ADAPTIVE CODING OF HYPERSPECTRAL IMAGERY

Glen P. Abousleman

Motorola, Systems Solutions Group
8201 E. McDowell Road, MD H1176
Scottsdale, AZ 85257
p26994@email.mot.com

ABSTRACT

Two systems are presented for compression of hyperspectral imagery. These systems utilize adaptive classification, trellis-coded quantization, and optimal rate allocation. In the first system, DPCM is used for spectral decorrelation, while an adaptive wavelet-based coding scheme is used for spatial decorrelation. The second system uses DPCM in conjunction with an adaptive DCT-based coding scheme. In each system, entropy-constrained trellis-coded quantization (ECTCQ) is used to quantize the transform coefficients. Entropy-constrained codebooks are designed for generalized Gaussian distributions by using a modified version of the generalized Lloyd algorithm. The wavelet-based system compresses an AVIRIS hyperspectral test sequence at 0.118 bits/pixel/band, while retaining an average peak signal-to-noise ratio (PSNR) of 41.24 dB. The DCT-based system achieves the same bit rate with an average PSNR of 40.72 dB.

1. INTRODUCTION

There has recently been increased interest in the field of remote sensing to perform precise recording of sensed energy in a number of narrow wavelength slices. Since various surface materials of interest have absorption features that are only 20 to 40 nm wide [1], the ability to discriminate among such features on the Earth's surface requires sensors with very high spatial, spectral, and radiometric sensitivity.

The next-generation of high-dimensional multispectral sensors are capable of collecting radiation in hundreds of distinct spectral bands in the visible and near-infrared wavelength region, with each band being on the order of 1000 × 1000 pixels. Combining these parameters with a radiometric sensitivity of 12 bits would produce a single *hyperspectral* image which comprises several hundred megabytes of digital information.

The algorithms developed in the present work for compression of hyperspectral imagery are based on DPCM used in conjunction with either the discrete wavelet transform (DWT), or the discrete cosine transform (DCT). In the wavelet-based coder, DPCM is used for spectral decorrelation, and each "error image" is coded using an adaptive 2-D DWT coder. In this coder, the error image is transformed using a 10-band octave split. Each subband is classified into one of J classes by maximizing the coding gain. All resulting sequences are quantized using entropy-constrained trellis-coded quantization (ECTCQ) [2].

In the DCT-based coder, the error image is divided into nonoverlapping 8 × 8 blocks and transformed using the DCT. Each block is classified into one of J classes by maximizing the coding gain. All resulting "like-coefficient" sequences are quantized using ECTCQ. In both coders, codebooks are optimized for different generalized Gaussian distributions. Codebook design uses a modified version of the generalized Lloyd algorithm in a training-sequence-based iterative scheme. Rate allocation is performed in an optimal fashion by an iterative technique that uses the rate-distortion performance of the various trellis-based quantizers.

2. ADAPTIVE HYPERSPECTRAL IMAGE CODERS

Consider the encoder configuration shown in Figure 1. Here, the DPCM loop operates on entire images rather than on individual sequences. Given an image, $\mathbf{x}_{n-1}$, the next image in the hyperspectral sequence can be estimated, and an "error image" can be formed from the difference $\boldsymbol{\epsilon_n} = \mathbf{x}_n - \tilde{\mathbf{x}}_{n|n-1}$. The error image (at each instant in time) is spatially correlated and can be quantized using any image coding scheme. Note that the error image must be decoded within the encoder loop so that the quantized image, $\hat{\mathbf{x}}_n$, can be constructed and used to predict the next image.

The prediction error images have much lower energy than the original bands and can be subjected to very coarse quantization. The bit rate chosen to encode each error image will be the asymptotic bit rate for the system. The first spectral band is encoded and transmitted (as the initial conditions for DPCM) at a total rate (including side information) of R_1 bits/pixel (bpp).

Testing of data obtained by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [3], revealed that the spectral correlation coefficient, ρ, for any pixel in the hyperspectral image, is approximately 0.95. Accordingly, this value of ρ was used in the DPCM loop.

