RATE-COMPLEXITY-DISTORTION OPTIMIZATION FOR QUADTREE-BASED DCT CODING

*Krisda Lengwehasatit**

PacketVideo Corp.
Core Engineering Division
10350 Science Center Drive
San Diego, CA 92121
krisda@packetvideo.com

Antonio Ortega†

Integrated Media Systems Center
Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, California 90089-2564
ortega@sipi.usc.edu

ABSTRACT

In this paper, we study rate-complexity-distortion (R-C-D) tradeoffs for video coding, where we focus on the complexity of computing the inverse DCT. A quadtree coding approach is used and the quadtree is optimized based on the constraints of not only a rate budget but also decoding complexity budget. We employ a variable complexity algorithm (VCA) for the IDCT in order to obtain better complexity result. The main novelty of this work is to demonstrate that VCA approaches combined with R-C-D optimization can provide better results (e.g., lower rate at same distortion and complexity) than approaches proposed in the past, which relied on fixed blocksize and fixed complexity techniques.

1. INTRODUCTION

Rate-distortion (R-D) based bit allocation techniques (e.g., [1],[2]) are tools to effectively achieve good decoded video quality for a given target bitrate. However, there are an increasing number of applications where it is also necessary to take into account the complexity of the codec and therefore the question of how to use bits effectively has to be attacked under constraints on the encoder or decoder complexity. While complexity has been considered explicitly by previous work, most of the proposed complexity optimization techniques are either ad hoc, in that they use simplified complexity models, or they do not provide sufficient computation scalability.

Rate-complexity-distortion (R-C-D) tradeoffs in the context of image coding have been addressed for example by Gormish [3] and Goyal et al. [4], where in both cases transform coding is considered. In order to select the level of

decoding complexity, the encoder can choose a given block size for the transform for all the blocks in the image. Then it is possible to reduce the decode complexity by using a smaller transform for all blocks, at the cost of some reduction in coding performance. While R-C-D tradeoffs are considered the proposed schemes provide very few levels of granularity in determining the decoder complexity.

Instead, in this paper, we present a coding scheme where the transform blocksize can be adaptively chosen based on a quadtree structure. Moreover, the IDCT is computed using a variable complexity algorithm (VCA) [5]. A VCA performs the required task in an input-dependent manner. The key idea is to classify the input space into several classes. For each class, a reduced algorithm obtained by pruning unnecessary operations is used. Therefore, different inputs require different amount of computations to finish the task. Specifically, in the IDCT case the decoder tests for the position of zeros in the data to be transformed, and skips the inverse transform computation for those coefficients that are zero. Given the cost of classification overhead, the worst case complexity of a VCA is worse than the fixed complexity case. However, through careful selection of the classification tests a VCA will have average case complexity that is no worse than that of its fixed complexity counterpart. Therefore our results will be shown to be better in terms of R-C-D performance than previously reported work which fixed complexity algorithms.

In this paper we extend our earlier work [6]. Our previous work considered fixed block size IDCTs, and computation scalability was achieved solely by increasing the coarseness of the quantization (i.e., coarser quantization leads to more zero DCT coefficients and therefore to faster inverse VCA transform.) Here, we extend this approach to allow for variable block sizes and we use a different VCA implementation of the IDCT. We employ a quadtree-based coding as in [7] and introduce a quadtree optimization that takes into account rate, distortion and complexity. Thus our goal will

be, for example, to minimize distortion for given constraints on rate and complexity. The Lagrange multiplier method [8] is used to obtain to set up the cost of each possible quadtree. Since there are two constraints, two Lagrange multipliers have to be used. The optimal tree pruning algorithm in [9] is used with this Lagrangian cost, and the search in the 2D space of Lagrange multipliers is done using the technique proposed in [10]. The resulting coding technique allows the encoder to select the decoder complexity with a fine level of granularity and therefore our results are superior, in R-C-D terms, to comparable work reported in the literature [4, 3].

2. VCA FOR IDCT

In our previous work [5], we proposed a VCA IDCT where a tree-structured classification was used to determine which DCT coefficients were zero and to skip operations involving zero-valued coefficients. This tree-structure was tightly coupled to the structure of the baseline IDCT algorithm [11] so that the sets of coefficients tested together were also used together in one of the steps of the IDCT calculation. This VCA approach was applied successively to rows and columns of each block in order to compute the 2D IDCT. The classification was optimized separately for rows and columns.

