HYPERSPECTRAL IMAGE COMPRESSION

Paper implementation of

“Satellite Hyperspectral Imagery Compression Algorithm Based on Adaptive Band Regrouping ”

by Zheng Zhou, Yihua Tan and Jian Liu

Syed Ahsan Ishtiaque

Ankush Khandelwal
Abstract:

“Compression of hyperspectral images holds importance because of bad transmission capability of satellites. The project is the implementation of the paper-“Satellite Hyperspectral Imagery Compression Algorithm Based on Adaptive Band Regrouping”. But during the course of the implementation many modifications have been made and some new concepts have been introduced. We have made the given regrouping process automatic which has made possible its use as a module before any prediction scheme. Introduction of our own prediction model and its comparison with the given model brought out some key points about hyperspectral images and possibility of achieving high compression rates by using object oriented approach to compress images.”

Introduction:

Hyperspectral Images are images with large number of spectral bands. These bands generally have good interband correlation (spectral redundancy) and intra band correlation (spatial redundancy). Traditional schemes for color images do not exploit inter band correlations. In this project a paper has been implemented which exploits inter band correlation to increase compression.

- Correlation Function

The following correlation function has been used throughout:

\[
Corr_{X,Y} = \frac{1}{M \times N} \sum \sum (X_{i,j} - \bar{X})(Y_{i,j} - \bar{Y})
\]

Where,

- \( Corr_{X,Y} \) = correlation value between band X and band Y.
- \( M \) = number of rows in a band.
- \( N \) = number of columns in a band.
- \( X_{i,j} \) = pixel value at \((i,j)\) position of the band X.
- \( \bar{x} \) = mean of the band X

This function measure the linear correlation between two random variables (which are band images in our case). It has been assumed that bands have linear correlation between them (illustrated by the paper).
• Why grouping of bands

Figure 1: Image showing average correlation of a band with every other band.

Above image shows the average of the correlation values of a band with every other band. Bands enclosed in elliptical regions have high value of average correlation and bands in square regions have bad average correlation. The bands in square regions are generally noisy bands. So if a band in square region predicts a good band (bands in elliptical regions) then large value of residuals will be obtained. This is the problem of traditional method of direct linear prediction (which takes all the bands together in prediction) which is avoided in grouping by keeping bands having good correlation among themselves into one group. Thus noisy band are generally left out and are not part of any group.
**Algorithm:**

This section describes the steps involved in the whole compression scheme. Here the description is in context with the algorithm proposed in the paper. This will set the basis of comparison which will make the analysis of the modifications and new introductions easy.

To start with we have a set of ungrouped bands which contains all the bands.

**STEP 1:** Selecting a *key band K*.

A band is picked from the set of ungrouped bands. Proposed algorithm picks it randomly.

**STEP 2:** Setting the *threshold T*.

Proposed algorithm chooses it after experimenting various values on different images.

**STEP 3:** *Grouping*.

After selecting the first *key band*, all the bands with correlation value more than T are taken into one group. Then again step 1,2 and 3 are repeated with the remaining bands and so on. After all groups are formed for each group a reference band is chosen. Proposed algorithm takes the band which is most close to the linear mean band of the group.

**STEP 4:** *Applying prediction models*.

After groups are formed, prediction models are applied on each group separately and residual images corresponding to each band are created. In the project we have used two prediction models, one has proposed in the paper and the second one has been introduced by us.

**STEP 5:** *Compressing* residuals

In this step residuals generated by prediction models are compressed. Paper uses state of art compression algorithm. In the project focused was only on first 4 steps, so in later sections only first four steps are discussed.
Methods/Implementation

In the light of above algorithm we have implemented two methods which also includes modifications brought by us.

STEP 1:

Instead of choosing the key band randomly we pick the band that has highest average correlation with all the other bands. This is the first modification that has been brought which stops a noisy band to become the key band which may lead to building of a bad group and generating large residuals.

STEP 2:

Here the second modification has been brought. Instead of depending on the experiments to set the threshold, we have made the threshold selection process automatic. This change brings a very significant change in the whole grouping scheme as it is now can be used as a module before any prediction scheme which was not possible before.

