OPEN JOURNAL SISTEMS
ISSN:2237-2202

Available on line at Directory of Open Access Journals

Journal of Hyperspectral Remote Sensing 02 (2015) 045-062

Journal of Hyperspectral Remote Sensing
www.ufpe.br/jhrs

Performance Measure-Based Band Group Selection for Agricultural Multispectral Sensor Design

Matthew A. Lee[1], Derek T. Anderson[1], John E. Ball[1], Daniel Reynolds[2], Lori M. Bruce[3], and Nicolas H. Younan[1]

[1]Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA, 39762
[2]Department of Plant and Soil Science, Mississippi State University, Mississippi State, MS, USA, 39762
[3]Professor and Dean of the Graduate School, Mississippi State University, Mississippi State, MS, USA, 39762

Received 17 September 2015; received in revised from 19 September 2015: accepted 20 September 2015

Abstract: Hyperspectral sensors are unfortunately plagued by relatively high financial cost and they require large amounts of storage and computation. Too often, these factors limit their applicability in precision agriculture and in particular applications requiring real-time signal processing. On the other hand, multispectral sensors are less expensive and require far less computational resources. If the task of the sensor or its platform is well known ahead of time, then it can be beneficial to design a multispectral sensor to accomplish the task as long as the performance is not degraded. In this article, we explore the design of a task specific multispectral sensor based on a new band grouping technique. Band grouping algorithms typically rely on a proximity measure to determine how similar (or dissimilar) the information contained in hyperspectral bands are to each other. Similar bands are then grouped. However, the proximity measure typically does not take into account the interactions of different band groups or how band groups will be used. The theory put forth in this article is unique because it makes global decisions for band groups by utilizing a performance measure to gauge the effectiveness of random partitionings at some given task, such as classification or anomaly detection. The band groups that are most correlated with good performance are then selected. Our technique is compared to the uniform partitioning technique using the Pecan1 data set. The results show that the overall accuracy is better with our technique when the same number of band groups are selected.

Keywords: Hyperspectral, band grouping, dimensionality reduction, herbicides.

Address all correspondence to: Matthew A. Lee, Mississippi State University, Electrical Engineering, 406 Hardy Road, Mississippi State University, United States, 39762; Fax: +1 662-325-2298; E-mail: lee@ece.msstate.edu

1 Introduction

The use of hyperspectral sensors in remote sensing is relatively new, as it has only been since the 1990s when AVRIS (AVRIS) and CASI (CASI) became publicly available. Ideally, hyperspectral sensors operate by gathering large amounts of data about their field of view (FOV) by sub-dividing the part of the electromagnetic spectrum that is often referred to as light into a relatively large number of evenly-spaced, narrow-wavelength bands. Often, there are typically hundreds or thousands of bands collected by a hyperspectral sensor, which typically cover ultraviolet, visible, near-infrared, and sometimes long-wave infrared (thermal). When a hyperspectral image is created, spectral band measurements are collected in the spatial region covered by each pixel. Thus, an extremely large amount of information is collected in a single hyperspectral image. One of the things that makes hyperspectral imaging particularly exciting, relative to past imaging sensors, is the ability to deeply analyze the results even at the pixel level. However, the large amount of information requires large amounts of resources to analyze.

The advantage of hyperspectral imagers is the flexibility derived from the many applications and algorithms that utilize the massive amount of information (or portions of it) to infer things about the objects within the FOV. However, this flexibility comes at a high cost in terms of money to manufacture hyperspectral imagers, digital storage requirements to store the images, and processing power to analyze the images. For well-defined, specific tasks, the flexibility provided by hyperspectral sensors may not be an asset, and therefore, the costs associated with hyperspectral sensing may be an unnecessary burden. In such cases, it would be useful to design a multispectral imager (with much fewer spectral bands and therefore cost) that does not compromise performance. The feasibility of maintaining performance is supported by a study that suggested that hyperspectral bands correlated in spectral frequency tend to have correlated information (Manolakis et al. 2008). In the pursuit of multispectral imager designs, hyperspectral sensors and band grouping strategies can be utilized to identify regions of high correlation in the spectrum, which are likely to be repeating the same information.

Band grouping is often used in classification tasks to collect the bands into smaller and more manageable groups that can be used with multiple classifiers and decision-fusion, or can be summarized to reduce the data dimensionality. When a multi-classifier decision-fusion (MCDF) approach is employed, a common practice is to use all of the bands from the same group as inputs to the same classifier (Kalluri et al. 2010; Prasad and Bruce 2008; West et al. 2009). Thus, each classifier makes a decision based on the information extracted from its associated band group. The decisions of all the classifiers are then combined to make a final decision. A relatively recent variation of the MCDF approach, proposed by Bruce, incorporates game theory principles of the Nash equilibrium and competitive rounds in addition to the mutual entropy metric to create band groups (Bruce 2013). It is useful to note that Bruce’s paper achieved an overall accuracy of 77% on the same Pecan1 (West et al. 2009) data sets used in this paper. When band groups are used to reduce dimensionality, it is typical to employ a function to summarize all the bands in the same group. This function can be a weighted average or be the parameters of a regression analysis (linear, polynomial, exponential, etc.). The papers by Lee et al. (2011) and Ren et al. (2011) are examples that use band grouping to reduce dimensionality. In the paper by Lee et al., the band groups were averaged to summarize the band groups, but in Ren et al., a single representative band was chosen for each band group.

Band groups are often determined using a proximity measure, e.g., a similarity measure or metric, because hyperspectral bands often contain redundant information. The proximity measure can be supervised, which requires training data, or unsupervised, which does not require training data. While there are a great number of possible proximity measures, Ball et al. (2014) presented a list of four commonly used similarity measures in the field of band grouping. These measures include the Bhattacharyya distance (Bhattacharyya 1943), correlation (Papoulis 1965; Su and Du 2014), Kullback-Liebler divergence (Kullback and Liebler 1951), and Jeffries-Matusita distance (Whitsitt and Landgrebe 1977). Additional similarity measures found in the literature include the (in)famous Euclidian Distance (Deza and Deza 2009) (a Lp-norm with p=2) and Mutual Information (Cover 2006; Ren et al. 2011). Additional similarity measures are sometimes obtained by combinations of these measures, such as the aggregation work of Ball et al (2014). The work by Lee et al. is another example of a variation where similarity was computed on spatial features derived from the bands in the hyperspectral image (Lee et al. 2011). There are many studies where combinations of measures have been used. Three examples from the literature that utilize a combination of measures are the product of Bhattacharyya Distance and Correlation and the product of Jeffries-Matusita and average mutual information (Cheriyadat and Bruce 2003; Prasad and Bruce 2008; and Venkataraman and Bruce 2005).