In the wavelet-based system, each error image is encoded using a scheme similar to that in [4]. The image is transformed using the 2-D DWT. A 10-band octave decomposition is used. The block size used for classification within each subband is adjusted so as to correspond to 16 × 16 blocks in the error image. Each subband is classified into J classes by maximizing the coding gain as discussed below. All $10J$ sequences are normalized by dividing by their respective standard deviations. The mean is subtracted from the lowest frequency subband.

The total side information to be transmitted consists of the mean of the lowest frequency subband and the standard deviations of all $10J$ coefficient sequences. These quantities are quantized using 16-bit uniform scalar quantizers for a total of $(10J+1) \times 16$ bits. The initial trellis state for each sequence must also be transmitted, which (for 4-state trellises) equals $20J$ bits. For $J = 4$, the total side information consists of 736 bits. This corresponds

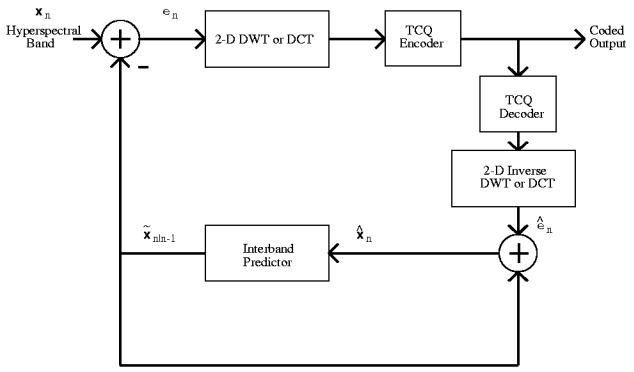

Figure 1: Hyperspectral image encoder.

to ≈ 0.0112 bpp for a 256×256 image, and ≈ 0.0028 bpp for a 512×512 image. In addition, the classification maps for the entire hyperspectral sequence are derived from the first spectral band and thus contribute only ≈ 0.00195 bits/pixel/band for a 40-band sequence. The first spectral band is encoded and transmitted (as the initial conditions for DPCM) at a total rate (including side information) of R_1 bpp.

In the DCT-based system, each image is divided into non-overlapping 8×8 blocks and transformed using the 2-D DCT. Prior to application of the DCT, each block is assigned to one of J classes by maximizing the coding gain. For each class, DCT coefficients corresponding to the same frequency within each block are grouped into sequences to be encoded using ECTCQ. All DCT coefficients are normalized by subtracting their mean (only the sequences corresponding to the DC transform coefficients have non-zero mean) and dividing by their respective standard deviations. Since each class contains 64 DCT coefficient sequences, the total number of sequences to be encoded would then be $64J$.

The side information required to encode each error image (and the first spectral band) consists of the means of the J DC sequences and the standard deviations of all $64J$ sequences. These quantities are quantized using 8-bit uniform scalar quantizers to yield $(64J + J) \times 8$ bits. In addition, the initial trellis state of each sequence requires 2 bits (for a 4-state trellis) which yields $128J$ bits. The total side information is then $648J$ bits which corresponds to $1296/(256)^2 = 0.0198$ bpp for a 256×256 image and $J = 2$, or $2592/(256)^2 = 0.0396$ bpp for $J = 4$. The first spectral band is encoded and transmitted (as the initial condition for DPCM) at a total rate (including side information) of R_1 bits/pixel. In addition, the classification maps for the hyperspectral sequence must also be transmitted. One map is used for every ten spectral bands. Each map requires 1024 bits for $J = 2$, or 2048 bits for $J = 4$. Averaged over ten spectral bands, this corresponds to 0.00156 bpp and 0.00313 bpp for $J = 2$ and $J = 4$, respectively.

The DWT or DCT coefficient sequences are assumed to have various generalized Gaussian statistics. Accordingly, codebooks were designed using sample sequences derived from generalized Gaussian pseudo random number generators as discussed below. Additionally, rate allocation is performed by using the algorithm in [5].

