M2.7

Transform Image Coding with a New Family of Models

Gregory W. Wornell and David H. Staelin

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

A set of adaptive transform image coding schemes is developed based on a family of composite block source models for imagery. An iterative Maximum Likelihood (ML) algorithm is developed for resolving model parameters from training set data. Both unconstrained (adaptive transform, adaptive quantization) and constrained (fixed transform, adaptive quantization) coders are obtained from the image model parameters. The resulting coders give excellent performance in coding test imagery at a variety of bit rates, and they consistently outperform the well-known Chen and Smith adaptive transform coders. For example, a 2 dB improvement over the Chen and Smith scheme is obtained with a constrained coder with 128 classes operating at 0.5 bits/pixel. Computational limitations inhibit the design of unconstrained coders with more than approximately 15 classes.

Introduction

This work concerns the design of transform-based image coding systems suitable for use with monochrome still-frame digital images partitioned into $N \times N$-pixel, square blocks.

In classical non-adaptive transform coding systems (see Fig. 1), image blocks are passed through an orthogonal transformation, afterwhich the resulting transform coefficients are separately quantized. Frequently, one is interested in optimizing the system parameters based on some training set data. In this case, system parameter design begins with a particular source model. Traditionally (and often implicitly), image blocks are modeled as originating from a jointly-Gaussian block source characterized by block mean $\mathbf{m}$ and covariance $\mathbf{\Lambda}$. In this case, Maximum Likelihood estimates of the appropriate source parameters $\mathbf{m}$ and $\mathbf{\Lambda}$ are obtained from a suitable training set of image blocks via the usual ensemble averaging. Having fully specified the source, the appropriate coder parameters (i.e., the transformation and quantizer bank) for coding with minimum mean square error distortion at a prescribed average bit-rate are well-known. These parameters correspond to the Karhunen-Loève transform, log-variance bit assignment, and Max's quantizer level placement (see, e.g., [4, Chapter 12]).

The goal of this recent research [6] was to develop a corresponding strategy for the design of *adaptive* transform image coding systems. The particular adaptive transform coder structure

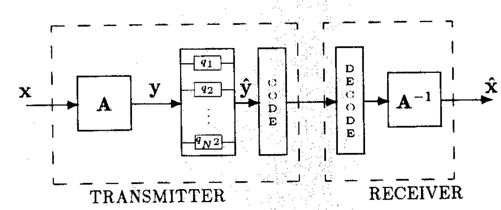

Figure 1: *Classical Non-Adaptive Transform Coding.*

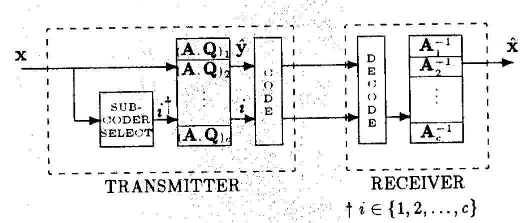

Figure 2: *Adaptive Transform Coding.*

of interest (see Fig. 2) is the switched-subcoder type, wherein a set of c subcoders (i.e., transform/quantizer-bank pairs) is made available for coding individual image blocks, and a suitable rule—the subcoder-select rule—is used to select an appropriate subcoder from the set for each block to be coded. Such a coder is termed a c-class coder.

In the subsequent discussion, it will prove convenient to adopt the following notation: let N^2-dimensional vectors represent $N \times N$-pixel image blocks (say, by a raster scan ordering of the pixels in a block). Then each block $\mathbf{x} = (x_1, x_2, \ldots, x_{N^2})$ corresponds to a point in *blockspace* $\mathbf{R}^{N^2}$.

The Model

As in the non-adaptive case, some model is required for the block process. In this case, the Composite Block Source Model (see Fig. 3) provides a rather general framework for such coding. This model consists of:

- a set of c distinct Gaussian subsources $s_1, s_2, \ldots, s_c$ each producing independent, identically-distributed image blocks

CH2561-9/88/0000-0777 $1.00 © 1988 IEEE

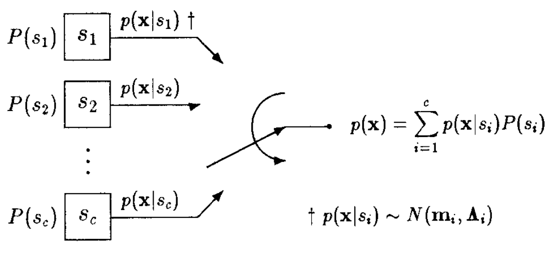

Figure 3: *The Composite Block Source Model.*

according to jointly Gaussian pdfs $p(\mathbf{x}|s_i)$ characterized by block mean $\mathbf{m}_i$ and covariance $\mathbf{\Lambda}_i$; and

- a memoryless, probabilistic switching mechanism characterized by $P(s_1), P(s_2), \ldots, P(s_c)$, the probabilities governing which subsource produces each particular block in an image.