The goal of this research is to develop a new way to identify band groups based on their correlation with performance on a desired task, and subsequently use those groups to reduce dimensionality of hyperspectral data. We use the arithmetic mean in our experiments to summarize the band groups because it permits the band groups to represent relatively broader bands present in common multispectral sensors, which is an attractive feature because of the lower costs associated with multispectral sensors in agricultural applications. In general, our band grouping method can be applied to any problem where performance can be measured. A few examples of performance measures are classification accuracies, false positive rate, area under the receiver operating characteristic curve, processing time, error estimates, etc. Thus, there are many problems that the proposed technique can be adapted for.

This work is novel in that it:

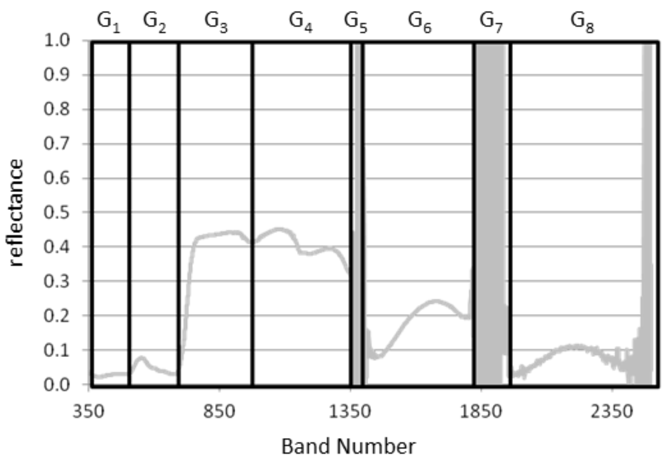
- Bases band selection on the performance at the intended task instead of a similarity or difference metric between the hyperspectral bands, which allows it to make decisions based on the global impact instead of local impact on perfomance.
- Can be applied to band grouping using the MCDF or dimensionality reduction approaches.
- Requires less a priori knowledge of optimal similarity or difference metrics.
- Has the ability to ignore band similarity when it does not lead to improved performance. This could be exceptionally useful in agriculture where the similarity in the bands may be present because all the training data is sampled from plants.

This paper is organized as follows. Section two discusses the algorithm. Section three describes data sets, section four presents the results, and section five discusses the results and conclusions. Finally, section six summarizes our findings and discusses future work.

2 Methodology

At a high level, the proposed band grouping technique strives to divide the spectrum into groups of bands and then it uses a function to summarize the resultant groups. This concept is illustrated in Figure 1. Based on the variations in spectral band choices for proposed solutions to different problems in the entirety of the hyperspectral and signal processing literature, it is almost certainly the case that there is no single global best (i.e., optimal) band grouping solution across all sets of problems. Variation arises due to the very nature of the band groups, which vary in terms of how the spectrum is partitioned and the function computed on the band groups.

When a characterization function (linear, polynomial, exponential, etc.) is used, the parameters of the best fit are used to characterize the band group. Another choice is to pick a function that produces a single scalar value from the bands in the group in order to significantly reduce the dimensionality. A few examples of these types of functions are the arithmetic mean, variance, standard deviation, weighted average (possibly determined using Fisher's linear discriminant) (Fisher 1936), integrals, or another appropriate aggregation operator (such as an ordered weighted average).



Possible Band Groups

Band group n: $\left(G_n, f^*_{G_n}\right)$
where

$$f^*_{G_n} \in \left\{ \begin{array}{l} \left.\begin{array}{c} linear \\ polynomial \\ classifier \\ \vdots \end{array}\right\} Characterization \\ \left.\begin{array}{c} mean \\ integral \\ \vdots \end{array}\right\} Dim.\,Reduction \end{array} \right.$$

Fig. 1 Illustration of the proposed band grouping strategy. The band grouping technique is characterized by parameters for the band groups and the function(s) computed on the band groups.

At the core of the proposed technique is the performance measure, which evaluates the quality of the band groups. The correlation coefficient between potential band groups and performance is then used to guide a search for the most effective set of band groups. The process has four primary steps. First, band groups are created by randomly partitioning the sensor's spectral range. Second, the performance of the partitioning is evaluated. The first two steps are repeated as needed to reduce noise in the correlation coefficients. Third, a correlation coefficient matrix (CCM) is constructed by computing the correlation coefficient between the presence of pairs of bands in the same group and the performance of the random partitioning. Last, the band groups are selected from the CCM(s). These steps are summarized in Figure 2, and details are given in the following subsections.

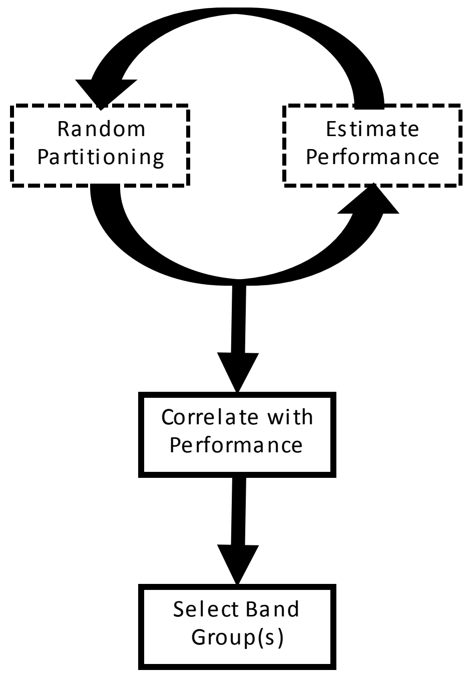


Fig. 2 Flow chart for the proposed algorithm.

2.1 Random Partitioning of the Spectral Range

Random partitionings are determined using Algorithm 1, which separates the spectral range of the sensor into contiguous band groups, and estimates the performance of the random partitionings.

Algorithm 1. Random Partition of Bands
1) for i = 1 to numberOfPartitions
2) partitions[i] = randomPartition(numberOfBands, numberOfGroups)
3) performance[i] = computePerformance(partitions[i],trainingData)
4) end

As Algorithm 1 shows, many random partitions are generated and their performance is evaluated. The number of random partitions is a parameter into the process. A larger number of random partitions results in a better quality CCM (as explained below), but takes longer to compute. The number of random partitions must balance the quality of the CCM and the computation time needed.

The function randomPartition() takes in the number of spectral bands in the hyperspectral data and the number of band groups to generate. The output is a set that indicates what band group each spectral band is assigned. For this procedure to work, it is critical that the random partitions be contiguous. By contiguous, we mean that if any two bands are members of the same group, then all the bands between them are also members of that same group. Let G_n be the n[th] band group, then, G_n is a contiguous band group. For groups with more than one band, the following condition holds:

Let b_i and b_j be any two bands from G_n and let $i < j$,

$$\{b_i, b_j\} \subseteq G_n \Rightarrow \{b_i, b_{i+1}, b_{i+2}, \cdots b_j\} \subseteq G_n. \quad (1)$$

The number of band groups is a parameter to the algorithm, which should be determined before partitioning begins. Preliminary investigations show that the number of band groups does influence the performance of the overall partitioning and the band groups that are eventually selected. Our methodology accounts for this influence by varying the average size of the band groups and develops CCMs for each average size of band groups. This solution does of course increase the computational complexity of the algorithm, but the computational requirements are not prohibitively high for offline processing when parallel, high performance computing solutions are utilized to generate and evaluate the partitioning.

