Optimal decomposition for quad-trees with leaf dependencies

Guido M. Schuster

U.S. Robotics Network Systems Div., 8100 N. McCormick Blvd.,
Skokie, Il 60076-2999, gschuste@usr.com

Aggelos K. Katsaggelos

Electrical and Computer Engineering Dept., Northwestern University
Evanston, Il 60202-3118, aggk@ece.nwu.edu

ABSTRACT

We propose a fast and efficient algorithm which finds the optimal quad-tree (QT) decomposition with leaf dependencies in the rate distortion sense. The underlying problem is the encoding of an image by a variable block size scheme, where the block size is encoded using a QT, each block is encoded by one of the admissible quantizers and the quantizers are transmitted using a first order differential pulse code modulation (DPCM) scheme along the scanning path. First we define an optimal scanning path for a QT such that successive blocks are always neighboring blocks. Then we propose a procedure which infers such an optimal path from the QT-decomposition and introduce a special optimal path which is based on a Hilbert curve. Then we consider the case where the image is losslessly encoded using a QT structure and propose a dynamic programming (DP) based multi-level approach to find the optimal QT-decomposition and the optimal quantizer selection. We then apply the Lagrangian multiplier method to solve the lossy case, and show that the unconstrained problem of the Lagrangian multiplier method can be solved using the algorithm introduced for the lossless case. Finally we present a mean value QT-decomposition example, where the mean values are DPCM encoded.

Keywords: Quad-tree decomposition, operational rate distortion theory, mean value decomposition, variable block size coding

1 INTRODUCTION

The problem we are studying in this paper is the encoding of a given frame by a variable block size scheme, where the block size is encoded using a QT, each block is encoded by one of the admissible quantizers for that block and the quantizers are transmitted using a first order DPCM along the scanning path. Several different schemes which are based on the QT-structure have been recently proposed.[1–4] All these approaches employ heuristic rules for finding the quantizers and the QT decomposition, whereas the following two approaches address the problem of optimal QT decomposition and quantizer selection.

In Ref.[5] a general scheme for tree-structured source coding and modeling is presented and one of the applications mentioned is the optimal QT decomposition in the rate distortion sense. In Ref.[6] a scheme is proposed, which is based on the Lagrangian multiplier method, to find the best QT decomposition in the rate distortion sense. Both schemes assume that the quantizers applied to the different blocks are encoded independent of each other and cannot be applied in the case where the quantizers are encoded using a DPCM scheme.

One of the basic reasons why natural images can be compressed efficiently is the existence of high correlation among neighboring pixels. Therefore there is also a high correlation among neighboring quantizers in the QT decomposition. This correlation can be utilized by employing a first order DPCM for encoding the quantizers of neighboring QT-blocks. The DPCM is along the scanning path and hence the selection of the scanning path is very important, as discussed in section 3. By utilizing the high correlation among neighboring blocks, the efficiency of the overall scheme can significantly be improved. For example, in Ref.[7] we represent the estimation of an inhomogeneous displacement vector field (DVF) using a QT, where the motion vectors (which can be considered quantizers) of the QT-blocks are encoded using first order DPCM. Clearly the motion vectors are highly correlated and hence a predictive encoding scheme is imperative for an efficient transmission of the DVF.

The paper is organized as follows. In section 2 we formulate the problem mathematically, introduce the necessary notation and present the underlying assumptions. In section 3 we introduce the definition of an optimal QT scanning path and a procedure to infer such an optimal path for any given QT-decomposition. In section 4 we consider the lossless case and propose an optimal and efficient solution based on a multi-level DP approach. In section 5 we propose an optimal solution for the lossy case which is based on the Lagrangian multiplier method and the multi-level DP approach introduced for the lossless case. In section 6 we introduce an example which is based on a mean value QT-decomposition and finally in section 7 we summarize the paper.

2 PROBLEM FORMULATION

The goal of the proposed algorithm is to optimally and jointly find the QT-decomposition and quantizer assignment for a QT-based image compression scheme. Clearly there is an inherent tradeoff between the rate and the distortion, in that a lower distortion requires a higher rate and a lower rate results in a higher distortion. The most common approach to solve this tradeoff is to minimize the frame distortion D subject to a given maximum frame rate R_{max}. This problem can be formulated in the following way,

$$\min_{x_0,\ldots,x_{N_{\mathcal{T}}-1}} D(x_0,\ldots,x_{N_{\mathcal{T}}-1}), \qquad \text{subject to:} \qquad R(x_0,\ldots,x_{N_{\mathcal{T}}-1}) \leq R_{max}, \tag{1}$$

where the state sequence $x_0,\ldots,x_{N_{\mathcal{T}}-1}$, which contains the choice of the QT-decomposition and the quantizers, is defined in the following paragraphs.

We assume that the frame is segmented by a QT, as in Fig. 1, where the black curve indicates the feature of interest. As can be seen in this figure, 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 a "1" (see Fig. 2).

