Complexity-scalable Transform Coding using Variable Complexity Algorithms

Wendi Pan and Antonio Ortega *
Integrated Media Systems Center and Signal and Image Processing Institute,
Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, CA 90089
Email: {dwpan , ortega}@sipi.usc.edu

Abstract

In applications where compression has to be performed under varying complexity constraints (e.g. with hardware having to operate in reduced power mode) it is beneficial to design compression algorithms that allow some degree of complexity scalability. In this paper we explore complexity scalability for transform coding algorithms. We show that a variable complexity algorithm (VCA), which uses energy thresholds to determine the number of coefficients to be computed for each input, is preferable to other alternatives such as a pruned transform, where the same number of coefficients is computed for the whole image. We show that the benefits include not only a higher degree of scalability, but also increased compression performance, as we take advantage of the energy classification that is needed for VCA operation and design quantizers that match each class. We provide expressions for the average complexity, as well as rate/distortion relations for a generic N-point VCA transform. For a two point case, we present closed-form relations describing the variance changes in two classes. In addition, rate-distortion-complexity relations are also empirically obtained. We apply VCA to eight-point KLT and 8×8 DCT in the JPEG framework and experiments show that the VCA approach is superior in rate/distortion performance at low rates as compared to the standard transform coding techniques.

1 Introduction

The need for compression algorithms that support some sort of complexity scalability is likely to increase as multimedia information is increasingly handled by portable devices where the ability to operate under low power conditions is a prime requirement. In this paper we study techniques that enable scalable operation of block transforms used in standard algorithms such as JPEG and MPEG.

Traditional transform image coding algorithms, such as the discrete-time Karhunen-Loeve transform (KLT) and the discrete cosine transform (DCT), have fixed computational complexity once the size of of the transform matrix is decided upon. Variable complexity transform coding algorithms, however, allow the complexity to change as a function of the input. In this case the relevant performance measure is the *average complexity* for a given input source. In this paper, we propose a VCA which uses energy thresholds to provide a trade-off in complexity, rate and distortion (RD). Since our approach is based on computation of a subset of the transform coefficients, our VCA approach yields both different computation time and RD performance than standard transform coders. In [2], variable complexity is achieved by classifying the DCT inputs and pruning the DCT computation accordingly. However, no trade-off in complexity, rate and distortion is available. Other

---

complexity-reduction schemes such as [3], use thresholds to select which DCT coefficients to compute, choosing, for example, whether to compute DC only, or compute 2×2, or 4×4, or all 64 coefficients. Our approach provides a much finer degree of granularity than [3] and other related techniques .

We will provide examples based on both the KLT and the DCT. Our objective is two-fold. First, we aim to demonstrate that lower overall complexity, with finer degree of granularity is possible with VCA approaches. Second, we demonstrate that the energy thresholds used to achieve variable complexity can also be used to improve the RD performance of the coders.

Our VCA transform coding is illustrated in Fig.1, where an eight-point transform is used as an example. Once the k-th coefficient Y_k is computed, a test (denoted by the circles in Fig.1) is performed so as to determine whether $\sum_{i=1}^{k} Y_i^2 \geq (T \cdot E)$, where T is the energy threshold, ranging from 0 to 1, and $E = \sum_{i=1}^{8} Y_i^2$, is the source signal energy. If the test passes (i.e. the first k coefficients have captured sufficient energy), then the algorithm stops, leading to a reduced computation for this input; otherwise, it proceeds to calculating the next coefficient. In the succeeding coding operation, all the coefficients that were computed are quantized, whereas those not computed are simply set to zero, with overhead being used by the encoder to indicate what coefficients are being sent. Note that this equivalent to the end-of-block (EOB) code that is used in traditional transform coding techniques. Thus in our approach we generate and quantize a variable number of coefficients per input.

This paper is organized as follows. In Section 2 we provide an analysis of various aspects of the performance of VCA approach, including the complexity, the probabilities and variances of each of the classes generated and the rate distortion performance. Section 3 presents our preliminary experimental results.
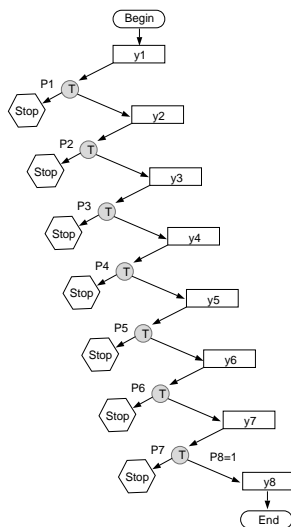

Figure 1: Flowchart of the VCA algorithm to compute a size-8 block transform.