In this paper, we follow a different approach and design a "dyadic" classification approach, where the test structure is chosen based on the observation that for typical video data the high frequency components are more likely to be quantized to zero. Our dyadic classification approaches takes an NxN DCT block and sequentially tests to determine whether (i) all coefficients are zero (all zero class), or (ii) only the DC coefficient is non-zero (DC-only class), or (iii) is such that all the coefficients outside of the lowest nxn frequency ones are zero (low-nxn classes, for $n = 2, 4, 8, \ldots N/2$.) The full-NxN class corresponds to having at least one non-zero coefficient outside of the lowest $\frac{N}{2}$x$\frac{N}{2}$ frequency coefficients. Since the classes listed above have increasing complexity (i.e., the all zero class leads to no computation being needed, whereas the full–NxN requires the normal DCT to be performed), we structure the test by testing for all-zero class before the DC-only class, which in turn comes before the low-2x2 class and so on. We start by using the all-zero test because we consider here a video coding application and motion compensated frames tend to have many all-zero blocks. The all-zero test may not be used in an image coding application. Note also that the proposed testing order follows the fact that for video coding classes with few non-zero coefficients are more likely.

In this paper, we use the algorithm from [11] as baseline. For each of the classes tested, a corresponding separable 1-D reduced IDCT (RIDCT) is applied along the nonzero rows and then all columns to get 2-D DCT. The RIDCT is obtained from pruning the baseline algorithm. Note that by using the algorithm in [11] the low-n RIDCT will have

the same complexity as size n IDCT. We will use the VCA IDCT based on dyadic classification for each block in the decoded image, where the size of the DCT used, i.e., N, is determined by the quadtree optimization discussed next.

3. QUADTREE OPTIMIZATION

The quadtree structure used here (see Figure 1) has, for simplicity, a limit on largest and smallest block size. The leaves (smallest blocks) correspond to a block of size 4x4, the middle nodes represent 8x8 blocks and the root of each tree is of size 16x16. We compute the motion estimation with a fixed block size so that we can assume that the motion vector overhead is fixed and not affected by the choice of DCT size. In coding, a DCT of the corresponding size is applied to each leaf of the quadtree structure. Then the DCT coefficients are quantized, zigzag scanned and entropy coded. The structure of the quadtree is sent as a side information.
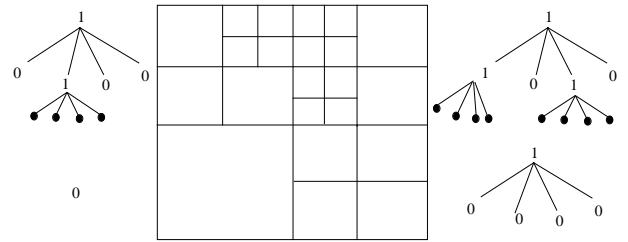

Fig. 1. Quadtree structures of four 16x16 regions and the corresponding representative bits.

In order to formalize the problem, we introduce notation similar to that in [7]. Let $X_{n,i}$ denote the i-th DCT block at level n. Let $n = 0, 1, 2$ represent block size 4x4, 8x8 and 16x16, respectively. If the parent node at level $n + 1$ is split, there are 4 children blocks, i.e., $X_{n,i}$ for $i = 0, 1, 2, 3$. Let $r_k(X_{n,i}), d_k(X_{n,i})$ and $t_k(X_{n,i})$ denote the bits, distortion and complexity (measured in terms of weighted number of operations) of $X_{n,i}$ in block k. The quadtree optimization is based on a bottom up approach [9], in which the decision is made on whether smaller blocks should be merged to achieve better R-C-D performance.

The goal of the optimization is to *minimize the total distortion* $\bar{D} = \sum_k d_k(X_2)$, *for all block k in the frame, such that the total rate* $\bar{R} = \sum_k r_k(X_2)$ *and complexity* $\bar{T} = \sum_k t_k(X_2)$ *are under their respective budget constraints R_b and T_b.* Using the Lagrange multiplier method, we can obtain the unconstrained problem, which consists of minimizing the following objective function

$$\min \sum_k d_k + \lambda_r \sum_k r_k + \lambda_t \sum_k t_k \qquad (1)$$

where $\lambda_r \geq 0$ and $\lambda_t \geq 0$ are Lagrange multipliers that control the R-D and C-D trade-offs, respectively.