Let $b_{l,i}$ be the 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 complete tree is denoted by $\mathcal{T}$ and a tree node is identified by the ordered pair (l,i), where again l is the level and i the number within that level (see Fig. 3). This ordered pair is called the index of the tree node. Each leaf of $\mathcal{T}$ represents a particular block in the segmented frame. For future convenience, let the leafs be numbered from zero to the total number of leafs in the QT $(N_{\mathcal{T}})$ minus one, from left-to-right and hence in increasing order of the measure $4^{l-n_0} * i$ (this ordering of the leafs is indicated by italic numbers in Fig. 3).

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 $s_{l,i} = [l, i, q_{l,i}] \in S_{i,l} = \{l\} \times \{i\} \times Q_{l,i}$ 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 $x = [l, i, q] \in X = \bigcup_{l=N}^{n_0} \bigcup_{i=0}^{4^{N-l}-1} S_{l,i}$ 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 $D(x_0, \ldots, x_{N_{\mathcal{T}}-1})$ of the reconstructed frame is the sum of the individual block distortions $d(x_j)$, that is,

$$D(x_0, \ldots, x_{N_{\mathcal{T}}-1}) = \sum_{j=0}^{N_{\mathcal{T}}-1} d(x_j). \tag{2}$$

Most common distortion measures, such as the mean squared error (MSE), a weighted MSE or the peak signal to noise ratio (PSNR $= 10 * \log_{10}(\frac{255^2}{MSE})$) fall into this class. Based on the first order DPCM assumption, the frame rate $R(x_0, \ldots, x_{N_{\mathcal{T}}-1})$ can be expressed as follows,

$$R(x_0, \ldots, x_{N_{\mathcal{T}}-1}) = \sum_{j=0}^{N_{\mathcal{T}}-1} r(x_{j-1}, x_j), \tag{3}$$

where $r(x_{j-1}, x_j)$ is the block rate which depends on the encoding of the current and previous blocks, since the rate required to encode the quantizer of the current block depends on the quantizer selected for the previous block.

3 OPTIMAL SCANNING PATH FOR A QT-DECOMPOSITION

We propose a procedure for inferring a scanning path from a given QT-decomposition. For this procedure, we derive a necessary and sufficient condition which guarantees that consecutive blocks along the scanning path are always neighboring blocks.

Since the QT decomposition is recursive in nature, a recursive version of the raster scan is commonly used to scan the QT in the predefined raster scan pattern which is the same for each level of decomposition. In Fig. 4 an example of this recursive raster scan is shown. Figure 4a through Fig. 4c show the scanning path at different levels and in Fig. 4d the resulting overall scanning path is displayed. Since the goal is to use a first order DPCM scheme along the scanning path, the correlation between successive blocks, and therefore the efficiency of the DPCM, strongly depends on the selected scanning path. The following definition explains how we interpret the concept of a "good" or optimal scanning path.

Definition 3.1. *A scanning path is optimal if it connects only neighboring blocks, i.e., blocks which share an edge, and if it visits each block once and only once.*

The motivation behind this definition comes from the observation that an optimal scanning path will lead to highly correlated successive blocks, since such blocks will always be neighboring blocks. Since the segmentation changes from frame to frame, no fixed scanning path can be employed. On the other hand, no additional information should have to be transmitted to indicate which scanning path was used. Therefore, a good scanning path must be inferable from the transmitted QT and it has to be optimal in the above defined sense.

Assume that the frame is completely segmented into blocks of the smallest size, i.e., $2^{n_0} \times 2^{n_0}$, as shown in Fig. 5. The QT representation of this segmentation is a complete tree, where all leafs are of level n_0. Further assume that the scanning path for this decomposition is known to the encoder and decoder. Let the scanning path for any other QT decomposition be defined recursively, by merging four consecutive blocks along the scanning path at the lower level to form the blocks at the higher level. Note that with this definition, the blocks need not be square. Clearly this recursive definition of the scanning path results in a scanning path which is inferable from the QT decomposition and the scanning path of the fully decomposed QT. The following lemma establishes a necessary and sufficient condition on the level n_0 scanning path, so that, using the above recursive definition of the scanning path for any QT decomposition, the resulting scanning path is always optimal.

LEMMA 3.2. *If the level n_0 scanning path is optimal, and the above recursive procedure is used for the generation of a scanning path for any QT decomposition, then the resulting scanning path is optimal.*

Proof: *Necessary condition:* If the condition is violated, then a completely decomposed tree (down to level n_0) would not result in an optimal scanning path since the level n_0 scanning path is not optimal.
Sufficient condition (by induction): The level n_0 scanning path is optimal. Consider an optimal scanning path. Pick any connected sub path visiting four blocks. Since the overall scanning path is optimal, the sub path is optimal too. By merging the four blocks of the sub path into a single block, the neighboring blocks of the four merged blocks are also neighboring blocks of the new single block. Hence the new scanning path is optimal. ■

