2011 International Conference on Computer Science and Information Technology (ICCSIT 2011)
IPCSIT vol. 51 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V51.125

An Improved Prediction Tree Compression Applied on Hyperspectral Image

Wang Lang[1+], Fan Shengli[2]

[1,2] Information branch Ningbo Institute of Technology of Zhejiang University, Ningbo, China

Abstract. This paper proposes a new recursive prediction tree method for hyperspectral image compression.The compression is achieved by using a novel four side neighborhood prediction tree for spatial and bidirectional interband prediction for spectral decorrelation followed by SPIHT algorithm for coding the resulting decorrelated residual images. The key feature of the new method is that it presents a recursive algorithm to reduce the interband correlation of hyperspectral image. This algorithm overcomes the drawbacks of one-sided prediction and coding noise accumulation in DPCM, and also overcomes the disadvantages of computation complexity in the CALIC and VQ-based methods. In addition, with this improved prediction tree algorithm, both spectral and spatial correlations inherent in hyperspectral images are exploited. The high compression performances are maintained. In order to evaluate the effectiveness of the proposed method, the computer simulations are conducted on NASA AVIRIS data. The experimental results indicate that the new method provides significant improvement over up to date method for hyperspectral image compression in both objective and subjective measures. While the average compression ratio (CR) is increase compared with SLSQ, 3D-CALIC and SPIHT. The effect of compression on classification is negligible compared with the results of the original image classification.

Keywords: hyperspectral image; prediction tree;

1. Introduction

With high spectral resolution of hyperspectral image, more detailed and subtle spectral signatures are provided and can be used to identify and quantify a large range of surface materials which cannot be identified by broadband or low spectral resolution imaging systems. A typical hyperspectral detector, such as the NASA JPL Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), has an operational spectral range spanning visible and near infrared regions (from 0.4 to 2.4m). Spectral information is quantized into 224 contiguous bands, of approximately 10nm each, with a spatial resolution of 20m at operational altitude. The amount of data generated by hyperspectral imaging spectrometers is enormous, and compression is of paramount importance to reduce storage costs and transmission time. In some applications the original quality of a hyperspectal image cannot be compromised, thus lossless compression is desired. The recent approaches to lossless compression of high spectral resolution images include transform-based, prediction based and vector quantization methods [1].

The lossless compression of multispectral imagery has performed successfully using prediction-based methods with low complexity. Studies of their suitability for hyperspectral images have also been performed. The lossless compression of multispectral image data based on prediction tree was given by Memon et al. [2]. The 3D CALIC (context-based adaptive lossless image codec) [2] was also proposed. Zhang and Yan [3] changed the definition of the four neighborhood model of prediction tree and gave side neighborhood minimum absolute weight prediction tree. Later, Wu and He [4] made an improvement on the method of prediction tree by setting up adaptive predictor aiming at multispectral images' statistic spectral redundancy. The prediction estimate uses two pixels in the previous bands in the same spatial position as the current pixel. By comparison, Rizzo et al. [5] gave an adaptive least-squares optimized prediction technique called spectrum-oriented least squares (SLSQ). The prediction coefficients can be computed using an offline procedure on training data. Recently there were some new compression methods, for example, Jarno [6] presented a method

+ Wang Lang
wanglang@nit.zju.edu.cn

using lookup tables (LUT) [7] for compression of hyperspectral image data in the band-interleaved-byline format with good performance.

To achieve a low algorithmic complexity and high compression ratios, we incorporate new multi-band prediction and prediction tree method into an error-compensated algorithm. Then an adaptive method is given to solve for the coefficients of predictors. After de-correlation of inter-and intra-band redundancy, the wavelet-based SPIHT [8] is used to encode error and residual image. Finally, experimental results on the data from AVIRIS JPL show that this algorithm outperforms 3D CALIC, SPIHT and SLSQ algorithms. The paper is structured as follows. Section Ⅱ presents design method. The experimental results are presented in section Ⅲ. Section Ⅳ gives the conclusion of the paper.

2. Design method

2.1. Definition of Prediction Tree

In consideration of the complexity and efficiency of new method, we define four side neighborhood prediction tree with minimum absolute weight and a backward adaptive technique[8] to transmit prediction tree's structure. The definition of four side neighborhood prediction tree is shown as Fig.1.

