

Lossless Predictive Compression of Hyperspectral Images

Hongqiang Wang and Khalid Sayood
Department of Electrical Engineering
University of Nebraska, Lincoln

1 Introduction

After almost three decades of successful data acquisition using multispectral sensors the first space based hyperspectral sensors were launched in 2000 on the NASA EO-1 satellite. However, airborne hyperspectral sensors such as AVIRIS, among others, have been generating useful data for many years. The advent of the space-borne ALI and Hyperion sensors as well as the successes of AVIRIS presage the development of many more hyperspectral instruments. Furthermore the success of multispectral imagers such as the Enhanced Thematic Mapper Plus (EMT+) on the LANDSAT-7 mission and the modestly named 36 band MODIS (Moderate Resolution Imaging Spectroradiometer) instrument aboard the *Terra* satellite and the *Aqua* spacecraft promises the deployment of significant numbers of other such instruments. The use of multispectral and hyperspectral sensors, while opening the door to multiple applications in climate observation, environmental monitoring, and resource mapping, among others, also means the generation of huge amounts of data that needs to be accommodated by transmission and distribution facilities that cannot economically handle this level of data. This means that compression, always a pressing concern [1], is now imperative. While in many cases the use of lossy compression may be unavoidable, it is important that the design always include the possibility of lossless recovery. Much effort usually has gone into the reduction of noise in the instruments. The voluntary addition of noise due to compression can be a bitter pill to swallow.

Compression is needed, and can be applied, in several different places from where the images are acquired to the end-user. At the point of acquisition compression may be required under several different scenarios. The satellite carrying the sensor may not be in continuous contact with ground stations. In the interval between contacts the data has to be stored on board. If these intervals are of any significant length the amount of data generated is likely to be very

high and compression becomes imperative. Any compression at this point has to be lossless. Even if there is relatively frequent contact between the satellite and the ground station if the portion of the bandwidth available to the instrument is less than the raw data rate again compression is required, and again this compression has to be lossless. Once the data is on the ground it needs to be archived. This is another point at which lossless compression may be needed. Finally, the data has to be distributed to end-users. Depending on the amount of data desired by a particular end-user compression may or may not be required. Furthermore, depending on the application of interest to the end-user this compression can be lossy or lossless. Thus, while it can be argued for certain applications that given a stringent enough distortion constraint lossy compression may be possible there are many scenarios in which only lossless compression will be acceptable. Various lossless compression schemes based on transform coding, vector quantization [2, 3], and predictive coding have been proposed. In this chapter we examine a number of lossless compression schemes based on predictive coding.

Most lossless compression schemes include prediction as part of the algorithm. Strong local correlations in the image allow prediction of the pixel being encoded. The difference between the pixel and the prediction, known as the prediction error or prediction residual, usually has a much lower first order entropy than the pixel itself and therefore, can be encoded with fewer bits. Well known predictive lossless image compression techniques developed for use with natural images include CALIC (*Context-based Adaptive Lossless Image Compression*) [4, 5] and LOCO-I (*Low Complexity Lossless Compression for Images*) [6] which is part of JPEG-LS, the ISO/ITU standard for lossless and near-lossless compression of images. These, as well as others, have their three dimensional counterparts. In recent years, various lossless hyperspectral image compression schemes using reversible integer wavelet transforms have been proposed. These include EZW (*Embedded Zerotree Wavelet*) [7] and SPIHT (*Set Partitioning in Hierarchical Trees*) compression [8]. These schemes takes advantage of the fact that when an image is decomposed based on its frequency the coefficients at higher frequency are closely related to the coefficients at lower frequency. Wavelet decomposition and predictive coding can be combined in two ways. The prediction error residuals can be encoded using a wavelet coder [9], or the decomposition of the images can be used as a preprocessing step prior to prediction.

In the following we describe some of the more popular predictive coding schemes as they have been applied to hyperspectral compression.

2 Predictive Coding - Overview

Compression algorithms depends on the underlying model of data we are dealing with. For speech data, which can be modeled by a mechanism that is fundamentally similar across individuals, it is relatively easy to obtain a model that generates the data in a standard way. This leads to the development of speech compression algorithms. However, for image data, the absence of a single generative mechanism means that it is very difficult to seek a single mode to be used for image compression. The problem can be circumvented to some extent by focusing on the sink or user, of the information, rather than the source. If images aims for human perception, they are likely to contain quasi-consent regions encoded by edges or regions of short transitions. We can model this as high frequency which consists of the edges combined with low frequency signal or the background. This view leads to approaches which decompose the image prior to further modeling, such as wavelet transform coding. A somewhat different approach is to focus on local relationships between pixels. This approach naturally leads to predictive coding algorithms which encode the difference between a prediction of the value of the pixel being encoded and the pixels itself.