Clearly there are many different optimal scanning paths, since there are many different optimal level n_0 scanning paths. One problem with the previous procedure is that an optimal level n_0 scanning path does not guarantee that the resulting blocks are squares for any QT decomposition. In this section we propose an optimal scanning path which results in square blocks, and hence in the kind of QT decomposition displayed in Fig. 1. The optimal scanning path we developed is based on a Hilbert curve. A Hilbert curve has a certain order D and the first order Hilbert curve can have one of four orientations, denoted by 0, 1, 2 and 3. Figure 7 shows the four first order Hilbert curves with the associated second order Hilbert curves drawn beneath it. Note that each second order Hilbert curve consists of four first order Hilbert curves connected by a dotted line and scaled by a factor of $\frac{1}{2}$. This relationship between first and second order Hilbert curves can be generalized using the following proposed algorithm. This algorithm generates a Hilbert curve of order N and shows the close relationship between a QT and a Hilbert curve. Let $O_{l,i} \in [0,1,2,3]$ be the orientation (scanning order of the blocks $b_{l-1,4*i+j}, j \in [0,1,2,3]$) of block $b_{l,i}$ and let $(\cdot)_4$ denote the modulo 4 operation. The recursion needs to be initialized by the desired orientation of the first order Hilbert curve, say $O_{N,0} = 0$ (Figure (a) in Fig. 7). Then we propose to create a Hilbert curve of order N in a Top-Down fashion by calling the function "orient(l,i)" below with the following parameters: orient(N,0).

orient(l, i)
  $O_{l-1,4*i+0} = (5 - O_{l,i})_4$; $O_{l-1,4*i+1} = O_{l,i}$; $O_{l-1,4*i+2} = O_{l,i}$; $O_{l-1,4*i+3} = 3 - O_{l,i}$;
  if $(l - 1 > 1)$
    orient$(l-1, 4*i+0)$; orient$(l-1, 4*i+1)$; orient$(l-1, 4*i+2)$; orient$(l-1, 4*i+3)$;

The resulting orientations at each QT level l, $(N \geq l \geq 1)$ correspond to a Hilbert curve of order $N - l + 1$, hence at level one, the orientations form a Hilbert curve of order N. In Fig. 8 the recursive generation of a third order Hilbert curve is shown. This curve was generated by using the following function call: orient(3,0).

Having defined the relationship between a Hilbert curve and a QT and discussed some properties of the Hilbert curve, it is clear that when the level n_0 scanning path is a Hilbert curve of order $N - n_0$, then the resulting scanning path is optimal, regardless of the QT decomposition and the resulting blocks are square. Besides resulting in square blocks, the Hilbert scan has also the advantage that it is spatially non-disruptive and hence tends to preserve local pixel correlations better than a raster scan.

4 OPTIMAL QT-DECOMPOSITION FOR LOSSLESS SCHEMES

In order to introduce the algorithm systematically, we first consider the problem when different lossless schemes are used to encode the blocks, instead of the quantizers. 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. This can be stated as follows,

$$\min_{x_0,\ldots,x_{N_{\mathcal{T}}-1}} R(x_0,\ldots,x_{N_{\mathcal{T}}-1}). \tag{4}$$

Since this is a lossless scheme, the blocks are not quantized, but encoded losslessly. With a slight abuse of notation, let $q_{l,i}$ represent different lossless encoding schemes for block $b_{l,i}$ (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 $q_{l,i}$'s quantizers in the following derivation.

Since we deal with a finite number of QTs and admissible quantizers, the above optimization problem can be solved by an exhaustive search. The time complexity for such an exhaustive search is extremely high. It can be shown that a lower bound for the number of different QTs is given by $2^{4^{N-n_0-1}}$, where $(N-n_0) \geq 1$. To find a lower bound on the total time complexity we consider the case when the QT is completely decomposed, i.e., the image is segmented into 4^{N-n_0} blocks of size $2^{n_0} \times 2^{n_0}$. We further assume that the cardinality $|S_{n_0,i}|$ of all admissible local states $S_{n_0,i}$ of the leafs is the same. Hence an exhaustive search, using only this single QT requires $|S_{n_0,i}|^{4^{N-n_0}}$ comparisons. Since we have only considered one QT, this is a lower bound on the time complexity for all QTs hence $\Omega_E(|S_{n_0,i}|^{4^{N-n_0}})$, where E stands for exhaustive search. We will show that the upper bound of the proposed algorithm is significantly smaller than this lower bound for an exhaustive search. 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.

The form of the objective function (see Eq. (3)) of the optimization problem of Eq. (4) suggests that DP should be used to find the optimal solution efficiently. To be able to employ forward DP (the Viterbi algorithm), a DP recursion formula needs to be established. A graphical equivalent of the DP recursion formula (for first order neighborhoods, i.e., first order DPCM) is a trellis where the admissible nodes and the permissible transitions are explicitly indicated. Consider Fig. 9 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 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 level and number within that level (which identifies the blocks size and its position in the frame), and its quantizer, each of the nodes contains the rate (and distortion, for the lossy scheme) occurring encoding the associated block with the given quantizer.