2   Analysis

In this section we first provide an expression of the average complexity as a function of the probability of falling into each class. Then, we provide closed-form expressions for these probabilities in a simple case. We also show that the pdf's of data in each class are different from those of the original data. Finally we combine these results to produce an RDC characteristic.

2.1 Complexity

Consider Fig.1, where P_i represents the probability of test_and_stop after computing Y_i, in other words, $P_i = \text{Prob.}\left[\sum_{k=1}^{i} Y_k^2 \geq (T \cdot E)\right]$. Obviously, the more coefficients are computed, the more likely our VCA will stop after testing. In general, we thus have $P_1 < P_2 < \ldots < P_8 = 1$. The average complexity can be found as:

$$C_{avg} = \sum_{i=1}^{8} Y_i - \left\{ \sum_{i=1}^{7} (P_i \cdot Y_{i+1}) + \left[\left(\sum_{i=1}^{6} P_i \right) - 7 \right] \cdot H \right\} + F, \qquad (1)$$

where, Y_i denotes the complexity of computing individual coefficient Y_i (for example computing each coefficient requires 7 adds and 8 multiplies in the KLT case), and H is the cost of testing after computing each coefficient Y_i. F represents the one-time cost of computing $T \cdot E = T \cdot \sum_{i=1}^{N} X_i^2$, the input signal energy scaled by the threshold T.

Note that for non-VCA, the complexity is $C_b = \sum_{i=1}^{8} Y_i$. Hence the complexity of VCA is less than that of non_VCA, by an amount S. It can be shown that the lower and upper bounds of S are:

$$S_{low} = P_1 \cdot C_b - (7 - 6P_1)H - F, \qquad S_{up} = P_7 \cdot C_b - (7 - 6P_7)H - F. \qquad (2)$$

And the tightness of the bounds is given by: $S_u - S_l = (P_7 - P_1)(C_b + 6H)$.

It can be seen from (1) that the cost of testing H, is an important factor in the average complexity of the VCA. Consider two transforms: First, the KLT has no structure since it depends on the autocovariance matrix of the input signal. Thus our VCA is useful within a KLT framework as it allows to reduce greatly its otherwise complex N^2 operations. The overhead of testing does not offset too much the complexity reduction gained by computing fewer coefficients. Second, for the DCT, however, there are numerous fast algorithms, many of which require only $O(NlogN)$ operations [1]. Although a VCA can still contribute to a reduction of computation, the relative overhead of the testing due to the VCA will be higher than in the KLT case. Therefore, the testing budget tends to be tight for variable complexity DCT. Some fast testing methods such as using absolute values rather than square(-multiplication) operations for finding the energy can be considered [2][3][4]. Moreover, if we use a larger energy threshold, it is more efficient to do the energy test only after several coefficients have been computed so that the overall testing overhead is reduced. Note that although the computational cost of testing could become a significant overhead to the overall complexity of a software system, it may not be a major concern in hardware implementation.

2.2 Class Probabilities

The VCA can be viewed as a pattern classifier in the sense that it classifies the source (coefficient sample space) into several disjoint cases. The classification criterion is the energy test controlled by the threshold use.

For a closed-form solution, we consider the 2-D scenario: two zero-mean/independent Gaussian random variables Y_1 and Y_2, with variances σ_1^2 and σ_2^2 respectively. The probability P of successful test_and_stop by using energy threshold T corresponds to the mass of the shaded region in Fig.2. It can be shown that

$$P = Prob\left\{ \frac{Y_1^2}{Y_1^2 + Y_2^2} \geq T \right\} = \frac{2}{\pi} \arctan\left(\frac{\sigma_1}{\sigma_2} \sqrt{\frac{1 - T}{T}} \right) = \frac{2}{\pi} \arctan(K \cdot r), \qquad (3)$$

where, $K = \sqrt{\frac{1}{T} - 1} = \tan \Phi$, and $r = \frac{\sigma_1}{\sigma_2}$, is the ratio of standard deviations. In the context of transform coding, r is an indicator of the energy compaction capability of the transform used.
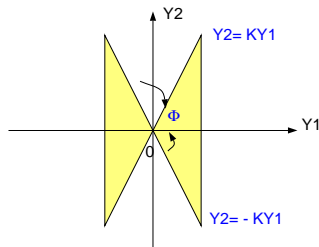
3

Figure 2: Partitioning of the Coefficient Sample Space by VCA. The shaded region contains all the inputs such that the first coefficient Y_1 contains a percentage of the energy above the threshold.

