ABSTRACT

The paper investigates the efficiency of spectral metrics when used in spectral screening of hyperspectral imagery. Spectral screening is the technique of selecting from the data a subset of representative spectra that can be used in further processing. The subset is formed such that any two spectra in it are dissimilar and, for any spectrum in the original image cube, there is a similar spectrum in the subset. The similarity can be described through various spectral distances and can be controlled by a threshold value. The spectral screening can be improved by associating to each spectrum in the subset a weighing factor proportional to the number of spectra in the original image that are similar to it. Following its generation, the screened subset is used in further computations instead of the full data. The resulting processing mappings are then applied to the data. The investigation has focused on the comparison between Spectral Angle (SA), Spectral Correlation Angle (SCA), Spectral Information Divergence (SID), and spectral gradient angle (SGA) in terms of accuracy of the results and speedup obtained. Spectral screening is performed prior to Principal Component Analysis. The PCA result is next extended to the full data. To quantify the accuracy we rely on unsupervised classification of the resulting processed data. Results from experiments on AVIRIS data show that no significant classification accuracy is recorded while the main processing was done on a subset representing only a very small fraction of the original data size.

INTRODUCTION

In remote sensing, data are usually collected as images (spectral images or spectral bands), with each image corresponding to intervals of wavelengths. Each element from the image (pixel) is associated with a certain area of the scene surveyed and with its spectral response (Lillesand, 2000). A collection of spectral images over several wavelength intervals for the same scene is called a multispectral (in case of few wide spectral bands) or hyperspectral (in case of many narrow spectral bands) image. Usually, hyperspectral sensors cover wavelengths from the visible range (0.4µm-0.7µm) to the middle infrared range (2.4µm) (Richards, 1999). Hyperspectral technology is rapidly evolving increasing the number of sensors, techniques, and areas of applicability (Bannon 2004, Lao 2001). Most of the techniques used in hyperspectral imagery processing are derived from similar applications based on multispectral imagery. When applied to large data sets, these techniques often tend to be less efficient and require significant computational power.

A key issue in hyperspectral imagery processing relies on measuring similarities between the spectra; i.e. the pixel vectors obtained by taking the pixel values of the same location in the data across the spectral bands. The main measures used, such as Spectral Angle and Euclidean Distance, have their origins in vector processing and have proven to be relatively reliable and are used in target detection and classification. In addition, new measures such as Spectral Information Divergence and Spectral Gradient were proposed to better reflect the nature of hyperspectral data. In our research we investigate the efficiency of various spectral distance measures with relation to data compression. Unlike a traditional approach where the spectral distances are used solely as distances in the classification process, we investigate their usefulness with regard to data reduction prior to processing. This technique, called spectral screening can be used as compression tool for the data when computing intensive processing is needed, as it is the case with most techniques.

In this paper, we performed spectral screening prior to Principal Component Analysis (PCA). The PCA result is next extended to the full data. Novel to this paper is the wide range of spectral measures considered and the quantification of the accuracy by performing unsupervised classification (k-means clustering) of the resulting processed data. The goal is to see if spectral screening affects the classification results. In the process we also
suggest improvements on the spectral screening by using weights associated to the spectra, to increase the representatives of the subset.

The paper is organized as follows. The first section provides an overview of the various spectral measures investigated as well as introduction of several modifications to them that would make them fit better the purpose of spectral screening. The next section describes in brief the spectral screening process and introduces the weighting variation. The experimental results presented next use the screening in connection to PCA and the classification and analyze the accuracy with respect to direct processing of the data and without PCA. Conclusions and references conclude the paper.

**Spectral Metrics**

**Spectral Angle (SA)**

Given two \( n \)-dimensional vectors \( \mathbf{x} \) and \( \mathbf{y} \), the spectral angle is defined as (Keshava 2004):

\[
SA(\mathbf{x}, \mathbf{y}) = \arccos \left( \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \right)
\]

where \( \langle \ldots, \rangle \) represents the dot product of the vectors and \( \|\| \) represents the Euclidean norm.

The spectral angle is invariant to scalar multiplication. The measure is used in hyperspectral data processing as a method for mapping the spectral similarity of image spectra to the reference spectrum (Kruse 1993). It was also employed as a distance measure for classification (various classes are viewed as being made of pixel vectors that make small angles) (Robila 2003).

**Spectral Correlation Angle (SC)**

Given two \( n \)-dimensional vectors \( \mathbf{x} \) and \( \mathbf{y} \), the spectral correlation coefficient is defined as (de Carvalho 2000):

\[
c(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y} - \bar{x} \cdot \bar{y}}{\|\mathbf{x} - \bar{x}\| \|\mathbf{y} - \bar{y}\|}
\]

where \( \bar{x} \) and \( \bar{y} \) are the expected values for the two vectors.

The coefficient is a derivative of the Pearsonian correlation coefficient. Compared to the spectral angle, the spectral coefficient has been shown to be able to distinguish between positive and negative correlations and to yield better estimates in some experiments (de Carvalho 2000). The spectral correlation coefficient provides a range of negative and positive values. Alternatively, we explore the design of a spectral correlation angle that provides results within the same range as the spectral angle. The spectral correlation angle can be defined through a formula such as (Robila 2004):

\[
SC(\mathbf{x}, \mathbf{y}) = \arccos \left( \frac{c(\mathbf{x}, \mathbf{y}) + 1}{2} \right).
\]
Spectral Information Divergence (SID)

Given two \( n \)-dimensional vectors \( x \) and \( y \), the Spectral Information Divergence is defined as (Chang 2003):

\[
SID(\, x, y \, ) = \sum_{i=1}^{n} \left( \frac{x_i}{\sum_{j=1}^{n} x_j} - \frac{y_i}{\sum_{j=1}^{n} y_j} \right) \left( \log \frac{x_i}{\sum_{j=1}^{n} x_j} - \log \frac{y_i}{\sum_{j=1}^{n} y_j} \right).
\]  

(4)

SID is derived from the Kullback-Leibler information measure:

\[
SID(\, x, y \, ) = D(x \parallel y) + D(y \parallel x)
\]  

(5)

The spectral information divergence has been shown to provide a relatively better quantification of similarity than the spectral angle. Additionally, combinations of the SA and SID (by using tan and sin functions) have been proposed and shown to increase the accuracy over both measures (Du 2004).

Spectral Gradient Angle (SGA)

Given a \( n \)-dimensional vectors \( x \), the spectral gradient is defined as (Angeloupolu 1999):

\[
SG(\, x \, ) = (x_2 - x_1, x_3 - x_2, x_4 - x_3, \ldots, x_n - x_{n-1})
\]  

(6)

Given two \( n \)-dimensional vectors \( x \) and \( y \), the spectral gradient angle is defined as:

\[
SGA(\, x, y \, ) = SA(abs(SG(x)), abs(SG(y)))
\]  

(7)

where abs(\( x \)) refers to the vector whose components are the absolute values of the components of \( x \).

Note that the spectral gradient angle is also invariant to illumination conditions. Compared with spectral angle, SGA provides the advantage of taking in consideration slope changes within the vector. This approach is partially answering to the need for takes in consideration not only the vector values but also the order or the values, a feature that we believe is important in hyperspectral data.

Euclidean Distance and other distances

Given two \( n \)-dimensional vectors \( x \) and \( y \), the Euclidean distance is given by the formula (Richards 1999):

\[
e(\, x, y \, ) = \sqrt{\langle x - y, x - y \rangle} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.
\]  

(8)

The measure is used in many classification techniques to estimate the distance between the spectrum to be classified and the representatives of the class. Additional distance measures between individual spectra and groups
SPECTRAL SCREENING

Given a distance measure \( d(\cdot, \cdot) \), and a threshold \( \beta \), we say that two spectra \( x \) and \( y \) are similar with respect to \( d \) and \( \beta \) if \( d(x, y) \leq \beta \). The two spectra are dissimilar otherwise. Fig. 1a presents the spectral screening algorithm. Start by considering \( S \) the set of all spectra available in the hyperspectral data. Take one of the spectra and place it in the subset \( S_1 \). Next, each spectrum \( x \) from \( S \) is compared to all the spectra already placed in \( S_1 \). If we find a spectrum \( x_1 \) in \( S_1 \) that is similar with respect to \( d \) and \( \beta \) to \( x \), we place \( x \) in \( S_1 \), otherwise we discard it. At the end of the process, the subset \( S_1 \) will contain spectra with the following two properties (Robila 2003):

\[
\forall x, y \in S_1, d(x, y) > \beta \tag{9}
\]

\[
\forall x \in S, \exists y \in S_1, d(x, y) \leq \beta \tag{10}
\]

The number of selected spectra depends on the distance used and on the threshold value \( \beta \). As \( \beta \) decreases, this number increases, since more and more spectra will be dissimilar from each other. Moreover, the number of selected spectra depends on the order they are processed.

