SEMIANNUAL RESEARCH PROGRESS SUMMARY

NASA GRANT NAGW-925

EARTH OBSERVATIONAL RESEARCH USING MULTISTAGE EOS-LIKE DATA

Principal Investigators

C. J. Johannsen
LARS/ Entomology Hall
Purdue University
West Lafayette, Indiana 47907
Phone 317-494-6305
Internet: johannsn@ecn.purdue.edu
NASAMail: CJohannsen

D.A. Landgrebe
School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907
Phone 317-494-3486
Internet: landgreb@ecn.purdue.edu
NASAMail: DLandgebe

For the research period October 1, 1989 to April 1, 1990

INTRODUCTION

Work under this grant is directed at preparation for the EOS era and, in particular, HIRIS and MODIS. The objectives of the work are (a) to advance the fundamental understanding of the manner in which soils and vegetative materials reflect high spectral resolution optical wavelengths and (b) to prepare suitable means for analyzing data from high spectral resolutions sensors such as HIRIS and MODIS. The work thus involves basic Earth science research and information system technique understanding and development in a mutually supportive way.

This summary is divided into two parts, an outline of *Earth Science Studies* work, following by that of *Information Systems Studies.*

EARTH SCIENCE STUDIES

During this report period, organic matter from four Indiana agricultural soils was extracted, fractionated, and purified, and six individual components of each soil were isolated and prepared for spectral analysis. The four soils, ranging in organic carbon content from 0.99% to 1.72%, represent various combinations of genetic parameters such as parent material, age, drainage and native vegetation. An experimental procedure was developed to measure reflectance of very small soil and organic component samples in the laboratory, simulating the spectral coverage and resolution of the HIRIS sensor. Reflectance data in 210 spectral bands were analyzed statistically to determine the regions of the reflective spectrum which provided useful information about soil organic matter content and composition.

Wavebands providing significant information about organic carbon content were located in all three major regions of the reflective spectrum: visible, near infrared and middle infrared. Although reflectance in the visible bands had the highest correlation ($r=-0.991$ or better) with organic carbon content among the soils with the same parent material and age, these bands also responded significantly to iron and manganese oxide content. For soils formed on different parent materials, five middle infrared bands (1955-1965 nm, 2215 nm, 2265 nm, 2285-2295 nm, 2315-2495 nm) gave the best correlation with organic carbon content ($r=-0.964$ or better).

Another study was directed at understanding the relationship of lower spatial resolution data with that at high spatial resolution such as that of HIRIS. Advanced Very High Resolution Radiometer Data (AVHRR) weather satellite images of Indiana taken 1987 and 1988 were processed to display different categories of vegetation stress. It was noted that stressed vegetation has different reflectances in all channels from healthy vegetation. Images were obtained on a 7-10 day interval so that differences as a function of the growing season could be evaluated.

This work was done in cooperation with the National Oceanic and Atmospheric Administration (NOAA) and EROS Data Center, in addition to researchers at the Agricultural Research Service Remote Sensing Unit at Beltsville, Maryland. In addition to mapping the extent of the drought areas, an Indiana soil association map was digitized with the locations of the 121 weather stations marked, so that this information could be added to the AVHRR images within a geographic information system. Future plans call for the development of techniques for separating pixels with predominantly row crops, forages and timber, testing models for predicting stress by vegetation categories and determining if one can correlate yield data with spectral responses.

INFORMATION SYSTEMS STUDIES

During past $3\frac{1}{2}$ years of this work, tasks have been pursued with the following specific objectives:

- **Feature Design.** Create a design procedure that would allow one to calculate what the best mission-specific spectral features would be for discriminating between a given set of Earth surface materials, given the location, time of season, and raw high resolution spectral samples to be available from a given sensor.
- **Analysis Algorithm Design.** Determine a set of analysis algorithms which are well matched to the EOS era high dimensional data to be available and a list of classes presumably larger in number and more detailed in character than have traditionally been possible to use.
- **System Simulation.** Create a capability to simulate an entire remote sensing system, including the ground scene, atmosphere, sensor system, and analysis procedure, so that it is possible to study the interrelated effects of various system parameter settings and noise sources across the entire system, including the functioning of the algorithms produced by the above research efforts. Here the definition of noise is taken to be any deleterious effect that occurs in such systems. The motivation for this study, which is more basic in character, stems from the fact that as the information to be derived from such systems becomes more detailed, the interrelated effects between various system parameter selections and degrading influences within such systems will need to become more fully understood if the full potential of such systems is to be realized. The simulator should also be useful for simulating data sets and analysis situations which are not yet available, but which will be in the future.
- **Analysis System Implementation.** Create a data analysis system implementation which has the power and flexibility needed in the future research environment, but which is economical to acquire and use and has greater emphasis on ease of use than has been the case in past implementations.

We shall briefly review the progress made to date in each.

Feature Design. Work to define a new feature design approach was initiated using the previous work of Wiersma[1] as a point of departure. Wiersma's basic concept was to view the spectral samples of a pixel as samples from the continuous function of spectral response, $f(\lambda)$, for that pixel. He provided a means to analytically determine a new signal space based upon *optimal* (but what were usually impractically complex) basis functions such that the ensemble of such spectral functions are represented with the minimum number of terms. During previous years' efforts under this grant we have defined and begun the evaluation of an extension of Wiersma's basic technique to analytically determine *practical* spectral features whose performance very closely approaches that of the optimal analytical spectral features, when probability of correct classification is used as the index of performance. The practical effect of this technique is that the new feature set results in very much lower dimensionality in the analysis process than that of the original raw data, and with no loss of information for the given classification task. Three papers [8,9,10][2] which more fully document this work have been prepared on this work. A paper describing a first test of this procedure in an Earth science context has also been prepared [11].

Analysis Algorithm Design. During the previous years' effort with regard to the data analysis process, a re-codification of an appropriate approach to analysis for this type of data has been defined [12]. The approach takes advantage of both spectral and spatial variations present in a given scene class in order to make the needed association with a defined class. This approach is also based upon a feature design approach with particular attention to high data rate, complex data, and would have the effect of requiring a much reduced dimensionality in the analysis process, and thus much less computation to carry out the classification.

We have also initiated work toward adapting the decision tree classifier (DTC) approach to the EOS era analysis process. DTC's have been under study for problems of a related nature for some time. There has previously been some work done on using DTC's for multispectral work[3,4], however none since the possibility of high spectral dimensionality has arisen. Utilizing funding from a related grant from another agency, a thorough review of the literature has been carried out, and a survey paper based on the literature review is in preparation[5]. The reformulation of a design technique compatible with the high data dimensionality, complex class structure environment of future systems has been begun as a part of this grant and and several papers on the results have been prepared [13,14,15,16].

System Simulation. It has become apparent that there may be substantial value in being able to study the integrated effects of the components of the entire remote

---

1 Wiersma, D.J., and D.A. Landgrebe, "Analytical Design of Multispectral Sensors," *IEEE Transactions on Geoscience and Remote Sensing,* Vol. GE-18, No. 2, pp. 180-189, April 1980.

2 Numbers in square brackets [ ] refer to references in the section of this proposal labeled Bibliography of Previous and Current Results.

3 C. Wu, D. Landgrebe, and P. Swain, "The Decision Tree Approach to Classification," PhD Thesis, Purdue University School of Electrical Engineering Report EE 75-17, 1975.

4 P. Swain and H. Hauska, " The Decision Classifier Design and Potential," IEEE Trans. Geoscience Electronics, Vol. GE-15, pp 142-147, 1977.

5 S. Rasoul Safavian and David Landgrebe, "A Survey of Decision Tree Classifier Methodology," (in preparation).

sensing information chain, and to be able to do so before such systems have been built and placed into operation. To study a system of this degree of complexity in a purely analytic fashion has been seen to be difficult. A simulation capability has thus be chosen as a more tractable approach at this time. As a result, jointly with the work of another funded grant, such a simulator has been constructed. The organizational nature for this system is shown in Figure 2.