The classification algorithm is similar to that presented in [4]. Consider a source X of length NL divided into N blocks of L consecutive samples, with each block assigned to one of J classes. If the samples from all blocks assigned to class i $(1 \leq i \leq J)$ are grouped into source X_i, the total number of blocks assigned to source X_i is N_i. Let σ_i^2 be the variance of X_i and p_i be the probability that a sample belongs to X_i (i.e., $p_i = N_i/N, 1 \leq i \leq J$). The algorithm in [4] is a pairwise maximization of coding gain and is repeated here for convenience:

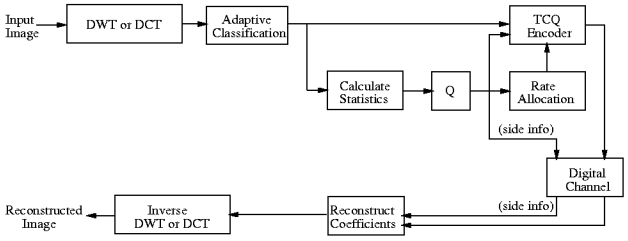

Figure 2: Adaptive image coder.

1. Initialize $N_1, N_2, \ldots, N_J$ to satisfy $\sum_{i=1}^{J} N_i = N, N_i > 0$ for $1 \leq i \leq J$. Let $j = 1$ and $\underline{N}_{prev} = [N_1, N_2, \ldots, N_J]'$.
2. Find N_j' and N_{j+1}' such that $N_j' + N_{j+1}' = N_j + N_{j+1}$ and $(\sigma_j^2)^{p_j'}(\sigma_{j+1}^2)^{p_j'}$ is minimized.
3. $N_j = N_j'$ and $N_{j+1} = N_{j+1}'$.
4. $j = j + 1$. If $j < J$, go to step 2.
5. Let $\underline{N} = [N_1, N_2, \ldots, N_J]'$. If $\underline{N}$ is equal to $\underline{N}_{prev}$, then **STOP**. Otherwise, $j = 1$. $\underline{N}_{prev} = \underline{N}$. Go to step 2.

Here, the average mean squared energy of a block (i.e., $E = (\sum_{i=1}^{L} x_i^2)/L$) is the criterion for classification.

2.1. Codebook Design

The probability distribution of each sequence to be encoded is modeled by the so-called *Generalized Gaussian Distribution* (GGD), whose probability density function (pdf) is given by

$$f_X(x) = \left[\frac{\alpha\eta(\alpha,\sigma)}{2\Gamma(1/\alpha)}\right] \exp\{-[\eta(\alpha,\sigma)|x|]^{\alpha}\} \tag{1}$$

where

$$\eta(\alpha,\sigma) \equiv \sigma^{-1}\left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{1/2}. \tag{2}$$

The shape parameter α describes the exponential rate of decay and σ is the standard deviation of the associated random variable [6]. The gamma function $\Gamma(\cdot)$ is defined as

$$\Gamma(n) = \int_0^{\infty} e^{-x}x^{n-1}dx. \tag{3}$$

Distributions corresponding to $\alpha = 1.0$ and 2.0 are Laplacian and Gaussian, respectively. Figure 3 shows generalized Gaussian pdfs corresponding to $\alpha = 0.5$, 1.0, 1.5, 2.0, and 2.5.

It can be shown that

$$E[X^4] = \int_{-\infty}^{\infty} x^4 f_X(x)dx = K\sigma^4 \tag{4}$$

or

$$K = \frac{E[X^4]}{\sigma^4} = \frac{\Gamma(5/\alpha)\Gamma(1/\alpha)}{\Gamma(3/\alpha)^2} \tag{5}$$

