Analysis of Inter-band Spectral Cross-Correlation Structure of Hyperspectral Data

AYMAN AHMED
National Authority for remote Sensing and Space Science, 23rd Josef tito street, Elnozha El gedida Cairo, 1564 Alf Maskan, Egypt, a_ymn2002@yahoo.com.

MOHAMED EL. SHARKAWY
Egypt-Japan University of science and technology, New Borg El Arab, Alexandria, Egypt

SALWA EL. RAMLY
Faculty of Engineering, Ain Shams University, 1 El Sarayat Street. Abbasaya, Cairo, Egypt.

Abstract: - Hyperspectral imaging has been widely studied in many applications; notably in climate changes, vegetation, and desert studies. However, such kind of imaging brings a huge amount of data, which requires transmission, processing, and storage resources for both airborne and spaceborne imaging. Compression of hyperspectral data cubes is an effective solution for these problems. Lossless compression of the hyperspectral data usually results in low compression ratio, which may not meet the available resources; on the other hand, lossy compression may give the desired ratio, but with a significant degradation effect on object identification performance of the hyperspectral data. Moreover, most hyperspectral data compression techniques exploit the similarities in spectral dimensions; which require bands reordering or regrouping, to make use of the spectral redundancy. In this paper, we analyze the spectral cross correlation between bands for AVIRIS and Hyperion hyperspectral data; spectral cross correlation matrix is calculated, analyzed, and by assessing the strength of the spectral matrix, we propose a new technique for bands regroup by finding the highly correlated groups of bands in the hyperspectral data cube based on "inter band correlation square".

Key-Words: -hyperspectral imaging, bands regrouping, edge detection, spectral correlation matrix, image compression

1 Introduction
As hyperspectral data contains a huge amount of spectral data distinctive in spectral resolution, which allows identification of each pixel based on its spectral footprint. On the other hand; this amount of data increases as spectral bands increase, usually satellite instruments that measures earth's illumination at specified spectral band has more dynamic range than visual images; typically ranges from 10 up to 16 bits per pixel per band; additionally, considering the swath width of the satellite imagery; hyperspectral imaging session of satellite may contain tremendous amount of digital data to be transmitted to ground station [1]; this limits the imaging session and spatial resolution in satellite imaging.

Many researches have been conducted to, efficiently and carefully, compress this amount of data without losing the main advantage of hyperspectral imaging, which is spectral resolution; two known compression approaches are usually investigated, lossy and lossless techniques; lossless compression is perfect for compression data and keeping the original information without distortion and in the same time allow further processing of the image to identify earth's objects accurately; unfortunately this approach of compression gives compression ratio ranging from 1 to 3 [2], [3]; that means the compressed data will have smaller volume down to 3 times less the original one; this in practical is not sufficient, and compressed data still represent a significant issue for satellite for transmission and storage [4].

On the other hand; lossy compression approach gives a great compression ratio, which may go to 40 times; this ratio, is achieved scarifying the low distortion rate; that means more losses will appear on the reconstructed data; which causes losses affecting the process of earth's object identification [5].

Another approach of compression is known as near lossless; this approach achieves relatively bigger compression ratio than achieved by lossless approach and smaller distortion less than that resulting from lossy compression approaches.

Meanwhile; researches are continuing to find optimum solution that can fit onboard satellites [6]; most researches go around exploitation of either spectral or spatial redundancy of hyperspectral data; spatial redundancy results from the fact that,
imaging certain territory will have similarity in spatial dimensions; these characteristics are exhaustively investigated; and as a result; we have discrete cosine transformation and wavelet transformation [3] that exploits the spatial redundancy in the images in different ways of implementation either in ground image processing software or onboard satellite instruments.

On the other hand; spectral redundancy is a relatively new dimension in hyperspectral imaging; many researches are trying to investigate the best way to deal with this redundancy [7], [8] and optimum techniques to exploit it; these fact lead to another branch of investigation and analysis of spectral structure of hyperspectral data [9], [10], this analysis will put the main outlines and answers on how to best utilize inter-band spectral correlation in hyperspectral data.