While the use of randomizations could imply the possibility that the algorithm could function differently depending on the random partitions, there is a precedent for using random feature selection to create ensembles of classifiers (Samiappan et al. 2013; Waske et al. 2010). Our work is similar to these works in that the means of the band groups determined by random partitionings are effectively randomly selected features. However, our work differs in that we attempt to identify the features that contribute to improved accuracy instead of fusing the decisions from an ensemble of classifiers. The advantage of our approach is that it produces a simpler classification scheme in exchange for the added upfront cost of identifying the band groups. One concern with our technique is whether the band group selection is repeatable over several runs. Fortunately, our experiments have found that the correlation coefficients that are computed from the random partitions appear to converge as more random partitions are used, which makes the algorithm consistent and repeatable as far as band group decisions are concerned. In addition, there are many other accepted algorithms that exploit random search with heuristics, such as genetic algorithms (Goldberg 1989), simulated annealing (Kirkpatrick et al. 1983), and particle swarm optimization (Kennedy and Eberhart 1995;., Wei et al. 2012). Furthermore, these techniques generally compute the utility of single randomizations, while our technique uses a large number of randomizations to estimate utility via the correlation coefficient.

The performance of the random partitions is computed using computePerfomance(). The function has parameters for a partition set and the training set, which consists of hyperspectral signatures (HS) and their corresponding label $\{HS, labels\}$, and outputs a scalar value indicating how well the partition performed. The scalar output values of computePerformance() are concatenated to from a set. There are several ways that performance of the partitioning can be measured. Ideally, performance should be measured in a way that reflects how the final band groups will be used. Often, it will be most natural to design a test that uses the random groups of the partitioning exactly the same way the final band groups will be used, and use the same performance metric on the random partitioning and final band groups. We use the overall accuracy obtained by a Bayes theoretic classifier trained using maximum likelihood estimation in our experiments. It is important to note that the overall accuracy can produce biased results when the number of samples in each class is grossly imbalanced. We mitigate this limitation in our experiment since we use the same number of samples in each class for training and testing. In problems where the number of samples in each class differs, there are other ways to measure performance, such as user's accuracy (Duda et al. 2006), which are less sensitive to unequal sample numbers.

2.3 Correlating band groups with performance

Our algorithm utilizes the correlation coefficient to estimate the utility of band groups from the random partitioning of the first step. The correlation coefficient is computed as

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}, \quad (2)$$

where X and Y are random variables. In our experiment, the Equation 2 is adapted as follows.

Let

$$\varphi(b_i, b_j) = \begin{cases} 1 \; if \; b_i \epsilon G_k \; and \; b_j \epsilon G_k, k \epsilon \{1, \ldots n\} \\ 0 \; else \end{cases} \quad (3)$$

indicate that bands i and j are in the same band group (Algorithm 2, line 4), where n is the number of bands in the hyperspectral data.

$$\rho_{\varphi(b_i,b_j),\hat{P}} = \frac{cov(\varphi(b_i,b_j),\hat{P})}{\sigma_{\varphi(b_i,b_j)}\sigma_{\hat{P}}} = \frac{E\left[\left(\varphi(b_i,b_j)-\mu_{\varphi(b_i,b_j)}\right)(\hat{P}-\mu_{\hat{P}})\right]}{\sigma_{\varphi(b_i,b_j)}\sigma_{\hat{P}}}, \quad (4)$$

where $\hat{P}$ is the array of performance values returned from computePerformance() in Algorithm 1. It is desirable to have large numbers of random partitionings with each band pair in the same group and in different groups since it allows for more combinations of band groups and generally results in more accurate estimates of the correlation coefficients. In order to compute correlation coefficients, the performance metric should produce a real scalar. The range of the value is not important, but larger values should indicate better (more desirable) results. Greater correlation coefficients indicate that it is useful to have the pair of bands in the same group. Additionally, since random partitioning produce random groups that are contiguous, the pair of bands indicates the beginning and ending bands of a group, and the correlation coefficient identifies the usefulness of that band group.

Algorithm 2. Computing Correlation Coefficients

```
1) for row = 1:numberOfBands
2)     for col = 1:numberOfBands
3)         for n = 1:numRandPartitions
4)             sameGroup[n] = group(row) == group(col)
5)         end
6)         CCM[row,col] = correlationCoefficient(sameGroup,performance)
7)     end
8) End
```

In order to visualize the correlation coefficients for the pairs of bands, it is useful to organize them into a matrix, which can then be displayed as a 2D image. The CCM ($\boldsymbol{M}$) is constructed from the correlation coefficients computed using Equation 4. It has the form

$$\boldsymbol{M} = \begin{bmatrix} \rho_{\varphi(b_1,b_1),\hat{P}} & \rho_{\varphi(b_1,b_2),\hat{P}} & \cdots & \rho_{\varphi(b_1,b_n),\hat{P}} \\ \rho_{\varphi(b_2,b_1),\hat{P}} & \rho_{\varphi(b_2,b_2),\hat{P}} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\varphi(b_n,b_1),\hat{P}} & \cdots & \cdots & \rho_{\varphi(b_n,b_n),\hat{P}} \end{bmatrix}. \quad (5)$$

The rows and columns correspond to the hyperspectral band numbers and each entry is a correlation coefficient for the co-grouping of the band pair and performance. Figure 3 shows an example of a CCM for the Pecan1 image.

The best band groups are identified by local maxima in the CCM. One important distinction of this algorithm from other band grouping algorithms is that the structure of the band groups appears different. Instead of squares and rectangles, which is typically seen, this algorithm produces local optima, and the number of local optima is heavily dependent on the average size of the band groups in the random partitions. Typically larger band group sizes result in fewer local optima. Since random partitionings have contiguous groups, the band group with the correlation coefficient value would stretch from the diagonal of the matrix to the coordinates of the matrix entry. Thus, according to Figure 3, a good band group ranges from band 107 to 128. In Figure 3, there is only one local optima with a positive correlation, which is caused by the rather large average band group size, but Figure 6 shows there are often more band groups at smaller average sizes. The CCM in Figure 3 was computed using tenthousands of random partitions and appears noise free, but if fewer random partitions are used the CCM will appear to have additive Gaussian random noise. Figure 4 shows a plot of the correlation coefficients for bands 107 through 128. The plot supports the selection of a band group covering the range [107, 128] because all partitions are contiguous, and the highest correlation with performance was observed when the band 128 was added to the group. As the plot shows, it may be beneficial to have more hyperspectral bands beyond the 128 physically on the camera since the correlation with performance seems to still be increasing. It is also useful to note that the band group running from [107, 108] has a negative correlation with performance and would be a bad choice for a band group.

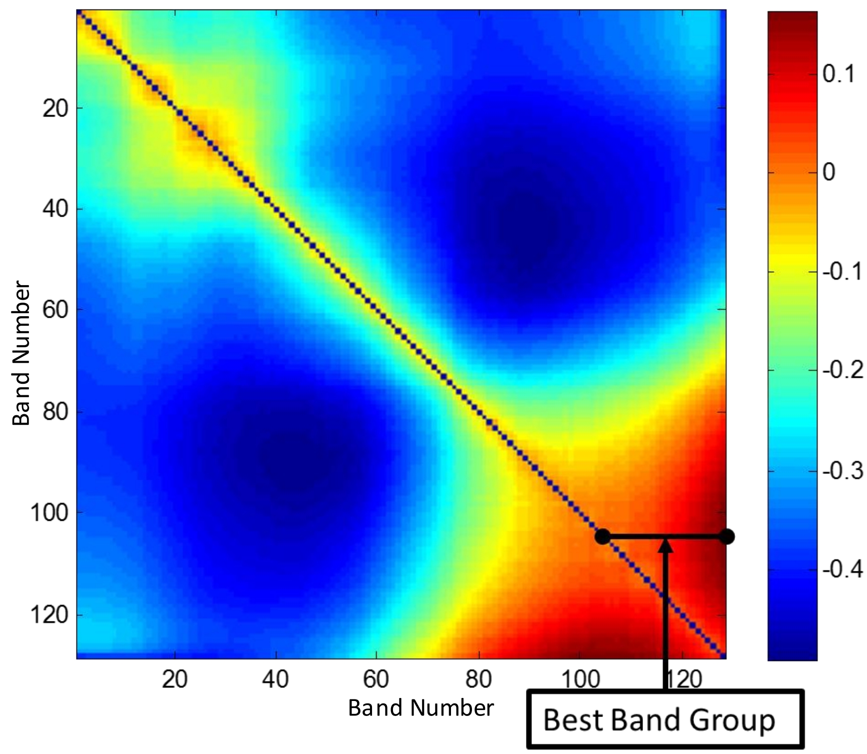


Fig. 3 CCM of Pecan1 with the average cluster size at half the available bands. The axes indicate the band numbers. The best band group runs from the diagonal to the global optimum since the band groups in the random partitions are contiguous.

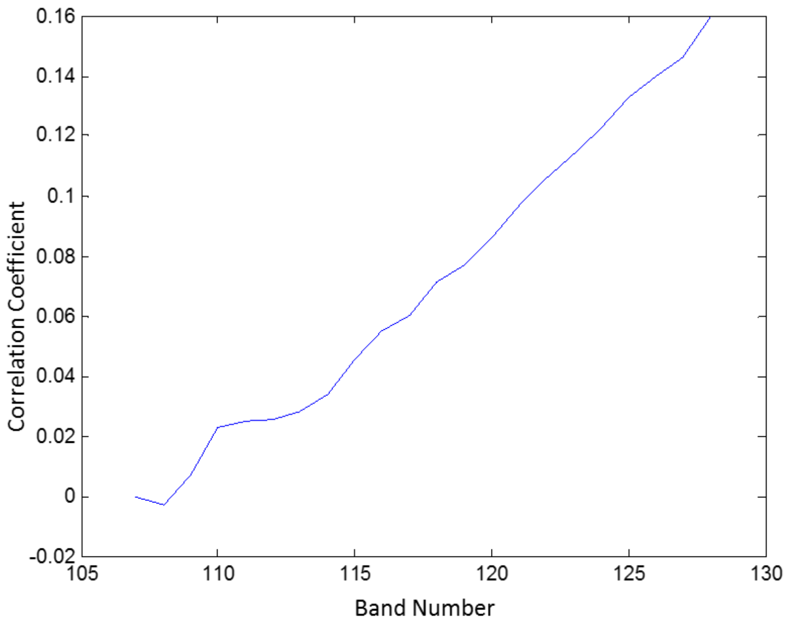


Fig. 4 Plot of correlation coefficients running from band 107 to band 128.

2.4 Selecting band groups based on correlation matrices

Algorithm 3. Selecting Band Groups

```
1)CCM = integrateCCM(CCM)
2)localMaxima[] = findLocalMaxima(CCM)
3)for i = 1:size(localMaxima)
4)  if (localMaxima[i].row < localMaxima[i].col)
5)     if (groupSize(bandGroup) < groupSize(localMaxima[i]))
6)           bandGroup = localMaxima[i]
7)     end
8)  end
9)end
```

An optional step in the process is integrating the CCM, which, in our experiments, produces a matrix that contains less noise. However, integrating the CCM may not be advantageous as the number of random partitions approaches infinity (where a noisy CCM is not an issue), and thus be a superior strategy in general. Since we limited the number of random partitions to 1000, it was advantageous in our experiments to integrate the CCM. The integral matrix ($\boldsymbol{S}$) is computed on the CCM ($\boldsymbol{M}$) using the following formula:

$$\boldsymbol{S}_{r,c} = \begin{cases} \sum_{i=r}^{c}\left(\boldsymbol{M}_{r,i}\right), & r < c \\ \sum_{i=c}^{r}\left(\boldsymbol{M}_{i,c}\right), & c < r \end{cases}. \quad (4)$$

where r is the row number and c is the column number. Figure 4 shows an integrated CCM. In this case, the integrated matrix also points to a band group range running from 107 to 128, but it is not always the case that both the raw CCM and integrated CCM agree with each other. Thus, it is useful to extract band groups using integrated and standard CCMs and choose the one that yields the best performance.

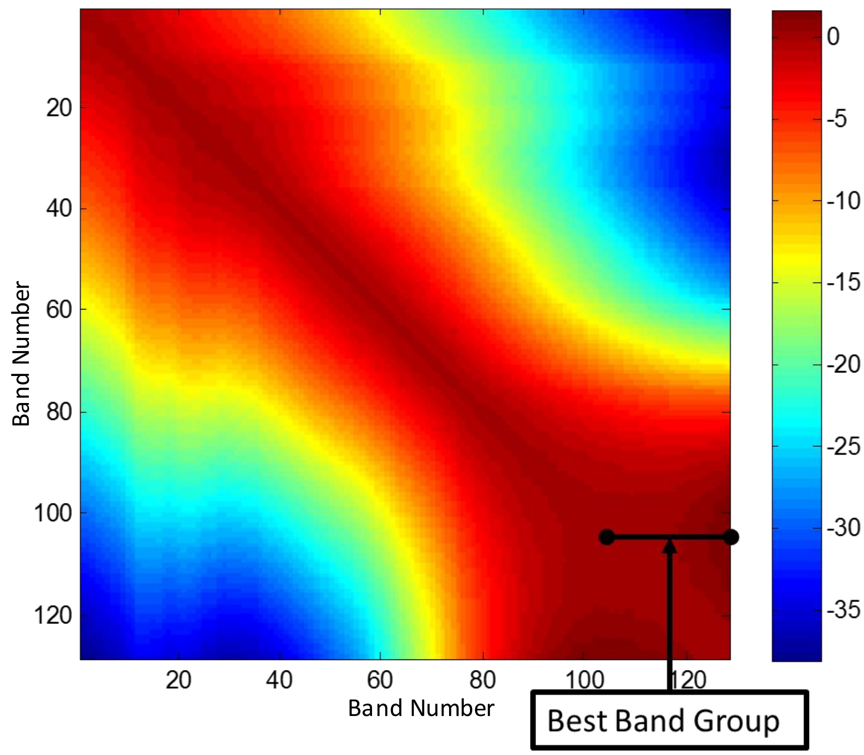
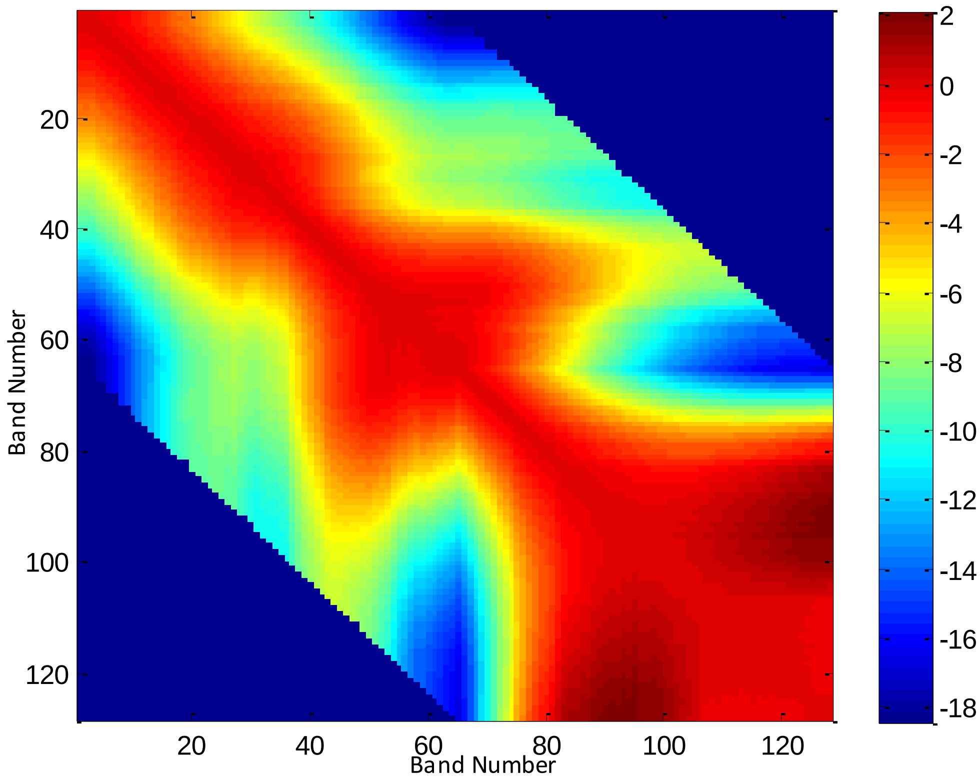


Fig. 5 Integral of the CCM for Pecan1 with the average group size at half of the available bands. The best band group runs from the diagonal to the global optimum since the band groups in the random partitions are contiguous.

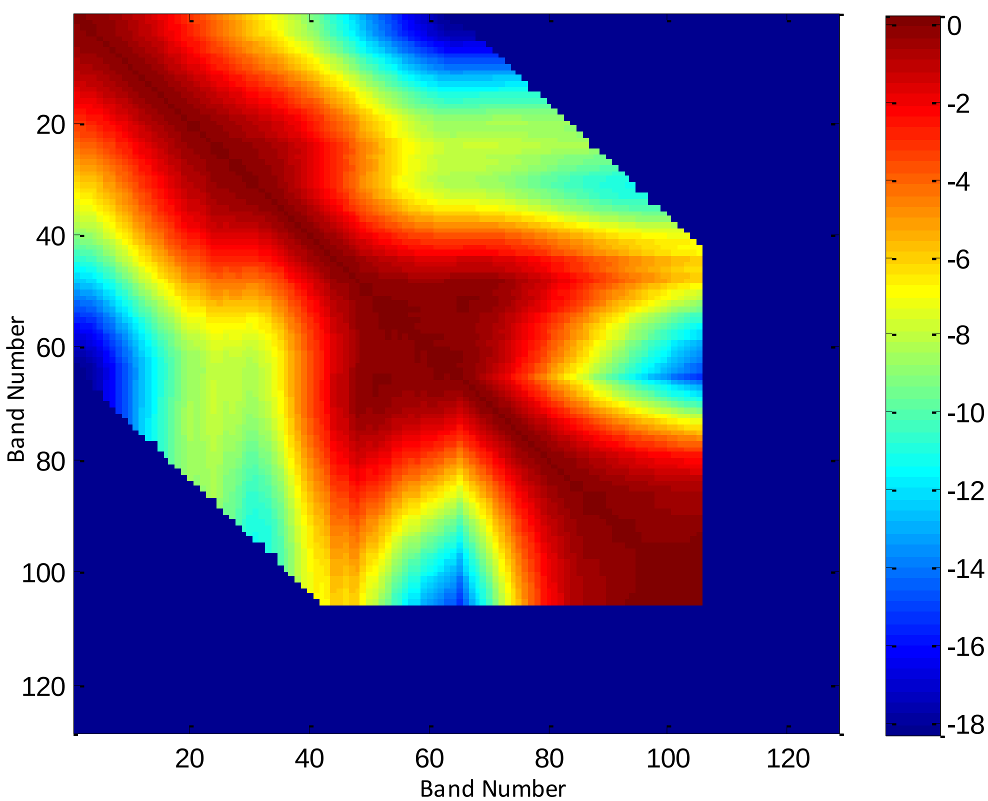
Based on our proposed random partitions approach, two factors need to be considered. First, sensitivity to the average number/size of band groups in the partitioning arises. Second, each time a band group is selected, the value of potential additional band groups changes. Thus, it may not be possible to "compute" the optimal band groups from the CCM or its integral, and instead, the optimal set of band groups needs to be "searched" for. One unfortunate aspect of this problem is that the topography of the search space changes as decisions are made. Thus, a valuable combination of two or more band groups may not be as valuable when an additional band group is added. This is not to say that the combination loses effectiveness, but it may not be as important to the performance of the whole set of band groups, and perhaps another subset of the band groups will be more advantageous once new band groups are added. This situation is similar to that encountered in feature selection, which is a non-polynomial hard problem. However, since the band groups can contain multiple bands each, the number of possible combinations is far greater. Additionally, if parts of the spectrum are not represented in the band groups, there will be even more possible combinations of band groups. Given the massive number of possible combinations, thus complexity of the search space, a heuristic is employed herein to search for good combinations of band groups.

