Lossless Compression of Hyperspectral Images Based on Searching Optimal Multibands for Prediction

Chengfu Huo, Rong Zhang, and Tianxiang Peng

Abstract—This letter presents a lossless compression algorithm for hyperspectral images, which is based on the strength of correlations between bands. First, a searching model is constructed using the tree structure. Second, multibands which have strong correlations to each chosen band are found out and are then used to predict the chosen band in a couple-group manner. Lastly, residual images are encoded using entropy coders. Experimental results show that our compression algorithm provides a competitive compression performance compared with most existing compression algorithms.

Index Terms—Couple-group manner, hyperspectral images, searching model, tree structure.

I. INTRODUCTION

HYPERSPECTRAL images, which contain not only spatial information but also spectral information, have been widely used in land management, meteorology, etc. However, the huge volume of data for hyperspectral images causes difficulties in transmission and storage. Undoubtedly, data compression is a necessary and effectual solving approach. In contrast to lossy compression, lossless compression is required in most applications of remote sensing such as spectral mixture analysis, and object identification and classification.

Lots of compression algorithms have been proposed recently, which can be generally classified into vector quantization [1], transform coding [2], and prediction coding [3] algorithms. Vector quantization algorithms usually require offline codebook training, which causes expensive computation. Transform coding algorithms always achieve good coding gains for lossy compression at low bit rates but poor coding gains for lossless compression. Prediction coding algorithms, seeking for the tradeoff between computational complexity and compression capability, are investigated widely by researchers. Laplacian pyramid vector quantization [4] is the optimal method based on vector quantization up to now. It applies partitioned vector quantization independently to each pixel and chooses variable size partitions adaptively. An adaptive least squares optimized prediction called Spectrum-oriented Least SQuares (SLSQ) is presented in [5]. Multiband context-based adaptive lossless image coder [6], which is based on context-based adaptive lossless image coding, is suitable for the compression of data in band-interleaved-by-line format. Correlation-based Conditional Average Prediction (CCAP) [7] explores spectral redundancy by a context-match method driven by the correlation between adjacent bands. Lookup table (LUT) [8] searches the previous band for a pixel of equal value to the pixel colocated to the one to be coded, and the LUTs are used to speed up the searching process.

In fact, most of the existing compression algorithms are based on single band prediction, which cannot fully take advantage of multiband prediction. To improve that, we present a lossless compression algorithm containing a searching model, a multiband predictor, and entropy coders. The algorithm tries to find the number of bands, which can achieve a good compromise between prediction performance and computational complexity. The context of this letter is organized as follows. Section II analyzes the correlations of hyperspectral images, and Section III describes a multiband prediction algorithm. Entropy coding and compression results are presented in Section IV. Finally, we conclude this letter in Section V.

II. CORRELATION ANALYSIS

By analyzing hyperspectral images acquired with the NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)’97 (http://aviris.jpl.nasa.gov/html/aviris.freedata.html), which are decomposed into 224 narrowbands, with approximately 10 nm wide each, ranging from 400 to 2500 nm, we find that there are two types of redundancy: spatial and spectral redundancies. Fig. 1 shows the spectral and spatial correlations of hyperspectral image set of Cuprite scene1, which are calculated by

\[
Cor_{r,c} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (x_{i,j,r} - \bar{x}_r)(x_{i,j,c} - \bar{x}_c)}{\sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} (x_{i,j,r} - \bar{x}_r)^2 \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{i,j,c} - \bar{x}_c)^2}}
\]

\[
Cor_{k} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (x_{i,j,k} - \bar{x}_k)(x_{i+1,j+1,k} - \bar{x}_k)}{\sum_{i=1}^{M} \sum_{j=1}^{N} (x_{i,j,k} - \bar{x}_k)^2}
\]

\(x_{i,j,p}\) and \(\bar{x}_p\) are the pixel value at location \((i,j)\) and mean pixel value of band \(p\), respectively. \(M\) and \(N\) are the width...
and height of each image. $\text{Cor}_{r,c}$ is the spectral correlation coefficient between neighboring bands $r$ and $c$. $\text{Cor}_k$ is the spatial correlation coefficient of band $k$.

Fig. 1 shows that the correlation coefficients between neighboring images are close to one for most bands, and spectral correlation is stronger than spatial correlation.

In terms of information theory, all compression techniques are reacted by reducing redundancy. In order to achieve an impactful prediction performance for hyperspectral images, a prediction algorithm should possess strong capability of decorrelation, using the characteristics shown in Fig. 1.