Image compression algorithms based on predictive coding generally use local modes to generate a prediction of the pixel being encoded. The input to the predictor are pixels available to both encoder and the decoder. The prediction error is then encoded with respect to a probability model. For instance, differential pulse code modulation (DPCM) schemes are usually employed for lossless compression of two-dimensional (2-D) image. This generally consists of a spatial prediction followed by an entropy coder for the residual images. Given the neighborhood pixels are causal, i.e. those pixels are already known for both encoder and decoder, a predictor can be designed such that the prediction is optimized in the sense of minimum mean squared error (MMSE) over the entire image. This can be done by using standard linear combination, or regression method. If the data is stationary, such prediction would be optimum. For speech signals, we can make an assumption that they are stationary

within a short segment, and it demonstrated that this assumption can result in very good results. Unfortunately, the assumption is generally not true for image data. As a consequence, it is not possible to figure out a single model which can capture all local structures in images, and most modern lossless image compression algorithms employ multiple predictors to capture the local relationships in different parts of the images, or explore the underlying data structure in a progressive manner. For instance, rather than use a uniform linear prediction model, ADPCM, updates the coefficients of predictors as long as new data is available. Another method is to switch between various predictors based on the local context or by combining the various predictions in some manner. This point has been demonstrated by the CALIC algorithm and the LOCO-I algorithm used in JPEG-LS standard.

For hyperspectral images, an additional source of richness is the existence of two types of correlation: spatial correlation between adjacent pixels in a band, and spectral correlation between adjacent bands. We can view this increase in the size of the neighborhood as an extension from 2-D prediction to 3-D prediction. However, this extension is not always straightforward and the direct extension of a 2D method to three dimensions may not always provide much benefit, and can at times be detrimental. As such, it is necessary to develop predictors specialized for the application 3-D hyperspectral images.

A significant improvement in coding performance of predictive coding algorithms used with hyperspectral images can be obtained if the bands are re-ordered to enhance the band-to-band predictability [10]. The motivation behind this idea is that the reflection intensity of the frequency components by materials in the earth generally are not monotonically increasing or decreasing. As the reordering has to be done offline this enhancement is not always possible.

3 3-Dimensional Adaptive Differential Pulse Coded Modulation (ADPCM)

Adaptive Differential Pulse Code Modulation for lossless compression of AVIRIS images was first proposed by Roger and Cavenor in 1996 [11]. The various predictors in the 3-D ADPCM algorithm exploited spectral (interband) correlation, spatial (intraband) correlations, and the combination of spectral and spatial correlation for the lossless compression of AVIRIS images. The residual images

obtained were then coded by using various variable length codes.

Similar to most lossless predictive coding schemes, the method has two stages, predictive decorrelation (which produces the residuals) and residual encoding. The residual image is obtained by taking the difference between the pixel values and their prediction values. The residual encoding primarily exploited Rice codes [12]. Two other forms of variable length coding were also evaluated.

3.1 Predictors for Image Decorrelation

Five predictors shown in Table 1 were used in this work. (The authors indicate that these five were selected from a set of 25 candidates). These predictors are causal, that is, the pixels values used for predictions are all available to both the encoder and the decoder. The predictors prefixed with “SA” are spatial predictors, or intraband predictors, and the predictors prefixed with “SE” are spectral predictors, or interband predictors. The predictors with a prefix of “SS” are spatial and spectral predictors which explore both spatial and spectral correlations.

Two of the five predictors have constant coefficients, while three predictors whose names contain an “o” have variable coefficients which are obtained by minimizing the variance of the prediction errors within each row. These coefficients are obtained by the least-squares minimization of $\sum_{j=1}^n (\hat{x}_{i,j,\lambda} - x_{i,j,\lambda})^2$, where n is the number of pixels in a row, i denotes the row, j denotes the column, and λ denotes the spectral band. The coefficients a , b , c , and d must be transmitted as header information to the decoder for decompression purpose. In the algorithm, a is quantized using 16 bits, and b , c , and d are quantized using 12bits for transmission.

3.2 Encoding of Residuals

The residual images are further entropy encoded. A simple variable length code, the Rice code, and two other variable length codes based on Huffman codes are evaluated. Rice codes have been shown to be the optimum codes for the information sources with a geometric distribution [12]. They are also very simple to implement in hardware. The residuals within each row are assumed

Predictor	Formula for Predictor $\hat{x}_{i,j,\lambda}$
SA-2RC	$(x_{i-1} + x_{i,j-1,\lambda})/2$
SS-1	$x_{i,j-1,\lambda} + x_{i,j,\lambda-1} - x_{i,j-1,\lambda-1}$
SE-o1B	$a + bx_{i,j,\lambda-1}$
SE-o2B	$a + bx_{i,j,\lambda-1} + cx_{i,j,\lambda-2}$
SS-o1	$a + bx_{i,j-1,\lambda} + cx_{i,j,\lambda-1} + dx_{i,j-1,\lambda-1}$

Table 1: Definition of Five Predictors

to have a geometric distribution and the Rice codes are then applied to each individual rows.