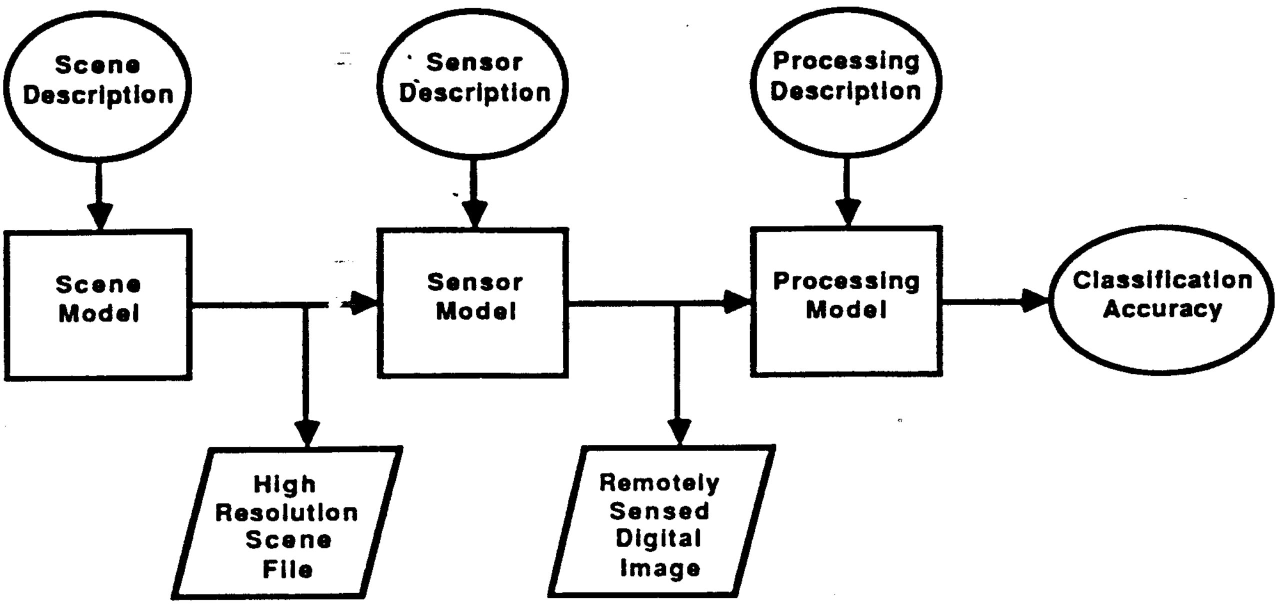

Figure 2. A Remote Sensing Information System Simulator

Several papers have been prepared to document the progress made in this area. [17,18,19,20] It is planned to use this simulator to evaluate various processing schemes and algorithms to determine the system's performance and rigor in various conditions, evaluations which would be expensive or impossible to carry out with real systems.

Other Approaches to Classifier Design. In addition to hierarchical schemes, as previously mentioned, other approaches to the analysis of high dimensional data have been explored to determine their potential. Several papers document this work [21,22,23,24,25,26]

Analysis System Implementation. Work is now nearing completion implementing a first draft of the LARSYS software system on a Macintosh II system. The original LARSYS was one of the first remote sensing multispectral data processing systems, originally created during the 1960's and resident upon an IBM mainframe. It supported a remote terminal system for research locally and with terminals across the country, one of the early applications of Telescience. A number of the systems in government laboratories, university research labs, and several commercially offered products are descendents of this system.

The original LARSYS was designed with data in mind that had of the order of 15 or less spectral bands and with the intention of discrimination between as many as 20 or so classes. The new implementation is intended to provide the same degree of interactiveness and ease of use, but as a first step toward dealing with the larger number of spectral bands (of the order of 100 to 200) of future sensor systems and a larger number of classes (of the order of 50 or more). The intent is to take advantage of the substantial and economical computing power now available in personal computer workstations and the favorable user interface available with such systems.

