

ADAPTIVE QUANTIZATION OF IMAGE SUBBANDS WITH EFFICIENT OVERHEAD RATE SELECTION

Youngjun Yoo and Antonio Ortega

Department of EE-Systems,
University of Southern California
Los Angeles, CA 90036, USA
yyoo,ortega@sipi.usc.edu

Bin Yu

Department of Statistics,
University of California
Berkeley, CA 94720, USA
binyu@stat.berkeley.edu

ABSTRACT

Recent subband image coding techniques owe much of their success to an effective use of adaptive quantization and adaptive entropy coding. It is often the case that adaptive quantization is achieved by defining a discrete set of quantizers from which one is chosen for a given set of coefficients. This type of forward adaptation thus requires that overhead information (the choice of quantizer) be sent to the decoder. Then, the quantized coefficients are transmitted using adaptive entropy coding, typically through backward adaptive arithmetic coding. In this paper we show that *a combination of forward and backward adaptation methods can be used to update the quantizers* thus reducing the overhead requirements while still providing good performance. Specifically, we present an algorithm where each coefficient is classified into several classes based on the past quantized data and where the quantizer to be used for each class can itself be adapted on the fly.

1. INTRODUCTION

Subband, or wavelet, coding has been shown in recent years to provide a very useful framework for image compression. It is well known that subband decompositions have the advantage of avoiding the annoying blocking artifacts typical of DCT based schemes while achieving good energy compaction and in fact recent developments in subband based image compression have greatly outperformed the DCT-based JPEG standard [1]. There are two major components in the emergence of wavelet based compression, which were both utilized in Shapiro's breakthrough work [2]. First, subband coders allow the use of *efficient data-structures*, such as the zerotree in [2] or the cross-band context information used by the arithmetic coder in [3]. These

techniques exploit local crossband correlations – a feature that has proven useful for many images.

A second component has been the use of *adaptive compression techniques*, i.e., those schemes which allow any of the three major building blocks of typical image coding, namely, transform or linear decompositions, quantization, and entropy coding, to change on an image by image basis or even locally within an image. In [2, 3] this is achieved using adaptive arithmetic coding. In other cases, the coder can select the subband decomposition for the given image [4] or both the subband decomposition and the quantization stepsize for regions of the image [5]. Finally algorithms such as those in [6] rely on assigning different quantizers to different regions of the image. In this paper we concentrate on the design of adaptive quantizers suitable for subband coding.

Note that arithmetic coding achieves adaptation by keeping track of past transmitted symbols, while most adaptive quantization schemes used for images select the quantizers based on the original data. Thus, these adaptive quantizers require overhead information to be sent to the decoder, whereas adaptive arithmetic coders do not require any overhead information. As an example, both in [5] and [7], from a discrete set of available quantizers the encoder can select the best quantizer for a given block or region of the image.

Our goal is to demonstrate that backward adaptive quantization schemes are also useful in image compression. As a starting point we consider the adaptive quantizer introduced in [8, 9] which estimates the distribution of the data based on the previously quantized information. It has been shown that for image subbands it is useful to separate the coefficients in a given subband into different classes so that different quantizers can be applied to each class. We propose a *novel approach to classification which relies only on the past quantized data*. For each class, a Laplacian model is fitted and the Laplacian parameter is transmitted as side information. We will also demonstrate how *the Laplacian parameter of a given class can itself be adapted on*

This work was supported in part by the National Science Foundation under grant MIP-9502227 (CAREER) (A. Ortega) and in part by the Army Research Office under grant DAAH04-94-G-0232 and the National Science Foundation under grant DMS-93-22817 (B. Yu).

the fly, again based on past quantized data.

Throughout this paper we will consider as a baseline a simple subband coder where each band is coded independently using a uniform threshold quantizer (UTQ) [10] and the same bit allocation is used in all the cases to be compared. Note that our goal is to demonstrate the potential gain in introducing these techniques. More sophisticated quantizers and schemes taking advantage of interband correlation can be incorporated into our proposed scheme and would give better results.

2. CLASSIFICATION

In [7] a method is proposed to classify blocks of subband coefficients into classes based on their “activity”. Each class is modelled as a Laplacian source with a different parameter and each block within a subband is thus assigned to a class, and its corresponding Laplacian model, and is quantized with a quantizer matched to the model. The classes are determined at the encoder and sent to the decoder as side information. In addition, for each block of coefficients the class index is also sent as side information. In order to keep the overhead at a reasonably low level, blocks, rather than individual coefficients, are classified. Moreover, the number of classes is also kept small in order to reduce the overhead per block.