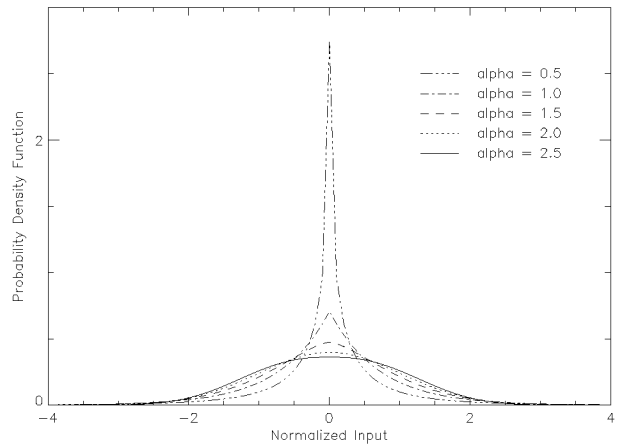

Figure 3: Probability density function for generalized Gaussian distributions with alpha values of 0.5, 1.0, 1.5, 2.0, and 2.5.

where K is the fourth central moment, or Kurtosis. Recall that the Kurtosis is a measure of the peakedness of a given distribution. If a pdf is symmetric about its mean and is very flat in the vicinity of its mean, the coefficient of Kurtosis is relatively small. Similarly, a pdf that is peaked about its mean has a large Kurtosis value.

The sample Kurtosis of any sequence can be calculated easily and used as a measure by which the distribution of the sequence can be determined. Since there is a one-to-one mapping between the shape parameter α and K, this mapping can be used to determine the appropriate α for a particular sequence.

Codebooks were designed for generalized Gaussian distributions with α values of 0.5, 0.75, 1.0, 1.5, and 2.0, using the algorithm in [2]. It was shown in [2] that for the Gaussian distribution, optimum codebooks do not yield significant MSE improvement over uniform codebooks at rates greater than 2.5 bits/sample. Experimentation revealed that this is also true for $\alpha = 1.5$ and $\alpha = 1.0$. However, for $\alpha = 0.75$, optimum codebooks are superior up to 3.0 bits/sample, while for $\alpha = 0.5$, optimum codebooks should be used up to 3.5 bits/sample. Accordingly, for α values of 2.0, 1.5, and 1.0, optimum codebooks were designed in one-tenth bit increments up to 2.5 bits/sample, while for $\alpha = 0.75$ and $\alpha = 0.5$, optimum codebooks were designed in one-tenth bit increments up to 3.0 and 3.5 bits/sample, respectively. Thereafter, uniform codebooks were designed in one-tenth bit increments up to 12 bits/sample. Training sequences consisted of 100,000 samples derived from generalized Gaussian pseudo-random number generators, each tuned to the appropriate α value.

2.2. Rate Allocation

Rate allocation is performed by using the algorithm presented in [5]. The overall MSE incurred by encoding the coefficient sequences using ECTCQ at an average rate of R_s bits/coefficient is represented by

$$E_s = \sum_{i=1}^{K} \alpha_i \sigma_i^2 E_{ij}(r_i) \tag{6}$$

where σ_i^2 is the variance of sequence i, $E_{ij}(r_i)$ denotes the rate-distortion performance of the j^{th} quantizer (i.e., the quantizer corresponding to the Kurtosis of sequence i) at r_i bits/sample, K is the number of data sequences, and α_i is a weighting coefficient to account for the variability in sequence length. For a 10-band decomposition and J classes, $K = 10J$. For 8×8 blocks and J classes, $K = 64J$.

The rate allocation vector $B = (r_1, r_2, \ldots, r_K)$ is chosen such that E_s is minimized, subject to an average rate constraint:

$$\sum_{i=1}^{K} \alpha_i r_i \leq R_s \text{ bits/coefficient.} \tag{7}$$

It is shown in [5] that the solution $B^*(r_1^*, r_2^*, \ldots, r_K^*)$ to the unconstrained problem

$$\min_{B} \left\{ \sum_{i=1}^{K} (\alpha_i \sigma_i^2 E_{ij}(r_i) + \lambda \alpha_i r_i) \right\} \tag{8}$$

minimizes E_s subject to $\sum_{i=1}^{K} \alpha_i r_i \leq \sum_{i=1}^{K} \alpha_i r_i^*$. Thus, to find a solution to the constrained problem of equations (6) and (7), it suffices to find λ such that the solution to equation (8) yields $\sum_{i=1}^{K} \alpha_i r_i^* \leq R_s$. Procedures for finding the appropriate λ are given in [5].