As can be seen in Fig. 9, 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 we will see 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. 9 is used. The transition costs between the trellis nodes would be the rate required to encode the DPCM 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 from 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 the presented multi level trellis, the nodes of the respective blocks hold the information about the rate (and in the case of a lossy scheme, the distortion) occurring when the associated block is encoded using the quantizer of the node. The rate needed to encode the DPCM encoding cost 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. These segmentation costs can be generated recursively in a Top-Down fashion by calling the function "seg(l,i)" below with the following parameters: seg(N,0). The recursion needs to be initialized with $A_{N,0}=1$.

seg(l,i)
  $A_{l-1,4*i+0}=A_{l,i}+4$; $A_{l-1,4*i+1}=0$; $A_{l-1,4*i+2}=0$; $A_{l-1,4*i+3}=0$;
  if $(l-1>n_0)$
    seg$(l-1,4*i+0)$; seg$(l-1,4*i+1)$; seg$(l-1,4*i+2)$; seg$(l-1,4*i+3)$;

Note that in the above function, the segmentation cost is distributed along the leftmost child. As mentioned before, any other assignment of the segmentation cost will lead to the same result. The recursion involved in the assignment of the encoding cost is illustrated in Fig. 10.

Having established the multi level trellis, the forward DP 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 (4). In Fig. 10, a QT of depth 4 is displayed and the optimal state sequence is indicated which leads to the segmentation shown in Fig. 11. Note that the resulting scanning path is optimal 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". The time complexity of DP is $O_{DP}(4^{N-n_0}*|S_{n_0,i}|^2)$ which is significantly smaller than the lower bound for the exhaustive search $\Omega_E(|S_{n_0,i}|^{4^{N-n_0}})$. Note that we have assumed that all the $S_{l,i}$ sets are of the same cardinality.

5 OPTIMAL QT-DECOMPOSITION FOR LOSSY SCHEMES

So far we have considered lossless schemes. In this section we will apply the algorithms derived for a lossless scheme to find the optimal solution to a lossy scheme. Clearly for a lossy scheme it does not make sense to minimize the frame rate R with no additional constraints, since this would lead to a very high frame distortion D. The most common approach to solve the tradeoff between the frame rate and the frame distortion is to minimize the frame distortion D subject to a given maximum frame rate R_{max}. This problem can be formulated in the following way,

$$\min_{x_0,\ldots,x_{N_\mathcal{T}-1}} D(x_0,\ldots,x_{N_\mathcal{T}-1}), \qquad \text{subject to:} \qquad R(x_0,\ldots,x_{N_\mathcal{T}-1}) \le R_{max}. \tag{5}$$

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. Since these solutions tend to be dense, this is a very good approximation. We solve this problem using the Lagrangian multiplier method.[8] First we introduce the Lagrangian cost function which is of the following form,

$$J_\lambda(x_0, \ldots, x_{N_\mathcal{T}-1}) = D(x_0, \ldots, x_{N_\mathcal{T}-1}) + \lambda * R(x_0, \ldots, x_{N_\mathcal{T}-1}), \tag{6}$$

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

$$[x_0^*, \ldots, x_{N_\mathcal{T}-1}^*] = \arg \min_{x_0, \ldots, x_{N_\mathcal{T}-1}} J_{\lambda^*}(x_0, \ldots, x_{N_\mathcal{T}-1}) \tag{7}$$

leads to $R(x_0^*, \ldots, x_{N_\mathcal{T}-1}^*) = R_{max}$, then $x_0^*, \ldots, x_{N_\mathcal{T}-1}^*$ is also an optimal solution to (5). When λ sweeps from zero to infinity, the solution to problem (7) 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[9] can be used to find λ^*. Therefore the problem at hand is to find the optimal solution to problem (7). The DP approach presented in the previous section can be modified to find the global minimum of problem (7). For a given λ, we replace the rate $r(x_{j-1}, x_j)$ by the following expression,

$$g(x_{j-1}, x_j) = d(x_j) + \lambda * r(x_{j-1}, x_j). \tag{8}$$

With this modification, the DP algorithm presented in section 4 leads to the optimal solution of problem (7).

6 EXAMPLE

In this section we use the mean value QT decomposition as an example of how the above algorithm finds the optimal quantizers and the optimal QT decomposition simultaneously. In the mean value QT decomposition, the frame is represented by differently sized blocks of a fixed luminance value. Usually it is implied that the luminance value is the quantized mean value of the original block. Conceptually, though, it could be any value, but clearly the distortion (we use the MSE) increases if another value than the mean is selected. Since these values are encoded using a DPCM scheme where small prediction errors are encoded using short code words, it might be beneficial, in a rate distortion sense, to select a value slightly different than the mean value, since this might result in a large rate drop relative to the small distortion introduced. In this example, we represent each luminance value as an integer between zero and 255. Therefore the set of admissible quantizers for each block is of size 256. On the other hand, if the value is not selected to be close to the actual mean, the resulting distortion is quite large, and hence the saving in rate cannot offset the increase in distortion. Let the mean of block $b_{l,i}$, rounded to the nearest integer, be denoted by $m_{l,i}$. We define the set of admissible quantizers for block $b_{l,i}$, $Q_{l,i} = \{m_{l,i} - 5, m_{l,i} - 4, \ldots, m_{l,i} + 5\}$. In other words, a block can be represented by its mean ± 5. Clearly for each of the different values a different distortion occurs.

