Multiclassifiers and Decision Fusion in the Wavelet Domain for Exploitation of Hyperspectral Data

Terrance West, *Student Member, IEEE*, Saurabh Prasad, *Student Member, IEEE,*
Lori Mann Bruce, *Senior Member, IEE,*

Department of Electrical and Computer Engineering
Mississippi State University
Starkville, MS, USA
trw10@ece.msstate.edu

***Abstract*—In this paper, the discrete wavelet transform (DWT) is employed as a preprocessing stage for a multiclassifier and decision fusion system for feature extraction and dimensionality reduction of hyperspectral data. As a result, both global and local spectral features can be exploited. Feature grouping is conducted according to wavelet decomposition levels, or scales. Each DWT decomposition level's detail coefficients are classified independently, creating a multiclassifer system. The resulting classifications are then fused using a simple majority voting scheme. The proposed target recognition system was applied to hyperspectral data for an agricultural applications, namely detecting the presence of the often devastating disease known as soybean rust in soybean crops. The proposed approach was compared to well-known hyperspectral dimensionality reduction methods, such as stepwise linear discriminant anlaysis (LDA). When using the DWT multiclassifier system, the overall classification accuracies ranged from the high 80's to the mid 90's. When using the stepwise LDA technique the overall classification accuracies ranged from the mid 60s to the mid 90's.
.**

Keywords-hyperspectral; dimensionality reduction; feature extraction; multiclassifiers; decision fusion; discrete wavelet transform

I. INTRODUCTION

Hyperspectral sensors have the ability to produce many spectral features per pixel and have been widely used in many remote sensing target recognition applications. However, with this increase in available features, hyperspectral systems often face the "curse of dimensionality", a problem encountered when the amount of labeled training data is not sufficient to support the number of potential features. A variety of dimensionality reduction and feature extraction methods have been investigated for hyperspectral systems. One of the more recent approaches is a supervised scheme that involves spectral band grouping, multiclassifiers, and decision fusion [1-2]. With this approach, the adjacent spectral bands are intelligently grouped in order to form lower dimensional subspaces. Then the spectral band groups are sent to a bank of classifiers, one classifier for each group. Next, the classifications made by the classifiers are fused using decision fusion to produce one final classification, e.g. target or non-target. The weights used in the decision fusion stage of the system typically takes into account the reliability of each group/classifier combination to accurately classify a pixel.

II. BACKGROUND

A major drawback of the spectral band grouping approach in hyperspectral automated target recognition (ATR) systems is its limited ability to extract large scale, or global, features from the hyperspectral signatures. Typically, the band grouping stage of the ATR system uses one of two approaches: (1) a sliding window of fixed size, or (2) a bottom up approach. For the latter approach, initially each spectral band is considered as a group. Then adjacent groups begin to be merged to form larger groups. Groups are allowed to grow across the spectrum as long as two criteria are met: (i) the merging of groups increases a pre-defined performance metric (e.g. class separation, classification accuracy, etc), and (ii) the group size does not grow larger than the training data can support. The second criterion is important, since the bands in a particular group are still considered as features, and the "curse of dimensionality" needs to be avoided. Since the size of the groups are limited, only small scale, or localized, features are produced. From previous research in wavelet analysis of hyperspectral signatures, we know that large scale, or global features, can be pertinent to particular target recognition problems. For example, the discrete wavelet transform (DWT) has been successfully used for extracting both local and global spectral features in several hyperspectral target recognition systems [3-8]. Bruce *et al.* used the DWT to successfully extract features from hyperspectral signatures [3-4], and interestingly, they found that oftentimes a combination of both small scale and large scale features were optimum. Zhang *et al.* used the DWT and linear discriminant analysis for feature reduction and optimization in hyperspectral soil texture classification [5]. Li *et al.* utilized the Haar DWT as a preprocessing stage to improve linear unmixing of hyperspectral signatures [6], where again a combination of both small scale and large scale features were optimum.