2 Inter-band spectral cross correlation and similarity measurement

Hyperspectral data can be viewed as a "Data-Cube"; this data cube has two spatial dimensions and one spectral dimension, spectral dimension represents the captured image in different spectral bands, usually in a successive manner and, typically, has around 250 bands. Spectral redundancy is based on the similarity between bands and each other’s; these similarity can be measured by spectral cross correlation [11], Conditional entropy, mutual information, Euclidian Distance, Maximum Absolute Distance, and Centered Euclidian Distance; these measures are well studied by researchers and compared to determine which one is best fit for reordering the bands for prediction based compression techniques; correlation is found to be the best for similarity measurement [8], [12], [13]; this fact is a good point to start the analysis of spectral structure of the hyperspectral data cube.

Correlation between spectral bands is named spectral correlation as it represents the correlation between two identical images in different spectral domain. The imaged piece of land by hyperspectral instrument is treated as unknown object, since the main objective of the satellite imager (spaceborne) is to provide data about these objects. Cross correlation is usually a standard measure of degree of similarity between two images (matrices); it is optimized for faster calculation and /or more accurate results, some techniques of cross correlation estimation was used to investigate the hyperspectral inter-band correlation [8], [9]; this process is time consuming and requires large computational power.

Cross correlation mainly depends on covariance calculation between the two bands; while normalized cross correlation uses variance of each band to make the value independent of variation of both brightness and contrast of the images.

Selection of estimation technique should be based on deterministic criteria; such as simplicity, speed, required resources of memory and computational power, and accuracy of calculation.

Fast normalized cross correlation is a very good technique for calculating the similarity between two images[14], [15]; as it is fast, accurate, and independent of images pixel's brightness and contrast values.

3 Spectral Cross correlation Matrix

When using fast normalized cross correlation technique, correlation between all hyperspectral bands in the data cube is estimated; for band i, correlation value is estimated with all other bands in the data cube j; using the Eq. (1).

\[
NCC(i,j) = \frac{\sum_{x,y} (iD(x,y) - \overline{iD}) (jD(x,y) - \overline{jD})}{\sqrt{\sum_{x,y} (iD(x,y) - \overline{iD})^2 \sum_{x,y} (jD(x,y) - \overline{jD})^2}}
\]

Where:
NCC(i,j): Normalized cross correlation between bands i and j, \(iD(x,y)\): intensity of pixel, (x, y) pixel indices within one band, \(\overline{iD}\): mean of pixel intensity values of band i.

This function is implemented in fast, optimized way in Matlab image processing tool box [16] based on "Fast Normalized Cross-Correlation" [14].

Fig.1. Comparing SCM computed using wavelet transform on bands and original ands, respectively, correlation 0.9137.

Estimation of the correlation value for the entire data cubes, is time consuming and needs powerful machine; we propose to use wavelet transform first to minimize dimensions of each band to the half in each axes; which results in band size quarter of the...
original one; at the same time resulted spectral correlation matrix will still almost the same, Fig. 1.; while the correlation between these two matrices is 0.9137; this value is accepted, since we will be interesting to find some region within the matrix itself, this will be more clear in the analysis of the data.

### 4 spectral correlation matrices estimation for data samples

Hyperspectral data cube can be classified according to inter-band correlation; to have "Weak Spectral correlation matrix" (WSCM), or "Strong correlation Matrix" (SSCM); this can be measured according to mean correlation value of the spectral correlation matrix as in Fig. 2.

Any spectral correlation matrix is symmetric around its diagonal, which is the nature of calculation method as shown in Fig. 3.

Estimating the inter-band cross correlation matrix for each hyperspectral data cube as stated in section 3; the resulted spectral correlation matrix for ten hyperspectral data samples, from both airborne and spaceborne instruments [17], [18], are illustrated in Fig. 4, Fig. 5, Fig. 6, Fig. 7, and Fig. 9.

As initially observed some data cubes have strong average correlation between bands even separated hundred bands away; while, on the other hand, some hyperspectral data cubes have a very weak average correlation values Fig. 4, Fig. 5.

Ten hyperspectral data samples are processed and used in this study; Table 1, Illustrates the details of each data samples, associated figure that reflects the image view of spectral correlation matrix and name of correlation matrix used in calculations.