Spectral screening is a time consuming process. It stays attractive as long as there is a need for fast results. In case of real time applications, collecting a screened subset and providing a fast result usually constitutes an acceptable solution even when the accuracy decreases. Spectral screening has been employed as an initial step in extracting endmembers for the linear mixing model (Bowles 2004). In addition, spectral screening using spectral angle has been used as a data reduction tool for real time processing of multispectral images when using PCA (Achalakul 2001). Fig. 2 shows how preprocessing through spectral screening can be envisioned. The spectral...
screened subset is first produced. Next, the subset is used in computing the processing parameters. These parameters are employed to process the entire data to yield the final result. Through these steps, it is hoped that the accuracy of the result has not decreased significantly compared to regular processing. In order to provide a better representation, we are investigating the use of weights. Each of the spectra in the subset will be processed with a weight that is proportional to the number of spectra from the full data that were determined to be similar to it (Fig 1b).

![Diagram](image)

Figure 2: Use of spectral screening for preprocessing hyperspectral data

Previous work has shown that spectral screening introduces significant processing speedup when connected to PCA or Independent Component Analysis (ICA) (Achalakul 2001, Robila 2003, Robila 2004). Moreover, the result’s variances (for PCA) or nongaussianity (for ICA) match closely the ones obtained when processing the full data. At the same time, it was desired to test how these results fare when used for classification or target detection. In the following we present our results for classification.

**EXPERIMENTS**

We tested the algorithms on an AVIRIS data set (MultiSpec©, 2003). The image was collected in June 1992 over the Indian Pine Test Site in NW Indiana (see Fig. 3a). It represents a scene with patches of forest (approximately one third of the image), various crops (approximately 2 thirds of the image), roads (highway and secondary), etc. and is organized as 220 spectral bands of 145x145 pixels (totaling 21025 pixels). In the experiments, from the original bands, 34 bands were discarded (due to sensor malfunctioning, water absorption, and artifacts not related to the scene) leaving 186 bands. Fig. 3b and 3c provide the ground data information available from the same source and displayed using the MultiSpec software (Langrebe 1998, MultiSpec©, 2003). Note that many of the classes correspond to vegetation. However, existent literature indicates that the vegetation areas have little canopy area and display variations within the fields. These factors lead us to believe that overall high accurate classification might not be possible.

The spectral screening experiments were performed in Matlab 6.5 with the classification done in ENVI 4.0. We first processed the data using PCA retaining the first 9 bands (composing 99.81% of the variance). The color composite image in Fig. 4 displays a RGB rendering of the first three obtained bands. Note that most of the features present in the scene are relatively well represented (such as the crops, roads, etc.) We also note that compared with ground data available, large parts of the forest area expose significant variability. In addition, many of the forest areas were marked unclassified in the ground data. We performed k-means clustering (using 16 classes) on both the full unprocessed data (184 bands) and on the first nine PCA bands. When compared with the ground truth the overall classification accuracy was 42% for the full data and 48.5% for the PCA data.

Next we screened the AVIRIS original data using four different distance measures (SA, SCA, SID, and SGA) and three different threshold values. Table 1. presents the threshold values and the resulting number of pixels in the subset for each of the measures. We note that the threshold values were chosen to produce relatively similar subset sizes ranging from one thousandth to close to one percent of the full data. Due to these small sizes, the overhead introduced by the screening is relatively small when compared to the regular processing. PCA processing of the full data took approximately 75 seconds while the screening followed by processing times ranged from 3.5 to
approximately 35 seconds with approximately 1.8 seconds (included in the totals) allotted for applying the resulting PC transform to the full data. We also note that SID takes the longest, most probably due to the repeated application of the logarithm. The data was randomly permuted before each run. Previous results (Robila 2003, Robila 2004) indicate that the spectral screened subset is not unique. However, the subset size has only relatively small variability.