As shown in Fig.3 (left figure), as expected, P decreases monotonically with increasing threshold T. If we fix T and increase r, then samples are more likely to fall into class 1. The reason is that a larger r means that $Y1$ contains more energy than $Y2$ on average and thus the probability of passing the energy test becomes higher. Fig.3 (right figure) shows the saving of complexity introduced by the VCA. If the savings drop below zero, it indicates that the VCA does not bring any gain in terms of complexity. This is because for those thresholds the complexity of testing is not compensated by savings in coefficient calculation, since with high probability both coefficients have to be computed.
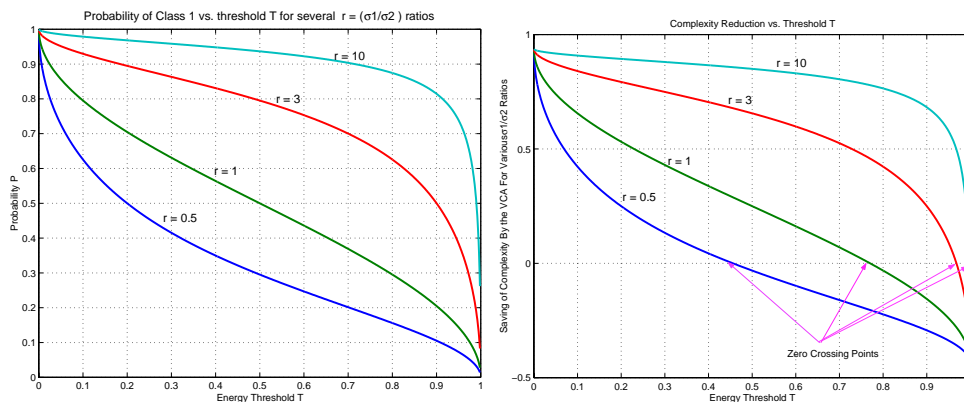

Figure 3: (Left) P as a function of r and T ; (Right) Complexity reduction as a function of r and T.

2.3   Variances within each Class

The energy threshold T controls the number of coefficients computed and thus the overall complexity. At the same time, the choice of T also affects both rate and distortion, given that the test determines which transform coefficients are quantized. For any given input vector of dimension N and a particular choice of T, there will be N classes, depending on how many coefficients are computed. For each class k, $(k = 1, 2, ..., N)$, k coefficients are computed and quantized, and the remaining $(N - k)$ are simply set to zero and not quantized. In practice, overhead will be used to inform the decoder that $(N - k)$ coefficients were not encoded. We assume that the probability of class k occurring is P_k. We can then represent the variances of the N coefficients *conditioned on* k, that is σ_{ki}^2 for $i = 1, \ldots, N$. See also TABLE 1. It is important to note that the variance will not be the same for each k, since through testing we are separating several different scenarios. For example, for a given T if k is small, i.e., a few coefficients contain most of the energy, then σ_{ki}^2 can be

4

expected to be relatively large for small i, and relatively small for large i, as compared to the original variances. Thus, in this example, if σ_i^2 is the original variance of the coefficient (when using the standard, non-VCA, transform), then we expect $\sigma_{ki}^2 \geq \sigma_i^2$ for $i < k$ and $\sigma_{ki}^2 \leq \sigma_i^2$ for $i > k$.

Table 1: Classes in an N-point VCA.

Class	Probability	Var of Coef's Comp.	Var of Coef's Not Comp.
1	P_1	σ_{11}^2	$\sigma_{12}^2, ..., \sigma_{1N}^2$
2	P_2	$\sigma_{21}^2, \sigma_{22}^2$	$\sigma_{23}^2, ..., \sigma_{2N}^2$
k	P_k	$\sigma_{k1}^2, ..., \sigma_{kk}^2$	$\sigma_{k(k+1)}^2, ..., \sigma_{kN}^2$
N	P_N	$\sigma_{N1}^2, ..., \sigma_{NN}^2$	

We take the two-point VCA as an example. By using the previous Gaussian model, we can find a closed-form expression of the variances of two classes in relation to the original source variance as follows:

Assume that $Y_i, i = 1, 2.$, are the jth coefficient belonging to class i. For brevity, only the pdf of Y_{11} of class I are given below :

$$f_{Y_{11}}(y) = \frac{2}{P}\left[\left(\frac{1}{\sqrt{2\pi}\sigma_2}\int_0^{K|y|} e^{-\frac{x^2}{2\sigma_2^2}}\,dx\right)\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{y^2}{2\sigma_1^2}}\right] \quad (4)$$

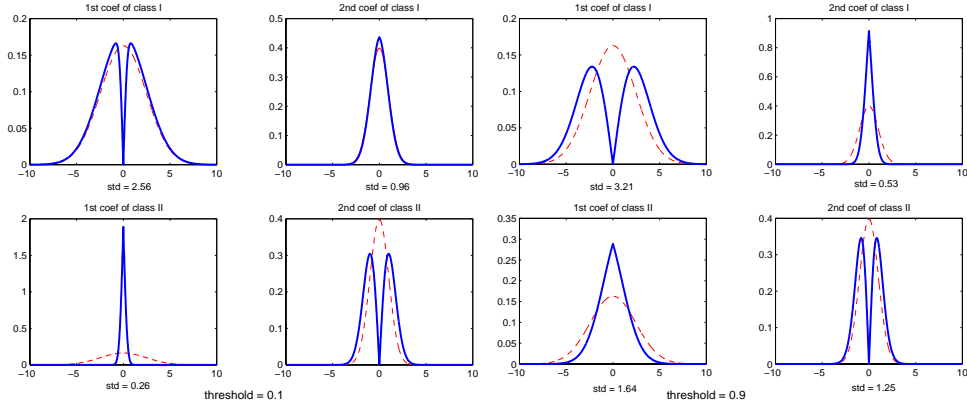where, P is the probability of class I occurring, as given in equation (3).

Figure 4: Distribution of Coefficients in Two Classes(solid curves):(Left) $T = 0.1$; (Right) $T = 0.9$. Dashed curves represent original normal distributions.

The distributions are plotted in Fig.4 for two thresholds. It can be observed that class II is a dual case to class I. Y_{11} splits into a bi-modal distribution from the original normal shape. As T is increased, the two peaks of Y_{11} are increasingly separated while Y_{12} shows an increasingly "peaked" distribution. This suggests that the variance of Y_{11} increases and that of Y_{12} decreases as a result.

It can be shown that class variances σ_{ij} are related to the original variances σ_i by two simple adjustment factors, called G_1 and G_2, as follows:

$$\sigma_{11}^2 = (1 + G_1)\sigma_1^2, \qquad \sigma_{12}^2 = (1 - G_1)\sigma_2^2, \quad (5)$$

$$\sigma_{21}^2 = (1 - G_2)\sigma_1^2, \qquad \sigma_{22}^2 = (1 + G_2)\sigma_2^2. \quad (6)$$

with $G_1 = \frac{2Kr}{P(K^2r^2+1)\pi} = \left(\frac{1}{P} - 1\right)G_2$, where K and P are as defined in Section 2.2, and $r = \sigma_1/\sigma_2$.

5

2.4 Rate Distortion Relations

In an N-point VCA, there exist N classes, such that in class k, only the first k coefficients are computed, and the remaining $(N - k)$ coefficients are simply set to zero. We introduce an optimal bit allocation by considering allocation at two levels. At the lower level, we consider coding within each class. There we can use standard bit allocation techniques to determine how to allocate B_k bits among the k coefficients coded in class k [5]. If coefficients computed are scalar quantized at high resolution, then the optimal distortion for the total number of bits B_k is

$$D_k = kH_k\rho_k^2 2^{-2\frac{B_k}{k}} + \sum_{i=k+1}^{N} \sigma_{ki}^2 \tag{7}$$

where, $H_k = \left(\prod_{i=1}^{k} h_{ki}\right)^{\frac{1}{k}}$, $\rho_k^2 = \left(\prod_{i=1}^{k} \sigma_{ki}^2\right)^{\frac{1}{k}}$ are the geometric means of the distortion constants and variances, respectively, which can be obtained from the probability distributions of the coefficients [5].

At a higher level, our goal is to allocate optimally bits to each of the classes, i.e., to find B_k for each class. This allocation has the aim of minimizing the total *average* distortion D under a constraint on total rate B, where:

$$D = \sum_{k=1}^{N}(P_k \cdot D_k), \quad \text{and} \quad B = \sum_{k=1}^{N}(P_k \cdot B_k) \tag{8}$$

Note that here the total rate and distortion are averaged with weights corresponding to the probabilities of each class. Conventional Lagrange multiplier techniques can be employed by introducing the cost function J and searching for the rates that minimize it for a given Lagrange multiplier $\lambda \geq 0$,

$$J = B + \lambda \cdot D = \sum_{k=1}^{N}(P_k B_k) + \lambda \cdot \left[\sum_{k=1}^{N} P_k \cdot \left(D_k' + \sum_{i=k+1}^{N} \sigma_{ki}^2\right)\right]. \tag{9}$$

We minimize J by taking the N derivatives: $\partial J/\partial B_k = 0$, with $k = 1, ..., N$. We obtain the following optimal bit allocation, expressed in average number of bits assigned to class k:

$$B_{k,opt} = \frac{B_k}{k} = \bar{B} + \frac{1}{2}\log_2 \frac{H_k\rho_k^2}{\left[\prod_{i=1}^{N}(H_i\rho_i^2)^{iP_i}\right]^{\frac{1}{P_t}}} \tag{10}$$

where, $P_t = \sum_{k=1}^{N} kP_k$, is the average number of coefficients computed for all N classes, and $\bar{B} = B/P_t$, is the average number of bits per coefficient. Note that the distortion of (7) corresponds to the best bit assignment to the coefficients within a class and has been used in estimating the distortion in (10). Thus, this two step procedure (allocation among classes first, then within classes) is also guaranteed to be optimal. Intuitively, (10) shows that the allocation to class k exceeds the average allocation $\bar{B}$ if the parameter $H_k\rho_k^2$ is greater than the class *probabilities weighted* geometric average of $H_k\rho_k^2$ for $k = 1, 2, \ldots, N$. Conversely, if $H_k\rho_k^2$ is less than this geometric mean, then the bit allocation is less than the average allocation $\bar{B}$. Note that the main change with respect to standard bit allocation results comes from the fact that the number of coefficients receiving bits varies from class to class. Thus, whereas in the standard geometric mean expression we had a $1/N$ term (assuming N inputs) and equal power for each of the terms, here we use P_t (the average number of inputs) instead of N and iP_i instead of constant weights for each class in the allocation. With this optimal allocation we can obtain the following R/D relations:

$$D = \sigma_t^2 + P_t \cdot \left[2^{-B} \cdot \prod_{k=1}^{N}(H_k\rho_k^2)^{\frac{kP_k}{2}}\right]^{\frac{2}{P_t}}, \quad \text{and} \quad \sigma_t^2 = \sum_{k=1}^{N} P_k\left(\sum_{i=k+1}^{N} \sigma_{ki}^2\right) \tag{11}$$

6

In summary, with the usual caveat that these results are only valid at high rates, we can see that the optimal allocation results are the same as for fixed complexity transform coding [5], except that in allocating bits across classes both the source variances *and* the probability of each class are taken into account.

3 Experimental Results

3.1 Two-point VCA

In our experiments, both the conventional two-point fixed complexity and the variable complexity KLT are applied to adjacent pixels of test image "Lenna". Uniform, midriser quantizers are used. We can vary rate and distortion by changing the quantization cell size. Entropy is measured as an estimate of the overall rate, and an energy threshold is used to control the complexity. Note that the advantage of the VCA approach comes from the fact that in the low rate scenario all the bits can be assigned to the low frequency coefficients whenever they have most of the energy. Because we can test for this condition at the encoder we can in effect use a simple vector quantizer as shown in Fig.5 (left), with different quantization characteristics depending on whether the test fails or not. More specifically, as derived in Section 2.3, we expect that (i) $\sigma_{11}^2 \geq \sigma_1^2$, $\sigma_{12}^2 \leq \sigma_2^2$ and (ii) $\sigma_{21}^2 \leq \sigma_1^2$, $\sigma_{22}^2 \geq \sigma_2^2$.

The experimental R/D/C surface is shown in Fig.5 (right), where rate denotes the total number of bits assigned to the two KLT coefficients, distortion is the MSE after the VC-KLT. Complexity is represented by the energy threshold T since complexity increases monotonically with T. It can be seen that for a given complexity an increase in rate results in decreased distortion, as was to be expected.

A more interesting result is that, for a given rate, the distortion varies as a function of the complexity and indeed for each rate there is an "optimum" complexity, that is the one corresponding to the threshold that guarantees minimal distortion at that rate. Distortion varies because the set of coefficients that are computed depends on the complexity. Thus for low threshold, many coefficients are not computed, and are effectively quantized to zero.

The curve of optimal distortion for each rate is plotted in Fig.5 (middle). Also shown in the same plot is the rate/distortion curve of the fixed complexity 2-point KLT. It is obvious that at low rates (total number of bits lower than 7), VCA does much better in rate/distortion sense. Note that one bit of overhead has been added to the VCA-KLT approach to notify the decoder of whether one or two coefficients are being transmitted. Clearly, when the rate becomes sufficiently high, the best approach would be to compute both coefficients so that the fixed complexity algorithm should be used.
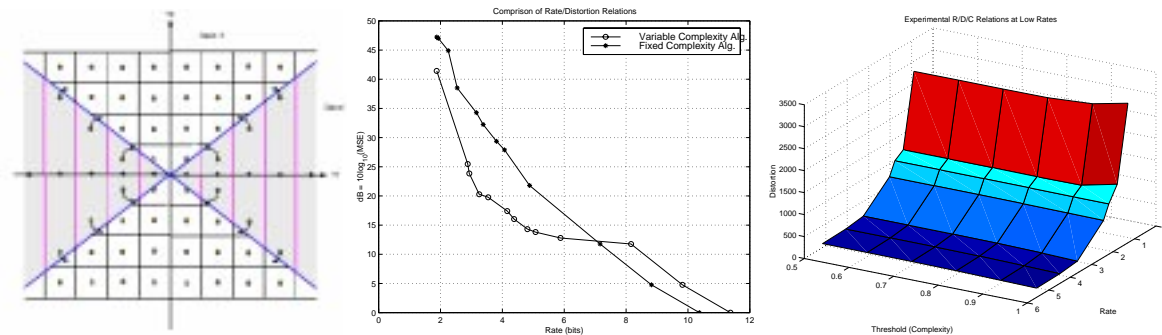

Figure 5: Two-point VC-KLT on "lenna". (Left) Quantization Scheme; (Middle) Comparison of R/D Relations; (Right) R/D/C surface.

3.2 Eight-point VCA

As in the previous experiments, we apply an eight-point variable complexity KLT to adjacent pixels of test image "Lenna". We study the variance of each of the eight possible classes and the result confirms our analysis in sec.2.3. The histogram of Y_{5k} in class k is shown as an example in Fig.6 (left). It can be observed that σ_{5k} rises until $k = 5$ and then drops with increasing k.

The rate/distortion curve is shown in Fig.6 (right). We test three schemes: (1) fixed complexity KLT (FCA) with one uniform quantizer for all eight coefficients. End_of_block (EOB) is introduced to represent the trailing zeros. EOB is also entropy coded; (2) VC-KLT with separate uniform quantizer for each of the eight classes; (3) Same as (2) except that a separate quantizer is used for each coefficient computed for each class, so that there are a total of 36 such quantizers.

It can be concluded that we can achieve better R/D performance with more quantizers tailored to the characteristics of each coefficient and of each class in the context of the VCA. At rates lower than 1.2bits/sample, the VCA scheme (2) outperforms fixed complexity ones by up to 2dB in MSE. This is consistent with what we have observed in the two-point experiment.
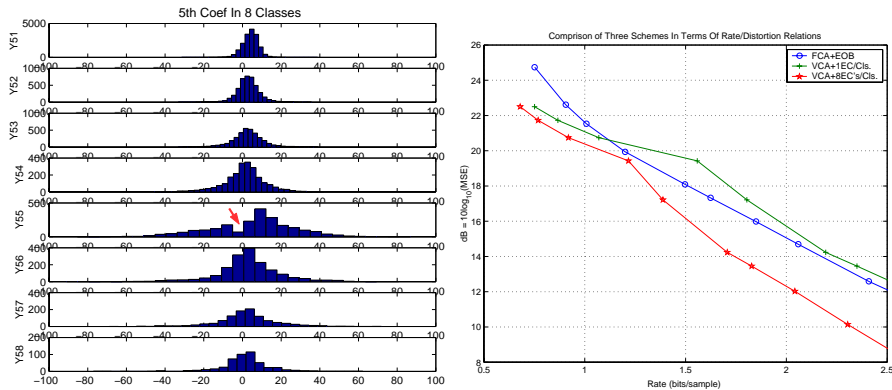

Figure 6: Eight-point VC-KLT on "lenna". (Left) Distribution of the 5th KLT coefficient in eight classes; (Right) R/D curves of FCA and VCA using different quantization schemes.