<table>
<thead>
<tr>
<th>Hyperspectral data sample</th>
<th>Details</th>
<th>Figure number</th>
<th>Spectral correlation matrix name</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;hawaii_sc01&quot;</td>
<td>AVIRIS instrument: 512 lines x 614 samples x 224 bands, instrument bit depth = 12 bits</td>
<td>Fig. 7</td>
<td>&quot;corr_mtxh&quot;</td>
</tr>
<tr>
<td>&quot;maine_sc10&quot;</td>
<td>AVIRIS instrument: 512 lines x 680 samples x 224 bands, instrument bit depth = 16 bits</td>
<td>Fig. 9</td>
<td>&quot;corr_mtxm&quot;</td>
</tr>
<tr>
<td>&quot;Aviris_sc0&quot;</td>
<td>AVIRIS instrument: 512 lines x 680 samples x 224 bands, instrument bit depth = 16 bits</td>
<td>Fig. 4</td>
<td>&quot;corr_mtx0&quot;</td>
</tr>
<tr>
<td>&quot;Aviris_sc3&quot;</td>
<td>AVIRIS instrument: 512 lines x 680 samples x 224 bands, instrument bit depth = 16 bits</td>
<td>Fig. 5</td>
<td>&quot;corr_mtx3&quot;</td>
</tr>
<tr>
<td>&quot;Aviris_sc10&quot;</td>
<td>AVIRIS instrument: 512 lines x 680 samples x 224 bands, instrument bit depth = 16 bits</td>
<td>Fig. 5</td>
<td>&quot;corr_mtxh10&quot;</td>
</tr>
<tr>
<td>&quot;Aviris_sc18&quot;</td>
<td>AVIRIS instrument: 512 lines x 680 samples x 224 bands, instrument bit depth = 16 bits</td>
<td>Fig. 6</td>
<td>&quot;corr_mtx18&quot;</td>
</tr>
<tr>
<td>&quot;T96070501&quot;</td>
<td>AVIRIS instrument: 512 lines x 680 samples x 224 bands, instrument bit depth = 16 bits</td>
<td>Fig. 6</td>
<td>&quot;corr_mtx11&quot;</td>
</tr>
<tr>
<td>&quot;EretaAle&quot;</td>
<td>Hyperion instrument: 3187 lines x 256 samples x 242 bands, instrument bit depth = 12 bits</td>
<td>Fig. 7</td>
<td>&quot;corr_mtxer&quot;</td>
</tr>
<tr>
<td>&quot;LakeMonona&quot;</td>
<td>Hyperion instrument: 3176 lines x 256 samples x 242 bands, instrument bit depth = 12 bits</td>
<td>Fig. 8</td>
<td>&quot;corr_mtxl&quot;</td>
</tr>
<tr>
<td>&quot;MtStHelens&quot;</td>
<td>Hyperion instrument: 3242 lines x 256 samples x 242 bands, instrument bit depth = 12 bits</td>
<td>Fig. 8</td>
<td>&quot;corr_mtxhel&quot;</td>
</tr>
</tbody>
</table>

Fig. 2. Illustrates some statistical values about each data samples; that includes, mean value, variance, standard deviation, median, minimum, and maximum values.

Mean values are calculated as the average value for each raw then averaging these values to get one mean value for the entire matrix, according to this "mean" the matrix is classified into WSCM or SCCM. Variance is the square value of the standard deviation; while median is median value of the medians for each raw.
Color mapping of the correlation values, is illustrated in Fig. 4; and as it is seen, red color means strong correlation with values around 0.8 or higher; passing through the color graduation until weak correlation represented by negative values with blue color.

5 Analysis of spectral cross correlation matrices

As the results are statistically analyzed in Fig. 2; it seems that correlation matrices can be classified into two categories; the first category, has strong correlation as visually seen, in image view of these matrices; at the same time, theses matrices have relatively large mean values, basically greater than 0.1(SSCM); while the second category, has small correlation less than 0.1(WSCM).

Hence; spectral correlation matrix that has mean value greater than 0.1(SSCM), has a group of correlated bands even if these bands are far from each other.

Based on this basic assumption, more analysis of each (SSCM) is performed to study how these "Group Of Bands" (GOB), if exists, are correlated.

"corr mtxl" and "corr_mtxer" represent the SSCMs; with a deep look at each of them, it can be noticed that for "corr_mtxer" there is a strong correlation between the groups of bands starting from approximately band number 10 till band number 55, and group of bands starting from approximately band number 180 till band number 220 (GOB (~10-55) and GOB (~180-220)).