STEP 3:

Grouping process also changes because of the modifications in step 1 and 2. For grouping one assumption has been taken. Image having n bands should have groups of sizes more than $\sqrt{n}$. Ideal case is having $\sqrt{n}$ groups having $\sqrt{n}$ number of bands in each group. This is to avoid both the extremes of having only one group with all the bands and on the other side have n groups with only one band in each group.
Pseudo code for the process above

\[ S = \{1,2,3,\ldots,n-1,n\} \]  \hspace{1cm} \text{% set of ungrouped bands (initialized with all the bands in the set)}

\[ T = 1.0 \]  \hspace{1cm} \text{% initialized}

\[ \text{CheckThreshold} = -1 \]  \hspace{1cm} \text{% holds the value of the threshold at which we get the first valid group.}

\text{Iteration:}

\[ \text{Group} = \text{Apply grouping scheme proposed in the paper with } S, T \]
\hspace{1cm} \text{% just changing the selection of } k \text{ by not being random to the way defined in step 1 of this method.}

\text{For all groups having size < predefined size}
\hspace{1cm} \text{Dissolve the group.}
\hspace{1cm} \text{% All the bands that are in surviving groups are taken out from the set } S.

\text{If any group survives and CheckThreshold == -1}
\hspace{1cm} \text{CheckThreshold = } T
\hspace{1cm} \text{If CheckThreshold – } T > 0.2 \text{ % termination condition}
\hspace{1cm} \hspace{1cm} \text{% (20\% limit on decrement has been put so that groups do not form at very low thresholds)}
\hspace{1cm} \hspace{1cm} \text{exit}
\hspace{1cm} \text{else}
\hspace{1cm} \hspace{1cm} T = T - 0.02
\hspace{1cm} \text{end}
\hspace{1cm} \text{% after leaving iteration we have groups with sizes more than predefined size, but there will be some bands which will not be able to form groups (noisy bands) as they have very bad correlation with other bands.}

\text{Create group with current } T \text{ and } S \text{ and this time without dissolving the groups.}

\text{% Generally after this step noisy bands make groups in which they alone are present. By this they are not able to predict any good band.}

\text{In this way there is no need to do experiments to find the threshold value. This method makes possible the use of regrouping as a module before any prediction technique.}
STEP 4:

In this method we have applied our prediction model. The model is as follows:

- **Prediction Model 1**:

  Key band is taken as the reference band of the group. Each band in a group is represented in terms of the reference band as

  \[ Y_{i,j} = a \times X_{i,j} + b \quad \text{for all } i, j \quad --- (1) \]

  Where,
  
  \( Y \) is the band whose prediction has been calculated in terms of the reference band \( X \).
  \( a \) and \( b \) are parameters. Each predicted band will have one value of \( a \) and \( b \).

  It means that a complete band can be linearly represented in terms of the reference band. The reason why this equation may work is that when this equation is used to replace \( Y \) for all \( i,j \) in the correlation functions then the correlation function gives the value 1 (which supports the linear relation between two bands). So it appears that the formula is worth testing.

  **Calculation of \( a \) and \( b \):**

  Adding the above equation for all \( i,j \) and dividing both sides by number of pixels in the image, we get following equation-

  \[ \bar{Y} = a \times \bar{X} + b \quad --- (2) \]

  Where,
  
  \( \bar{Y} \) and \( \bar{X} \) are mean of the bands

  Similarly when the equation 1 is used to replace each \( Y_{i,j} \) in terms of \( X_{i,j} \), we get the following equation-

  \[ V_y = a^2 \times V_x \quad --- (3) \]

  **Using 2 and 3 \( a \) and \( b \) can be calculated.**

  Then using this model residuals of the bands are produced. Residual image of a band is the difference between the original band and its prediction.
Prediction Model 2

In this method the process till group formation is same as of method 1. But after this step one more step is performed and that is of Band Ordering (requirement of the prediction model). Ordering of bands of a group has been done using concept of Minimum Spanning Trees.

In this model, for a band, reference band is the band which is just behind it in the ordering.

The equations for the second prediction model are:

\[ Pr_{r,c,b} = \text{floor}(A_{r,c,b} \times P_{r,c,b-1} + \frac{1}{2}) \]

where

\( Pr_{r,c,b} \) is the value at position \( r,c \) in the prediction image of band \( b \).

\( P_{r,c,b-1} \) is the value at position \( r,c \) in the band \( b-1 \) (reference band)

and

\[ A_{r,c,b} = \frac{P_{r-1,c,b} + P_{r-1,c,-1,b} + P_{r-1,c+1,b} + P_{r,c-1,b}}{P_{r-1,c,b-1} + P_{r-1,c,-1,b-1} + P_{r-1,c+1,b-1} + P_{r,c-1,b-1}} \]

Causal neighborhood

So each band will be predicted using the band which is behind it in the ordering. The first band will not be predicted and will be kept as it is. Moreover the boundary pixels will not be predicted because they don’t have causal neighborhood and thus will be retained as it is in the residuals (they act as overhead).
Results and Analysis:

The major part of the project was the analysis of the two prediction models. Their comparison brought out some good points.

