Predictive encoding techniques for lossless hyperspectral compression

Dmitry Pertsev, Belarussian State University of Informatics and Radioelectronics (01.11.2012, prof. Alexander Doudkin, United Institute of Informatics Problems)

Abstract
The article describes a structure of hyperspectral images, their features that are most often used in compression algorithms. Also we review the general view of predictive coding techniques. The results of compression for various predictive encoders are provided at the end of the article.

1. Introduction
Hyperspectral images are development in the field of a spectral imaginary. It represents a 3D cube of data comprising spatial cover information in various spectral areas. A hyperspectral image has a large volume of data. That's why it is difficult to transmit from the satellite to Earth for the subsequent analysis and processing. At the time hyperspectral images are often used in environmental monitoring, climate research, gas leak detection, etc.

In this article we describe the general compression scheme of hyperspectral images and the most perspective class of algorithms in the field of lossless compression (a predictive coding) is considered. Lossless compression was chosen to guarantee possibility of the most exact image recovery to process it on the Earth.

2. Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)
We choose the AVIRIS data, because the data of the spectrometer are publicly available[1] and absolutely free for experimental research. It is an optical sensor that delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral channels (bands) with wavelengths from 400 to 2500 nanometers with a nominal 10 nm sampling. Each pixel covers an area approximately 20 meters square on the ground, thus yielding a ground swath about 11 kilometers wide. When all of this data is processed and stored on the ground, it yields approximately 140 Megabytes for every 512 lines of data. Each 512 line set of data is called a "scene", and corresponds to an area about 10 km long on the ground.

Publicly available hyperspectral images are usually introduced in one of two format – BIP and BIL. Pixel information stores band by band for each line of the image in the BIL (Band Interleaved by Line) format. The diagram illustrates it is introduced on fig.1.

![Band Interleaved by Line format](image1)

BIP (Band Interleaved by Pixel) format is similar to the BIL format, except that the data for each pixel is written band by band. The diagram illustrates it is introduced on fig.2.

![Band Interleaved by Pixel format](image2)
3. General scheme of data compression

The common view of compression algorithm is introduced on fig.3 [2].

![Fig.3. Image compression algorithm](image)

The first step is image analyze. Band reordering is one of the most efficient image prepare algorithms. The main idea is to find optimal band order to minimize redundancy. As usual all algorithms in the way based on correlation. High correlated bands can be expressed from each other. There are two correlation types in hyperspectral images – spatial and spectral correlation. High spatial correlation is based that nearby places consist of one material. A spectral correlation is a correlation between spectral bands. The majority compression algorithms focus attention on spectral correlation, because it is always stronger than spatial correlation.

A spectral (1) and a spatial (2) correlation equations [3] are

\[ c_{u,v} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} \tilde{x}_{i,j,u} \cdot \tilde{x}_{i,j,v}}{\sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} \tilde{x}_{i,j,u}^2 \cdot \sum_{i=1}^{M} \sum_{j=1}^{N} \tilde{x}_{i,j,v}^2}} \]  

\[ c_k = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} \tilde{x}_{i,j,k} \cdot \tilde{x}_{i+1,j+1,k}}{\sum_{i=1}^{M} \sum_{j=1}^{N} \tilde{x}_{i,j,k}^2} \]  

where \( \tilde{x}_{i,j,k} = x_{i,j,k} - \bar{x}_k \), \( x_{i,j,k} \) and \( \bar{x}_k \) represents the pixel value at the spatial location of \((i,j)\) and mean value of band \(k\) respectively, \(M\) and \(N\) are the width and height of each band.

A spectral and a spatial correlation example for Yellowstone calibrated scene 11 image are introduced on fig.4 and fig.5 correspondingly.

![Fig.4. Spectral correlation (Yellowstone, scene 11)](image)

![Fig.5. Spatial correlation (Yellowstone, scene 11)](image)

The second step is image coding. There are three types of data compression – lossy, lossless, and near-lossless. The decoded and the original images for lossless algorithm will be the same (for lossy algorithm will be not). Near-lossless algorithms are combination of lossy and lossless algorithms. Data compression algorithm will be near-lossless if one of the next ideas will be true:

- the encode noise is less then sensor noise[4];
- the distortion is bounded[5].

Also the most algorithms can be applied to two big groups: transform coding, prediction coding, vector quantization.

The most transform techniques are lossy compression (for example, Principal Component Analysis, Karhunen-Loeve Transform, etc.). We can found some lossless algorithms, but it is not efficient.
Vector quantization techniques require codebook training and an effective quantization-index searching algorithm. It makes them computationally expensive and not always well suited for real-time application.

