lambda * R_k(x_0,\ldots,x_{N_\mathcal{T}-1}), \quad (5)$$

where $\lambda \ge 0$ is called the Lagrangian multiplier. It is well known, that if there is a λ^* such that

$$[x_0^*,\ldots,x_{N_\mathcal{T}-1}^*] = \arg \min_{x_0,\ldots,x_{N_\mathcal{T}-1}} J_{\lambda^*}(x_0,\ldots,x_{N_\mathcal{T}-1}) \quad (6)$$

leads to $R_k(x_0^*,\ldots,x_{N_\mathcal{T}-1}^*) = R_{max}$, then $x_0^*,\ldots,x_{N_\mathcal{T}-1}^*$ is also an optimal solution to (4). When λ sweeps from zero to infinity, the solution to problem (6) traces out the convex hull of the rate distortion curve, which is a non-increasing function. Hence bisection or the fast convex search we presented in [7] can be used to find λ^*.

Therefore the problem at hand is to find the optimal solution to problem (6). The original DP approach can be modified to find the global minimum of problem (6). For a particular λ_p, the DVF cost and the segmentation cost in the multilevel trellis are multiplied by λ_p. In addition, at each node, the DFD distortion is added to the DFD rate which is weighted by λ_p. With these changes, the Viterbi algorithm results in the optimal solution of problem (6). Note that the proposed algorithm is symmetric in the rate and the distortion and hence, it can also be used to find the smallest rate for a given maximum distortion.

5 IMPLEMENTATION AND EXPERIMENTAL RESULTS

In this section we present a specific implementation of the proposed theory, where the lossy encoding of the DFD is accomplished in the DCT domain. In fact the DCT, quantization, run length and entropy coding are exactly the same as in test model four (TMN4) [8] of the H.263 standard. For this implementation we selected $N = 8$ and $n_0 = 3$. Note that the optimal coder writes a bit stream which is uniquely decodable

by our decoder. Hence the listed bit rates are the effective number of bits used. TMN4 was used to encode every 4^{th} frame of the first 200 frames of the QCIF sequence "Mother and Daughter" with a fixed quantizer step size $QP = 10$. This leads to an encoded frame rate of 7.5 frames/second. Since the optimal coder implementation encodes only the luminance component (Y) of the sequence, TMN4 was slightly changed so that it also encodes only the Y channel. The employed distortion measure is the PSNR. The goal of this experiment is to compare the optimal coder with TMN4 in the case where their frame distortions are matched. This can be achieved by setting D_{max}, the maximum frame distortion, equal to the frame distortion of TMN4. Clearly D_{max} changes from frame to frame, following the distortion profile of the TMN4 run. The optimal coder will lead to the smallest number of bits needed to encode a given frame for the given maximum distortion D_{max}. For an average distortion of 33.0 dB PSNR, TMN4 requires an average bit rate of 22.9 kbits/second, whereas the optimal coder uses only 17.1 kbits/second, which is an improvement of over 25%.

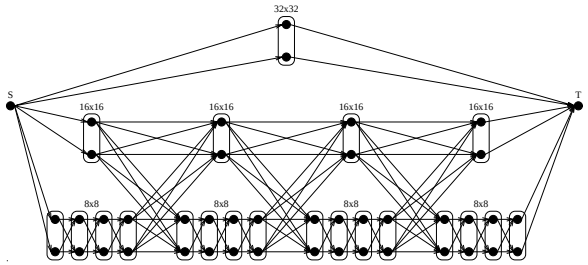

Figure 1: The multilevel trellis for $N = 5$ and $n_0 = 3$

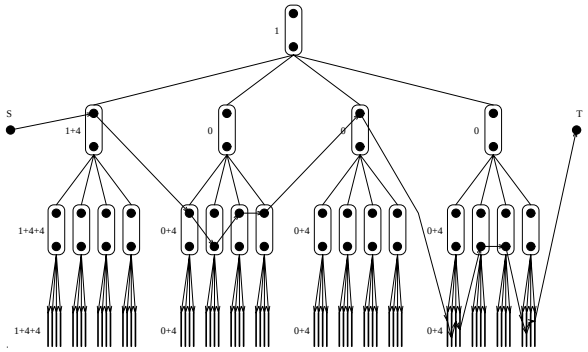

Figure 2: The recursive distribution of the encoding cost

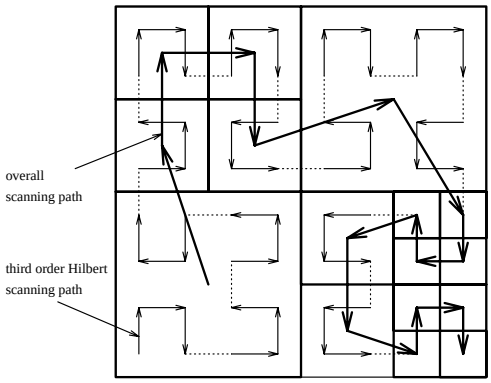

Figure 3: QT decomposition corresponding to Fig. 2

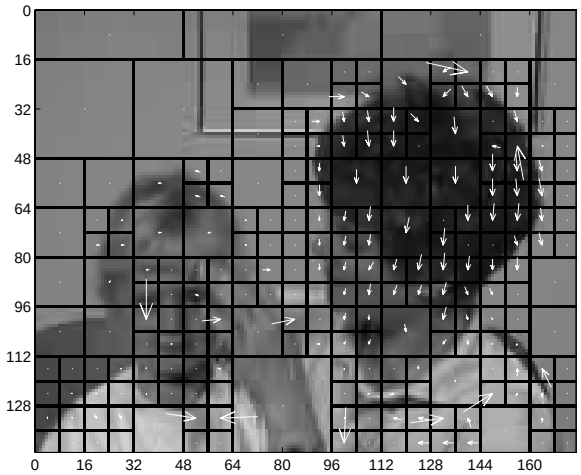

Figure 4: Optimal QT and DVF for the 16^th frame

References

[1] J. Lee, "Optimal quadtree for variable block size motion estimation," in *Proc. ICIP*, vol. 3, pp. 480–483, Oct. 1995.

[2] P. Strobach, "Tree-structured scene adaptive coder," *IEEE Trans. on Com.*, vol. 38, Apr. 1990.

[3] G. M. Schuster and A. K. Katsaggelos, "A video compression scheme with optimal bit allocation between displacement vector field and displaced frame difference," in *Proc. ICASSP*, May 1996.

[4] F. S. Hill, *Computer graphics*. Macmillan Publishing Company, 1990.

[5] G. M. Schuster, *A video compression scheme with optimal bit allocation among segmentation, motion and residual error*. PhD thesis, Northwestern University, Evanston, Illinois, USA, June 1996. Department of Electrical Engineering and Computer Science.

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

[7] G. M. Schuster and A. K. Katsaggelos, "Fast and efficient mode and quantizer selection in the rate distortion sense for H.263," in *Proc. VCIP*, pp. 784–795, SPIE, Mar. 1996.

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