3.3 8×8 VCA

We extend our VCA to the 8×8 DCT employed extensively in most existing international standards in image and video compression. VCA divides the 8×8 DCT coefficients into 64 disjoint classes according to the actual number of DCT coefficients computed after the energy test. We adopt the zigzag order used in JPEG [6] to index these 64 DCT coefficients in an 8×8 block.

Fig.7 (left) illustrates the number of blocks classified into each class by VCA. Obviously, the higher the energy threshold is used in the test, the more blocks are assigned to the classes of higher index. Fig.7(right) illustrates the variances of the DC(1st),second,third and tenth DCT coefficients in each class. A peak in variance always exists which corresponds to the ith coefficient in class i. This feature is caused by the energy test of VCA and can lead to coding gains if properly utilized.

We then apply the VCA onto JPEG baseline algorithm. The approach taken in our preliminary experiment is to perform the energy test of the 8×8 DCT coefficients in a block by following the zigzag order. Once the energy threshold is satisfied, we set the remaining coefficients in the block to zero. This procedure can be readily embedded into the JPEG encoder. Note that this is similar to the approach of thresholding [7][8], with the

difference being that here coefficients are not even computed if the energy of the vector is deemed to be sufficient already, whereas in [7] this decision was made after the coefficients were computed.

The rate of JPEG can be varied by changing the scaling factor of the quantization table. We then obtain the rate (bits/sample) and distortion (PSNR) curves with/without the proposed energy test, see Fig.8 (left) for the comparison of JPEG versus VCA. It can be seen that VCA provides a better R/D performance for a certain range of thresholds than that obtained by merely decreasing the scale factor in JPEG. This effect is even more pronounced at rates below 0.2bits/sample, which agrees well with our findings in two-point and eight-point cases.
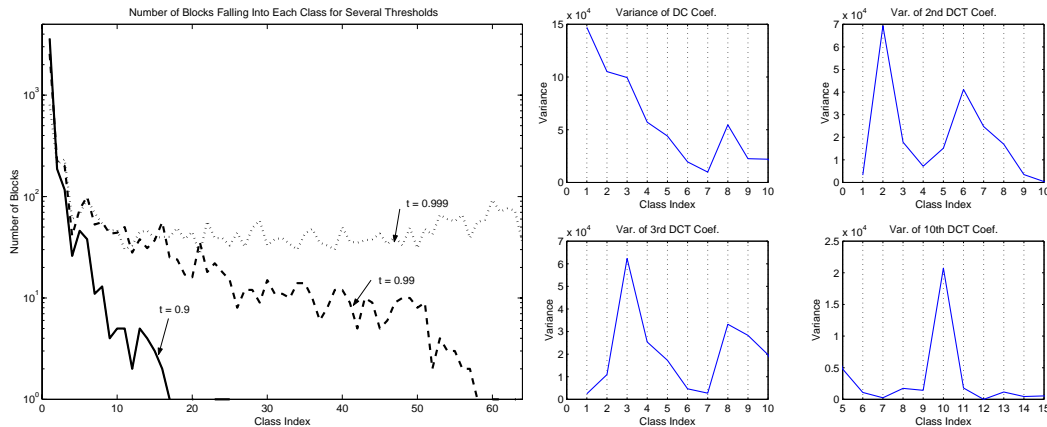

Figure 7: 8 × 8 VC-DCT (Left)Number of blocks falling into each class for image "Peppers"; (Right)Variances of DCT coef's for classes.

Note that the threshold we have been using so far is relative in the sense that it represents the *percentage* of the total vector energy we want to retain. From a rate/distortion tradeoff perspective, however, an *absolute* energy threshold may be more desirable in some situations. We introduce such an absolute threshold scheme in the JPEG experiment, i.e., we keep computing DCT coefficients until the energy of the coefficients not yet computed drops below an absolute threshold. For these residual coefficients not computed, we simply set them to zero, exactly as we have done in relative threshold approach. The absolute threshold, as a control parameter, represents the distortion (MSE) of the remaining coefficients, and thus enables us to have direct control over the larger distortion. By contrast, in relative threshold approach, we can not distinguish DCT coefficient blocks of different total energy. Therefore the distortion caused by the coefficients which are not computed varies from block to block and is not directly related to the threshold itself as in the absolute threshold case. In general, we thus expect that the absolute threshold approach to perform much better than the relative one on images of high inter-block energy variations.