Predictive coding algorithms are the most popular and effective. Also their computational cost and complexity are usually less than vector quantization methods. The most popular predictive coding algorithms in hyperspectral compression are lookup tables (LUT)[6], CALIC (3D-CALIC and M-CALIC)[7], Spectral-oriented Least Squares (SLSQ)[8], Context-based Conditional Average Prediction (CCAP)[9] and their combinations.

4. Predictive coding

The common scheme of a predictive coding algorithm is introduced on fig.6. Predictive coding approaches usually employ spatial, spectral, and hybrid predictors to decorrelate the image, followed by entropy coders.

A predictor is a function

\[ V = f(x, y, n, p) \]  

where \((x, y)\) is a pixel coordinates, \(n\) is a vector of the next pixels values, \(p\) is image parameters.

Step by step algorithm:
1. \(v_i\) is an original pixel which is suited as coded;
2. for each pixels in image:
   a. calculate a new pixel value \(p_i = f(x_i, y_i, n, p)\), where vector \(n\) contains all encoded pixels;
   b. calculate prediction error \(e_i = v_i - p_i\) and mark it as coded.

A predictive function is choosing based on minimizing relative frequency (fig.7).

There are some predictor types:
- based on parameters using:
  - a static predictor. There are no parameters, all the time used the same predictor;
  - an adaptive predictor. The parameters are changing during the coding process.
- based on model type:
  - a linear predictor. The predictor based on linear equation.
  - a non-linear predictor. The predictor based on non-linear equation (for example, neural and genetic)

4.1. Linear and non-linear predictive coding

A linear predictive coding based on the next equation 4:

\[ p(i) = \sum_{j=i-n}^{j=i+1} k_j v(j) \]

where \(p(i)\) is a predictive value for \(i^{th}\) pixel, \(v(j)\) is a value of pixel \(j\), \(k_j\) is a coefficient, \(n\) is encoded pixels count.

The main problem of this type is coefficient \(k_j\), because it can be optimal to some images, but very bad for others. It is main reason why the linear predictive coding never used alone, but often as part of adaptive predictor.

Non-linear predictor is more complex. There are the predictor tries to analyze current pixel neighboring and based on the result try to find optimal value of new pixel. The examples of non-linear predictors are Paeth, DARK, median edge detection predictor, gradient-adjusted predictor, etc.

4.2. Adaptive predictor

There are two different class of the adaptive predictor. The first class is
\[ p(i) = \sum_{j=1}^{N} \alpha_j p_j(i) \]  

(5)

where \( p(i) \) is a predictive pixel value, \( p_j(i) \) is a pixel value predicted by a \( j^{th} \) predictor, \( \alpha_j \) is a coefficient.

The coefficient \( \alpha_j \) is calculated based on “penalize” predictor method\[10\]. The penalty term is calculated at first:

\[ G_j = \sum_{k=1}^{L} | v(k) - p_j(k) | \]  

(6)

where \( L \) is considered number of analyzed pixels, \( v(k) \) is a \( k^{th} \) pixel value, \( p_j(k) \) is a \( j^{th} \) pixel value predicted by a \( j^{th} \) predictor.

The coefficient \( \alpha_j \) is calculated based on equation 7:

\[ \alpha_j = \frac{1}{G_j} \sum_{i=1}^{N} \frac{1}{G_i} \]  

(7)

The second class is the same as equation 4, but the coefficient \( k_j \) is dynamically calculated. There are often used ordinary least squares for it.

All predictive algorithms are based on the considered scheme and differ by an applied predictor and the scheme of pre-processing. So, for example, SLSQ algorithm is based on a spectral linear predictor, CCAP consist of inter-band/intra-band prediction and try to remove the redundancy of residual images. LUT algorithm at first searches the previous band for a pixel whose value is equal to the one co-located to the current pixel, and then predicts the current pixel using the one in the current band whose spatial position is the same as the found.

5. Conclusion and future work

In comparison with lossy compression, a lossless technique provides low compression ratio. Partially it is connected with that bands reordering analyze demands essential computing resources that make it unsuitable for satellite application.

In future work it is planned to analyze Spectral-oriented Least Squares (SLSQ) algorithm in a combination with a Context-based Conditional Average Prediction (CCAP) since these algorithms mathematically are rather simple and can be applied as compression system on the satellite.

Bibliography

[1] AVIRIS Home Page. [Online]. Available: Title of paper, information about editor e.g.:  

Authors:  
MSc. Dmitry Pertsev
Belarusian State University of Informatics and Radioelectronics  
P. Browki Str., 6  
220013 Minsk  
tel. + 375 17 293-80-39  
email: DmitryPertsev@gmail.com