Table 1 shows a comparison between different mean value only encoding schemes, for the first luminance frame of the QCIF (176×144) "Mother and Daughter" sequence. The schemes in the first three rows segment the frame into 8×8, 4×4 or 2×2 blocks of which only the mean values, rounded to the nearest integer, are transmitted. In the "PCM rate" column, the resulting bit rate is stated when 8 bits per mean value are used. In the "DPCM rate" column, a first order DPCM scheme is employed, which results in a smaller bit rate. The proposed optimal mean value QT decomposition with DPCM encoding of the luminance values finds the QT decomposition and the luminance values which result in the smallest distortion for a given bit rate. We apply this algorithm to the mean value QT decomposition, where the biggest block size is 16×16 and the smallest block size is 2×2. Therefore the original QCIF frame is represented by 99 QTs, since there are 99 different 16×16 blocks. In the experiment displayed in row four of Table 1, we set the distortion equal to the distortion of the 4×4 fixed block size implementation. Using the optimal bit allocation algorithm, the distortion can be exactly matched and the required rate is 5049 bits, which is 44% fewer bits than the fixed block size implementation using the same DPCM encoding (9080 bits) and 60% fewer bits than the fixed block size implementation using pulse code modulation (PCM) encoding (12672 bits). This clearly shows the efficiency of the proposed algorithm.

With a quantizer step size equal to 10, H.263 uses 15000 bits to encode the above frame. Using the optimal mean value QT scheme and fixing the rate to be equal to the H.263 rate, the resulting distortion is 29.52 dB PSNR. In Fig. 12 the result of this experiment is displayed and the segmentation of the frame is shown in Fig. 13. In Fig. 12, the frame appears somewhat blocky, since we do not use a reconstruction filter as the ones employed in,[1,2,4] which can improve the reconstructed quality tremendously. Nevertheless, for a simple mean value scheme, the result is quite impressive.

7 SUMMARY

In this paper we presented an optimal and efficient algorithm for jointly selecting the QT decomposition and the block quantizers for a QT-decomposition with leaf dependencies. The algorithm is based on a multi-level DP approach and the Lagrangian multiplier method. We introduced a procedure for inferring optimal QT-scanning paths and proposed a special optimal scanning path which is based on a Hilbert curve. This algorithm is very general in nature and has been applied to the very efficient motion estimation,[7] the development of motion compensated video coders[10] and in this paper to a very efficient mean value QT-decomposition.

8 REFERENCES

[1] P. Strobach, "Quadtree-structured recursive plane decomposition coding of images," *IEEE Transactions on Signal Processing*, vol. 39, pp. 1380–1397, June 1991.

[2] J. Vaisey and A. Gersho, "Image compression with variable block size segmentation," *IEEE Transations on Signal Processing*, vol. 40, pp. 2040–2060, Aug. 1992.

[3] X. Zhang, M. C. Cavenor, and J. F. Arnold, "Adaptive quadtree coding of motion compensated image sequences for the use on the broadband ISDN," *IEEE Transactions on Circuits and Systems*, vol. 3, pp. 222–229, June 1993.

[4] E. Shusterman and M. Feder, "Image compression via improved quadtree decomposition algorithms," *IEEE Transactions on Image Processing*, vol. 3, pp. 207–215, Mar. 1994.

[5] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Optimal pruning with applications to tree-structured source coding and modeling," *IEEE Transactions on Information Theory*, vol. 35, pp. 299–315, Mar. 1989.

[6] 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.

[7] G. M. Schuster and A. K. Katsaggelos, "An optimal quad-tree-based motion estimator," in *Proceedings of the Conference on Digital Compression Technologies & Systems for Video Communications*, pp. 50–61, Oct. 1996.

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

[9] 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.

[10] G. M. Schuster and A. K. Katsaggelos, "A very low bit rate video codec with optimal trade-off among DVF, DFD and segmentation," in *Proceedings of the European Signal Processing Conference*, (Trieste, Italy), pp. 439–442, Sept. 1996. invited paper.

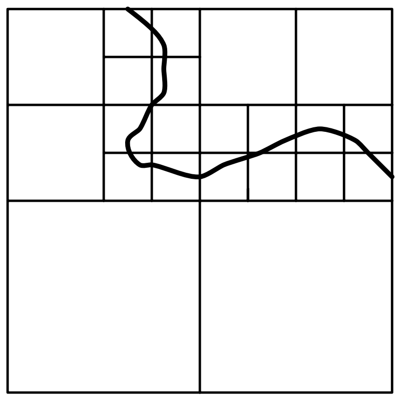

Figure 1: Frame segmented by a quad-tree.

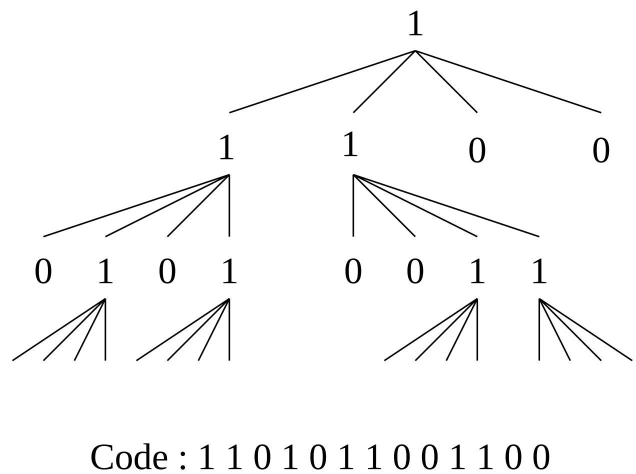

Figure 2: Quad-tree representation of the frame.