Figure 3. AVIRIS data scene a) color composite image, b) ground data for the scene, c) description of the classes.

The obtained subsets were then processed using PCA. In this case, we have looked at both the regular screening and the weighted screening. Fig. 5 presents the variance for the first five bands obtained using various techniques. In all images, the dotted line represents the variance of the bands obtained with the original PCA. The blue, red and green correspond to values obtained using different threshold (in increasing order respectively). Note that the weighted version consistently outperforms the non-weighted screening by closely matching the original PCA bands. Fig. 6 presents the same results grouped by thresholds for all 8 runs. Blue corresponds to SA, red to SCA, green to SID and magenta to SGA. The dotted lines correspond to the regular screening version while the solid line corresponds to the weighted version. We note a slight decrease of accuracy as the subset size decreases. However, the spectral correlation angle and the spectral information divergence tend to remain accurate (in the weighted version) even when one in thousand pixels is employed.

The first bands of the eight resulting data sets were then used for unsupervised classification using the same k-means clustering approach as for the regular data and the PCA processed data. Fig. 7 presents the results. We note that the classification accuracy stays higher for the weighted versions while the SGA (47.2%) and the SID (45%) tend to provide slightly better results. Compared to the classification originally performed, all the screening
approaches have performed well. However, a weighted approach presents a reduced decrease in accuracy. Overall, the experiments indicate that spectral screened subsets do not significantly affect the classification accuracy. This is important since our experiments have used only few spectra as compared to over 20,000 composing the full image.

CONCLUSIONS

Spectral screening has proven to be a useful tool when speedup in hyperspectral processing is desired. Nevertheless, previous work has focused only on one main measure (spectral angle) and has looked at only how the results fare with regard to some general processing techniques (PCA, and ICA). Our work has presented a comprehensive investigation of various spectral distances and various approaches to screening (weighted, or non-weighted). In addition, we tested the results by assessing the unsupervised classification accuracy. Overall, all measures performed well with no clear differentiation among them being possible. The spectral measures seem to provide results that match closely the processing on full data and may be considered as a substitute for the spectral angle in some applications. In addition, weighting the screened data has clearly improved the accuracy over non-weighting without increasing the overhead. Overall, the techniques we investigated prove to be an efficient tool for data reduction. The results, accurately match the ones obtained after processing the full data at significantly lower computational cost. It is desired an extension of the experiments to other data sets and other techniques. This is a topic that warrants further investigation.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spectral Angle (SA)</strong></td>
<td>threshold value</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>subset size (pixels)</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>execution time (seconds)</td>
<td>3.50</td>
<td>4.19</td>
</tr>
<tr>
<td><strong>Spectral Correlation Angle (SCA)</strong></td>
<td>threshold value</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>subset size (pixels)</td>
<td>24</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>execution time (seconds)</td>
<td>4.65</td>
<td>5.18</td>
</tr>
<tr>
<td><strong>Spectral Information Divergence (SID)</strong></td>
<td>threshold value</td>
<td>0.005</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>subset size (pixels)</td>
<td>22</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>execution time (seconds)</td>
<td>9.71</td>
<td>11.88</td>
</tr>
<tr>
<td><strong>Spectral Gradient Angle (SGA)</strong></td>
<td>threshold value</td>
<td>0.2</td>
<td>0.175</td>
</tr>
<tr>
<td></td>
<td>subset size (pixels)</td>
<td>23</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>execution time (seconds)</td>
<td>4.44</td>
<td>5.10</td>
</tr>
</tbody>
</table>

Table 1: Threshold values used in the experiments. Resulting subset sizes and execution times including subset selection and transform applied to the full data.
Figure 5: Variance of the first five bands a) SA, b) SA with weights, c) SCA, d) SCA with weights, e) SID, f) SID with weights, g) SG, h) SG with weights.
Figure 6: Variance of the first five bands using all eight screening techniques and grouped by threshold values (in increasing order of the threshold) a) first (largest) threshold value – smallest subset sizes, b) second threshold value, c) third (smallest) threshold value – largest subset sizes.

Figure 6: Classification accuracy results on the data when spectral screening is involved. The thresholds are ordered from the largest to the smallest (smallest subset size to the largest subset size)
REFERENCES

Robila S.A. (2004), "Using spectral distances for speedup in hyperspectral image processing", manuscript submitted for publication.