Two other variable length codes based on Huffman codes are evaluated. The first one is semi-static Huffman coding with truncation. The advantages of these codes is that the coding allows for an uncertain range of residuals. However it needs codebooks to be constructed to match the data's statistics. This necessarily means that the coding is a two-pass scheme. The second one is dynamic Huffman coding in which the Huffman code tables are updated as the compression proceeds. To improve the compression performance, eight codebooks are used in the algorithm and the residuals for each row are assigned to a codebook according to their statistics.

The authors provide compression ratios obtained by using 5 predictors, 10 AVIRIS hyperspectral images, and three variable length codes. It demonstrated that the predictor SA-2RC that only uses spatial correlations is the poorest, even though among the (old) JPEG predictors [13] it performs the best for 2D natural images. The SS-1 predictor which exploits both spectral and spatial correlation works much better than the SA-2RC. The best performance is obtained by using the three optimized predictors.

The overall compression ratio for the typical hyperspectral images are in the range of 1.6 -2.0:1. The methods can provide 1.0-1.8 bits/pixels more compression than the best of the JPEG predictors. They also perform better than the prediction tree approach of Memon et al. [14].

4 Interband Context-based, Adaptive, Lossless Image Codec (CALIC)

CALIC operates by exploiting local structure in images in an exhaustive fashion both for prediction and for context coding of the residuals. Since its introduction in 1995, CALIC has remained at the forefront in terms of compression performance. Two dimensional CALIC, was extended to 3-dimensional hyper-spectral images by Wu and Memon [15]. The three dimensional CALIC was then improved upon by Magli et al. [16]. We present how the prediction in CALIC is extended from 2-D to 3-D and how spectral correlation is used to improve the prediction.

4.1 Intraband CALIC

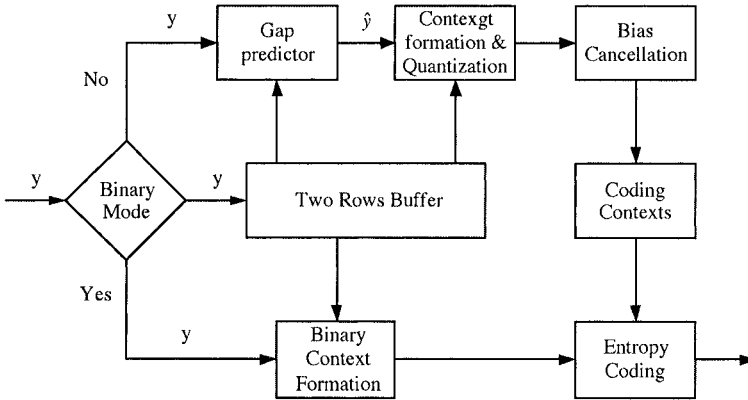


Figure 1: Schematic description of the CALIC image coding system

CALIC operates in two modes: a binary mode and a continuous-tone mode as shown in Fig.1. In the continuous-tone mode, predictive coding is used, therefore, it is the continuous tone mode that is of interest to us. In this mode the encoder obtains a prediction which is adjusted based on the local gradients. The prediction is then further adjusted based on the context in which the pixel occurs. The performance of the predictor in similar contexts is monitored and any bias in the prediction computed. This bias is then removed from the prediction. The prediction error is then encoded using a context based variable

length coder, either a Huffman coder or an arithmetic coder.

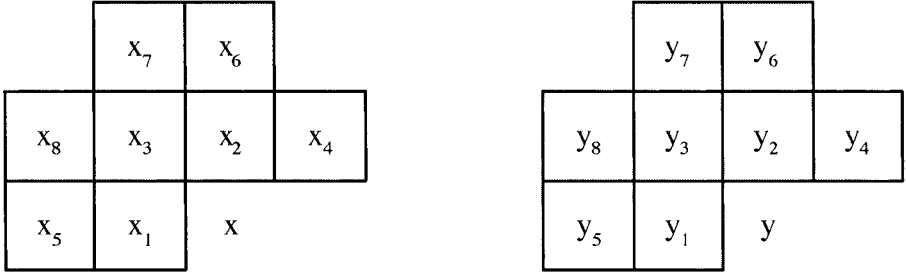


Figure 2: Labelling of neighboring pixels used in prediction and modeling.

4.2 Interband CALIC

The predictor in 3D CALIC incorporates the 2D CALIC predictor in a correlation-driven adaptive prediction framework. For a particular pixel y , we trace back to its counterpart pixel x at the same spatial location in the reference band, as shown in Fig. 2. With a causal system we assume all the pixels in the pixel set $Y = [y_1, y_2, \dots, y_n]^T$ and $X = [x_1, x_2, \dots, x_n]^T$ are available to both the encoder and decoder, except the pixel y , which we are trying to predict. The correlation between the two data sets is measured by the correlation coefficient which is defined as follows

$$\rho = \frac{\sum_{i=0}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_{i=0}^n [(x_i - \bar{x})^2 (y_i - \bar{y})^2]}} \quad (1)$$

where \bar{x} is the mean of x_i , and \bar{y} is the mean of y_i , $i = 0, 1, \dots, n$.