In terms of our Euclidean blockspace $\mathbf{R}^{N^2}$, the model represents the decomposition of an arbitrary block pdf into a weighted superposition of N^2-dimensional jointly Gaussian pdfs. Physically, the subsources may be interpreted as a set of c "texture generators" for the imagery.

Parameter Resolution

In order to choose appropriate values for the coder parameters, the corresponding model parameters must be known. Suitable values for these parameters can be obtained via Maximum Likelihood parameter estimation techniques. Specifically, let

$$\mathcal{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \tag{1}$$

be a training set of n image blocks, and let Ξ be a many-dimensional vector of the model parameters. Then the model parameter estimates are those for which the likelihood function

$$\mathcal{L} \triangleq p(\mathcal{X}|\Xi) = \prod_{k=1}^{n} p(\mathbf{x}_k|\Xi) = \prod_{k=1}^{n} \sum_{i=1}^{c} p(\mathbf{x}_k|s_i)P(s_i) \tag{2}$$

is maximized subject to three obvious sets of constraints:

$$\sum_{i=1}^{c} P(s_i) = 1, \tag{3}$$

$$P(s_i) > 0, \; i = 1, 2, \ldots, c, \tag{4}$$

$$\mathbf{\Lambda}_i > 0, \; i = 1, 2, \ldots, c. \tag{5}$$

From this formulation, a set of implicit equations for the parameters can be derived in estimator-update form [6]. The estimator equations are (in terms of the intermediate quantities $\hat{P}(s_i|\mathbf{x}_k, \hat{\Xi})$):

$$\hat{P}(s_i) = \frac{1}{n} \sum_{k=1}^{n} \hat{P}(s_i|\mathbf{x}_k, \hat{\Xi}), \tag{6}$$

$$\hat{\mathbf{m}}_i = \frac{\sum_{k=1}^{n} \hat{P}(s_i|\mathbf{x}_k, \hat{\Xi})\mathbf{x}_k}{\sum_{k=1}^{n} \hat{P}(s_i|\mathbf{x}_k, \hat{\Xi})}, \tag{7}$$

$$\hat{\mathbf{\Lambda}}_i = \frac{\sum_{k=1}^{n} \hat{P}(s_i|\mathbf{x}_k, \hat{\Xi})(\mathbf{x}_k - \hat{\mathbf{m}}_i)(\mathbf{x}_k - \hat{\mathbf{m}}_i)^T}{\sum_{k=1}^{n} \hat{P}(s_i|\mathbf{x}_k, \hat{\Xi})}. \tag{8}$$

The update equation is:

$$\hat{P}(s_i|\mathbf{x}_k, \hat{\Xi}) = \frac{|\hat{\mathbf{\Lambda}}_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_k - \hat{\mathbf{m}}_i)^T \hat{\mathbf{\Lambda}}_i^{-1}(\mathbf{x}_k - \hat{\mathbf{m}}_i)} \hat{P}(s_i)}{\sum_{j=1}^{c} |\hat{\mathbf{\Lambda}}_j|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_k - \hat{\mathbf{m}}_j)^T \hat{\mathbf{\Lambda}}_j^{-1}(\mathbf{x}_k - \hat{\mathbf{m}}_j)} \hat{P}(s_j)}. \tag{9}$$

These equations are intuitively satisfying as they correspond to weighted sample means and sample covariances. The weights $\hat{P}(s_i|\mathbf{x}_k, \hat{\Xi})$ correspond to the probability that training set block $\mathbf{x}_k$ was generated by subsource s_i. Moreover, these equations define a steepest ascent iterative algorithm for the parameter estimation [2,3].

Empirically satisfying initial estimates for the algorithm are obtained using an ac-energy based partitioning of the training set based on the work of Chen and Smith [1]. Essentially, the ac-energy of each training set block (i.e., the average squared deviation from the block average) is used to partition the training set into c equally populated classes. Over each class, block means and covariances are estimated to provide subsource parameter initial estimates. Evidently, the switching probabilities are initially identical.