The R/D curve when applying the absolute energy threshold in JPEG is shown in Fig.8 (right). The curves obtained when using relative thresholds are also shown in the same plot for comparison. It can be seen that absolute threshold approach can always outperform the relative one (up to about 1dB). Consequently, the absolute threshold approach does better than standard JPEG within a wider range of rates than the relative threshold method.

4 Conclusions

This paper presents a variable complexity algorithm (VCA) for transform coding. VCA uses energy thresholds to provide flexible tradeoffs among computational complexity, rate and
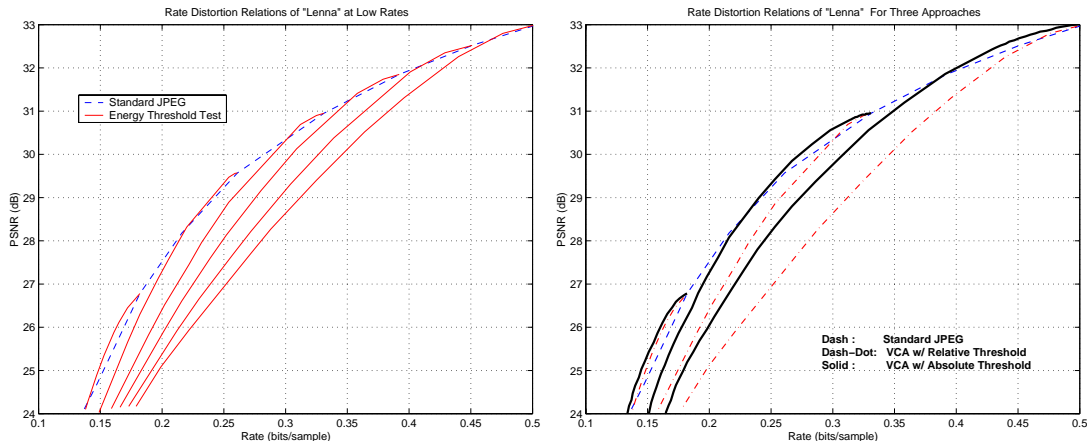
Figure 8: Comparison of R/D relations between JPEG and VCA's using two threshold approaches. (Left) VCA with relative thresholds. The dashed curve is for the JPEG baseline, and solid lines stand for the curves obtained by using different energy thresholds when we fix the scale factor of the quantization table.; (Right) VCA with absolute thresholds.

distortion. We view VCA as a pattern classifier and carry out extensive statistical analysis on the average complexity, as well as rate/distortion relations for the generic N-point VCA. By introducing a standard model in a two point case, we derive closed-form relations which describe the variance variation after classification. In addition, rate-distortion-complexity relations are also empirically obtained. In the eight-point VCA study, We show that by quantizing each coefficient within each class separately, we can achieve better rate/distortion performance. Finally, we apply VCA to the JPEG baseline algorithm, where both the relative and absolute threshold method are used and compared. Experimental results show that our VCA approaches can provide better rate/distortion tradeoff, especially at low rates than the JPEG techniques.

References

[1] K.R.Rao and P.Yip, *Discrete Cosine Transform—Algorithms, Advantage, Applications*. Academic Press, 1990.

[2] K.Lengwehasatit and A.Ortega, "DCT computation based on variable complexity fast approximations," in *Proc. of Intl' Conf. of Image Proc., ICIP'98*, (Chicago), Oct. 1998.

[3] I.-M. Pao and M.-T. Sun, "Modeling DCT coefficients for fast video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 608–616, June 1999.

[4] B.Girod and K.Stuhlmller, "A content-dependent fast DCT for low bit-rate video coding," in *Proc. of Intl' Conf. of Image Proc., ICIP'98*, (Chicago), Oct. 1998.

[5] A.Gersho and R.M.Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[6] W.B.Pennebaker and J.L.Mitchell, *JPEG:Still Image Data Compression Standard*. Van Nostrand Reinhold, 1992.

[7] K. Ramchandran and M. Vetterli, "Rate-distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility," *IEEE Trans. on Image Proc.*, vol. 3, pp. 700–704, Sept 1994.

[8] M. Crouse and K. Ramchandran, "Joint thresholding and quantizer selection for transform coding: Entropy-constrained analysis and applications to baseline JPEG," *IEEE Trans. on Image Proc.*, vol. 6, pp. 285–298, Feb. 1997.