There are a number of possible heuristics that can be used to search for good combinations of band groups. A few examples are genetic algorithms (Goldberg 1989), simulated annealing (Kirkpatrick et al. 1983), particle swarm optimization (Kennedy and Eberhart 1995; Wei et al. 2012), and hill climbing (Russel and Norvig 2003). In this paper, we use the CCMs to guide a hill climbing algorithm. The selection algorithm relies on computing a series of CCMs where the average size of the random partitions is varied. The largest band group identified by each CCM defines the possible choices the hill climbing algorithm can choose. In this experiment, the algorithm always chooses the largest possible band group. This behavior is supported by the observation that selecting band groups defined by the local maxima in the CCM has minimal effect on the correlation coefficient values of other potential band groups relative to each other. Furthermore, choosing the largest group results in fewer band groups, which reduces the computational time required to assign all the hyperspectral bands to a group, and provides for greater dimensionality reduction. This is illustrated in Figure 5, which shows how the selection of the band group with the highest correlation coefficient from Figures 3 and 4 affects the integrated CCM. As indicated in section 3.1 (and apparent in the comparison of Figures 3 and 5-A), the average size and number of band groups affects the correlation coefficient values, and therefore, necessitates the consideration of several CCMs computed with different band group sizes. Choosing the largest groups first is well suited to the constraints we placed on the solution, but there may be a strategy that discovers better solutions that has not yet been considered.

After a band group is chosen, the CCMs are recomputed by analyzing random partitioning where the selected band groups are locked into each partitioning. Finally, a new band group is selected from the CCMs. This process continues until either a predetermined percentage of the hyperspectral bands are placed in groups or a predetermined number of bands groups are selected (user or application specified parameters). Our experiments utilized the prior option with almost all the hyperspectral bands placed in band groups in order to reveal the trend of performance as more band groups were added.


Band Number
Band Number
20
40
60
80
100
120
2
0
-2
-4
-6
-8
-10
-12
-14
-16
-18

(A)


Band Number
Band Number
20
40
60
80
100
120
0
-2
-4
-6
-8
-10
-12
-14
-16
-18

(B)

Fig. 6 Comparison of the integrated correlation coefficients before (A) and after (B) the selection of a band group including bands from 107 to 128. The maximum group size was half of the number of bands available, which resulted in areas in the top-right and bottom-left corners where the correlation coefficient was not a number. Limiting the maximum group size generally reveals more structure, which simplifies visual comparison.

3 Data and Experiments

In order to test our method, we utilized the Pecan1 data set and a single classifier. The data set is from an agricultural scene, and was obtained from an airborne sensor. The classifier used is Bayes classifier, but the method does not imply that one classifier must always be use. For example, one could use a classifier like a support vector machine (SVM) (Melgani and Bruzzone 2004). As well as demonstrating the effectiveness of the technique, these experiments can be used as an example of how it can be implemented to solve a problem.

3.1 Data

The Pecan1 data set was collected by a SpecTIR airborne hyperspectral sensor with 128 spectral bands covering the range of 400-994 nm (SpecTIR). At the altitude the aircraft flew during collection, the spatial resolution of each pixel is approximately 1x1 m^2. The image contains mature corn plants that have been sprayed with one of seven concentrations of glyphosate herbicides (the active ingredient in Round-up herbicide). The spray concentrations controlled by diluting the herbicide to 1 (no dilution), $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}$, and 0 (pure water) of the recommended dosage. Figure 6 shows a color image from the Pecan1 data set.

Pecan1

Fig. 7 Contrast enhanced color image of the Pecan1 data sets. Pecan1 is named after the field it was collected over. It is an approximately one meter spatial resolution image of a corn field that was sprayed by varying concentrations of herbicide, which created the striped pattern.

3.2 Experiments

In our experiments, the function computed on all of the band groups is the mean of the bands in that group. The mean function is chosen because it can be used to approximate the wider bands common on many multispectral imagers, which allows the band groups to reveal potentially optimal bands in a multispectral imager for the task.

A uniform random variable was used to determine the size of the random partitions. The maximum size of the groups was limited to $\left\{\frac{1}{1},\frac{1}{2},\frac{1}{3},\frac{1}{4},\frac{1}{5},\frac{1}{6},\frac{1}{7},\frac{1}{8},\frac{1}{9},\frac{1}{10}\right\}$ of the total number of bands in the image. Thus, the average size of band groups is $\left\{\frac{1}{2},\frac{1}{4},\frac{1}{6},\frac{1}{8},\frac{1}{10},\frac{1}{12},\frac{1}{14},\frac{1}{16},\frac{1}{18},\frac{1}{20}\right\}$ of the spectral bands for the sensor. For each of these band group sizes, 1000 random partitionings are generated. The number of partitioning is determined by examining the CCMs generated in step 3, and it was found that 1000 provided a good compromise between computation time and CCMs with lower noise levels.

The overall accuracy using the Bayes decision theoretical classifier trained using the maximum likelihood estimation procedure was used to evaluate the band groups. This classifier has been selected because it is well established, easy to implement, and it requires relatively little execution time.

The number of band groups determined by the algorithm cannot be easily controlled. Thus, we used brute force (Russel and Norvig 2003) to select the band groups when the number of band groups was reduced for testing and comparison. Brute force involves testing every possible combination of the band groups and then selecting the best combination. This technique is possible because the number of band groups to choose from is relatively small. As the number of band groups increases, this selection technique may become intractable. Fortunately, there are a number of feature selection techniques in the literature that can be adapted for band group selection (Cormen et al. 2009; Goldberg 1989).

The overall accuracy was compared to band groups created by uniformly partitioning the available spectrum. Uniform partitioning was used to generate baseline plots of 1 to 40 band groups both with and without Fisher's linear discriminant analysis (LDA) (Fisher 1936). Fisher's LDA is used to find a projection in which the samples of both classes have the best between-class-variation divided by within-class-variation. The Fisher's LDA is given by

$$\bar{w} \propto (\Sigma_0 + \Sigma_1)^{-1}(\bar{\mu}_0 - \bar{\mu}_1), \quad (5)$$

where Σ_0 and Σ_1 are the covariances of the two classes, and $\bar{\mu}_0$ and $\bar{\mu}_1$ are the means of both classes. In all of the experiments (including the ones using the band grouping algorithm), the number of training and testing samples was limited to 50 randomly selected sample for all seven classes. The random selection insured that the set of training samples and the set of testing samples were mutually exclusive with each other. The number of training samples was chosen to mirror the number of plants that a smaller scale agricultural scientist or engineer can feasibly obtain by growing plants in the field or greenhouse for analysis. At these numbers, Hughes phenomenon is typically evident in hyperspectral data (Hughes 1968). Hughes phenomenon is caused by errors accumulating in the distribution estimates when more dimensions are added. The accumulated error eventually is greater than the benefit gained by adding more dimensions, and causes the classification accuracy to degrade when more dimensions are added. Obviously, more accurate estimates of the distributions can be obtained by using more training data. However, one of the problems with hyperspectral sensors is that they often produce higher dimensionality than the number of samples available can support. Thus, dimensionality reduction techniques, such as Fisher's LDA or this proposed band grouping technique, are useful.