Coder Design

The corresponding adaptive transform coder for this model assigns one subcoder to each Gaussian subsource. Since subsource parameter estimates $(\hat{\mathbf{m}}_i, \hat{\mathbf{\Lambda}}_i)$ are available, the appropriate transform for each subcoder—the Karhunen-Loève transform—is obtained via the usual diagonalization of $\hat{\mathbf{\Lambda}}_i$. It is important to remark that, by construction, the subcoders operate "efficiently," i.e., they code Gaussian sources, so that the transform coefficients undergoing scalar quantization are rendered statistically independent by the chosen transforms.

The resulting sets of independent Gaussian transform coefficients are to be quantized so as to achieve minimum mean square error coding distortion subject to a constraint on the allowable average bit rate. Let the variance of the j^{th} transform coefficient from subcoder i be σ_{ij}^2, and let R_{ij} be the number of bits assigned to code this coefficient. Then the average mean square error quantization distortion D is given by

$$D = \sum_{i=1}^{c} P(s_i) \frac{1}{N^2} \sum_{j=1}^{N^2} \sigma_{e_{ij}}^2, \tag{10}$$

where the quantizer error variance $\sigma_{e_{ij}}^2$ is given approximately by [4]

$$\sigma_{e_{ij}}^2 \approx \epsilon_*^2 2^{-2R_{ij}} \sigma_{ij}^2. \tag{11}$$

(with ϵ_*^2 a constant appropriate to the minimum mean square error quantization of Gaussian random variables). The average bit rate constraint may be expressed as:

$$\sum_{i=1}^{c} P(s_i) \frac{1}{N^2} \sum_{j=1}^{N^2} R_{ij} \leq R. \tag{12}$$

In this case, minimum distortion is achieved when the R_{ij} are of the form

$$R_{ij} = \kappa + \frac{1}{2}\log_2 \sigma_{ij}^2, \tag{13}$$

with κ some constant chosen to attain the desired average bit rate R. Hence, a log-variance bit assignment rule like that used in non-adaptive transform coding applies. Iterative schemes (see, e.g., [4]) exist for obtaining non-negative, integer-valued R_{ij} obeying this rule.

Given the R_{ij}, the appropriate Gaussian pdf-optimized minimum mean square error distortion quantizers are designed as per Max [5].

The subcoder-select is designed to distribute incoming blocks among the c subcoders in such a way that each subcoder codes data obeying the block pdf for which it was designed. To accomplish this, pseudorandom assignment is employed. In particular, for each image block $\mathbf{x}$ to be coded, a weighted pseudorandom selection among the c subcoders is made according to the weights

$$\hat{P}(s_1|\mathbf{x},\hat{\Xi}), \hat{P}(s_2|\mathbf{x},\hat{\Xi}), \ldots, \hat{P}(s_c|\mathbf{x},\hat{\Xi}).$$

These weights correspond to the estimates of the probabilities that $\mathbf{x}$ came from each of $s_1, s_2, \ldots, s_c$. In terms of distributing blocks among the subcoders appropriately, this procedure is as good as knowing the exact subsource from which each block originates.

A Fast Coder

Frequently, one is interested in working with a restricted class of adaptive transform coders, perhaps due to computational complexity constraints. For example, it is reasonable to want to replace all the subcoder transforms with a single Discrete Cosine Transform (DCT). The DCT has the computational advantage of having both a separable form and a fast algorithm, yet at the same time it possesses good decorrelation properties when used with image data. In general, the approach taken to designing constrained coders of this type is to appropriately constrain the underlying model so that the resulting coder has the desired properties. For the present example, this is accomplished rather simply. What is required is that the training set blocks be pre-transformed by the DCT, and that the ML estimator-update equations be applied to the transformed training set data *with diagonal covariance matrices.* The resulting model leads immediately to a coder with DCT transforms for all subcoders. In addition to generating a computationally faster coder, the corresponding ML design procedure is much accelerated, too, due to the constrained covariance structures involved.

Experiments

The experimental test data consisted of the monochrome, 8-bit/pixel, 512×480-pixel image LENA (see Fig. 4), partitioned into 8×8-pixel blocks.

The coding experiments involve comparisons among the following three types of coders:

The ML Coders The coders designed from the ML parameters of the unconstrained model.