2.1. Context-based classification

We now show that if the quantized data for classification is used then these two shortcomings can be overcome. Classification can then be done on a coefficient by coefficient basis and, if needed, an increased number of classes can be used. This idea has been used for lossless compression of images, as in [11], and is also analogous to the context-based adaptation used in some arithmetic coders, see for example [3].

While many scan orders are possible (column-wise, Peano scan, etc.), we assume a row-wise raster scan without loss of generality. Consider the neighboring quantized coefficients of the current coefficient $X_{i,j}$, for example $\{\hat{X}_{i-1,j}, \hat{X}_{i-1,j-1}, \hat{X}_{i,j-1}, \hat{X}_{i/2,j/2}^c\}$, where $\hat{X}_{i/2,j/2}^c$ is the quantized coefficient in the coarser level subband of same orientation (thus it is possible to exploit crossband correlation). The encoder can establish a rule to classify each coefficient $X_{i,j}$ as a function of the neighboring quantized coefficients. In this work we ignore the crossband effects and consider the average magnitude in a three coefficient neighborhood, i.e., our classification is based on $\frac{1}{3}(|\hat{X}_{i-1,j}| + |\hat{X}_{i-1,j-1}| + |\hat{X}_{i,j-1}|)$. To motivate the performance of this scheme consider Figs. 1 and 2. Fig. 1 depicts an example where four classes are used and clearly indicates that the high energy areas can be correctly classified. Fig. 2 shows



Figure 1: Classification map obtained by using a four-coefficient context comprising pixels from the LH band and the LH2 band (band of same orientation at coarser resolution). Each coefficient is classified into one of 4 classes. Black indicates the least active regions and white depicts the more active.

the resulting histograms of unquantized data in each of the four classes and also indicates that classification based on the quantized context has separated the pixels into four classes with varying degrees of activity, i.e., higher activity results in higher variance.

The encoder classifies the coefficient data into classes and then estimates the Laplacian parameter for each class as follows. For each class, we have data Y_1, \dots, Y_n and the Laplacian model assumes that Y_1, \dots, Y_n are i.i.d. from a Laplacian density, which is equivalent to having $|Y_1|, |Y_2|, \dots, |Y_n|$ as i.i.d. data from an exponential source $f_\lambda(y) = \lambda e^{-\lambda y}$ for $y \geq 0$. Then the log-likelihood of parameter λ given the data is $l(\lambda) = n \log \lambda - \lambda \sum_i |Y_i|$, which, when maximized with respect to λ , gives the maximum likelihood estimator

$$\hat{\lambda} = n / \sum_i |Y_i|.$$

The encoder explicitly sends to the decoder the Laplacian parameter for each class, which is used by encoder and decoder to define the quantizer to be used. As compared to a block-based method we have the advantage of achieving classification regions with arbitrary shapes. This will prove particularly useful because each region can then be allocated a different number of bits.

2.2. Classification thresholds selection

The above examples have shown how classification based on the quantized past provides a good localization of the various types of coefficients. We now introduce a method to determine good thresholds. We first quantize the subband coefficients with a uniform quantizer

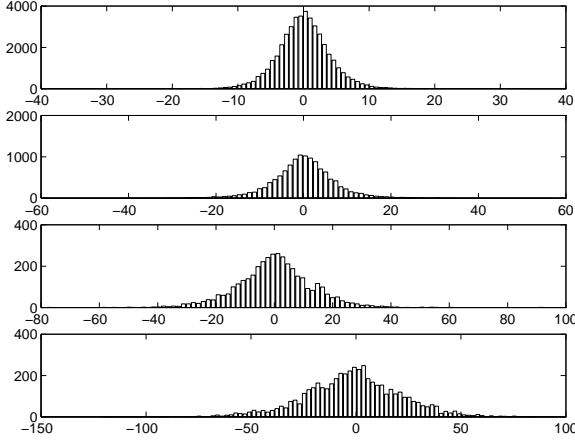


Figure 2: Histograms of the actual coefficient values in each of the 4 classes obtained from the context based classification, from low activity (top) to high activity (bottom).

with quantization step close to what we eventually expect to use. Note that this is a major simplification since the quantization stepsize of the UTQ is not known a priori, and will be different for each class of data. A better technique would involve an iterative procedure where, first, given the thresholds the quantizers are designed and, second, given the quantized data a new classification is performed.