If the correlation is very high this means that we can use the pixels in the previous band around the spatial location of the pixel being encoded to form

the prediction. In particular, let

$$A = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} \quad (2)$$

and assume that x and y satisfy a linear relationship as

$$y = \alpha x + \beta \quad (3)$$

where α is the gain and β is the offset. Using standard linear regression techniques, the system that minimizes the error (in the sense of mean square error) can be obtained by

$$[\alpha, \beta] = (A^T A)^{-1} A^T Y \quad (4)$$

where A^{-1} is the inverse matrix of A , and A^T is the transpose of the matrix A .

Thus, $\hat{y} = \alpha x + \beta$ can be used as a predictor of y if the correlation coefficient ρ is sufficiently high. A variation of the predictor as follows is used when there is a sharp edge at y :

$$\hat{y} = \frac{|x - x_2| \hat{y}_h + |x - x_1| \hat{y}_v}{|x - x_2| + |x - x_1|} \quad (5)$$

where

$$\hat{y}_h = y_1 + \alpha(x - x_1) \quad (6)$$

$$\hat{y}_v = y_2 + \alpha(x - x_2) \quad (7)$$

This predictor is based on horizontal and vertical gradients computed in the reference band. As the various bands are imaging the same physical location it is very likely that an edge in the reference band will occur at the same location as an edge in the current band.

Because the reflectance of different materials can be strikingly different in different spectral bands, the correlation between the bands can exhibit significant variation across the image. In the regions where the correlation between neighboring bands is low it is much better to use an intraband predictor. Switching between interband and intraband predictors can be performed by monitoring the local correlation. As this correlation is available to both encoder and

decoder there is no need for any side information. The threshold for switching in this work was set at 0.5.

As in the case of the original CALIC algorithm, the three dimensional extension also includes a bias cancellation step. The bias in the predictions occurring in various quantized contexts is computed on an ongoing basis and subtracted from the prediction. The number and size of the contexts in the 3D version is different from the 2D version. However, the contexts are used in the same way. The prediction residual is encoded using context based arithmetic coders.

The interband CALIC is compared with the intraband CALIC and LOCO-I on various multispectral images and it outperforms both of them by an appreciable margin. More than a 20% compression gain can be obtained for typical multi-spectral images.

Magli et al. [16] extend this work by noting that the 3D CALIC proposed in [15] does not make full use of the spectral resolution available in hyperspectral images. They show that the performance of 3-D CALIC can be significantly improved by taking account of the fact that in hyperspectral images spectral prediction can be enhanced by taking into account multiple bands. They do so by creating a reference band in which the $(i, j)^h$ pixel of the reference band used to predict the pixels in the k^h spectral band is given by

$$y_{i,j}^k = \gamma_1 + \gamma_2 x_{i,j}^{k-1} + \gamma_3 x_{i,j}^{k-2}$$

where $x_{i,j}^{k-1}$ and $x_{i,j}^{k-2}$ are the pixels in the same spatial location in the previous two bands. The coefficients γ_i are obtained by minimizing $\|x_{i,j}^k - (\gamma_1 + \gamma_2 x_{i,j}^{k-1} + \gamma_3 x_{i,j}^{k-2})\|_2^2$. The reference band pixel $y_{i,j}^k$ is then refined by adding to it the average difference between the previous band and the previous reference band. They further optimize the parameters of 3D CALIC for hyperspectral images and show significant gains over the original 3D CALIC proposal.

5 Optimum Linear Prediction

Given a sample x_k and a set of “past” samples $\{x_{k-1}, x_{k-2}, \dots, x_{k-M}\}$, we can obtain the optimum linear prediction \hat{x}_k for x_k as

$$\hat{x}_k = \sum_{i=1}^M \alpha_i x_{k-i}$$

where α_i are solutions of the Yule-Walker equations

$$\alpha = R^{-1}P$$

where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_M]^T$, R is the autocorrelation matrix with

$$R[i, j] = E[x_{k-i}x_{k-j}] \quad i, j = 1, 2, \dots, M$$

and P , known in the filtering literature as the steering vector is given by

$$P[i] = E[x_k x_{k-i}] \quad i = 1, 2, \dots, M$$

The problem with applying this approach to hyperspectral image compression is that the images are statistically nonstationary. One way to get around this problem is to partition the hyperspectral images into sets which are quasi-stationary. Aiazzi et al [17] use a fuzzy clustering scheme to partition the hyperspectral data into clusters. They then compute optimal predictors for each cluster. The final prediction is obtained as a weighted sum where the weights correspond to membership in the clusters. Mielikainen and Toivanen [18] obtain the clustering by using the LBG algorithm [19] and then obtain the coefficients for a quasi-linear predictor. Rizzo et al. [20] use a membership function to separate the pixels into two groups, a group for which intraband prediction is used and a group for which interband prediction is used.

5.1 JPEG-7 Prediction

In this section we present a very simple lossless compression scheme which takes advantage of the similarity of local structure of spectral bands in a much more ad-hoc manner than in the schemes described previously. This method

was originally proposed for the prediction of color band in lossless video compression [21]. Before we describe the hyperspectral compression scheme, we review the JPEG lossless compression schemes.

The previous JPEG lossless still compression standard [13] provided eight different predictive schemes from which users can select. The first scheme makes no prediction. The next seven are listed below. Three of the seven are one-dimensional predictors, and others are 2-D prediction schemes. Here $I(i, j)$ is the (i, j) th pixel of the original image and $\hat{I}(i, j)$ is predicted value for the (i, j) th pixel.

$$1 \quad \hat{I}(i, j) = I(i - 1, j) \quad (8)$$

$$2 \quad \hat{I}(i, j) = I(i, j - 1) \quad (9)$$

$$3 \quad \hat{I}(i, j) = I(i - 1, j - 1) \quad (10)$$

$$4 \quad \hat{I}(i, j) = I(i, j - 1) + I(i - 1, j) - I(i - 1, j - 1) \quad (11)$$

$$5 \quad \hat{I}(i, j) = I(i, j - 1) + (I(i - 1, j) - I(i - 1, j - 1))/2 \quad (12)$$

$$6 \quad \hat{I}(i, j) = I(i - 1, j) + (I(i, j - 1) - I(i - 1, j - 1))/2 \quad (13)$$

$$7 \quad \hat{I}(i, j) = (I(i, j - 1) + I(i - 1, j))/2 \quad (14)$$

Different images can have different structures that can be best exploited by one of the eight modes of prediction. If the compression is performed in offline mode, all eight predictors can be tried and the one that gives the best compression is used. The mode information is stored as header information in the compressed file for the decoding purpose.

The JPEG-7 predictors can be extended to 3-D hyperspectral image compression. As we have discussed before, strong correlations exist between adjacent bands. However, strong correlation between two bands does not mean that the pixel values between two bands are similar. In addition, the computation cost is very high if we calculate the correlation coefficient for each windowed local region. Due to these two factors, we might look at the problem from another perspective and pursue a relatively simple way to deal with the similarity. If strong correlation exists, the relationship between the pixel and its surrounding pixels are similar. Therefore, rather than use the pixels in the reference band to estimate the pixel in the current band, we take the pixels in the reference band to select an appropriate predictor among all these seven JPEG predictors, and the predictor is used for the prediction of the corresponding pixel in the current band. The selection criteria is the predictor that minimizes

the prediction error in the reference band. This leads to a simple adaptive predictive coding schemes. Surprisingly, it this prediction schemes works very well when the dynamic range of the pixels is relatively small. For example, the average residual band entropy for the cuprite89 image using this very simple scheme is around 4.6 bits/pixel, which compares favorably with some of the more complex schemes.

5.2 Correlation-based Conditional Average Prediction (CCAP)

The optimal estimate (in the sense of MMSE) of a random variable X given a set of observation Y_i is known to be the conditional expectation of X given Y_i

$$E[x|Y_i] = \sum xP[X = x|Y_1 = y_1, y_2, \dots, y_N] \quad (15)$$

Therefore, the optimal predictor of the value of a pixel is

$$E[X_{i,j} | \{X_{i-l,j-m}\}_{(l,m)=(1,1)}^{i,j}]$$

the conditional expected value. In practice we can assume that the pixel $X_{i,j}$ is conditionally independent of pixels that are some distance from it and hence the conditional variables can be limited to pixels in the causal neighborhood, or causal context, of $X_{i,j}$. For jointly Gaussian processes the conditional expectation can be expressed as a linear combination of the observation. However, for the non-Gaussian case the computation of the conditional expectation requires the availability of the conditional probability density function. In the case of image pixels it would be difficult to assume that the process under consideration is Gaussian. Slyz and Neuhoff [22] reduce the size of the problem by replacing the conditioning variables with their vector quantized representation. However, vector quantization of the neighborhoods leads to an apriori ad-hoc partitioning of the conditioning space which can result in significant loss of information.

In the area of text compression Cleary and Witten [23] developed a blending approach for estimating the conditional probabilities in their development of the prediction-with-partial-match (PPM) algorithm. This approach implicitly relies on the fact that the textual information contains many exact repeats. As this situation is not duplicated in natural images the algorithm used in PPM cannot be applied directly to the problem of generating predictions. Fortunately, while we do not have exact repeats as in textual data our objectives

are also not the same. We are interested in an expected value which can be estimated using a sample mean.

Given a pixel $x_{i,j}$, let $C_{i,j}$ be the set of pixels in the causal context of $x_{i,j}$. For convenience, we put an ordering on these pixels so we refer to them as $x_1^{i,j}, x_2^{i,j}, \dots, x_k^{i,j}$. Given a particular set of value $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$, define

$$C_k(\bar{\alpha}) = \{x_{l,m} : x_1^{l,m} = \alpha_1, x_2^{l,m} = \alpha_2, \dots, x_k^{l,m} = \alpha_k\} \quad (16)$$

Then we can estimate $E[X_{i,j} | x_{l,m} : x_1^{l,m} = \alpha_1, x_2^{l,m} = \alpha_2, \dots, x_k^{l,m} = \alpha_k]$ by the sample mean

$$\hat{\mu}_{X|\alpha} = \frac{1}{\|C_k(\bar{\alpha})\|} \sum_{x \in C_k(\bar{\alpha})} x \quad (17)$$

where $\|\cdot\|$ denote the cardinality.

Before we use this method in practice, we need to address several issues. We need to decide on the size and composition of the causal context. We need to decide on how large $\|C_k(\bar{\alpha})\|$ should be to make $\hat{\mu}_{X|\alpha}$ a good estimate and we need to decide what to do when $\|C_k(\bar{\alpha})\|$ is not a large enough for $\hat{\mu}_{X|\alpha}$ to be a valid estimate.

To address these issues, first we define the MED predictor used in JPEG-LS to be the default predictor for the first band. Given the nomenclature of Fig.2 the prediction using the MED algorithm is as follows:

$$\hat{X} = \begin{cases} \min(x_2, x_1), & \text{if } x_3 \geq \max(x_2, x_1); \\ \max(x_2, x_1), & \text{if } x_3 \leq \min(x_2, x_1); \\ x_1 + x_2 - x_3. & \text{otherwise.} \end{cases} \quad (18)$$

Second, we use the JPEG 7 prediction as the default prediction if the number of valid estimates is not sufficient, i.e., $\|C_k(\bar{\alpha})\|$ is small or even zero. As the set $\|C_k(\bar{\alpha})\|$ is built using the past history of the image, the information about the size of the set is available to both the encoder and decoder. The decoder can perform exactly the same procedures as the encoder was doing. The size of $\|C_k(\bar{\alpha})\|$ required for a valid estimate was determined empirically by conducting experiments with a test set of images.

The question about the size and composition of the context, and the definition of context match is somewhat more difficult. As might be expected using a large context generally leads to a better prediction. However, when a larger

context is used it is less likely to have been encountered in the history of the image. It is somewhat reasonable to use a larger context first and if there are not sufficient matched to this context in the history we shift to a smaller context, similar to the way the PPM performs context match. However, it is found that contexts of sizes greater than 4 or contexts less than 4 give only marginal gains over the smaller contexts and it also shows that there is no much improvement if we use dynamical context composition and size, rather than fixed contest composition and size. Therefore, context size is fixed to 4.

5.3 Algorithm Parameters

The definition for a context match is a critical part of the algorithm. There are two methods available. Given a sequence of pixels $Y = y_1^{i,j}, y_2^{i,j}, \dots, y_k^{i,j}$, which take on the set of value $\beta = \beta_1, \beta_2, \dots, \beta_k$, we declared the pixel $y_{l,m}$ to be a member of $C_k(\bar{\alpha})$ if:

- $|\alpha_i - \beta_i| \leq T_1, i = 1, 2, \dots, k$, or
- $\wp < T_2$, where \wp is the correlation coefficient between Y and β .

Note that neither of these matches partition the space of conditioning contexts into disjoint sets as would be the case if we used a vector quantizer to reduce the number of contest. Ideally, the context match based on the first definition should give a good match performance, given the data is stationary. And experiments show that it did work greater than GAP predictor of CALIC and most other prediction methods for stand-alone 2-D still image compression, in terms of entropy. However, when applied to hyperspectral images, context matches based on the first definition did not provide any improvements in compression. This can be explained by noticing that although the correlation can be very strong between two bands, the difference between corresponding pixels values are not small. In other words, if we use the linear mode described in Section 4.2, either the gain factor α or the offset β is typically very large for hyperspectral images. As a consequence, the context match according to the first criteria does not work as well as we had expected from experiments conducted using single band natural images. Therefore, we use the second definition to perform the context match. Next a context search area needs to be specified. It is usually beneficial if we search a particular context within a windowed search

area from its adjacent bands, as shown in Figure 3. Alternatively, we can use the pixel in the same spatial location in the reference bands to perform the context match, rather than use a search window which contains more pixels. It would be reasonable to assume that this should work as each individual pixel location represents the same material on the earth and they should have strong correlation. However, our experiments show that the most correlated pixels in reference bands are not always in the same location as the pixel to be predicted. As a result, it is reasonable to use a search window instead of a single pixel.