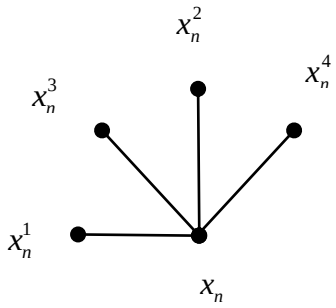

Fig. 1.The definition of four side neighborhood prediction tree

We assume that the size of image to be predicted is M×N with all band number is L. The predicted value of pixel $x_{n,i,j}$ (ith row, jth column of nth band) is $x_{n,i,j}$ with its abbreviation x_n .

Where $\hat{x}_n = x_n^{\bar{k}}$

k is defined by (n-1)th band with prediction tree method, and it must satisfy the equation

$$\left| x_{n-1} - x_{n-1}^{\bar{k}} \right| = \min_{k=1,2,3,4} \left| x_{n-1} - x_{n-1}^{k} \right| \tag{1}$$

2.2. Optimal Linear Predictor of Band

In consider of high spectral resolution of hyperspectral image, there is more high correlation between bands. One side prediction is always used [4], but there is still some redundancy ignored and coding noise accumulation in DPCM. So, to minimize error created between the predicted and actual value, we design a bidirectional m-order DPCM-like predictor from two dimension sources. So we present a method of using previous m bands to predict the current one, which is different from the former algorithm.

Let $a_n = \left(a_n^{(1)}, a_n^{(2)}, \cdots, a_n^{(m)} \right)$ be coefficients of the predictor for the current band predicted by m reference bands.

The predicted value is

$$\begin{aligned} \hat{x}_n &= a_n^{(-m)} x_{n-m} + \ldots + a_n^{(-1)} x_{n-1} + a_n^{(1)} x_{n+1} \cdots + a_n^{(m)} x_{n-m} \\ &= \sum_{\substack{t=-m \\ t \neq 0}}^{m} a_n^t x_{n+t} \end{aligned} \tag{2}$$

2.3. A Novel Hybrid Error Compensated Algorithm

The error created by prediction tree is $\varepsilon = x_{\mathrm{n}} - x_{\mathrm{n}}^{k}$

Here, we use the approximations, and from equation (2), we can get

$$\begin{aligned} & x_{\mathrm{n}} - x_{\mathrm{n}}^{k} \approx \hat{x}_n - \hat{x}_n^{k} \\ & = \left(a_n^{(-m)} x_{n-m} + \cdots + a_n^{(m)} x_{n+m}\right) - \left(a_n^{(-m)} x_{\mathrm{n\text{-}m}}^{k} + \cdots + a_n^{(m)} x_{\mathrm{n+m}}^{k}\right) \\ & = a_n^{(-m)}\left(x_{n-m} - x_{\mathrm{n\text{-}m}}^{k}\right) + \cdots + a_n^{(m)}\left(x_{n+m} - x_{\mathrm{n+m}}^{k}\right) \end{aligned}$$

Where k= {1, 2, 3, 4}.

To minimize predictive error， we assume $k = \bar{k}$, which $\bar{k}$ should satisfy equation (1).

Namely the predicted pixel of nth band is

$$\begin{aligned} \hat{x}_n &= a_n^{(-m)}\left(x_{n-m} - x_{\mathrm{n\text{-}m}}^{k}\right) + \cdots + a_n^{(m)}\left(x_{n+m} - x_{\mathrm{n+m}}^{k}\right) + x_n^{\bar{k}} \\ &= \sum_{\substack{t=-m \\ t\neq 0}}^{m} a_n^{t} x_{n+t} + x_n^{\bar{k}} \end{aligned} \tag{3}$$

It should be noted that in equation (3), we use the same coefficients a_n of band predictor with x_{n} to predict x_{n}^{k}. And the precondition mentioned above is that we assume x_{n} 's neighboring region is stationary. The assumption should be tenable [9] [10].

Equation (3) means that error created by method of prediction tree can be compensated by $a_n^{t}\left(x_{n-t} - x_{n-t}^{\bar{k}}\right)$.The idea for prediction makes use of correlation of statistics, framework and space.