In this study, the authors investigate the use of the discrete wavelet transform (DWT) as a preprocessing stage for a multiclassifier and decision fusion system for hyperspectral data. That is, the spectral band grouping stage will be replaced with a DWT stage. As a result, both global and local spectral features can be utilized. After these features are extracted, they are then reduced based on the number of features the training data can support. Then each set of detail wavelet coefficients at each scale is sent to a classifier, and the classifications are fused to form a single, final output label for the hyperspectral signature. The authors compare the proposed system to existing, hyperspectral analysis methods, such as stepwise linear discriminant analysis (LDA).

III. METHODOLOGIES

A. Wavelet Decomposition

The DWT is a well known feature extraction and mathematical analysis method. The DWT decomposes a signal by projecting it onto dilated (or scaled) and translated versions of a prototype wavelet function known as the mother wavelet [11]. The DWT can be implemented using a dyadic filter tree, as shown in Figure 1 [11]. This implementation is utilized due to its computational efficiency, and is implemented using a two-channel filter bank which are low-pass and high-pass. Each level of the filter tree corresponds to a dyadic scale (2^j) of the wavelet decomposition. At each scale, the wavelet approximation coefficients are produced by the low-pass filter, and the wavelet detail coefficients are produced by the high-pass filter. In this study, the mother wavelet of choice is the Haar wavelet because if its simplicity and because it has been shown to be the optimum choice in other hyperspectral wavelet analyses [3-4,6]. The maximum level of decomposition is dependent upon the number of spectral bands (2151 in our study) and the mother wavelet (Haar). As a result, the wavelet decomposition was restricted to 10 levels in our experimental analysis.

B. Dimensionality Reduction

Each set of wavelet detail coefficients, along with the final set of approximation coefficients, is considered a feature vector, similar to a spectral band group as in [1-2]. The feature vector should next be reduced/optimized. As in the spectral band grouping approach, we use LDA for this dimensionality reduction and feature optimization. However, LDA is a statistical, supervised method that can also face the “curse of dimensionality”. Thus, LDA might not be directly applicable to the lower levels of the wavelet decomposition. This will be the case if the number of small scale, fine detail coefficients is too large (i.e. cannot be supported by the amount of training data). In this study, we use a 3-to-1 rule of thumb to determine which levels of the decomposition could have LDA directly applied. That is, if the number of detail coefficients is one-third the number of training samples (N/3), then we assume that LDA can be accurately applied. However, if the number of detail coefficients is too large (>N/3), then the detail coefficients are grouped into contiguous sets of size N/3, and each set is then reduced/optimized via LDA.

C. Classification and Decision Fusion

Each reduced/optimized feature vector (output of LDA) is independently classified. In this study, we use a maximum likelihood classifier in each case. However, it should be noted that other classifiers could be utilized, such as non-parametric nearest neighbor classifiers, neural networks, etc. The resulting classifications are then fused using a simple majority vote scheme, given by

$$w = \underset{i \in \{1,2\ldots C\}}{\arg\max}\, N(i) \tag{1}$$

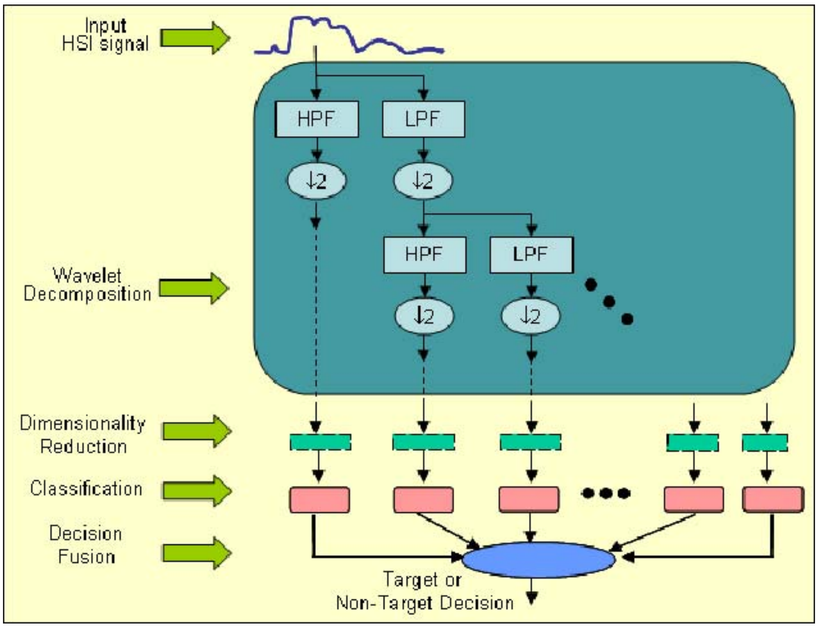

Figure. 1 Proposed hyperspectral target recognition system.

where w is the class label from one of the C possible classes for the test pixel, and $N(i)$ is the number of times class i was detected in the bank of classifiers. The result is a single, final classification for the input hyperspectral signature.

IV. CASE STUDY

The proposed ATR system is applied to experimental hyperspectral data for an agricultural application, namely the early detection of a disease known as soybean rust (*Phakopsora pachyrhizi)* in soybean crops [7]. The ability to rapidly detect soybean rust onset is critical to the US economy, and agencies such as the U.S. Department of Agriculture (USDA) and Department of Homeland Security (DHS) are very interested in this particular application. Soybean rust, which is caused by *Phakopsora pachyrhizi,* is a windborne pathogen which can be transmitted over large areas in a matter of weeks [9]. In 2002/2003, Brazil suffered an estimated loss in soybean crop of 3.4 million tons and a $600 million estimated cost for fungicide sprays. The USDA estimates an economic loss of $640 million to $1.3 billion in the first year of a widespread soybean rust invasion in the United States [9].

The hyperspectral data was collected using the Analytical Spectral Devices (ASD™) Fieldspec Pro handheld spectroradiometer [10]. The ASD has a spectral range of 350 – 2500 nm, spectral resolution of 3 nm @ 700 nm and 10 nm @ 1400/2100 nm, and uses a single 512 element silicon photodiode array for sampling 350 - 1000 nm and two separate, graded index Indium-Gallium-Arsenide photodiodes for the 1000 - 2500 nm range [10]. The signatures in this experiment were collected over a two week period in a green house outside the city of Encarnacion, Paraguay, in 2005 with the humidity at 100% and the temperature kept close to 80 - 85 F.

For this study, 678 samples were used for evaluation, 320 samples of the control soybean and 358 samples of the inoculated soybean. For each class, 160 signatures were used to train the system, while the remaining 160 samples of the control soybean and 198 samples of the inoculated soybean were used for testing the accuracy of the system.

V. RESULTS AND DISCUSSIONS

Figure 2 shows the overall accuracies of the DWT multiclassifier and decision fusion approach and the stepwise LDA approach over a 4 date period and the results for combining all of the dates. It is clear from Figure 2 that the DWT target recognition system outperformed the stepwise LDA method in every category. The global and local features provide by the DWT technique seemed to have a stronger impact on classification when data was organized by date. Note that on date 3 (only 3 days after the plants have been inoculated with the disease), the DWT and stepwise LDA approaches result in an overall accuracy of approximately 92% and 68%, respectively. The proposed DWT approach represents a significant improvement in ability to discriminate between healthy soybean crops and those inoculated with soybean rust, particularly in the critical early stages of the disease. When all dates where combined, both stepwise LDA and the DWT technique could not accurately determine whether the hyperspectral signature represented a healthy soybean crop or one inoculated with the disease. The overall accuracy for the separated dates for the DWT technique range in the high 80's to the mid 90's. The overall accuracies for the separated dates for the stepwise LDA method ranges from the high 60's to the mid 90's. Note that the DWT technique performed its best on the separate dates than on the combined dates. Also, one can notice that the classification accuracies per date are larger for the last two dates. This result is not surprising because one would expect the soybean pathogen to have a more significant impact on the spectral reflectance of the plant as time progressed.

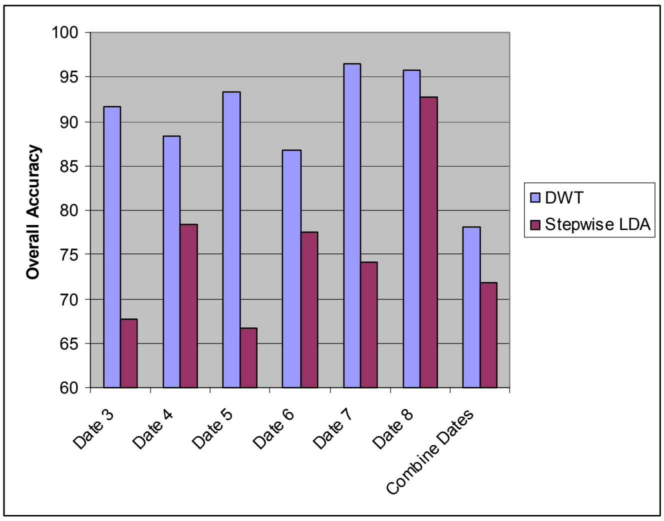

Figure 2 Overall classification accuracies for the proposed DWT approach versus stepwise LDA method for two-class problem of soybean with and without inoculation of soybean rust pathogen.

VI. CONCLUSION

Effectively exploiting hyperspectral signatures is a challenging task, particularly when only a limited amount of training data is available. Dimensionality reduction and feature extraction play a critical role. In this paper, the authors propose a hyperspectral ATR system that utilizes the multiresolution analysis (DWT) combined with multiclassifiers (bank of maximum likelihood classifiers) and decision fusion (simple majority vote). The DWT allows for the exploitation of both global and local spectral features. The multiclassifier and decision fusion approach enables the utilization of statistical methods like LDA for feature optimization and maximum likelihood for classification when only small numbers of training data are available. The proposed ATR system was applied to experimental hyperspectral data for an agricultural application, namely discriminating between healthy soybean crops and those containing the soybean rust pathogen. When compared to traditional methods like stepwise LDA, the proposed DWT approach consistently resulted in significantly higher accuracies, sometimes increasing the overall accuracy by as much as 20%.

Acknowledgments

The authors would like to acknowledge the Department of Homeland Security (DHS) research funding. The authors would like to thank Wade Givens and Lewis Wasson for hyperspectral data collection.

References

[1] S. Venkataraman, L.M. Bruce, "Hyperspectral Dimensionality Reduction via Localized Discriminant Bases," *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Seoul, Korea, July 25-29, 2005.

[2] A. Cheriyadat, L. M. Bruce, "Decision Level Fusion with Best- Bases for Hyperspectral Classification", *Proc. IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, October 2003.

[3] L.M. Bruce, C. Morgan, S. Larsen, "Automated detection of subpixel targets with continuous and discrete wavelet transforms," *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2217-2226, 2001.

[4] L.M. Bruce, C.H. Koger, J. Li, "Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction," *IEEE Trans. Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331- 2338, 2002.

[5] X. Zhang, N. H. Younan, C.G. O'Hara, "Wavelet domain statistical hyperspectral soil texture classification," *IEEE Trans. Geoscience and Remote Sensing*, vol 43, no. 3, pp. 615-618, March 2005.

[6] J. Li, L. M. Bruce, "Improving the Accuracy of Linear Pixel Unmixing via Appropriate Endmember Dimensionality Reduction", *Proc. IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, October 2003.

[7] R. L. King, "A Comprehensive Threat Management Framework for a Crop Biosecurity National Architecture," available online at http://www.lars.purdue.edu/seminar/

[8] S. Kaewpijit, J. L. Moigne, T. El-Ghazawi,"Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Anaylsis," *IEEE Trans. Geoscience and Remote Sensing,* vol. 41, no. 4, pp. 863 – 871.

[9] M. Livingston, R. Johansson, S. Daberkow, M. Roberts, M. Ash, and V. Breneman,"Economic and Policy Implications of Wind-Borne Entry of Asian Soybean Rust into the United States," available online at http://usda.mannlib.cornell.edu/usda/ers/OCS/2000s/2004/OCS-04-27-2004_Special_Report.pdf

[10] Analytical Spectral Devices Fieldspec Pro Spectroradiometer specifications. Available http://asdi.com/products_specifications-FSP.asp

[11] S. Burrus, R. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer,* Englewood Cliffs, NJ: Prentice-Hall, 1998.