III. PREDICTION ALGORITHM

More and more research works on lossless compression of hyperspectral images have been focused on the improvement of decorrelation algorithms, since encoding techniques can encode a source optimally close to its first-order entropy. Spatial redundancy can easily be exploited by general compression techniques which are developed for gray images. However, solutions on how to efficiently exploit spectral redundancy are still not well established.

The prediction algorithm proposed in this letter is based on a searching model, shown in Fig. 2, where each node represents a band.

According to the searching model, we can decorrelate the strong correlations between bands and remove spectral redundancy as much as possible. The basic idea is to predict each father node using its son nodes. It has been proven indirectly in [9] that the performance of prediction improves as $n$ increases, while computation becomes much more complex. Our primary work is to find the best value $n$, which achieves compromise between prediction performance and acceptable computational complexity, and to chase down $n$ son-node images for each father node.

A. Searching Scheme

The searching scheme, in the case of $n$ being an arbitrary even number, is specified as the following four steps.

1) Choose a band randomly as the first father node of an $n$-ary tree, and set a threshold denoted by $T$.

2) Calculate the correlation coefficients between the father node and other bands using (1).

3) Select the maximum $n$ correlation coefficients using (3), and compare them with $T$

$$
\text{Cor}_{r_1,c} = \left\{ \begin{array}{l}
\text{Cor}_{r_1,c} | r_1 = \arg \max_{r=1,2,...,s} (\text{Cor}_{r,c}) \\
\text{Cor}_{r_2,c} | r_2 = \arg \max_{r\neq r_1} (\text{Cor}_{r,c}) \\
\vdots \end{array} \right.
$$

$$
\text{Cor}_{r_n,c} = \left\{ \begin{array}{l}
\text{Cor}_{r_n,c} | r_n = \arg \max_{r\neq r_1,...,r_{n-1}} (\text{Cor}_{r,c}) \\
\end{array} \right. 
$$

(3)

In (3), $s$ is the number of bands remained currently, which have not been selected as father nodes. Supposing that $\text{Cor}_{r_1,c}, \text{Cor}_{r_2,c}, ..., \text{Cor}_{r_n,c},$ and $\text{Cor}_{r_n,c}$ are correlation coefficients between band $c$ and bands $A_1, A_2, ..., A_{n-1}$, and $A_n$, respectively, the growth of $n$-ary tree can be presented as follows.

If $\text{Cor}_{r_v,c} \geq T$ for $v = 1,2,...,n$, then
a. treat $A_1, A_2, ..., A_{n-1}$, and $A_n$ as son nodes.

b. set $a_v = \text{Cor}_{r_v,c}$ as corresponding weight.

c. set $A_n$ as a new father node.
d. repeat step 2) and step 3).

Else
a. the tree stops growing.
b. choose a new band, that has never been selected as any node, as the first father node of a new tree.
c. repeat the step 2) and step 3).
Notice that, in the aforementioned searching course, any node except the father node of each layer should be included in the searching range of the following layers. For example, for \( \nu = 1, 2, \ldots, n \) in Fig. 2, \( B_{\nu} \) could be any band except \( A_{\nu} \). The reason of setting \( A_{\nu} \) as the new father node is that, for \( \nu = 1, 2, \ldots, n - 1 \), \( \text{Cor}_{r_{\nu},c} \leq \text{Cor}_{r_{\nu+1},c} \). It means that \( A_{\nu} \) has the minimal correlation with the father node in the range of son bands, and there would be a maximum probability of getting a more ascendant prediction performance in the following layers.

4) Process the remaining nodes that have never been signed as any father node in each tree.

Because of the strong correlations between bands, most of the nodes in a layer are signed as father nodes in the following layers, and we only need to process a few remained nodes.

The adopted scheme to process these remained nodes in each tree is setting them as new nodes of the same tree, as shown in Fig. 2. The suffixes \( i \) and \( k \) may be any number from zero to \( n - 1 \). If the suffix is zero, it means that those nodes in the corresponding layer can entirely be signed as father node in their following layers.

**B. Predicting Scheme**

For each layer of an \( n \)-ary tree, multiband prediction can be actualized as Fig. 3, where \( s_0 \) denotes the correlation coefficient \( \text{Cor}_{S_{i,j,F}} \) calculated by (1) for \( \nu = 1, 2, \ldots, n \).