2.4. Calculation for coefficients of the predictor

From equation (2), predictive square error is

$$\varepsilon_n^2 = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left(x_{n,i,j} - \hat{x}_{n,i,j}\right)^2 = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left(x_{n,i,j} - a_n^{(1)} x_{n-1,i,j} - a_n^{(2)} x_{n-2,i,j} - \cdots - a_n^{(m)} x_{n-m,i,j}\right)^2$$

If the determination of coefficient a_n is based on the principle of minimum mean square error, it should satisfy $\dfrac{\partial \varepsilon_n^2}{\partial a_n^{(t)}} = 0$, namely is

$$\begin{cases} \dfrac{\partial \varepsilon_n^2}{\partial a_n^{(1)}} = -2[r(n,n-1) - a_n^{(1)} r(n-1,n-1) - a_n^{(2)} r(n-2, \\ n-1) - \cdots - a_n^{(m)} r(n-m,n-1) - a_n^{(m+1)} u(n-1)] = 0 \\ \dfrac{\partial \varepsilon_n^2}{\partial a_n^{(2)}} = -2[r(n,n-2) - a_n^{(1)} r(n-1,n-2) - a_n^{(2)} r(n-2, \\ n-2) - \cdots - a_n^{(m)} r(n-m,n-2) - a_n^{(m+1)} u(n-2)] = 0 \\ \vdots \\ \dfrac{\partial \varepsilon_n^2}{\partial a_n^{(m+1)}} = -2[u(n) - a_n^{(1)} u(n-1) - a_n^{(2)} u(n-2) - \\ \cdots - a_n^{(m)} u(n-m) - a_n^{(m+1)}] = 0 \end{cases} \tag{4}$$

Where $r(k,l) = r(l,k) = \dfrac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} x_{k,i,j} x_{l,i,j}$, and, $u(k) = \dfrac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} x_{k,i,j}$

In matrix form, it is

$$\begin{vmatrix} r(n-1,n-1) & r(n-2,n-1) & \cdots & r(n-m,n-1) & u(n-1) \\ r(n-1,n-2) & r(n-2,n-2) & \cdots & r(n-m,n-2) & u(n-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r(n-1,n-m) & r(n-2,n-m) & \cdots & r(n-m,n-m) & u(n-m) \\ u(n-1) & u(n-2) & \cdots & u(n-m) & 1 \end{vmatrix} \cdot \begin{vmatrix} a_n^{(1)} \\ a_n^{(2)} \\ \vdots \\ a_n^{(m)} \\ a_n^{(m+1)} \end{vmatrix} = \begin{vmatrix} r(n,n-1) \\ r(n,n-2) \\ \vdots \\ r(n,n-m) \\ u(n) \end{vmatrix}$$

Certainly, we could solve this matrix equation to get the coefficient vector a_n, but, as we can see, the computational complexity of a_n is o(n^2), which is of high computational cost.

With the observation of equation (2), we can get $x_n \approx \hat{x}_n = \sum_{\substack{t=-m \\ t\neq 0}}^{m} a_n^t x_{n+t}$. Because $x_n \neq 0$, hence

$$1 \approx \frac{\hat{x}_n}{x_n} = \sum_{\substack{t=-m \\ t\neq 0}}^{m} a_n^t \frac{x_{n+t}}{x_n} \tag{5}$$

Then we can find out from equation (5), if $|x_n - x_{n-i}|$ is smaller, the correlation between x_n and x_{n-i} is stronger and the value of coefficient $a_n^{(t)}$ is bigger relative to the other coefficients. So we can evaluate $a_n^{(t)}$ by means of absolute difference value $|x_n - x_{n-t}|$ between x_n and x_{n-t}. In hyperspectral image, correlation coefficient between two bands images maybe is small, but statistical correlation between local regions with same spatial location between different bands maybe is big. In the light of above thinking, we consider a correlation-driven adaptive predictor whose parameters are uniquely determined by the previously coded pixels.

We adopt the eight neighboring pixels around x_n and eight neighboring pixels around x_{n-t} as respectively illustrated by Fig.2. The pixels around $x_{n,i,j}$ and $x_{n-t,i,j}$ which belong to the same spatial location but in different bands (the current band and the reference band).

	$x_{n+t,i-2,j-1}$	$x_{n+t,i-2,j}$	
$x_{n+t,i-1,j-2}$	$x_{n+t,i-1,j-1}$	$x_{n+t,i-1,j}$	$x_{n+t,i-1,j+1}$
$x_{n+t,i,j-2}$	$x_{n+t,i,j-1}$	$x_{n+t,i,j}$	

Fig. 2. Labeling of neighboring pixels used in prediction

As shown in Fig.2, when the pixel predicted $x_{n,i,j}$ is located in the first, second row and column or in the last column, all the eight pixels needed for prediction cannot be available. In this case, we can search as many pixels as possible to calculate.

We establish evaluation function $q_{t,i,j} = \sum_{I,J} |x_{n,i,j} - x_{n-t,i,j}|$, where $I \in \{i-2, i-1, i\}$, $J \in \{j-2, j-1, j, j+1\}$.

We assume that the estimate for a_n^t is

$$\hat{a}_n^t = \frac{1}{q_{t,i,j}} = \frac{1}{\sum_{I,J} |x_{n,i,j} - x_{n-t,i,j}|} \tag{6}$$

When $q_{t,i,j}$'s value is smaller, the correlation between $x_{n,i,j}$ and $x_{n-t,i,j}$ is stronger and its specific weight in predictor of $x_{n,i,j}$ is more significant. Then we set up normalization coefficient Q in order to let $\hat{a}_n^t$ satisfy $\sum_{t=1}^{m} \hat{a}_n^t = 1$, because a_n^t means weight of $x_{n-t,i,j}$ acting in the predictor.

Namely, that is

$$\left(\sum_{t=1}^{m} \frac{1}{q_{t,i,j}} \right) Q = 1 \tag{7}$$

$$Q = \frac{1}{\sum_{t=1}^{m} \frac{1}{q_{t,i,j}}}$$

From equation (7), we can reach , and we modify the estimate of a_n^t to get

$$\hat{a}_n^t = \frac{Q}{q_{t,i,j}} \tag{8}$$

Putting equation (8) into equation (3), so we reach final predictor equation

$$\hat{x}_{n,i,j} = x_{n,i,j}^{\bar{k}} + \sum_{\substack{t=-m \\ t\neq 0}}^{m} \frac{Q}{q_{t,i,j}} \left(x_{n+t,i,j} - x_{n+t,i,j}^{\bar{k}} \right) \tag{9}$$

There are two situations in the prediction for the current band

I. The number of the current band to be predicted is from 2 to m+1

The number of current band is n, and n<m, so there is no enough m bands available to predict x_n. Then n-order predictor is adopted.

Namely, the prediction equation changes into

$$\hat{x}_n = x_n^{\bar{k}} + \sum_{\substack{t=-n \\ t\neq 0}}^{n} \frac{Q}{q_{t,i,j}} \left(x_{n+t} - x_{n+t}^{\bar{k}} \right) \tag{10}$$

Specially, the compression of the first band image uses intra-band prediction tree method.

II. The number of the current band to be predicted is m+1 behind

According to equation (9), m-order predictor is to be carried out.

3. Formation of entropy coding

The algorithm proposed for entropy coding in detail is:

(1) The first band is transformed with 5/3 integrated wavelet, and encoded with SPIHT algorithm.

(2) To nth band (n =2…m) of the image, we solve the coefficients $a_2^{(1)}$ according to equation group (8). Then $\hat{x}_{n,i,j}$ will be gotten according to equation (9), and be taken half integers to get $\lfloor \hat{x}_{n,i,j} \rfloor$.

(3) From the (m+1)th band to the Lth band image, we can get the estimate value $\hat{x}_n$ according to equation (10).

(4) We calculate error $\varepsilon_{n+1,i,j} = x_{n+1,i,j} - \hat{x}_{n+1,i,j}$ and transform residual image with 5/3 integrated wavelet, and encoded with SPIHT algorithm.

(5) Let n=n+1,if n<L, then turn to step (4)

(6) End.

4. Experimental results

The data for experiment are from NASA JPL AVIRIS' radiance data [11], which has 16-bit resolution. All the images contain 224 bands. After calculation, we can get Fig. 3 show that the performance is depended on m.

