Reprinted from

Eighth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Crop Inventory and Monitoring

July 7-9, 1982

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

SSG-4 -- AN AUTOMATED SPRING SMALL GRAINS PROPORTION ESTIMATOR

T.B. DENNIS, R.B. CATE, M.M. SMYRSKI,
T.C. BAKER

Lockheed Engineering and Management
Services Company, Inc.
Houston, Texas

C.V. NAZARE

Intergraph Corporation
Houston, Texas

ABSTRACT

This paper describes a totally automated system for estimating spring small grains acreages within 5- by 6-nautical-mile sample segments as recorded in Landsat data. This procedure was developed for and tested in the fiscal year 1981 U.S./Canada Spring Small Grains Pilot Experiment conducted at the Lyndon B. Johnson Space Center as part of the Foreign Commodity Production Forecasting project of the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing program. The system was derived from attempts to model some of the human functions performed in the image analysis of Landsat data which was routinely carried out during the Large Area Crop Inventory Experiment.

I. INTRODUCTION

The classification procedures employed in the Large Area Crop Inventory Experiment (LACIE) all required that a human analyst provide labeled training samples. The procedures for labeling samples were labor intensive and involved many subjective decisions on the part of individual analysts. This paper describes an automated proportion estimation procedure which has been derived from the early field-labeling procedures used in LACIE. This procedure was developed for and tested in the U.S./Canada Spring Small Grains Pilot Experiment conducted at the Lyndon B. Johnson Space Center (JSC) in Houston, Texas, as part of the fiscal year (FY) 1981 Foreign Commodity Production Forecasting (FCPF) project of the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) program. This procedure attempts to mimic some of the basic steps followed by analyst-interpreters in locating and labeling training fields for earlier Landsat proportion estimation procedures. At the same time, it attempts to quantify the subjective decisions that were part of those earlier procedures. Additionally, this procedure represents a first attempt to estimate the effect of acquisition history on labeling accuracy and to use such an error estimate to correct final proportion estimates.

*Under Contract NAS 9-15800 at the National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, Houston, Texas 77058.

II. AUTOMATING COLOR PERCEPTION

The procedure begins by quantifying the notion of color. The automated procedure applies a transformation to the data which is similar to that used in making film products for an analyst. The analysts have traditionally used color film products in the identification of small grains; and in this procedure, the computer classifies the data using the same numbers that are used in generating the film products. The identification of small grains is made by multitemporally comparing the observed color with an expected response based on crop calendar information. In the automated version, a discrete "spectral appearance" or "color" crop calendar is predicted from meteorological data. This crop calendar is used to select acquisitions for processing. The ideal is to select one acquisition when the majority of the spring small grains are in the jointing-to-heading stage, one when they are beginning to ripen or turn, and one when they have been harvested or are completely ripe but when summer crops continue to grow. To an analyst with the usual simulated color infrared film product, the classic spring small grains fields would appear red on the first acquisition, anywhere from red to golden brown or green on the second, and something other than red on the last acquisition. The automated procedure then has been designed to look for fields and to classify them according to their multitemporal color pattern.

A. FIELD EXTRACTION

Quasi fields are first located within each individual acquisition. In one acquisition, the presence or absence of vegetation is the most striking contrast between fields. For a single acquisition then, the quasi fields are defined to be maximally connected regions of vegetation (those pixels which appear red) and nonvegetation (those pixels which do not). These quasi fields are adjusted to ensure that each field satisfies some basic topological properties before the acquisitions are combined. The final set of target fields for labeling is obtained by taking intersections of the quasi fields from all three acquisitions. The final set of fields to be labeled are now connected regions of pixels which follow a distinctive pattern of vegetation/nonvegetation through the set of acquisitions.

B. BIOSTAGE DETERMINATION

At this point, each field is considered a prospective small grains field, and as such it is either in the planting, jointing, turning, or harvested biostage on each acquisition. The biostage of each field is defined and calculated in terms of the distribution of the color of pixels within the field. For example, a field is in the jointing-to-heading stage when the pixels in the field are predominantly red. The fields are classified as spring small grains or as other by noting the temporal sequence of biostages through which the field progresses on the observed acquisitions. In general, a field is classified as small grains if it follows a temporal development pattern consistent with that for small grains. The deviation of the given acquisition history from the ideal placement of acquisitions defined by the crop calendar allows the procedure to select acceptable biostage sequences from a larger set of possible biostage sequences for small grains.

C. ERROR ESTIMATION

Because of the field-to-field variability, not all small grains fields will have the same appearance on a given acquisition. Thus, not all spring small grains fields will be identifiable even given an ideal set of acquisitions. The less ideal the placement of an acquisition becomes, the greater the error in identification becomes because of the overlap between crop stages. Because of this tendency, this procedure has been designed to be conservative in labeling and to try to compensate for the error in proportion estimation by predicting the rate of omission error due to misplacement of acquisitions. Also, the tradeoff between requiring good acquisition histories and wanting a high rate of processability had to be considered in the design of the procedure. The approach taken here was to estimate the amount of omission as a function of the lateness of the first acquisition. The lateness of this acquisition will cause some spring small grains fields to show no sign of vegetation in the selected acquisitions. The amount of omission is estimated from an empirical model derived from development data. The omission estimate is combined with a direct estimate obtained by enumerating pixels in those fields labeled as small grains to obtain a segment-level proportion estimate.