Let us consider a simple searching model in the case of \( n = 2 \), which is a binary tree structure. Supposing that \( x_F \) denotes current father node, \( x_{S_1} \) and \( x_{S_2} \) denote couple son nodes of \( x_F \), and \( x_{i,j,F} \) denotes an arbitrary pixel at location \( (i,j) \) of \( x_F \), we use optimal linear prediction scheme and get the prediction residual value as follows, respectively:

\[
\hat{x}_{i,j,F} = \alpha_1 \cdot x_{i,j,S_1} + \alpha_2 \cdot x_{i,j,S_2} \quad (4)
\]

\[
\varepsilon_{i,j} = x_{i,j,F} - \hat{x}_{i,j,F} \quad (5)
\]

In (4), \( \alpha_1 \) and \( \alpha_2 \) are the prediction coefficients, which can be calculated as follows, where \( \varepsilon^2 = (1/M \cdot N) \sum_{i=1}^{M} \sum_{j=1}^{N} \varepsilon_{i,j}^2 \):

\[
\begin{align*}
\frac{\partial^2 \varepsilon^2}{\partial \alpha_1} &= 0, \\
\frac{\partial^2 \varepsilon^2}{\partial \alpha_2} &= 0. \\
\end{align*}
\]

Using the global minimum-mean-square-error criterion, prediction coefficients \( \alpha_1 \) and \( \alpha_2 \) can be calculated as follows, where \( \varepsilon' = (\sum_{i=1}^{M} \sum_{j=1}^{N} x_{i,j,S_1} x_{i,j,S_2})^2 - \)

\[
\begin{align*}
\text{Fig. 4.} & \quad \text{Entropy of Cuprite scene1.} \\
\end{align*}
\]

\[
\begin{align*}
\sum_{i=1}^{M} \sum_{j=1}^{N} (x_{i,j,S_1}^2) \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{i,j,S_2}^2) & \text{ and } \text{ref}_{uv} = \\
\sum_{i=1}^{M} \sum_{j=1}^{N} x_{i,j,u} \cdot x_{i,j,v} &
\end{align*}
\]

\[
\begin{align*}
\alpha_1 &= \frac{\text{ref}_{S_1} \cdot \text{ref}_{S_2} - \text{ref}_{S_1} \cdot \text{ref}_{S_2} \cdot \text{ref}_{S_2} \cdot \text{ref}_{S_2}}{\nu} \\
\alpha_2 &= \frac{\text{ref}_{S_1} \cdot \text{ref}_{S_2} \cdot \text{ref}_{S_2} \cdot \text{ref}_{S_2} \cdot \text{ref}_{S_2} \cdot \text{ref}_{S_2}}{\nu}. \\
\end{align*}
\]

When \( n > 2 \), a simplified method, named couple-group manner, is adopted. It is to divide the \( n \) nodes into \( n/2 \) couples and use (7) to calculate a pair of prediction coefficients for each couple of nodes. Supposing that a set of interim prediction images \( \hat{x}_{F}^{(i)} \) for \( \nu = 1, 2, \ldots, n/2 \) is calculated by (4), then the following is used to calculate the final prediction image \( \hat{x}_{F}^{(i)} \), where \( \xi = \sum_{n=1}^{n} s_n \) and \( c_0 \) represents the sum of correlation coefficients corresponding to each couple of nodes which are used to calculate the interim prediction image \( \hat{x}_{F}^{(i)} \):

\[
\hat{x}_{i,j,F} = \sum_{\nu=1}^{n/2} \frac{c_0}{\xi} \hat{x}_{i,j,F}^{(i)}. \\
\]

For those nodes without brother nodes, such as \( A_1 \) and \( C_k \) shown in Fig. 2, and those bands that have not been signed as any node in any tree, we just keep them without applying the predicting scheme.

We have validated our algorithm with \( n = 2, n = 4, \) and \( n = 6 \), corresponding to binary, four-ary, and six-ary trees, respectively. The entropy of each type of tree structure is calculated and shown in Fig. 4, where \( n = 0 \) represents the original bands.

Fig. 4 shows that the proposed multiband prediction algorithm can decorrelate redundancy of hyperspectral images efficiently. Prediction effect improves distinctly when \( n \) changes from two to four but slightly when \( n \) changes from four to six; the same thing happens as \( n \) continues to increase. However, for each time that \( n \) increases by two, a computation of four additions, three multiplications, and one division for each pixel is added in predicting scheme. Thus, \( n = 4 \) is chosen as the optimal value of our predicting scheme. In other words, \( n = 4 \) gives the best tradeoff between prediction performance and computational complexity.

To isolate the effect of our searching scheme, we have predicted each band of Cuprite scene1 using \( n = 4 \), previous bands in nature order. The difference entropies of residual images
Fig. 5. Difference entropy between our method and the nature-order-prediction method.

Fig. 6. Spectral and spatial correlations of residual images of Cuprite scene1.

between our method and the nature-order-prediction method have been shown in Fig. 5. It shows that our method achieves a more efficient prediction performance for most bands. The reduction of average entropy by our scheme is 0.1682 bit/pixel.

IV. ENTROPY CODING AND COMPRESSION RESULTS

Spectral redundancy can be decorrelated effectively for most bands by the predicting scheme. Fig. 6 shows the spectral and spatial correlations of residual images. It is obvious that spatial redundancy has also been decorrelated in predicting scheme but not drastically. If adaptive arithmetic coding is applied to encode residual images directly, the best compression effect cannot be reached.

A. Entropy Coding

In order to wipe off spatial redundancy drastically for some residual images, both the JPEG-LS [10] and adaptive arithmetic coder are applied to encode residual images from predicting scheme, and the one that achieves lower bit rates is chosen as the entropy coder for the current residual image. Fig. 7 shows the differences between the JPEG-LS and adaptive arithmetic coder bit rates of Cuprite scene1. For most of the residual images, JPEG-LS achieves lower bit rates.

B. Compression Results

As the complexity of the proposed algorithm is an important factor that must be considered, in this letter, we adopt two approaches for amelioration in searching scheme. The first one is limiting the searching range in bands adjacent to each father node, which is based on the fact that a band always has stronger correlations with its neighbors than with others. The second one is taking only one pixel for every four pixels to calculate the correlation coefficients. Experimental results show that these simplifications reduce the computational cost without affecting the prediction accuracy.

In our experiment, AVIRIS’97 image sets are compressed using the proposed compression algorithm in the case of $T = 0.90$. The image sets include Cuprite, Jasper, and Lunar scenes; each image consists of $512 \times 614$ pixels; and each spectral component is represented by a 16-bit precision.

Table I shows compression ratios of the proposed algorithm. Compared with several other compression algorithms, our algorithm achieves a competitive compression performance. Clustered differential pulse code modulation, spectral relaxation-labeled prediction, and spectral fuzzy-matching pursuit [11] achieve higher compression ratios than ours but at the cost of expensive computations that are hardly acceptable in some applications. Specifically, these algorithms are executed with predictors of length 20, which is five times of ours, and the computation of prediction coefficients is also more complex because of clustering and partitioning.

Our proposed method can be divided into searching, predicting, and coding stages. Compared with the algorithms listed in Table I, it is not much more complex in the searching and coding stages but in the predicting stage. For the predicting stage, our method needs about eight additions, six multiplications, and two divisions for each pixel, while SLSQ requires four multiplications and six additions for each pixel, and CCAP
requires 11 additions, two multiplications, and two divisions for each pixel.

As analyzed earlier, we draw the conclusion that the proposed algorithm achieves a competitive compression performance despite with a little more computational complexity.

Table II shows some compression results of algorithms based on LUT [8]. It seems that these algorithms have much better performance than those listed in Table I; however, these compression results are achieved with hyperspectral image sets of Cuprite, Jasper, and Lunar scenes of size $2206 \times 614$, $2586 \times 614$, and $1431 \times 614$, respectively. LUT and locally averaged interband scaling (LAIS)-LUT [12] are not proper for images of size $512 \times 614$ since the amounts of memory required by LUTs are about 10% and 20% of the memory used by images, respectively. The developed algorithms LAIS-quantized LUT (QLUT)-heuristic and LAIS-QLUT-optimal [13] reduce the size of the LUTs by using a uniform quantization and have the classic compression efficiency. Inspired by LUT, we develop a novel idea that is combining our searching scheme with LAIS-QLUT. Specifically, instead of searching the previous band for aim pixels, we first search for aim pixels in each son band. There will be four predictors, in the case of four-ary tree, for each pixel in the corresponding father band. The average of the four predictors can be used as a final predictor. We sense that this idea, which will be our next research subject, would have a satisfied compression performance.

V. CONCLUSION

Based on the characteristics of hyperspectral images, this letter proposes a compromise compression algorithm by constructing a searching model of four-ary tree structure. It simplifies the prediction scheme by decomposing multiband prediction into several couple-group manner predictions. Experimental results reveal that the proposed algorithm yields a competitive performance compared with other existing compression algorithms.

REFERENCES