This correlation appears in a square manner, dashed square in Fig. 7, Fig. 8, are called "Inter Band Correlation Squares" (IBCS); almost all correlated GOBs have IBCS pattern in spectral correlation matrix.

IBCS is not symmetric around spectral correlation matrix; otherwise, it is inter-band correlation triangle (IBCT).

Checking of the existence of the IBCS or IBCT, allows determining the correlation between bands in hyperspectral data cube, this helps in band reordering techniques used in compression of hyperspectral data [12], [13].

Determining the group of bands, that are highly correlated, is one of the interesting subjects [11], [12], [13] that would increase the efficiency of using compression coders, especially video codec [19], which depends on exploiting the maximum redundancy between frames, band in this case, to achieve higher compression ratio [13][20], [21].
Fig. 5. Image view of SCM for "corr mtx3", and "corr mtx10", respectively.

Fig. 6. Image view of SCM for "corr mtx18" and "corr mtx1", respectively.

Fig. 7. Image view of SCM for "corr mtxer" and "corr mtxh", respectively.

Fig. 8. Image view of SCM for "corr mtx1l" and "corr mtxhel", respectively.
6 Inter-band correlation square

Inter-band correlation square is a pattern of spectral correlation between bands in hyperspectral data; finding this square(s) refers to the location of the group(s) of bands that are highly correlated, and usually these GOBs are far away from each other.

Edge detection is a concept of image processing that helps to locate edges in the processed image; some algorithms are used in this area; such as, Sobel Method, Prewitt Method, Roberts Method, Laplacian of Gaussian Method, Zero-Cross Method, and Canny Method.

The interest here is to find the algorithm that determines the location(s) of the IBCS in the (S)SCM; the algorithm should, at least, be able to determine the location of the biggest IBCS; or, if many similar exist, to determine at least one of them.

Sobel, Prewitt, and Roberts methods find edges using the corresponding approximation to the derivative, and return edges at those points of maximum gradient. The Laplacian of Gaussian method finds edges by looking for zero crossings after filtering the matrix with a Laplacian of Gaussian filter. Zero-cross method finds edges by looking for zero crossings after filtering matrix with a selected filter.

The Canny method finds edges by looking for local maxima of the gradient. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise and more likely to detect true weak edges [22].

Comparative Study of these techniques [23] have been carried out; also performance analysis of each of them [24], recommends that Canny Method has better performance while detection of edges and less sensitivity for noise.

Using "Canny method" to estimate the edges of the IBCS and IBCT; we got the results in Fig. 10 and Fig. 11.

Finding the highest IBCS, requires an iterative process, beginning with highest threshold in the matrix; that corresponds to highest possible correlation values between bands, the iteration step resolution will invoke more iterations; it is recommended to examine the matrix strength, as indicated earlier, by its mean value, i.e. (SSCM or WSCM); most probably, WSCM will have no IBCS or if it has, it will be small.

Fig. 12. Illustrates the process of finding the first (strongest IBCS) using canny method staring from threshold 0.95 and step of .05, for correlation matrix "corr_mtxhel".

It was shown that at threshold of 0.85; the strongest IBCS has occurred, while further smaller threshold will give unwanted edges to the results; we should notice that the strongest IBCS has not the highest value in the matrix, since the regions near the diagonal should have higher values.

If we tried the same iteration for WSCM, "corr_mtxm", we will have the result in Fig. 13; and as it is seen, we needed more iterations to find a small IBCS, which is also detected by incomplete square, further iterations behind this, resulted in the unwanted edges.
7 Conclusion

In this paper, we have studied the spectral cross correlation structure of hyperspectral data cube for ten data samples; spectral cross correlation matrices have been constructed for each data sample; looking at the image view of these matrices; it was noticed that usually spectral correlation matrices can be classified into two categories; strong and weak correlation matrix, identified by matrix mean value. we proposed a new concept of groups of bands correlation, called inter band correlation square; it defines the manner in which, groups of bands are correlated in strong spectral correlation matrices; for almost all strong spectral correlation matrices, there are group of bands correlated to another group of bands far away from each other; this fact is identified by inter band correlation square; canny method for edge detection was used to locate inter band correlation square in the correlation matrix; the process of locating the square is iterative and recommended for strong spectral correlation matrix only.

References:


