Lossless Hyperspectral Image Compression Using Context-based Conditional Averages[1]

Hongqiang Wang[1], S. Derin Babacan [2], and Khalid Sayood[1]
[1]Department of Electrical Engineering
University of Nebraska-Lincoln
{hqw, ksayood}@eecomm.unl.edu
[2]Department of Electrical Engineering
Northwestern University
derin.babacan@gmail.com

Abstract

In this paper, we propose a compression algorithm focused on the peculiarities of hyperspectral images. The spectral redundancy in hyperspectral images is exploited by using a context matching method driven by the correlation between adjacent bands of hyperspectral spectral images. The method compares favorably with recent proposed lossless compression algorithms in terms of compression, with significantly lower complexity.

1 Introduction

Concerned with the effects of global climate change NASA initiated the *Mission to Planet Earth* enterprise in 1991. Many of the efforts of NASA and other national space agencies are coming to fruition. Among these efforts has been the development of remote sensing instruments with high levels of spatial and spectral resolutions. The products generated by these sensors promise to revolutionize our understanding of climatology, meteorology, and land management. As the amount of data generated by these sensors is enormous and as the number of sensors continues to grow it is clear that the role of data compression will be crucial in this development. At different points in the path from the sensor to the end-user the compression needs will be different and both lossy and lossless compression approaches will be needed. Our focus in this paper is on lossless compression.

Lossy compression schemes have been an enabling technology for the multimedia revolution of the past decades. These techniques have been successful in part because the human visual system is insensitive to certain kinds and levels of distortion. Thus selective loss of information can be used to significantly enhance the compression performance. However, hyperspectral images are not solely intended to be viewed by human beings. They are usually processed for various applications such as feature extraction, classification, target detection, and object identification. Any distortion inroduced during the compression process can get amplified by the processing and have a significant deleterious effect on the application. As such, lossy compression is not always acceptable. At a minimum lossless compression is needed if the data is to be stored for some period of time on the acquisition platform and for transmission from the acquisition platform to the ground station. Lossless compression is also needed for data archiving.

Compression schemes, and in particular lossless compression schemes rely on statistical structure in the data. In hyperspectral images there are, loosely speaking, two kinds

Proceedings of the 2005 Data Compression Conference (DCC'05)
1068-0314/05 $ 20.00 **IEEE**

of correlations that can be exploited. One is spatial correlation between adjacent pixels in a spectral band, and the other is correlation between pixels in adjacent bands. The spatial correlation can be easily exploited using compression method developed for standard gray-scale or color images. However, how to efficiently make use of the redundancy between adjacent bands for high resolution hyperspectral images is still an open question, and will largely determine how much further compression gain can be obtained.

Most compression techniques proposed for lossless hyperspectral image compression fall into two categories: vector quantization and predictive coding. An early example of the use of vector quantization for lossless compression of multi-spectral images is the mean-removed vector quantization algorithm described in [7]. A more recent algorithm, with substantially better performance, is the locally optimal partitioned vector quantization (LPVQ) algorithm [13] In this algorithm hyperspectral images are compressed by applying partitioned VQ to the spectral signatures. The predictive coding techniques proposed for hyperspectral image compression generate a prediction of the pixel being encoded by using spatial predictor, a spectral predictor, or a hybrid predictor, followed by an entropy coder for the residual images. Examples of these techniques can be found in CCSDS recommendations for lossless data compression [1, 2], and in [5], [4], [15], [10], [9], and [14]. The order in which bands are used for prediction turns out to be important and a techniques to obtain an optimum band orders for prediction is developed in [3].

In [5], an adaptive differential pulse coded modulation (ADPCM) is presented in which spatial, spectral, as well as hybrid predictors, are used to capture the local structure, and an optimum linear predictor in the minimum mean square error sense (MMSE) is obtained by using a standard linear combination. If the data was stationary, prediction based on a global model would be optimum. Unfortunately, a stationarity assumption is generally not valid for image data. As a consequence, no single model can capture all local structure in images. Therefore, most modern lossless image compression algorithms employ multiple predictors to capture the local relationships in different parts of an image, or explore the underlying data structure in a progressive manner. In [10] a 3-dimensional context based adaptive lossless image coder (Interband CALIC) is developed. It is an extended version of the well known 2-dimensional CALIC and works between two modes: intra-band and inter-band. The algorithm switches between the two based on the strength of the correlation between two adjacent bands.

In the work we present a compression scheme which also uses multiple predictors. However the predictors are generated using local statistics and are used based on a switching algorithm which relies on the correlation between bands.

This paper is organized as follows. The algorithm which we call correlation-based conditional average prediction is presented in Section 2. The results are given in Section 3, while in Section 4 we summarize the compression scheme.

2 Proposed Algorithm

The algorithm works in two modes. One is called JPEG-7 prediction and uses the fact that different bands of a hyperspectral image are imaging the same physical location albeit from different spectral viewpoints. The second mode uses a correlation-based conditional

average prediction. We discuss these two modes in Section 2.1 and Section 2.2 respectively.

2.1 JPEG-7 Prediction

Here we present a very simple lossless compression scheme which takes advantage of the similarity of local structure of spectral bands. The original JPEG standard had a little known section on lossless compression. This lossless compression standard provided eight different predictive schemes from which users could select. The first scheme makes no prediction. The next seven are listed below. Three of the seven are 1-D predictors, and others are 2-D prediction schemes. Here $x(i,j)$ is the (i,j)th pixel of the original image and $\hat{x}(i,j)$ is predicted value for the (i,j)th pixel.

$$1 \quad \hat{x}(i,j) = x(i-1,j) \tag{1}$$

$$2 \quad \hat{x}(i,j) = x(i,j-1) \tag{2}$$

$$3 \quad \hat{x}(i,j) = x(i-1,j-1) \tag{3}$$

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

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

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

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

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 can be 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. Although strong correlation exists between adjacent bands this does not mean that pixel values in the two bands are similar. This means that we cannot always use the pixels in one band to predict the pixel value in the neighboring band in a linear fashion. However, because the structures imaged in both bands is the same, the relationship between a pixel and its surroundings in one band will likely be similar to the relationship of a co-located pixel with its surroundings in the neighboring band. Therefore, rather than use the pixels in the reference band to estimate the pixel in the current band, we use the pixels in the reference band to select an appropriate predictor among the seven JPEG predictors for use in the prediction of the corresponding pixel in the current band. The seven predictors are used in the reference band to obtain a prediction for the pixel co-located with the pixel being encoded. The predictor that minimizes the prediction error in the reference band is selected for use in the current band. This leads to a simple and efficient predictive coding scheme. Our results show that this prediction schemes works very well for images with low dynamic range. For instance, the average band entropy for cuprite89 can be as low as to 4.6bits/pixel, which corresponds to a compression ratio of 3.5:1.

2.2 Correlation-based Conditional Average Prediction (CCAP)

2.2.1 Conditional Expectation

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

$$E[x|Y_i] = \sum xP[X = x|Y_1 = y_1, y_2, \cdots, y_N] \tag{8}$$

Therefore, the optimal predictor of the value of a pixel is $E[X_{i,j}|\{X_{i-l,j-m}\}^{i,j}_{(l,m)=(1,1)}]$, 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 this is not available. Slyz and Neuhoff [11] 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 [12] 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 which would have no meaning in the context of text compression. We apply the ideas behind PPM in the following manner.

Given a pixel $x_{i,j}$, we can define a set of pixels which are in the causal neighborhood of $x_{i,j}$ as its causal context. The exact composition of this set will vary depending on our method of scanning (and hence the operational meaning of "causal") and our definition of neighborhood. For convenience, we put an ordering on these pixels so we refer to them as $x_1^{i,j}, x_2^{i,j}, \cdots, x_k^{i,j}$. Given a particular set of value $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_k)$, define

$$C_k(\bar{\alpha}) = \{x_{l,m} : x_1^{l,m} = \alpha_1, x_2^{l,m} = \alpha_2, \cdots, x_k^{l,m} = \alpha_k\} \tag{9}$$

That is, the set $C_k(\bar{\alpha})$ consists of all pixels whose causal neighborhoods take on the values $\alpha_1, \alpha_2, \cdots, \alpha_k$. Then we can estimate $E[X_{i,j}|x_1^{i,j} = \alpha_1, x_2^{i,j} = \alpha_2, \cdots, x_k^{i,j} = \alpha_k]$ by the sample mean

$$\hat{\mu}_{X|\alpha} = \frac{1}{\|C_k(\bar{\alpha})\|} \sum_{x \in C_k(\bar{\alpha})} x \tag{10}$$

where $\|\|$ 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.

2.2.2 JPEG-7 Prediction Mode

We use the JPEG-7 prediction as the default prediction if the number of valid estimates is not sufficient, i.e., for the case where $\|C_k(\bar{\alpha})\|$ is small or even zero we use the JPEG 7 prediction. 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.

2.2.3 Algorithm Parameters

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 reasonable to use a larger context first and if there are not sufficient matches to this context in the history we shift to a smaller context, similar to the way the PPM performs context match. We have found that for hyperspectral image compression contexts of sizes greater than 4 give only marginal gains over the smaller contexts. Therefore, for this application the context size is fixed to 4.

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

- $|\alpha_i - \beta_i| \leq T_1$, $i = 1, 2, \cdots, k$;

or

- $\wp < T_2$.

where $\wp$ is the correlation coefficient between β and α, $i = 1, 2, \cdots, k$ and T_1 and T_2 are two empirically determined thresholds.

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 if the data is stationary. And our experiments show that this approach does work better than JPEG-7 prediction and most other prediction methods for general purpose image compression. However, when applied to hyperspectral images, context matches based on the second definition provide for significantly better performance. 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. As a consequence, the context match according to the first criteria is not as good as what we have expected from 2-D context match. As such, 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 Fig. 1. Alternatively, we can use the pixel in the same spatial location in reference bands to perform the context match, rather than use a search window which contains more pixels.

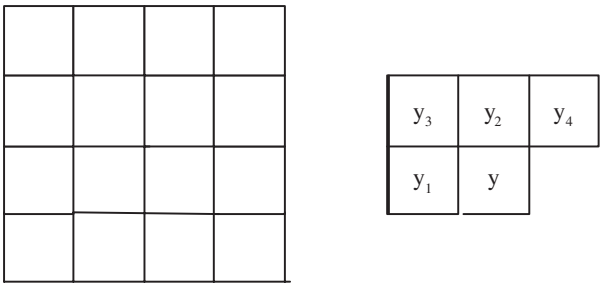

Figure 1: Context Search Window

Figure 2: Pixel Naming for Correlation-based Context Match

This should work as each individual pixel location represents the same material on the earth and they should have strong correlation. Due to nonidealities in the sensors, however, the same pixel location in various bands may not represent the identical location in the earth. Indeed, our experiments show that the most correlated pixels in reference bands are not always in exactly 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.

If such a match is found, what we need to do is to obtain the prediction. Because of the widely varying dynamic ranges in different bands of the hyperspectral images we have found that linear prediction does not work well. Instead, we generate a prediction by using a scaled neighboring pixel as follows:

$$\hat{y} = \begin{cases} \frac{x}{x_1} y_1, & \text{if } x_1 \neq 0 \\ y_1 & \text{otherwise} \end{cases} \tag{11}$$

Therefore, for each pixel to be predicted, the algorithm searches the adjacent 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.11 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 10.

To find the value of $C_k(\bar{\alpha})$ for which the estimate $\hat{\mu}_{X|\alpha}$ is valid we ran a series of experiments by varying the correlation coefficient threshold T_2 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 residual entropy is obtained if we take 5 bands, constrain the search window to 3×3 pixels, and set the value of T_2 to 0.95.

Prediction	band	Image				
		Cuprite	Jasper	Low Alt.	Moffett	Lunar
Previous Best JPEG-7		4.64	6.8	7.7	7.5	7.4
CCAP	5	4.53	6.26	5.69	6.22	4.64
	4	4.55	6.41	5.71	6.51	4.66
	3	4.76	7.70	5.93	6.83	5.28
	2	4.90	7.89	6.23	7.13	5.63

Table 1: Average Band Entropy Obtained From JPEG-7 Prediction and CCAP

	JPEG-LS	JPEG2000	LPVQ	SLSQ-Opt	CCAP
Cuprite	2.09	1.91	3.18	3.24	3.53
Jasper	2.00	1.80	2.88	3.15	2.56
Low Alt.	2.14	1.96	2.94	3.04	2.81
Moffett	1.91	1.78	3.00	3.21	2.57
Lunar	1.99	1.82	3.28	3.14	3.21

Table 2: Comparison in terms of compression ratios with other schemes

3 Results

The average band entropy of various residual images by using JPEG-7 prediction and correlation-based conditional average prediction to AVIRIS images shown in Table 1. As we can see, the entropy difference between both methods for the image of cuprite97 is marginal. However, for other images, we can see that the performance of CCAP is much better than that of JPEG-7 prediction. If we examine the cuprite89 image, we find that the intensity variation between adjacent bands is very small, and the dynamic range of pixels is much narrower compared to other images. For all these images, most correlation coefficients between neighboring bands are close to 1. However, cuprite97 has a small gain factor and offset between adjacent bands and JPEG-7 prediction work much more efficiently in this situation. The reflectance property is solely determined by the materials in the earth that were sensed, and the situation where the dynamic ranges are very large is common for hyperspectral images. The CCAP approach is able to capture such spectral structures much better than JPEG-7 prediction.

The residual images can be encoded by any entropy coder, such as an arithmetic coder. Similar to CALIC, the residual entropy can be further reduced by bias cancellation as well as band reordering process. As compared with general entropy coder, the final compression ratio can be improved by using a context-based arithmetic coder.

Finally, we compare the proposed approach to other schemes in the literature in Table 2. The data for these were obtained from [15]. Clearly, the proposed approach outperforms both JPEG-LS and JPEG2000. This is not surprising as these are general purpose image compression algorithms. We bring up this point just to point out the importance of developing application specific compression techniques. The comparison with techniques developed for hyperspectral compression presents a more mixed picture. While the

proposed technique outperforms both LPVQ and the optimized SLSQ algorithms for the *Cuprite* image, it substantially under performs both of them for *Jasper* and *Moffett* while being competitive for *Lunar*.

4 Summary

Here is the summary of how the algorithm performs prediction.

1. JPEG-LS prediction is used for the first band.
2. In the following bands, search for context matches in the previous bands and calculate the correlation coefficient for each pixel in the search window.
3. Determine whether JPEG-7 prediction or CCAP is to be used based on the correlation coefficient, and the number of valid predictions.
4. If JPEG-7 prediction is used, the best predictor based on previous band is used for the prediction of the current pixel.
5. Otherwise, the CCAP is used for prediction.
6. Entropy of the residual images can be further reduced by exploiting bias cancellation.
7. The final residual images can then be encoded by using a context-based entropy coder.

References

[1] "Lossless data compression, CCSDS recommendation for space data systems standards," Washington, DC, Blue Book 120.0-B-1, 1997.

[2] "Lossless data compression, CCSDS report concerning space data systems standards," Washington, DC, Green Book, 120.0-G-1, 1997.

[3] S. Tate, "Band ordering in lossless compression of multispectral images," in Proc. 1994 Data Compression Conf., Mar. 1994, pp. 311-320.

[4] N.D. Memon, K. Sayood, S.S Magliveras, " Lossless compression of multispectral image data," Geoscience and Remote Sensing, IEEE Transactions on, Volume: 32, Issue: 2, pp. 282-289, March 1994

[5] R.E. Roger, M.C. Cavenor, "Lossless compression of AVIRIS images," Image Processing, IEEE Transactions on, Volume: 5, Issue: 5, pp. 713-719, May 1996

[6] M. J. Weinberger, G. Seroussi, and G. Sapiro, "LOCO-I: A low complexity context-based lossless image compression algorithm," in Proc. 1996 Data Compression Conf., Mar. 1996, pp. 140-149.

[7] M.J. Ryan, J.F. Arnold, "The lossless compression of AVIRIS images by vector quantization," Geoscience and Remote Sensing, IEEE Transactions on, Volume: 35, Issue: 3, pp. 546-550, May 1997

[8] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," IEEE Trans. Commun., vol. 45, pp. 437-444, Apr. 1997.

[9] B. Aiazzi, P. Alba, L. Alparone, S. Baronti, "Lossless compression of multi/hyperspectral imagery based on a 3-D fuzzy prediction," Geoscience and Remote Sensing, IEEE Transactions on, Volume: 37, Issue: 5, pp. 2287-2294, Sept. 1999

[10] Xiaolin Wu, N. Memon, " Context-based lossless interband compression-extending CALIC," Image Processing, IEEE Transactions on, Volume: 9, Issue: 6, pp. 994-1001, June 2000

[11] 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.

[12] 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.

[13] Motta G.,Rizzo F., J.A. Storer,"Compression of hyperspectral imagery," Data Compression Conference, 2003. Proceedings. DCC 2003, pp. 333 - 342, 25-27 March 2003

[14] J. Mielikainen, P. Toivanen, "Clustered DPCM for the lossless compression of hyperspectral images," Geoscience and Remote Sensing, IEEE Transactions on, Volume: 41, Issue: 12 pp. 2943-2946, Dec. 2003

[15] F. Rizzo, B. Carpentieri, G. Motta, J.A. Storer, "High performance compression of hyperspectral imagery with reduced search complexity in the compressed domain," Data Compression Conference, 2004. Proceedings. DCC 2004, pp.479-488, 23-25 March 2004