Consider the region enclosed in the red square of the image

![Image](image.png)

The selected region contains two types of object, land and water. Now let us see the across band variations of the pixels of any of the rows of the selected region.

(Across Band Variation: It is just variation of pixel value corresponding to a single pixel across all bands e.g. checking the values of pixel (20,20) across all the 224 bands of a sample image)
These images are in increasing order of the pixel in a row. As you can see there are dominantly only two types of templates that are appearing, one corresponds to land and the other one to water. What if the image contain only one type of type template (only one object), modeling would become so much efficient!!!

Now let us see the histograms of the residuals created by both methods.
These residuals are showing peak at zero which shows that most of the values have been correctly predicted. In general if the residual histogram shows peak then the model has worked fine. If peak is not at zeros that means the value of b has been incorrectly calculated. So in all method 2 has given satisfactory results.

Now let’s see the histograms of the image generated from method 1:
It is clearly visible that some of the histograms have two peaks and some have one peak. Moreover the histograms are scattered. The reason for these observations lies in the prediction models. Our model produced bad results because we give same value of slope (variable a in the prediction equation) to both the templates that we observed in the across band variations of the pixels. But their model gives the value of a locally (by using causal neighbourhood). The value of a differs for both the templates as both vary differently across bands. Following image shows the variation of both the templates together.

Across band variation of positions (2,15) and (2,30)
Thus model proposed in the paper is more robust to images having multiple objects.

Now let's remove one of the objects and see how both the models change their behaviour.

Consider the selected region of the following image:

The only template that is present in this image is
Now let's see the histograms produced from both the methods.
The response of the method 2 has not changed significantly. It is still giving single peak and that too at zero.

Now the histograms of the method 1:

Histograms by method 1
Now, one can easily see the change in histograms of method 1. It is now giving only one peak and that too at zero. So this makes one thing clear that our method can work well with images having single objects. So if we have an image containing only one image then our method can be a better option because it is computationally less expensive than model 2 and moreover does not have the overhead of the boundary pixels which are passed as it is in the residuals.

Till now the analysis was subjective. To get the numerical comparison between the methods, we have implemented step 5 with **Huffman coding** and we have given each image a number which indicates about its compression.

**Scheme:**

- Each band has a code of residual band, parameters (a and b in case of method1), dictionary associated with it.
- Byte has been taken as the unit.
- A variable which is of type double will occupy 8 units of space.
- A code of the band will occupy ceil (n/8) units of space where n is the length of the code.
- For dictionary with n symbols we will have n*8 units for storing the symbols and then space to store the code of each symbol.
- In this way space occupied by the compressed image is calculated. Dividing the space occupied by original image by the space of compressed image will give us the value of the compression obtained by method.

After using the above scheme, it is clear that we can achieve high compression rates if single objects are present in the image. So these observations led us to conclusion that object oriented approach of compression can bring significant increase in the compression rates. Moreover, use of single object images can increase the efficiency of prediction models. This make possible eliminating the residuals from the process and just storing the parameters for the model s and using predicted images instead of original images.
Future Work:

*Lossy Compression scheme*

A small experiment showed some encouraging results.

If in an image only the visual appearance is important and not the exact content then we need not store any residual images and the overhead of compressing them, what we can do is just store the reference bands and store the prediction model which may be directly fed into applications which decompresses it, so by this we can achieve great compression ratio as we are only storing reference bands and according to our concept on an average there are $\sqrt{n}$ groups so there are only $\sqrt{n}$ reference bands so we theoretically achieve a compression ratio of $\sqrt{n}$ where n is number of bands.

One can see that there is only change in the brightness of the prediction image. This loss is unacceptable in case of image where the values of the pixels are important. But images like that of a city, this can be bearable because for those images only subjective quality is important. So there exists a possibility of achieving high compression ratios on some type of images.
Conclusion:

All the analysis done in the project has opened a new way for compressing hyperspectral images. Object oriented way of compressing an image in itself breaks into two. One is keeping the subjective quality of the image safe (for image of a city) and other one is keeping the numerical quality (variations among the pixels of an object as they can be indicators of some physical property of the object. eg. depth of water, chlorophyll content, etc.) of the image safe. All this will be the part of the second phase of the work. Moreover the correlation function used will also be put to test and different correlation functions will be tested with proper understanding of their behavior.