4 Results and Discussion

The overall accuracy versus the number of bands plots for both data sets are shown in Figures 7 and 8. As shown in the Pecan1 results, 6 band groups out-performed band groups chosen using correlation or uniform partition without Fisher's LDA, and performed as well as uniform partition with Fisher's LDA when 17 band groups were used (the point when performance no longer improved as features were added).

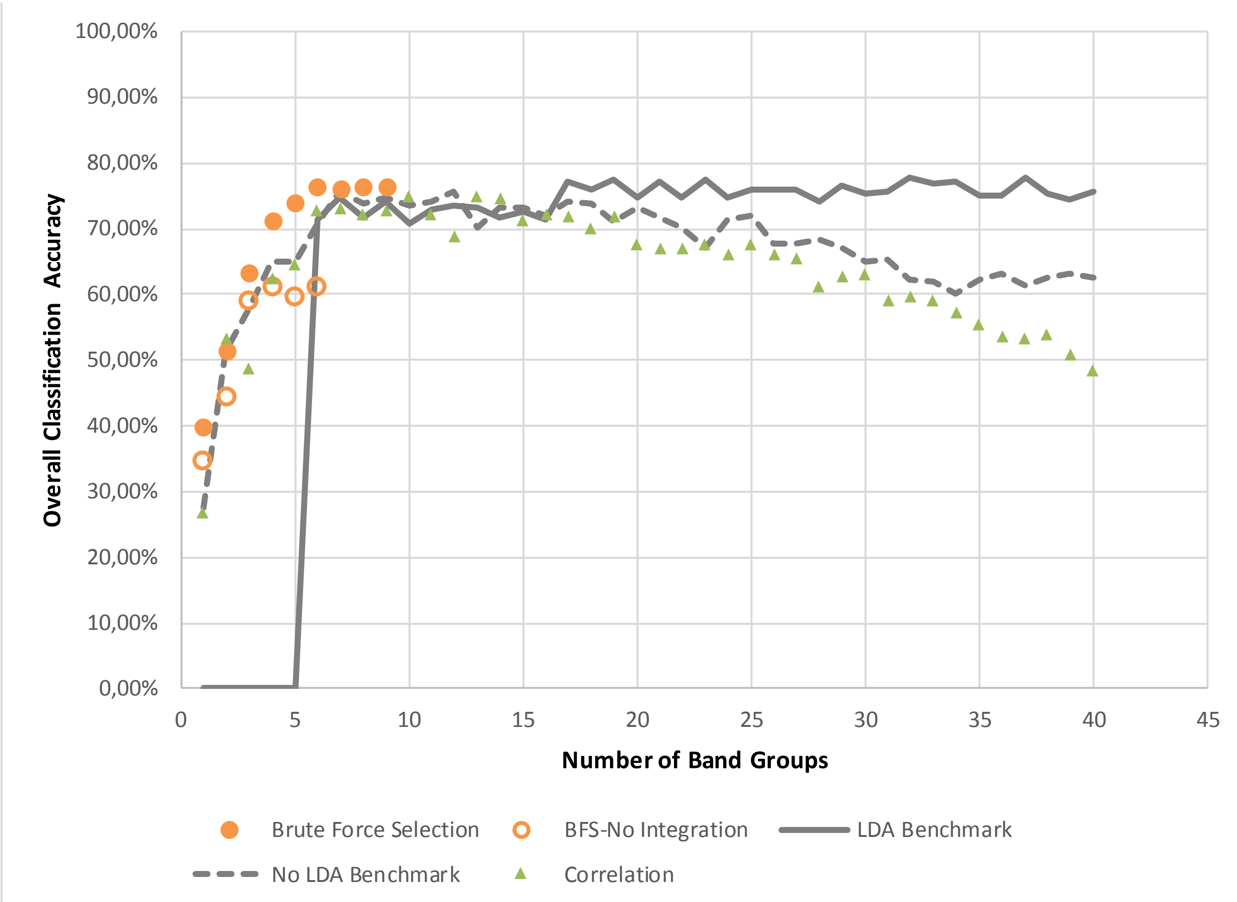
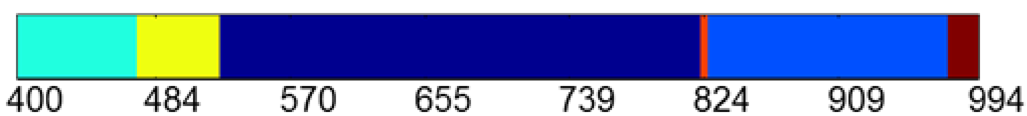
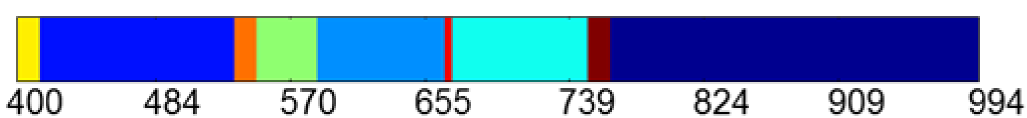


Fig. 8 Overall accuracy vs. number of band groups.



Pecan1 Raw CCM



Pecan1 Integrated CCM

Fig. 9 Band groups selected using the raw and integrated CCMs for Pecan1. The colors indicate band groups, while x-axis indicates the wavelength in nm.

One trend that is evident from the results is that integrating the CCM almost always yielded the best results (see Figure 7). This is likely because not integrating the CCM produces fewer band groups, which was an unexpected result (see Figure 8). Integrating almost always moves the peaks in the CCM further away from the diagonal, which is the reason why it was expected to result in larger and fewer groups. However, the raw CCM is more sensitive to random noise caused by too few random partitionings. Most likely the algorithm chooses larger initial groups with the raw CCM in response to a noise spike in the CCM since it is keying on local optima in the search for the largest group. This can be compensated for by using more random partitioning to estimate the CCM values. Of course, increasing the number of random partitioning requires more computational resources. Each time a group is selected, the noise level decreases since there are fewer possible band group combinations left and the number of random partitioning remains the same. This means the algorithm will eventually start choosing smaller groups, as expected.

The best overall accuracy was achieved when using the integrated CCMs. The fact the results show that the overall accuracy obtained with 6 band group features is not matched until 17 uniform features were used with LDA demonstrates the dimensionality reduction capabilities of band grouping. This is more significant since the band group features do not use LDA for classification, which indicates that the same performance can be achieved with fewer features and a simpler classifier. In addition, the plot for the LDA benchmark plateaus after 17 features. Based on this, we suspect that the six band group features are just as good as using all of the 128 hyperspectral bands as features given the amount of training data used. The plot also shows that the performance of band groups chosen using correlation closely follows those chosen by uniform partitioning.