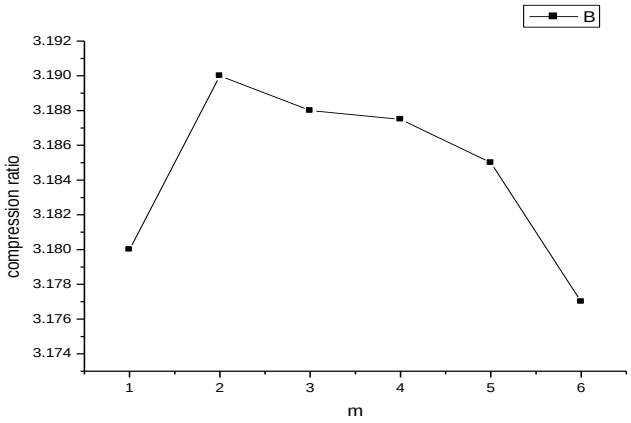

Fig. 3.Compression ratios with order of m from Jasper Ridge

As Fig.3 shows that too many bands to be predicted cannot achieve best performance. There is strong spectral correlation between bands in hyperspectral image; however with the increase of interval between bands, the correlation correspondingly decreases. Therefore, the selection of 4 order predictor outperforms the other in the Jasper Ridge. During the code, not only code streams but also the structure of the prediction tree and coefficient $a_n = \left(a_n^{(1)}, a_n^{(2)}, \cdots, a_n^{(m)}\right)$ must be transmitted. Because we adopt the backward adaptive technique, the storage costs for structure of prediction tree are much reduced. And if a floating point number is represented by four bytes, 32(m+1)L bit need transmitting. It is so small storage space needed that can be almost ignored.

Finally, the experimental results are compared with 3D CALIC, SPIHT and SLSQ algorithms, with results show as follows. Some of the results are from the literature. Table .1 shows compression ratio for AVIRIS images from four areas: Jasper Ridge, CA; Lunar Lake, NV; Cuprite NV; Moffett Field, CA. The four images' height are 2586,1431,2206,2031 pixels respectively, and their width is 512 pixels.

Table. 1. Comparison of compression ratio achieved

	SLSQ	SPIHT	3D-CALIC	NEW METHOD
Jasper Ridge	3.15	2.839	2.98	3.195
Lunar Lake	3.15	3.108	3.01	3.249
Cuprite	3.15	2.938	2.97	3.205
Moffett Field	3.14	2.787	3.17	3.153

5. Conclusion

In this paper, we have proposed a novel method for lossless compression of hyperspectral image. We establish a correlation-driven adaptive bidirectional predictor whose parameters are uniquely determined by the previously coded pixels without any side information. Facing different image, the algorithm proposed in this paper can give correspondent order m to get best compression performance. And the method is matched to an efficient entropy coder. In the end, the experimental results show that our algorithm gives the highest compression ratio for AVIRIS image. The results indicated that the method based on improved prediction tree is more suitable for lossless compression of hyperspectral images than single linear prediction or SPIHT algorithm with low computation complexity without high storage capacity.

6. References

[1] Memon N. D., Sayood K., Magliveras S., "Lossless Compression of Multispectral Image Data," IEEE Trans. on Geosci. & Remote Sensing, 32(2): 282 -289 (1994).

[2] Wu X., Memon N., "Context-Based Lossless Interband Compression-Extending CALIC," IEEE Trans. on Image Processing, 9(6): 994 -1001 (2000).

[3] Rong Zhang, Yan Qing, "A Prediction Tree-based Lossless Compression Technique of Multispectral Image Data," Journal of Remote Sensing, 2(3): 171-175 (1998).

[4] Zheng Wu, Mingyi He, "Lossless Compression of Multispectral Imagery by Error Compensated Prediction Tree," Journal of Remote Sensing, 9(2): 143-147 (2005).

[5] F. Rizzo, B. Carpentieri, G. Motta, and J. Storer, "Low-complexity lossless compression of hyperspectral imagery via linear prediction," IEEE Signal Process. Lett., vol. 12, no. 2, pp. 138–141, Feb. 2005.

[6] Jarno. Mielikainen, "Lossless compression of hyperspectral images using lookup tables", IEEE Signal Processing Letter, 2006,13(3): 157 – 160

[7] E. Magli, G. Olmo, and E. Quacchio, "Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC," IEEE Geosci. Remote Sens. Letter., vol. 1, no. 1, pp. 21–25, Jan. 2004.

[8] A. Said ,W. A. Pearlman, "A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees," IEEE Trans. Circuits System Video Technology, 1996 (6) :243-250 (1996).

[9] B. Aiazzi, P. S. Alba, L. Alparone, S. Baronti, and P. Guameri, "Reversible inter-frame compression of multispectral images based on a previous-closest-neighbor prediction," Proc. ICARSS, pp. 460-462 (1996)

[10] X. Li, M. T. Orchard, "Edge-Directed Prediction for Lossless Compression of Natural Images," IEEE Trans. Image Processing, 10(6):813-817 (2001).

[11] Free AVIRIS Standard Data Products. [Online]http://aviris.jpl.nasa.gov/html/aviris.freedata.html.