Given the quantized data we can determine to which class each (unquantized) coefficient belongs. We start by considering a large number of classes N and select a set of monotonically increasing thresholds T_0, \dots, T_{N-1} such that each of the resulting classes has roughly the same number of coefficients. For simplicity, we assume that each of the classes is approximately Laplacian. The Laplacian parameter for each class can be estimated from the data as described in Section 2.1. The algorithm operates iteratively by removing at each stage the two adjacent (in the sense of corresponding to consecutive thresholds) classes which have the smallest classification gain. As in [7] we estimate the classification gain using the variance of the data in each class, say σ_i^2 and σ_{i+1}^2 , and comparing it with the variance σ^2 if the two classes were merged. Given that n_i and n_{i+1} are the number of occurrences of the two classes, the estimated Laplacian parameter λ of the two classes put together is $\lambda = (n_i + n_{i+1}) \cdot (n_i/\lambda_i + n_{i+1}/\lambda_{i+1})^{-1}$. Under the Laplacian assumption the variances are $\sigma_i^2 = 1/\lambda_i^2$ and thus the classification gain is

$$G = \frac{\sigma^2}{\sigma_i^2 \sigma_{i+1}^2}$$

where $p_i = n_i/(n_i + n_{i+1})$, can be written as function of the Laplacian parameters. We use a greedy algorithm which merges the two classes that have the smallest

classification gain until the desired number of classes is reached.

3. ADAPTIVE QUANTIZATION

The above described scheme can be directly extended to include quantizer adaptation within each class. Assume that we initialize the quantizer for one class based on a Laplacian distribution of parameter λ , which is sent to the decoder as overhead information. This parameter reflects the overall characteristics of the class. In order to adapt to local variations one can keep track of the past quantized coefficients in the class. The local Laplacian parameter can be estimated from the quantized data using the following method. With the same assumptions and notations as in Section 2.1 and for an L -level quantizer with decision levels $0 = b_0 < b_1 < b_2 < \dots < b_L < \infty$, let \hat{p}_k be the frequency of occurrence of quantized data with levels $1, 2, \dots, k$.

For any given k , we can get an estimator of λ , called $\hat{\lambda}_k$. Denote by $F_\lambda(y)$ the cumulative distribution function (cdf) of the exponential random variable with parameter λ . That is, for $y > 0$,

$$F_\lambda(y) = \int_0^y \lambda e^{-\lambda t} dt = 1 - e^{-\lambda y}. \quad (1)$$

\hat{p}_k is a good empirical estimate, based on n data points, of $F_\lambda(b_k)$, which is the proportion of data falling into levels 1 through k if we have an infinite amount of data. To be more precise, by Law of Large Numbers,

$$\hat{p}_k \rightarrow F_\lambda(b_k),$$

as $n \rightarrow \infty$ and in probability.

Solve equation (1) for λ and we get

$$\lambda = \log(1 - F(y))/y.$$

Hence a natural estimate of λ based on $F_\lambda(b_k)$ is

$$\hat{\lambda}_k = -\log(1 - \hat{p}_k)/b_k.$$

The question remains how to choose k . Let $p_k = F_\lambda(b_k) = 1 - e^{-\lambda b_k}$, then

$$\text{Var}(\hat{p}_k) = p_k(1 - p_k)/n,$$

since $n\hat{p}_k$ is Binomial (n, p_k) . By the δ -method, we can obtain an approximate variance of $\hat{\lambda}_k$ as

$$\frac{1}{nb_k^2} \frac{p_k(1 - p_k)}{(1 - p_k)^2} = \frac{p_k}{nb_k^2(1 - p_k)} = \frac{e^{\lambda b_k} - 1}{nb_k^2}.$$

It follows that for $\hat{\lambda}_k$ to have a small variance, $(e^{\lambda b_k} - 1)/b_k^2$ should be small. However, as $b_k \rightarrow 0$ or ∞ ,

$(e^{\lambda b_k} - 1)/b_k^2 \rightarrow \infty$. Hence there is a value of b_k somewhere for the approximate variance to be at its minimum. To know the precise value of this minimum, we need to know λ , which is unknown.

What we can do is to pick an initial estimate λ_0 which corresponds to some b_{k_0} in the middle of the range of the quantizer decision levels and then find k^* such that

$$k^* = \operatorname{argmin}_k (e^{\lambda_0 b_k} - 1)/b_k.$$

Or we can just use λ_0 . It seems reasonable to pick k_0 such that \hat{p}_{k_0} is around 1/2.