For a given λ, the solution to the unconstrained problem is obtained by minimizing each term of the sum in (8) separately. If S_j is the set of allowable rates for the j^{th} quantizer and r_i^* is the i^{th} component of the solution vector B^*, then r_i^* solves

$$\min_{r_i \in S_j} \{\alpha_i \sigma_i^2 E_{ij}(r_i) + \lambda \alpha_i r_i\}. \tag{9}$$

3. RESULTS AND CONCLUSIONS

Coding simulations were performed using a 140-band, 8-bit hyperspectral image sequence of Cuprite, Nevada, obtained by the AVIRIS system. The bands were 256×256 pixels and the performance of the coder is reported using the peak signal-to-noise ratio.

The first band in the sequence is quantized at $R_1 = 0.75$ bits/pixel and is used as the initial condition for the spectral DPCM. This rate was chosen so that the PSNR of the coded first band did not significantly deviate from the average PSNR of the sequence, when encoded at an asymptotic rate R_s, of 0.10 bits/pixel/band (b/p/b).

Figure 4 shows the PSNRs obtained by encoding bands 30 through 69 of the hyperspectral sequence using the DWT-based system with $J = 4$ classes, and the DCT-based system with $J = 2$ classes. For comparison, Figure 4 also shows the performance curves of the 3-D DCT and hybrid coders presented in [7], the ECPTCQ coder presented in [8], and the robust wavelet coder presented in [9]. All coders were operated at an asymptotic bit rate[1] of $R_s = 0.1$ b/p/b. The average PSNR of the adaptive-DWT-coded sequence is 41.24 dB, while the average PSNR of the adaptive-DCT-coded sequence is 40.72 dB. The average PSNRs of the 3-D DCT, hybrid, ECPTCQ, and robust wavelet systems are 40.75 dB,

[1] It should be noted that the asymptotic rates of the adaptive-DWT, adaptive-DCT, and hybrid systems include the side information required to encode each error image, since the contribution of the side information to overall rate will remain constant regardless of sequence length.

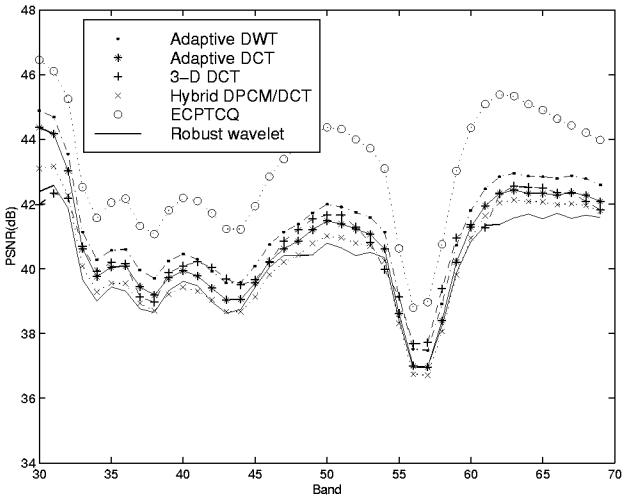

Figure 4: Performance of encoding hyperspectral sequence at $R_s = 0.10$ b/p/b.

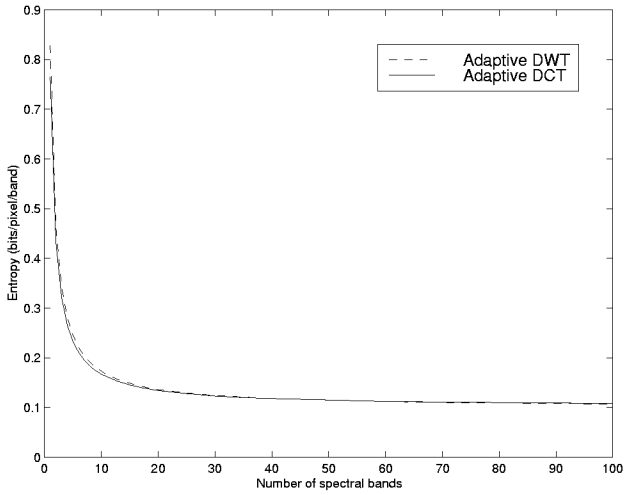

Figure 5: Overall rate versus number of spectral bands.