5 Summary and Future Work

The result of the proposed research is extremely encouraging. Herein, we demonstrated that it is feasible to reduce dimensionality using a band grouping technique, while maintaining performance. Furthermore, our band grouping framework can be utilized in conjunction with most existing advanced signal processing systems and allow for simpler or better solutions to many problems.

In the future, we intend to run several experiments aimed at improving the overall algorithm. These include

- Varying the number of random partitions to determine sensitivity.
- Testing the algorithm on different agricultural data sets.
- Testing the algorithm with different sensors.

The goal of these additional tests would be to show that the algorithm is flexible and generalizes well.

6 Acknowledgement

This effort was partially sponsored by the Engineering Research & Development Center under Cooperative Agreement number W912HZ-15-2-0004. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Engineering Research & Development Center or the U.S. Government. Material presented in this paper is also a product of the CREATE-GV Element of the Computational

Research and Engineering Acquisition Tools and Environments (CREATE) Program sponsored by the U.S. Department of Defense HPC Modernization Program Office. This effort was sponsored under contract number W912HZ-13-C-0037.

References

AVIRIS system specifications. Available: http://aviris.jpl.nasa.gov/

Ball, J. E., Anderson, D. T., & Samiappan, S. (2014). Hyperspectral band selection based on the aggregation of proximity measures for automated target detection. *Proceedings of SPIE 9088, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery*, 1(90), 908814, doi:10.1117/12.2050638

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by probability distributions. *Bulletin of Calcutta Mathematical Society*, 35, 99-109.

Bruce, L. M. (2013). Game theory applied to big data analytics in geosciences and remote sensing. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 4094-4097, doi: 10.1109/IGARSS.2013.6723733

CASI 550 - VNIR Spectrographic Imaging System. Available:http://www.itres.com/products/imagers/casi550/

Cheriyadat, A., & Bruce, L. M. (2003). Decision Level Fusion with Best- Bases for Hyperspectral Classification. *Proceedings of IEEE Geoscience and Remote Sensing Symposium Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, 399 – 406, doi: 10.1109/WARSD.2003.1295221

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms, 3rd Ed*. Boston, MA: Massachusetts Institute of Technology.

Cover, T. (2006). *Elements of Information Theory, 2nd ed*. Hoboken, NJ:Wiley.

Deza, M. M., & Deza, E. (2009). *Encyclopedia of Distances*. Berlin, Germany: Springer.

Duda, R., Hart, P. E., & Stork, D. G. (2006). *Pattern Classification, 2nd ed*. Chicester, U.K.: Wiley.

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2), 179–188, doi:10.1111/j.1469-1809.1936.tb02137.x. hdl:2440/15227

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley Publishing Company, Inc.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1), 55-63, doi: 10.1109/TIT.1968.1054102.

Kalluri, H., Prasad, S., & Bruce, L. M. (2010). Decision Level Fusion of Spectral Reflectance and Derivative Information for Hyperspectral Classification and Target Recognition. *IEEE Transactions on Geoscience and Remote Sensing*, 48(2), 4047-4058, doi: 10.1109/IGARSS.2013.6723733

Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. *Proceedings of IEEE International Conference on Neural Networks*, 4, 1942–1948, doi:10.1109/ICNN.1995.488968

Kirkpatrick, S., Gelatt Jr,, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, 220(4598), 671–680, doi:10.1126/science.220.4598.671

Kullback, S., & Liebler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1), 79-86.

Lee, M. A., Bruce, L. M., & Prasad, S. (2011). Concurrent spatial-spectral band grouping: Providing a spatial context for spectral dimensionality reduction, *Proceedings of IEEE Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1-4, doi: 10.1109/WHISPERS.2011.6080949

Manolakis, D., Lockwood, R., & Cooley, T. (2008). On The Spectral Correlation Structure of Hyperspectral Imaging Data. *Proceedings of IEEE Internatinal Geoscience and Remote Sensing Symposium (IGARSS)*, II-581-II-584.

Melgani, F. & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on Geoscience and Remote Sensing, 42, 1778–1790.

MultiSpec: A Freeware Multispectral Image Data Analysis System, (1994). Available: https://engineering.purdue.edu/~bieh/MultiSpec/hyperspectral.html

Papoulis, A. (1965). *Probability, Random Variables, and Stochastic Processes*. New York: McGraw Hill.

Prasad, S., & Bruce, L. M. (2008). Decision Fusion With Confidence-Based Weight Assignment for Hyperspectral Target Recognition. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 1448-1456, doi: 10.1109/TGRS.2008.916207

Ren, J., Kelman, T., & Marshall, S. (2011). Adaptive clustering of spectral components for band selection in hyperspectral imagery. *Proceedings of Strathclyde's Second Annual Academic Hyperspectral Imaging Conference*, 90-93.

Russel, S. J., & Norvig, P. (2003). *Artificial Intelligence: A Modern Apporach, 2nd Ed.* UpperSaddle River, NJ: Pearson Education, Inc.

Samiappan, S., Prasad, S., & Bruce, L. M. (2013). Non-Uniform Random Feature Selection and Kernel Density Scoring With SVM Based Ensemble Classification for Hyperspectral Image Analysis. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 6(2), 792-800.

SpecTIR Advanced Hyperspectral & Geospatial Solutions. Available: http://www.spectir.com/technology/hyperspectral-imaging/

Su1, H., & Du1, P. (2014). Multiple Classifier Ensembles with Band Clustering for Hyperspectral Image Classification. *European Journal of Remote Sensing*, 47, 217-227.

Venkataraman, S., & Bruce, L. M. (2005). Hyperspectral Dimensionality Reduction via Localized Discriminant Bases. *Proceedings of IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, 2, 1245-1248, doi: 10.1109/IGARSS.2005.1525344

Waske, B., Linden, S., Benediksson, J. A., Rabe, A., & Hostert, P. (2010). Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 2880–2889, 2010.

Wei, W., Du, J., & Younan, N. H. (2012). Fast Supervised Hyperspectral Band Selection Using Graphics Processing Unit. *Journal of Applied Remote Sensing*, 6(2), 061504, doi: 10.1117/1.JRS.6.061504

West, T., Prasad, S., Bruce, L.M., Reynolds, D., & Irby, T. (2009). Rapid Detection of Agricultureal Food Crop Contamination via. Hyperspectral Remote Sensing. *Proceedings of IEEE Internatinal Geoscience and Remote Sensing Symposium (IGARSS)*, IV-889 – IV-892.

Whitsitt, S. J., & Landgrebe, D. A. (1977). *Error estimation and separability measure in feature selection for multiclass pattern recognition*. School Elect. Eng., Purdue Univ., West Lafayette, IN, Tech. Rep. 77-34.