Functions of LARSYS which have been implemented to date include the Image Display, Histogram, Class Statistics, Clustering, Feature Selection, and Classifier Processors. This new software runs on all varieties of Macintosh II systems. It has been given preliminary tests at dimensionalities as high as 200 and is being operated routinely on several Macintosh's in researchers' offices, and on nine Macintosh II's with 2 Mb of memory and 13" color screens which are available to graduate students. A separate version of the system will also run on Macintosh SE's and Macintosh Plus's. Macintosh's at the various sites across the campus have been interconnected where there are researchers participating in this research via the Engineering Computer Network/Purdue University Computing Center Ethernet backbone. Alpha versions of the software have also been made available to other researchers at sites across the country and abroad. As a part of this work, a paper was prepared on means for properly displaying images on CRT screens [27]

The goal for this first draft of the program is to provide an effective tool for analyzing multispectral data of moderate dimensionality (≈ 10 to 20 features) and moderate numbers of classes (≈ 10 to 20). Upon completion of that task, the next stage will be to incorporate appropriate of the new algorithms now under research so as to provide a capability effective for the larger number of features and classes of future sensor systems.

BIBLIOGRAPHY OF PREVIOUS AND CURRENT RESULTS

Following is a bibliography of papers, reports, theses, published abstracts, and presentations produced previously by work supported at least in part by this grant.

[1] Wu, You, *Quantitative Assessment of Landsat TM Data for Detailed Soil Mapping,* Master of Science Thesis, 1988, Agronomy Department, Purdue University, W.Lafayette, IN.

[2] Fernandez, R.N., D.F. Lozano-Garcia, and C.J. Johannsen, *Development of Micro-computer Based Georeferenced Information Systems - Case Studies from Argentina and Mexico,* Interciencia, Vol. 14, No. 5, pp. 247-253, Sept-Oct, 1989.

[3] D.F. Lozano-Garcia, R.N. Fernandez and C. J. Johannsen, *Assessment of Regional Biomass-Soil Relationships using Vegetation Indexes.* Accepted for Publication in the IEEE Transactions on Geoscience and Remote Sensing, 7 pp. 1990.

[4] Johannsen, C.J. *Remote Sensing and GIS Applications: Mapping Land Use Changes as Influenced by Land Resources,* Geological Society of America Abstracts Vol. 21: Abstract 029594, 1989.

[5] Johannsen, C.J. R.N. Fernandez, D.F. Lozano-Garcia. *Use of Remotely Sensed Data for Studying the 1988 Drought,* Soil and Water Conservation Society Annual Meeting Abstracts, p 22, 1989.

[6] Lozano-Garcia, D.F., R.N. Fernandez, P.J. Wyss, C.J. Johannsen. *Assessment of the 1988 Drought in Indiana*, 1989 American Society of Agronomy Abstracts, p. 18.

[7] T.L. Henderson, et al, *Use of High-Dimensional Spectral Data to Evaluate Organic Matter-Reflectance Relationships in Soils*, (Master's Thesis/Agronomy) LARS Technical Report 013090, Purdue University, January 1990.

[8] C-C Thomas Chen and D. A. Landgrebe, *Spectral Feature Design for Data Compression in High Dimensional Multispectral Data*, Proceedings of the 1987 International Geoscience and Remote Sensing Symposium (IGARSS '87), pp 685-690, May 1987.

[9] C-C Thomas Chen and D. A. Landgrebe, *A Spectral Feature Design System for High Dimensional Multispectral Data*, Proceedings of the 1988 International Geoscience and Remote Sensing Symposium (IGARSS '88), Edinburgh, Scotland, September 12-16, 1988, pp 891-894

[10] C.C. Thomas Chen and David A. Landgrebe, *A Spectral Feature Design System For The HIRIS/MODIS Era*, IEEE Transactions on Geoscience and Remote Sensing, Vol 27, No. 6, pp 681-686, November 1989.

[11] Tracey L. Henderson, Andrea Szilagyi, Marion F. Baumgardner, Chih-Chien Thomas Chen, and David A. Landgrebe, *Spectral Band Selection Algorithm for Classification of Soil Organic Matter Content*, Soil Science Society of America Journal, Vol. 53, No. 6, pp 1778-1784, Nov-Dec 1989.