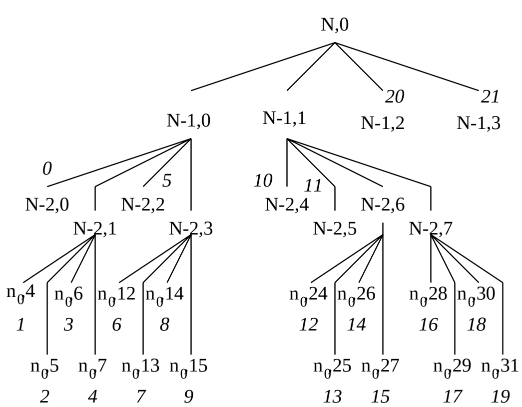

Figure 3: Quad-tree notation. Each node is identified by the ordered pair (l, i), where l is the level and i the number within that level. Also the leafs are numbered from zero to the total number of leafs in the QT minus one.

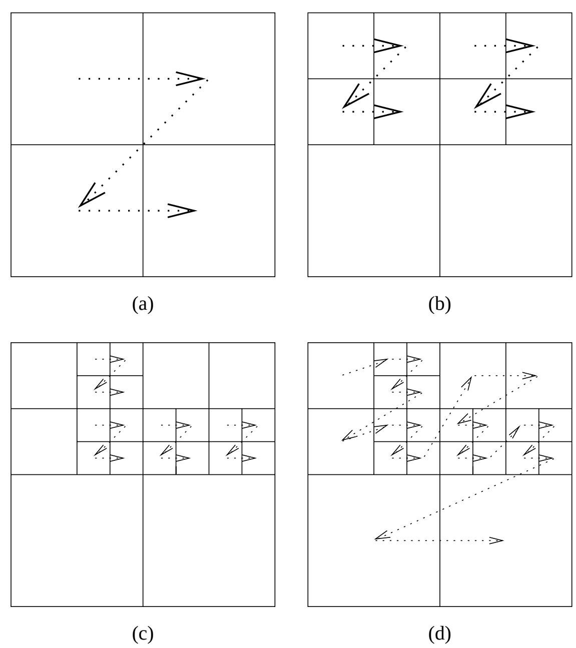

Figure 4: Recursive raster scan for a QT decomposition. Figures (a) through (c) show the raster scan at different levels of the QT and figure (d) shows the resulting overall scanning path.

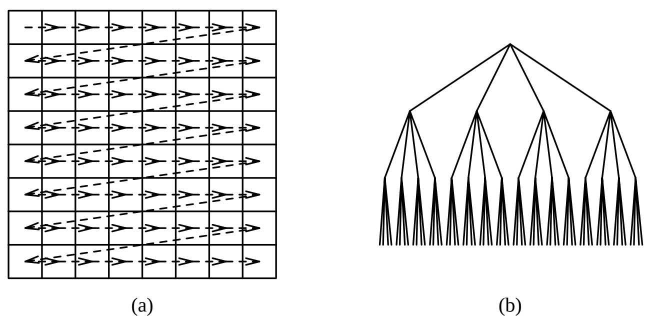

Figure 5: Completely decomposed quad-tree. Level n_0 is scanned with a raster scan.

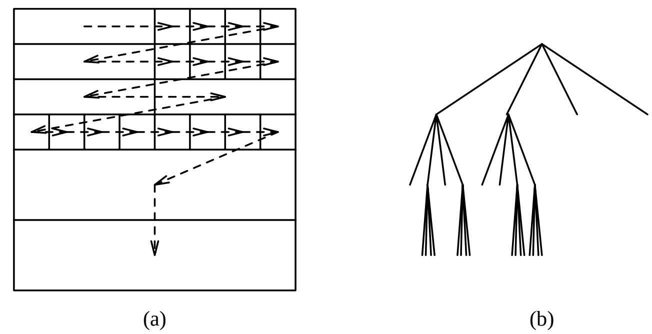

Figure 6: Recursive definition of the scanning path.