The FML Coders The fast coders designed from the ML parameters of the covariance-constrained model.

The C&S Coders The adaptive transform coders developed by Chen and Smith [1].

Figure 4: *The 8-bit/pixel 512×480-pixel test image LENA.*

In the first set of experiments, the three coders are each trained with LENA, then used to code LENA at 0.5 bits/pixel. The number of classes for each coder (i.e., the order of the coder) is varied from one to 128. In all cases, the ML algorithms were observed to converge satisfactorily within 10 iterations. Fig. 5 shows the SNR performance of the resulting coding systems. The following observations can be made:

- The C&S coder performance "saturates" at approximately 16 classes, while FML coder performance saturates at approximately 100 classes. The FML coder gives consistently better performance, and the asymptotic gain over the C&S coder is approximately 2 dB.
- The ML coder gives consistently better performance than the FML coder over the range of c for which comparisons may be made. Computational limitations prevent completion of the ML coder performance curve, so the ultimate performance potential of the ML coder remains unresolved.

In the second set of experiments, the three coders are trained with LENA, then used to code LENA with 8 classes. The coders are operated at average bit rates varying between 0.25 and 1.0 bits/pixel. The results, shown in Fig. 6, show strongly increasing performance with bit rate for all coders. Moreover, the performance margins of the ML over the FML coder, and the FML over the C&S coder also widen.

Finally, Fig. 7 shows the test picture LENA coded at 0.5 bits/pixel with the C&S (with 15 classes) and FML coders (with 128 classes). The scaled absolute coding errors are also displayed. Clearly, the FML coder offers a substantial improvement over the C&S coder both in terms of subjective performance and in terms of SNR.

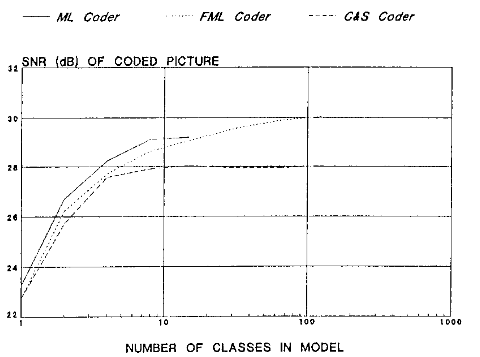

Figure 5: *Coder performance in coding LENA at 0.5 bits/pixel as a function of the number of classes.*

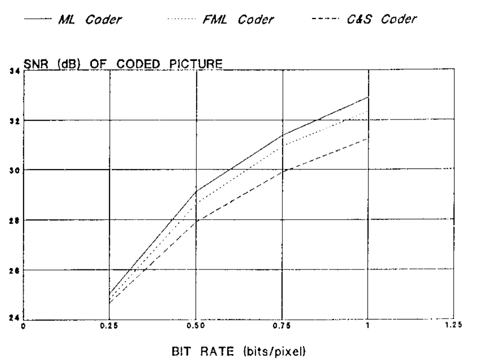

Figure 6: *8-class coder performance in coding LENA as a function of bit rate.*

Acknowledgements

This work was supported, in part, by a 1967 Science and Engineering Scholarship from the Natural Sciences and Engineering Research Council of Canada, which is presently held by the first author.

References

[1] W.H. Chen and C.H. Smith, "Adaptive Coding of Monochrome and Color Images," *IEEE Transactions on Communications,* Vol. COM-25, No.11, pp.1285-1292, Nov. 1977.

[2] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis,* New York, NY : Wiley-Interscience, 1973.

[3] V. Hasselblad, "Estimation of Parameters for a Mixture of Normal Distributions," *Technometrics,* Vol.8, No.3, pp.431-446, Aug. 1966.

[4] N.S. Jayant and P. Noll, *Digital Coding of Waveforms,* Englewood Cliffs, NJ : Prentice-Hall, 1985.

[5] J. Max, "Quantizing For Minimum Distortion," *IRE Transactions on Information Theory,* Vol. IT-6, No.1, pp.7-12, Mar. 1960.

[6] G.W. Wornell, "Transform Image Coding with Composite Block Source Models," S.M. Thesis, M.I.T., Cambridge, MA, June 1987.

Figure 7: *Top: C&S coded LENA, 0.5 bits/pixel, 15 classes; Middle: FML coded LENA, 0.5 bits/pixel, 128 classes; Bottom: Absolute Errors × 30 for C&S (left) and FML (right) coders.*