[12] Hassan Ghassemian and David Landgrebe, *An Unsupervised Feature Extraction Method for High Dimensional Image Data Compaction*, Proceedings of the 1987 International IEEE Conference on Systems, Man, and Cybernetics, Alexandria, Virginia, Oct. 20-23, 1987. Also appeared in IEEE Control System Magazine, vol.8, no.3, pp 42-48, June 1988.

[13] Byeungwoo Jeon and David Landgrebe, *A New Supervised Absolute Classifier*, Proceedings of the International Geoscience and Remote Sensing Symposium, Washington, D.C., May 20-24, 1990.

[14] Byungyong Kim and David Landgrebe, *Hierarchical Classification in High Dimensional, Numerous Class Cases*, Proceedings of the International Geoscience and Remote Sensing Symposium, Washington, D.C., May 20-24, 1990.

[15] Byungyong Kim and David Landgrebe, *Prediction of the Optimal Number of Features*, Proceedings of the International Geoscience and Remote Sensing Symposium, Washington, D.C., May 20-24, 1990.

[16] Chulhee Lee and David Landgrebe, *A Fast Multi-Stage Gaussian ML Classifier*, Proceedings of the International Geoscience and Remote Sensing Symposium, Washington, D.C., May 20-24, 1990.

[17] John Kerekes and David Landgrebe, *A Noise Taxonomy for Remote Sensing Systems*, Proceedings of International Geoscience and Remote Sensing Symposium, Ann Arbor, Michigan, May 18-21, 1987.

[18] John P. Kerekes and D. Landgrebe, *Simulation of Optical Systems for Earth Resource Analysis*, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Edinburgh, Scotland, September 12-16, 1988, pp 1211-1214.

[19] John P. Kerekes and David A. Landgrebe, *Remote Sensing System Research Using A System Simulation*, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS'89), Vancouver, B.C., July 10-14, 1989.

[20] John P. Kerekes and David A. Landgrebe, *Simulation of Optical Remote Sensing Systems*, IEEE Transactions on Geoscience and Remote Sensing, Vol 27, No. 6, pp 762-771, November 1989.

[21] Hakil Kim and P.H. Swain, *Combining Multispectral and Ancillary Data in Remote Sensing Using Interval-Valued Probabilities*, Presented at the World Conference on Information Processing and Communication, Seoul Korea, June 13-16, 1989

[22] Hakil Kim and P.H. Swain, *Multisource Data Analysis in Remote Sensing and Geographic Information Systems Based on Shafer's Theory of Evidence*, Presented at the International Geoscience and Remote Sensing Symposium, Vancouver, B.C., July 10-14, 1989.

[23] S. Rasoul Safavian and David A. Landgrebe, *Use of Robust Estimators in Parametric Classifiers*, 1989 IEEE International Conference on Systems, Man, and Cybernetics, Cambridge, MA, November 14-17, 1989, Vol. 1, pp 356-7.

[24] Jon A. Benediktsson, P.H. Swain, O.K. Ersoy, and D. Hong, *Classification of Very High Dimensional Data Using Neural Networks*, Proceedings of the International Geoscience and Remote Sensing Symposium, Washington, D.C., May 20-24, 1990.

[25] Rasoul Safavian and David Landgrebe, *Predictive Density Approach to Parametric Classification*, Proceedings of the International Geoscience and Remote Sensing Symposium, Washington, D.C., May 20-24, 1990.

[26] Hakil Kim and P.H. Swain, *A Method for Classification of Multisource Data Using Interval-Valued Probabilities and Its Application to HIRIS Data*, Presented at the IAPR TC7 Workshop on Multisource Data Integration in Remote Sensing, College Park, Maryland, June 14-15, 1990

[27] Qian Lin and Jan Allebach, *Displaying Multispectral Images on Video Terminals in RGB Color*, Proceedings of the International Geoscience and Remote Sensing Symposium, Washington, D.C., May 20-24, 1990.