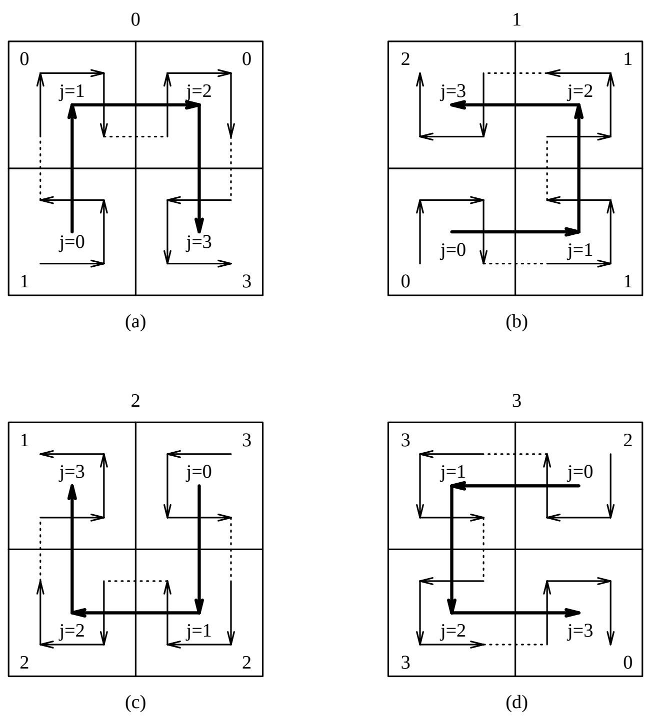

Figure 7: The four possible first order Hilbert curves (-) with the respective second order Hilbert curves underneath (-). The orientations $O_{l,i}$ are shown on top of the squares and the orientations $O_{l-1,4*i+j}$ are shown in the corners of the smaller squares.

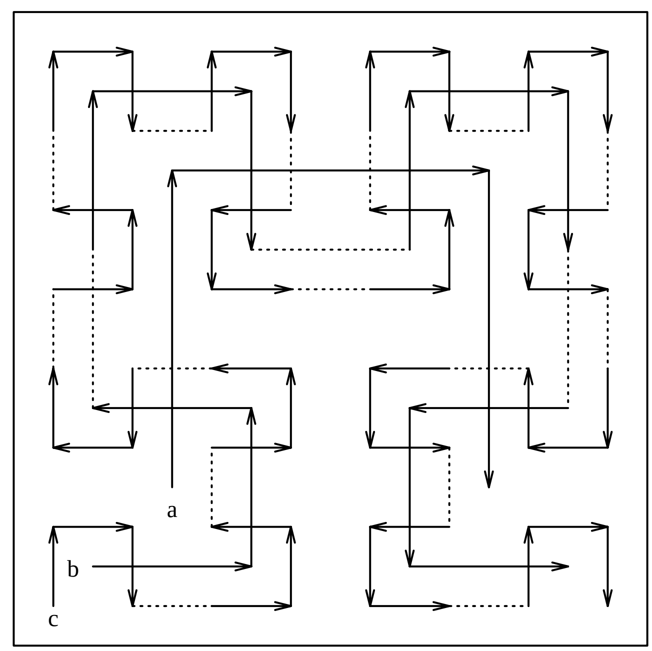

Figure 8: Recursive Hilbert curve generation: a) first order Hilbert curve of orientation 0, b) second order Hilbert curve, c) third order Hilbert curve

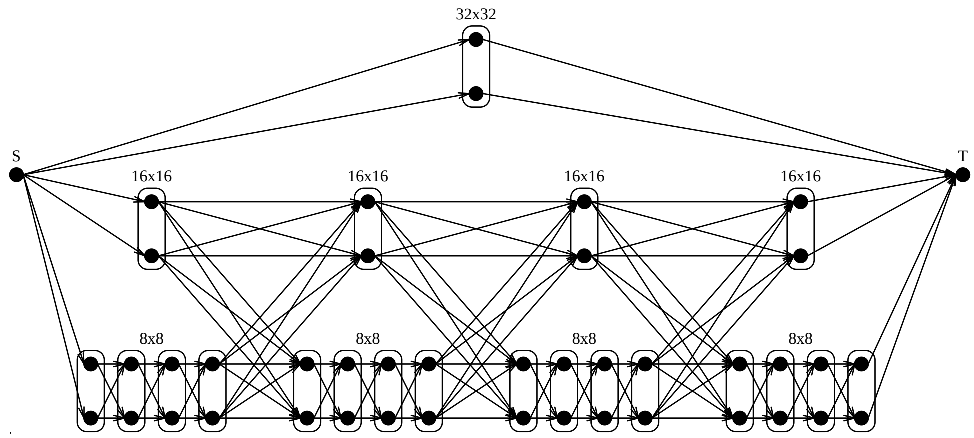

Figure 9: The multilevel trellis for $N = 5$ and $n_0 = 3$.

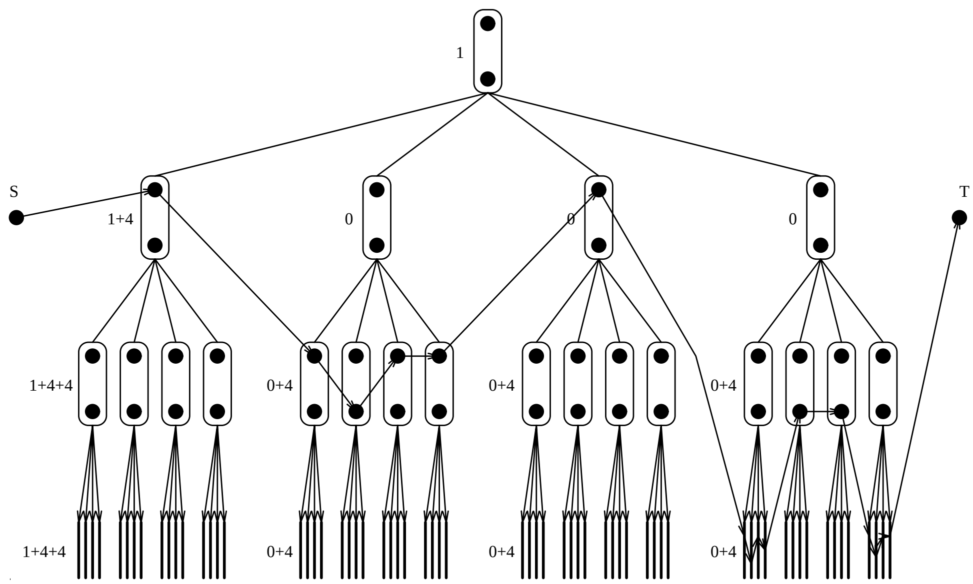

Figure 10: The recursive distribution of the quad-tree encoding cost among the trellis nodes and an optimal path through a quad-tree of depth 4.

Block size	PCM rate (bits)	DPCM rate (bits)	Distortion (PSNR)
8×8	3168	2560	23.5
4×4	12672	9080	26.3
2×2	50688	32573	30.2
$16 \times 16 \to 2 \times 2$	NA	5049	26.3

Table 1: Comparison between mean value only encoding schemes for different, block sizes, using the first luminance frame of the QCIF (176×144) "Mother and Daughter" sequence.

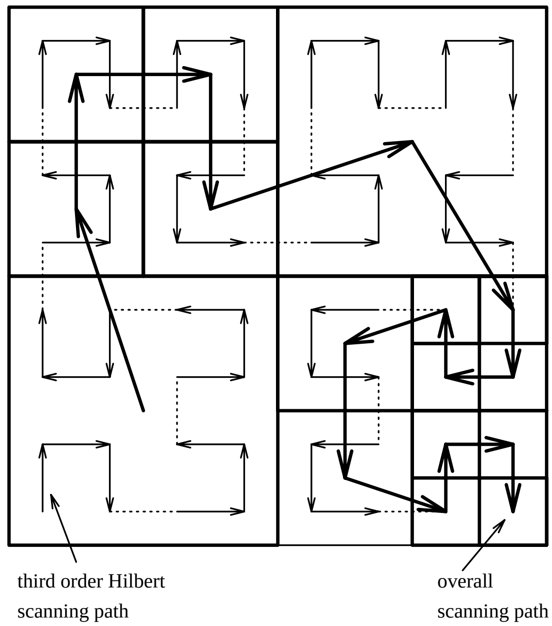

Figure 11: Quad-tree decomposition corresponding to the optimal state sequence

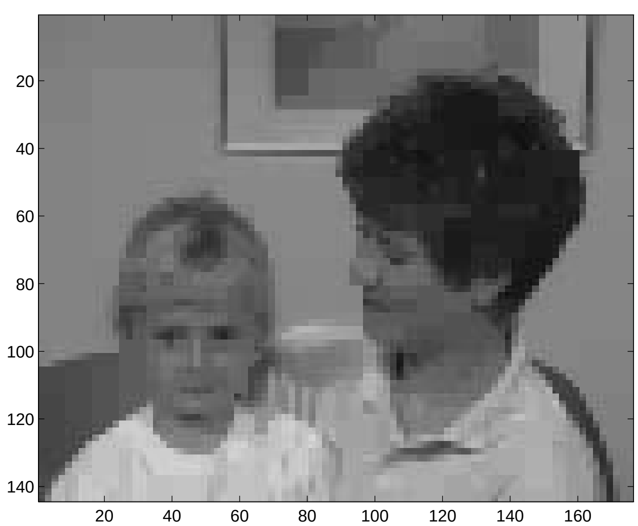

Figure 12: First luminance frame of the "Mother and Daughter" sequence encoded by the optimal mean value QT decomposition. The resulting bit rate is 14996 bits and the resulting distortion is 29.52 dB PSNR.

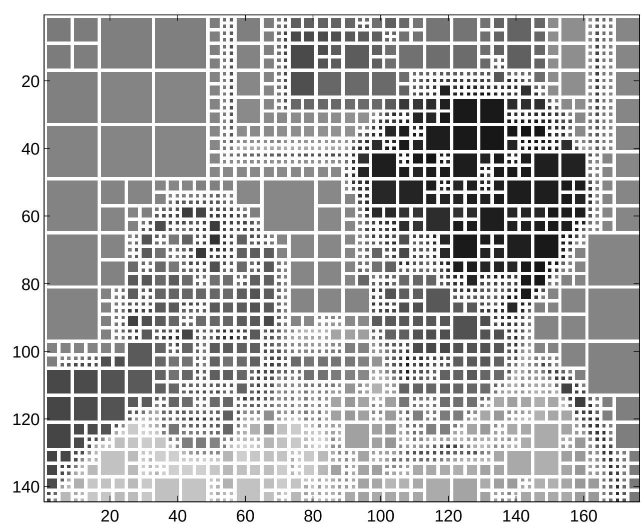

Figure 13: Segmentation of the first luminance frame of the "Mother and Daughter" sequence encoded by the optimal mean value QT decomposition.