NASA-CR-201804

IN-43-CR
10 CIT.

SECOND ANNUAL STATUS REPORT

NASA GRANT NAGW-3924

Research on Methods for Analysis of Multispectral Earth Observational Data

Principal Investigator

David A. Landgrebe
School of Electrical & Computer Engineering
Purdue University
West Lafayette, Indiana 47907-1285
Phone 317-494-3486
FAX 317-494-3358
Internet: landgreb@ecn.purdue.edu

Cognizant NASA Technical Officer

Dr. D. E. Wickland
Code YS, NASA Headquarters

Year 2 of 3

For the research period February 1, 1995 to January 31, 1996

Summary

The work pursued under this grant has as its goal the definition of means for effectively analyzing multispectral and hyperspectral image data based upon principles of signal theory and modern signal processing technology in addition to physical and biophysical ones. This work is a logical extension of work done under NASA Grant NAGW-925.

In our studies of methods for the analysis of multispectral and hyperspectral data over the last several years, the following not widely perceived fundamental concepts about the analysis of such data have become increasingly both apparent and pivotal.

- *As the number of features to be used in discriminating between ground cover types grows beyond three, the second order variations in the spectral response of a cover class becomes ever increasingly dominant in importance over first order effects.* That is, the shapes of the class distributions in feature space become increasingly more significant relative to discrimination between classes than is their location in feature space. This may appear at first to be counter-intuitive. For example, two distributions may have the same mean value, an yet be separable. However, upon examination, there do appear to be both geometrical and biophysical explanations for this observation. It is an extension of the fact that a single point in feature space, i.e., a single spectral response curve is not an adequately complete quantitative representation of a ground cover class.

- *Related to this, as the number of features to be used in any given analysis is increased, the importance of increasingly precise quantitative characterization of the classes of ground cover of interest grows as well.* This property has been known to signal processing specialists for some time, but has not been commonly recognized by Earth scientists. It points to the importance of accumulating as many design samples per class as possible, while at the same time reducing the number of features to be used to the smallest set which can provide the desired accuracy.

- *The immediate goal for the discriminator design process for any given case is to establish an adequately precise quantitatively defined list of classes which are simultaneously exhaustive, separable, and of informational value to the analyst.* The implication here is that, though one is frequently interested in only one or a small number of classes in the scene, for the highest performance, all ground scene classes, i.e., the distribution of the entire data set in feature space, must be adequately modeled quantitatively.

Previous work has focused upon devising means for optimal feature extraction [7][1] to provide the smallest, yet most effective feature sets, upon improving the precision of the individual class models for a given sized design set [2, 9], and the degree to which the (presumably exhaustive) set of class models accurately model the entire data set. During this reporting period, the research effort has been focused upon Projection

1 Numbers in [ ] refer to papers and presentations listed below.

Pursuit [3,6,10] as a means for further improving the extraction of case-specific features, upon testing of previously reported feature extraction techniques [8], and upon means for improving the generalization capability of classifiers. Further initial work was undertaken toward simplifying the process for labeling of training samples [8]. A method for decision tree classifier design was also examined as a means for focusing the feature design process on pairs of features rather than attempting optimality over the entire list of classes.

Remaining major obstacles to a well designed analysis process for use by Earth scientists and applications specialists include a) devising a clustering algorithm which is sensitive to second order class variations, b) improving the generalization capabilities of classifiers, c) means for simplifying the labeling of design samples, and d) integrating the algorithms which form the analysis process into a straightforward and easy to use process that non-signal processing specialists will find attractive.

MultiSpec Distribution

In addition to the research itself, the intent is to deliver the knowledge gained and algorithms created in this research to the Earth science (user) community by as rapid and effective means as possible. Thus an integral part of the work is to provide to this user community a capability to learn and evaluate the algorithms defined via an easy to acquire and easy to use software package. This software package is called MultiSpec©, and it has been under development in this mode and for this purpose for the last several years. The package has been copyrighted to protect against misuse, but it is freely distributed to requesters by the least expensive and most convenient means available. User interest has been significant. In previous years, several hundred requests for copies were received, and copies were mailed to these requesters. A summary of registered users up to May 31, 1994 is available at:

http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/Final_Report_table1.html

Although a more current summary would be much larger, this list reasonably illustrates the types of organizational units which are downloading copies during the current reporting period.