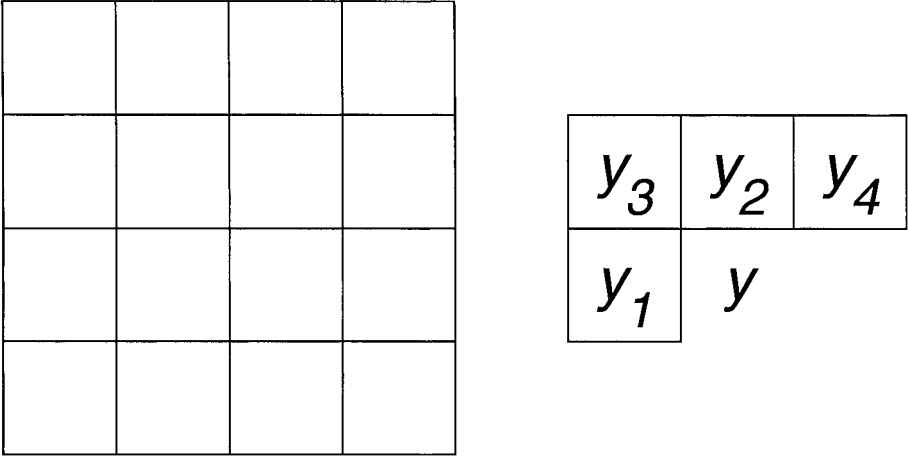


Figure 3: Context Search Window

If such a match is found, what we need to do now is to obtain the prediction. It looks very straightforward if we use the linear model to get the prediction. However, experiment shows that in terms of entropy of residual images, this method is inferior to the simple predictor by just taking a scaled neighboring pixel as follows:

$$\bar{y} = \frac{x}{x_1} * y_1 \quad (19)$$

Therefore, for each pixel to be predicted, the algorithm goes back to the previous bands and calculates the correlation coefficient for each pixel within the window. If the correlation coefficient is greater than a threshold, we then use Equation 19 to obtain a prediction value.

Considering the nonstationary property of hyperspectral images, only one prediction value is not sufficient. As such, we extend the search area and find

more than one such context matches and their predictions, and take the average of those prediction values as the prediction value, as described by Equation 17.

To find the value of $C_k(\tilde{\alpha})$ for which the estimate $\hat{\mu}_{X|\alpha}$ is valid we ran a series of experiments by varying the correlation coefficient threshold \bar{T} for a valid context match, the number of bands and the window size to be searched. Experiments show that the best performance in terms of average band entropy is obtained if we take 5 bands, constrain the search window as 3×3 pixels, and set the value of T_2 to 0.95.



Figure 4: pixel labelling for correlation-based context match

5.4 Results

The average band entropy of various residual AVIRIS images by using the JPEG7 predictors and conditional average images is shown in Table 2. As we can see, the entropy difference between both methods for the cuprite89 image is marginal. However, for other images, we can see that the performance of CCAP is much better than that of JPEG7. The reason for this is that the intensity variation between adjacent bands in the cuprite89 image is very small, and the dynamic range of pixels is within a much narrower range, as compared with other images. In all these images, most of the computed correlation coefficients between neighboring bands are close to 1, that is, the bands are strongly correlated. However, cuprite89 has a small gain factor and offset between adjacent bands and JPEG7 works much better in this situation. The reflectance property is solely determined by the materials in the earth that were sensed situations in which the dynamic ranges are very large are common for hyper-spectral images. the CCAP is able to catch such spectral structures much better than JPEG7 predictors.

The residual images can be encoded by any entropy coder, such as arith-

Image	Cuprite89	JASPER92	LA92	MF92	SUN 92
JPEG7	4.64	6.8	7.7	7.5	7.4
CCAP	4.55	6.26	5.69	6.55	4.64

Table 2: Average band entropy obtained from JPEG7 and CCAP

metic codes. Similar to CALIC, the residual entropy can be further reduced by bias cancellation as well as band reordering. The final compression ratio can be improved by using a context-based arithmetic coder.

5.5 Summary

Here is the summary of how the algorithm performs prediction.

- JPEG-LS prediction is used for the first band.
- In the following bands, search for the context match in the previous bands and calculate the correlation coefficient for each pixel in the search window.
- Determine whether JPEG7 or CCAP is used according to the correlation coefficient and its threshold, and the number of valid predictions.
- If JPEG7 is used, 7 predictors based on previous band are used for the prediction of the current pixel.
- Otherwise, the CCAP is used for prediction.
- Entropy of the residual images can be further reduced by exploiting bias cancellation.
- The final residual images can then be encoded by a context-based entropy coder, such as the arithmetic coder.

6 Conclusion

We have presented a number of different predictive coding schemes for the compression of hyperspectral images. While the schemes differ in the details of their implementation their outline is essentially the same. Each algorithm tries to approximate the ideal case of stationary data for which optimal predictors can be computed. The approximations can be viewed as attempting to partition the data space into sets within which the stationarity assumption can be applied with some level of plausibility. The extent to which these algorithms function or fail depends upon the validity of their assumption. There is clearly much more work to be done before we can claim that the problem of predictive lossless compression has been solved.

References

- [1] K. Sayood. Data Compression in Remote Sensing Applications. *IEEE Geoscience and Remote Sensing Newsletter*, (84):7–15, September 1992.
- [2] M.J. Ryan and J.F. Arnold. The Lossless Compression of AVIRIS Images by Vector Quantization. *IEEE Transactions on Geoscience and Remote Sensing*, Vol 35:546–550, March 1997.
- [3] G. Motta, F. Rizzo, and J.A. Storer. Compression of Hyperspectral Imagery. In *Proceedings of the Data Compression Conference, DCC '03*. IEEE, 2003.
- [4] X. Wu, N.D. Memon, and K. Sayood. A Context Based Adaptive Lossless/Nearly-Lossless Coding Scheme for Continuous Tone Images. ISO Working Document ISO/IEC SC29/WG1/N256, 1995.
- [5] X. Wu and N.D. Memon. CALIC - A Context Based Adaptive Lossless Image Coding Scheme. *IEEE Transactions on Communications*, May 1996.
- [6] M. Weinberger, G. Seroussi, and G. Sapiro. The LOCO-I Lossless Compression Algorithm: Principles and Standardization into JPEG-LS. Technical Report HPL-98-193, Hewlett-Packard Laboratory, November 1998.

- [7] J.M. Shapiro. Embedded Image Coding Using Zerotrees of Wavelet Coefficients. *IEEE Transactions on Signal Processing*, SP-41:3445–3462, December 1993.
- [8] A. Said and W.A. Pearlman. A New Fast and Efficient Coder Based on Set Partitioning in Hierarchical Trees. *IEEE Transactions on Circuits and Systems for Video Technologies*, pages 243–250, June 1996.
- [9] A.C. Miguel, A.R. Askew, A. Chang, S. Hauck, R.E. Ladner, and E.A. Riskin. Reduced Complexity Wavelet-Based Predictive Coding of Hyperspectral Images for FPGA Implementation. In *Proceedings of the Data Compression Conference, DCC '04*. IEEE, 2004.
- [10] S.R. Tate. Band Ordering in Lossless Compression of Multispectral Images. *IEEE Transactions on Computers*, pages 477–483, April 1997.
- [11] R.E. Roger and M.C. Cavenor. Lossless Compression of AVIRIS Images. *IEEE Transactions on Image Processing*, pages 713–719, May 1996.
- [12] R. F. Rice, P. S. Yeh, and W. Miller. Algorithms for a very high speed universal noiseless coding module. Technical Report 91-1, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California., February 1991.
- [13] G. K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34:31–44, April 1991.
- [14] N.D. Memon, K. Sayood, and S.S.. Magliveras. Lossless Compression of Multispectral Image Data. *IEEE Transactions on Geoscience and Remote Sensing*, 32:282–289, March 1994.
- [15] X. Wu and N.D. Memon. Context Based Lossless Intraband Adaptive Compression - Extending CALIC. *IEEE Transactions on Image Processing*, 9:994–1001, June 2000.
- [16] E. Magli, G. Olmo, and E. Quacchio. Optimized Onboard Lossless and Near-Lossless Compression of Hyperspectral Data Using CALIC. *IEEE Geoscience and Remote Sensing Letters*, 1:21–25, January 2004.
- [17] B. Aiazzi, P. Alba, L. alparone, and S. Baronti. Lossless Compression of Multi/Hyper-spectral Imagery Based on 3-D Fuzzy Prediction.

- IEEE Transactions on Geoscience and Remote Sensing*, 37:2287–2294, September 1999.
- [18] J. Mielikainen, P. Toivanen, and A. Kaarna. Linear Prediction in Lossless Compression of Hyperspectral Images. *Optical Engineering*, 42:1013–1017, April 2003.
- [19] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantization design. *IEEE Transactions on Communications*, COM-28:84–95, Jan. 1980.
- [20] F. Rizzo, B. Carpentieri, G. Motta, and J.A. Storer. High Performance Compression of Hyperspectral Imagery with Reduced Search Complexity in the Compressed Domain. In *Proceedings of the Data Compression Conference, DCC '04*. IEEE, 2004.
- [21] N.D. Memon and K. Sayood. Lossless Compression of Video Sequences. *IEEE Transactions on Communications*, Vol 44:1340–1345, October 1996.
- [22] M.J. Slyz and D.L. Neuhoff. A Nonlinear VQ-based Predictive Lossless Image Coder. In *Proceedings of the Data Compression Conference, DCC '94*. IEEE, 1994.
- [23] J. G. Cleary and I. H. Witten. Data compression using adaptive coding and partial string matching. *IEEE Transactions on Communications*, 32(4):396–402, 1984.

Hyperspectral Data Compression

Motta, G.; Rizzo, F.; Storer, J.A. (Eds.)

2006, XII, 418 p., Hardcover

ISBN: 978-0-387-28579-5