For each 16x16 block, the quadtree optimization follows the same procedure as [7], where the optimal quantizer for each node is found such that (1) is minimized and the structure of the quadtree is optimized by bottom-up merge decision. In order to find the decoding time $t_k(X_n)$ for each block, the dyadic classification is assumed to be used at the decoder. The process continues until the root node is reached and the optimal quadtree of every region in a frame of image is computed. Then the Lagrange parameters λ_r and λ_t are adjusted, and the whole process repeats until the bit budget and the complexity budget constraint are met. There are several methods for adjusting the Lagrange multipliers. In this paper, we use the linear approximation algorithm [10] in which each Lagrange multiplier is shrunk (by a factor $\gamma < 1$) or expanded (by a factor $1/\gamma$) depending on whether its corresponding constraint (rate or complexity) has been met or exceeded, respectively.
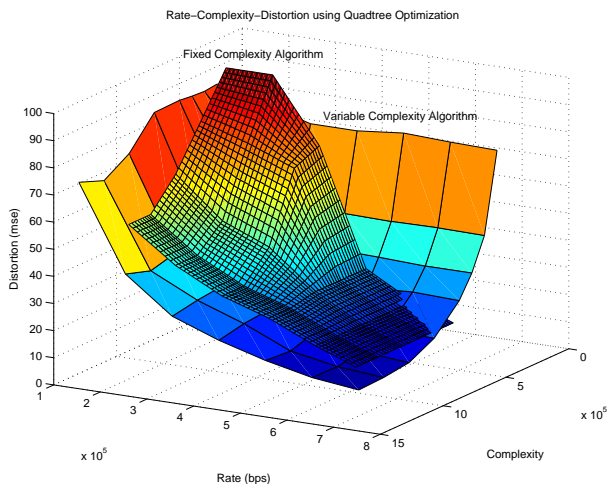
4. EXPERIMENTAL RESULTS

Fig. 2. Rate-complexity-distortion surface using Quadtree optimization framework comparing VCA and fixed complexity IDCT. Note that in general using VCA gives better R-C-D results.

In our experiments, we apply the above quadtree optimization to the first 30 INTER frames of the "Miss America" sequence with Huffman coding of the residue. We perform block based motion estimation at fixed block size of 8x8 and code the residue. As in [7], we do not take the bits necessary for motion vector coding into consideration since it can be regarded as a fixed overhead expense and depends on the motion estimation. We use a uniform quantization matrix with stepsize 2 for all DCT coefficients. There are 5 choices of QP to be selected which are 4, 10, 16, 22 and 28, so that the one with minimal Lagrange cost function is selected. The unit of complexity used in this paper is the

weighted number of operations, i.e., the various operations are weighted according to the average clock cycle spent on them. Operations considered are not only addition and multiplication but also conditional operation and memory access. We use the set of these weighting factors as in [6].
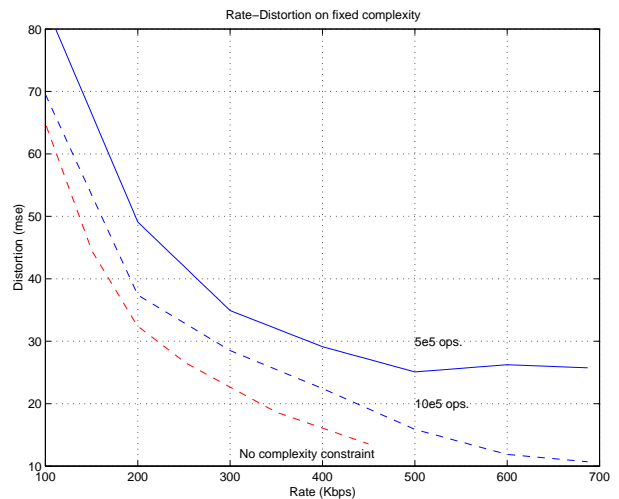

Fig. 3. Constant-complexity R-D curves compared at different complexity constraints. The unconstrained case corresponds to the approach in [7].

Figure 2 shows the R-C-D surface obtained by minimizing D given an R and C budget using the optimization proposed earlier. One can see the tradeoffs between C-D, R-D and R-C in the convex hull of the surface follow the convexity theorem established for Gauss-Markov sources in [4]. In the same figure we also show the comparison between using the VCA IDCT versus using fixed complexity IDCT (where quadtrees are used, but the IDCT for each block of a given size has fixed complexity.) We can see that the difference between the two approaches is smaller in the region of high rate and low complexity budgets. As rate budget is more flexible, smaller quantizers are selected which in turns yields more non-zero DCT coefficients, thus reducing the complexity savings due to the VCA approach. Furthermore as the complexity constraint becomes tighter, the quadtree will tend to use smaller block sizes, thus resulting in smaller VCA complexity savings. Therefore, in Fig. 2, it can be seen that in high rate-low complexity region the VCA is worse.