An unexpected past outcome was that a substantial number of requests for MultiSpec came from teachers of the K-12 level. As a result of this type of interest, a distribution license has been granted to The Consortium for Mathematics and Its Applications (COMAP), a National Science Foundation funded program based in Lexington, Massachusetts, for use in their ARISE (Applications/Reform in Secondary Education) program. This is a 5-year project to generate new mathematics curricula for grades 9-11. They report having incorporated MultiSpec into their new curricula as a motivating tool. Many requests for MultiSpec have also come directly from secondary school teachers.

During the current reporting period, new versions were issued periodically with substantial new capabilities. A Windows 3.1/Windows 95/Windows NT compatible version of MultiSpec is also in process of being prepared. This was motivated by the NASA/NOAA/NSF GLOBE program, a program motivated by Vice President Gore to have K-12 children involved in collection of environmental data. More than 2800 schools across the U.S. and 30 foreign countries have signed up to participate.

As a result of the growing demand for copies of MultiSpec, rather than distribute copies by mail, a MultiSpec web site has been established in such a fashion that the several versions of the application program, its documentation, and examples of its use, including the data used in the examples, can be directly downloaded by anyone having access to the WWW. The URL involved is:

http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/

Persons downloading copies are requested (but not required) to register their copies so that we know who has them. From February 14, 1995 when the web site went on line through the end of this reporting period, 586 copies were registered in this fashion. A survey of the requesters indicates that approximately 40% were associated with GLOBE, ARISE, and other K-12 educational institutions. The other 60% were primarily from the remote sensing research, educational, and application community in the U.S and around the world, although a small number went into hospitals and other health care organizations.

Vegetation and Soils Field Research Data Summary

On a related note, we receive requests periodically for data from our Vegetation and Soils Field Research Data Base developed in the mid '70's to early '90s. The requesters are looking for spectral reflectance measurements of vegetation and soil canopies under contolled conditions that are well documented. We have made a significant portion of this data available now to anyone having access to the WWW. The URL is:

http://dynamo.ecn.purdue.edu/~frdata/FRData/

The web pages contains summaries of the experiments for which spectral and agronomic data were collected. The data files for many of the experiments can be downloaded along with the documentation of the data formats.

Publications

The following papers and presentations were given during the period of this report and on the subject of this research.

[1] David A. Landgrebe, "On the Analysis of Multispectral and Hyperspectral Data," SPIE Imaging Spectrometry at AeroSense '95, Orlando, Florida, April 17-21, 1995.

[2] Joseph Hoffbeck and David Landgrebe, "Covariance Estimation For Classifying High Dimensional Data," Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'95), Florence Italy, July 10-14, 1995

[3] Luis Jimenez and David Landgrebe, "Projection Pursuit For High Dimensional Feature Reduction: Parallel And Sequential Approaches," Presented at the International Geoscience and Remote Sensing Symposium (IGARSS'95), Florence Italy, July 10-14, 1995.

[4] David Landgrebe, "Multispectral Data Analysis: An Overview - Past, Present, & Future, Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'95), Florence Italy, July 10-14, 1995.

[5] David Landgrebe, "Land Satellite Information Extraction: Past, Present, And Future, Proceedings of the American Society of Photogrammetry and Remote Sensing Symposium, "Landsat Satellite Information in the Next Decade," Vienna, VA, September 25-28, 1995.

[6] Luis Jimenez and David Landgrebe, "Projection Pursuit in High Dimensional Data Reduction: Initial Conditions, Feature Selection and the Assumption of Normality," IEEE International Conference on Systems, Man, and Cybernetics, Vancouver, Canada, October 22-25, 1995.

The following papers have been submitted with publication pending as indicated.

[7] Chulhee Lee and David A. Landgrebe, "Decision Boundary Feature Extraction for Neural Networks," *IEEE Transactions on Neural Networks*, submitted March 5, 1993. Accepted for a special issue May 11, 1996.

[8] Hoffbeck, Joseph P. and David A. Landgrebe, "Classification of Remote Sensing Images having High Spectral Resolution," *Remote Sensing of Environment*, submitted August 4, 1994, accepted for publication April 11, 1995.

[9] Hoffbeck, Joseph P. and David A. Landgrebe, "Covariance Matrix Estimation and Classification with Limited Training Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, submitted December 19, 1994, accepted for publication April 26, 1995.

[10] Jimenez, Luis, and David Landgrebe, "Supervised Classification in High Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data," *IEEE Transactions on System, Man, and Cybernetics*, submitted December 5, 1995.