40.29 dB, 43.10 dB, and 40.13 dB, respectively. The dip in PSNR around bands 56 and 57 is indicative of high sensor noise that is clearly evident upon visual examination.

Despite having lower PSNR performance, the adaptive DCT system gains advantages over the adaptive DWT coder in terms of computational simplicity, and over the ECPTCQ system in terms of memory requirements. The adaptive coders require only two bands at once to encode an image sequence, while the ECPTCQ system uses the entire transformed sequence. When compared to the 3-D DCT coder, the adaptive DCT coder achieves comparable PSNR performance while offering lower complexity and memory requirements. Both adaptive coders outperform the DCT-based hybrid system, and the robust wavelet system. However, the robust wavelet system is a non-entropy-coded design which gives it several advantages as outlined in [9].

Figure 5 shows the overall coding rate (including all side information) of the adaptive systems as a function of sequence length. Note that for encoding short subsequences (i.e., less than 10 bands), the overall rate of either system is dominated by the rate required to code the first spectral band. The overall rate for encoding the 40-band sequence using either system is ≈ 0.118 b/p/b when R_s = 0.1 b/p/b. If all 140 bands were coded, the overall rate would be 0.106 b/p/b. It is evident that at least 73 bands are required such that the overall rate is within 10% of the asymptotic rate (when R_1 = 0.75 bpp). Note that for short sequence lengths, the adaptive wavelet system exhibits slightly higher overall rate for a given asymptotic rate.

Finally, we would like to note that many compression systems discussed in the literature are image class dependent. That is, the codebooks for each system must be optimized for a specific image class. For example, codebooks optimized for urban imagery may perform poorly when used on rural imagery. An alternative, which is used by other systems, is to design one set of codebooks using a wide variety of training data. Our systems do not utilize image training data, but rather statistical models which are scene independent, thus affording robustness to varying terrain.

4. REFERENCES

[1] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, pp. 1147–1153, June 1985.

[2] M. W. Marcellin, "On entropy-constrained trellis coded quantization," *IEEE Trans. Commun.*, vol. 42, pp. 14–16, Jan. 1994.

[3] W. M. Porter and H. T. Enmark, "A system overview of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Imaging Spectroscopy II,* G. Vane, Editor, Proc. SPIE 834, pp. 22–29, 1987.

[4] R. L. Joshi, T. R. Fischer, and R. H. Bamberger, "Optimum classification in subband coding of images," *Proc. International Conference on Image Processing*, pp. 883–887, Nov. 1994.

[5] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 36, pp. 1445–1453, Sept. 1988.

[6] N. Farvardin and J. W. Modestino, "Optimum quantizer performance for a class of non-gaussian memoryless sources," *IEEE Trans. Inform. Th.*, vol. 30, pp. 485–497, May 1984.

[7] G. P. Abousleman, M. W. Marcellin, and B. R. Hunt, "Compression of hyperspectral imagery using the 3-D DCT and hybrid DPCM/DCT," *IEEE Trans. Geoscience and Remote Sensing*, vol. 33, pp. 26–34, Jan. 1995.

[8] G. P. Abousleman, M. W. Marcellin, and B. R. Hunt, "Hyperspectral image compression using entropy-constrained predictive trellis coded quantization," *IEEE Trans. Image Proc.*, vol. IP-6, pp. 566–573, Apr. 1997.

[9] G. P. Abousleman, "Wavelet-based hyperspectral image coding using robust fixed-rate trellis-coded quantization," *Algorithms for Multispectral and Hyperspectral Imagery IV,* S. Shen, Editor, Proc. SPIE 3372, pp. 74–85, 1998.