In Figure 3, we show the R-D curves obtained for given constant complexity, where the result of [7] corresponds to the case when there is no complexity constraint. As can be seen the R-D performance deteriorates as the complexity constraints become more stringent. For example, if the complexity budget is tight the coder is forced to use smaller block size and coarser quantization, which affect the R-D performance.
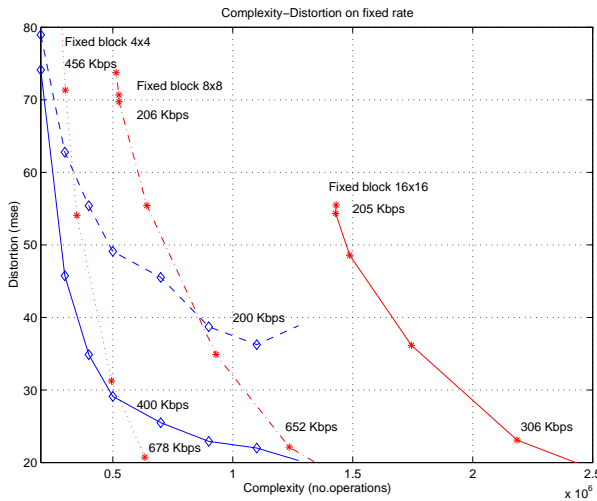
Fig. 4. Constant-rate complexity-distortion curves ('◇') at 200 Kbps ('- -') and 400 Kbps ('–'). These are compared with the best C-D performance with no rate constraints ('∗') but at fixed block size at 4x4 ('· · ·'), 8x8('-·') and 16x16 ('–'), respectively.
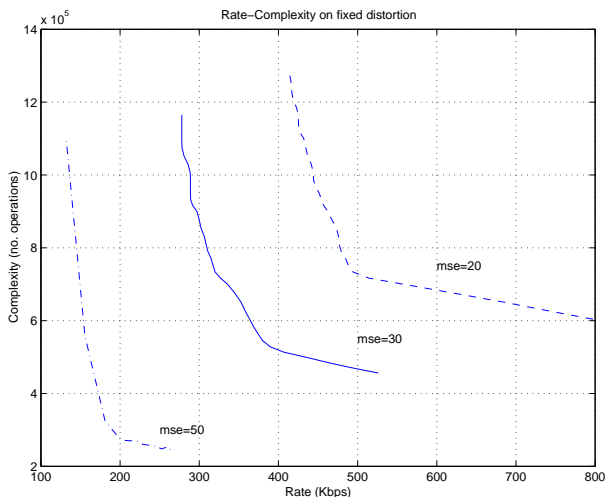

Fig. 5. Constant-distortion rate-complexity curves at MSE = 20, 30 and 50.

In Figure 4, constant-rate C-D curves are shown. It can be explained that as the rate budget is tighter, larger block size and/or larger QP tend to be chosen. As a result, complexity increases, and the larger QP implies more distortion. We also show the results of the C-D tradeoffs as in [6], i.e., the distortion minimization is subject to only a complexity constraint and performed for a fixed block size. As denote in Fig. 4, the rate varies at different C-D operating points. Smaller block size gives getter C-D than a larger one but with the cost of higher rate. It can also be seen that in general the quadtree optimization yields better C-D tradeoffs at the same rate.

The quadtree allows more degree of freedom in controlling the R-C tradeoffs as seen in Fig. 5 where constant-distortion R-C curves are shown. As distortion requirements are more demanding, smaller QP and/or larger block size are likely to be selected. Consequently, higher rate and higher complexity are unavoidable.

5. REFERENCES

[1] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Proc. Magazine*, vol. 15, no. 6, pp. 23–50, November 1998.

[2] G. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Proc. Magazine*, vol. 15, no. 6, pp. 74–90, November 1998.

[3] M. J. Gormish, *Source Coding with Channel, Distortion and Complexity Constraints*, Ph.D. thesis, Stanford University, Mar. 1994.

[4] V. Goyal and M. Vetterli, "Computation distortion characteristics of block transform coding," in *Proc. of ICASSP'97*, Munich, Germany, Apr. 1997.

[5] K. Lengwehasatit and A. Ortega, "DCT computation with minimal average number of operations," in *Proc. of VCIP'97*, San Jose, CA, Feb. 1997, vol. 1, pp. 71–82.

[6] K. Lengwehasatit and A. Ortega, "Distortion/decoding time tradeoffs in software DCT-based image coding," in *Proc. of ICASSP '97*, Munich, Germany, April 1997, pp. 2725–2728.

[7] G. J. Sullivan and R. L. Baker, "Efficient quadtree coding of images and video," in *Proc. of ICASSP'91*, 1991.

[8] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. on Signal Proc.*, vol. 36, no. 9, pp. 1445–1453, Sept. 1988.

[9] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Optimal pruning with applications to tree-structured source coding and modeling," *IEEE Trans. on Info. Th.*, 1989.

[10] G. Cheung, "Optimal bit allocation strategy for joint source/Channel coding of scalable video," M.S. thesis, Electrical Eng. and Computer Sci.n, U.C. Berkeley, 1998.

[11] W.-H. Chen, C. H. Smith, and S. C. Fralick, "A fast computational algorithm for the discrete cosine transform," *IEEE Trans. on Comm.*, vol. COM-25, no. 9, pp. 1004–1009, September 1977.