For each of the classes the encoder transmits to the decoder the thresholds to be used in the classification as well as the best Laplacian parameter that can be matched to each of the classes. This parameter serves to initialize the adaptive quantizer. The adaptation algorithm estimates an updated Laplacian parameter based on the past quantized data in each class. Since the initial Laplacian parameter is a fairly good representation of the data in the class we in fact use an average between that quantizer and the newly obtained (local) one. This system could be further improved by using overhead to help in determining the centroids (cf. [12]), deciding on a block-by-block basis whether to adapt the quantizer, etc.

4. RESULTS AND DISCUSSION

In our experiments we consider a 3-level wavelet decomposition using the Daubechies D4 filters of [13]. As already mentioned the bit allocation is the same for all the schemes considered and schemes differ only in terms of the quantization. The first-order entropy is used to compare our results. We compare UTQ with (i) a classification based method with four classes (C-UTQ), where each class is coded with a separate UTQ quantizer, and (ii) an adaptive version (AC-UTQ) where the Laplacian parameter is locally adjusted. The overhead involved in each case is restricted to the three class thresholds and the four Laplacian parameters for each class. For example, with four classes, the maximum overhead involved would be of the order of 0.02 bits/coefficient for the coarser bands (assuming 16 bits per parameter). The small overhead involved will allow us to consider an increased number of classes. It can be seen from Table 1 that classification results in a significant increase in performance (about 1dB PSNR) and that further adaptation within each band is also useful. We emphasize that similar gains can be potentially achieved with more sophisticated quantization schemes, since our method essentially discriminates among classes of data and the alternative quantizing methods could then be trained on each class. Also, we believe that further improvements are possible in the classification process. These, as well as the

Rate	UTQ	C-UTQ	AC-UTQ
0.5bpp	32.64dB	33.58dB	33.72dB
1bpp	36.43dB	37.35dB	37.50dB

Table 1: Performance comparison in PSNR (dB)

extension to other quantization environments and the selection of larger number of classes, will be the subject of future work.

5. REFERENCES

- [1] W. Pennebaker and J. Mitchell, *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold, 1994.
- [2] J. M. Shapiro, "Embedded image coding using zero-trees of wavelet coefficients," *IEEE Trans. on Signal Proc.*, vol. 41, pp. 3445–3462, Dec. 1993.
- [3] D. Taubman and A. Zakhor, "Multirate 3-d sub-band coding of video," *IEEE Trans. on Image Proc.*, pp. 572–588, Sept. 1994.
- [4] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. on Image Proc.*, vol. 2, pp. 160–175, Apr. 1993.
- [5] Z. Xiong, K. Ramchandran, and M. Orchard, "Space-frequency quantization for wavelet image coding," *IEEE Trans. on Image Proc.*, 1995. Submitted.
- [6] R. L. Joshi, V. J. Crump, and T. R. Fischer, "Image subband coding using arithmetic coded trellis coded quantization," *IEEE Trans. on Circ. and Sys. for Video Tech.*, vol. 5, pp. 515–523, Dec. 1995.
- [7] R. Joshi, T. Fischer, and R. Bamberger, "Optimum classification in subband coding of images," in *Proc. of ICIP'94*, vol. 2, (Austin, Texas), pp. 883–887, Oct. 1994.
- [8] A. Ortega and M. Vetterli, "Adaptive quantization without side information," in *Proc. ICIP'94*, vol. 3, (Austin, Texas), pp. 856–860, Nov. 1994.
- [9] A. Ortega and M. Vetterli, "Adaptive scalar quantization without side information," *IEEE Trans. on Image Proc.*, 1996. Accepted for publication.
- [10] N. Farvardin and J. W. Modestino, "Optimum quantizer performance for a class of non-gaussian memoryless sources," *IEEE Trans. on Info. Th.*, vol. IT-30, pp. 485–497, May 1984.
- [11] M. J. Weinberger, J. J. Rissanen, and R. B. Arps, "Applications of universal context modeling to lossless compression of gray-scale images," *IEEE Trans. on Image Proc.*, 1995. Submitted.
- [12] B. Yu, "A statistical analysis of adaptive quantization based on causal past," in *IEEE Int'l Symp. on Info. Th., ISIT'95*, 1995.
- [13] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. on Pure and Appl. Math.*, vol. 41, pp. 909–996, Nov. 1988.