III. RESULTS

In the FCPF spring small grains pilot, this procedure was applied to 331 segments from the northern Great Plains (North Dakota, South Dakota, Minnesota, Montana, and Saskatchewan, Canada). Of those segments, 220 were processable to completion for a processability rate of 67 percent. The results of the processings on 169 ground truth blind sites are summarized in table 1. These statistics are broken down by year and compared with the manual procedures used operationally in the corresponding phase of LACIE. This table shows the procedure to be less biased than the manual procedures but slightly more variable, while maintaining a comparable rate of processability. An error characterization study performed as part of the pilot experiment has attempted to address the deficiencies of the procedure. In this study, the 20 worst segment errors were examined in detail to determine if any trends were evident in the errors that could easily be corrected to improve the procedure. It was concluded that nine of the worst segment errors were caused by clerical or software implementation errors that were easily correctable, four of the errors would require an improvement in the biowindow designations, and seven of the errors would require additional procedure development to correct. Thus, with a minimal development effort, some gain in proportion estimation accuracy can be expected while maintaining total automation and a high rate of processability.

The chief advantage of automation is efficiency. Efficiency data for the area estimation component of pre-AgRISTARS technologies were incomplete and did not provide good visibility to procedure affordability. As a result, improvements in technician support - including data entry, data handling, and status and tracking - cannot be directly compared. However, most of these technician functions are potentially automatable. Where a direct comparison is possible, a significant improvement is obtained in analyst time-line efficiency. Table 2 shows the gains in efficiency in terms of analyst time and central processing unit (CPU) usage that are provided by the automation of the SSG4 procedure.

As part of the pilot experiment, the segment-level proportions for each crop year were aggregated. Table 3 compares the aggregation results obtained from SSG4 estimates with the results obtained from aggregating ground truth proportions. The four-state area estimates and the observed coefficient of variation (CV) are plotted for each crop year as percentages of the U.S. Department of Agriculture (USDA) standard acreage estimates. In comparing SSG4 with other current technologies used in the pilot, it was concluded that the aggregation results for all procedures corresponded closely to USDA estimates for all years. Each procedure appeared equally good in 1977 and 1978 and equally poor in 1976 and 1979. The main advantage observed with SSG4 was the low CV that it attained because of its high processability rate.

Table 1. Pilot Proportion Estimation Results.

Statistic	All years	1976	1977	1978[†]	1979	Procedure
Bias	-1.74*	-4.48*	1.95	-2.21	-2.14	Automated
Standard deviation	11.51	10.68	11.65	10.34	13.53	
n[‡]	169	36	38	61	34	
Bias		-5.51*	-6.10*	-4.0, -2.9	-3.5*	LACIE operational
Standard deviation		8.52	5.40	7.4, 7.36	5.9	
n		35	45	38, 15	35	

[*]Indicates that the bias is significant at the 10% level.

[†]The first and second values given under this column for the LACIE operational procedure are for the United States and Saskatchewan, Canada, respectively.

[‡]Represents the number of segments.

Table 2. Comparison of Procedure Efficiency.

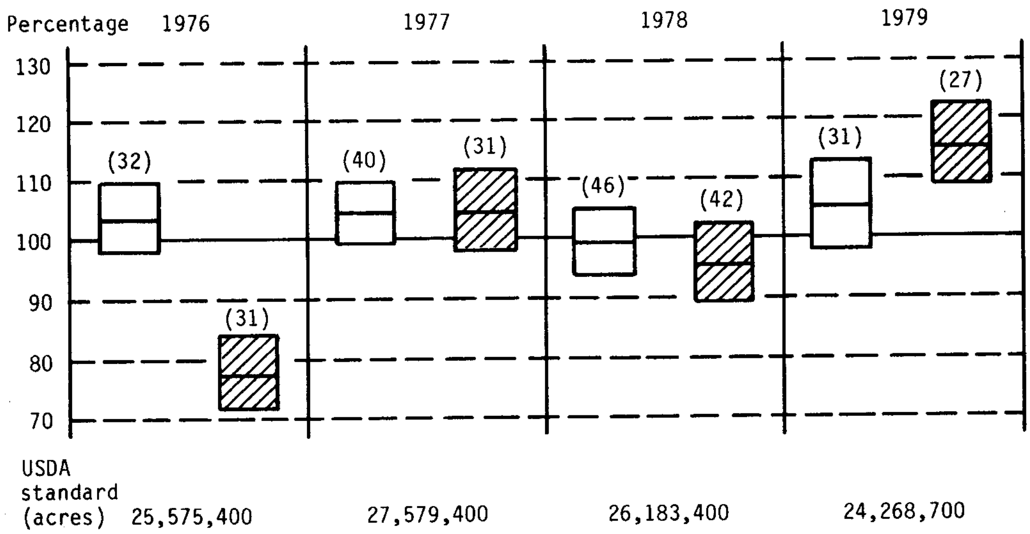
Function		Time, min.	
		Pre-AgRISTARS	SSG4
Data base preparation	Manual		
	Analyst	Not available	15
	Technician	Not available	140
	Computer		
	CPU	Not available	5
Procedure execution	Manual		
	Analyst	240	0
	Technician	Not available	5
	Computer		
	CPU	60	8.1

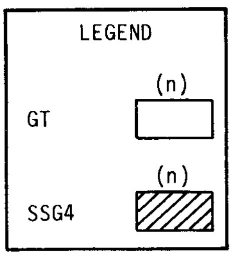
IV. CONCLUSIONS

In summary, it has been demonstrated that the spatial/color-based proportion estimation procedure provides the agricultural remote-sensing community with the basic tools to develop unbiased and highly efficient procedures for obtaining crop area estimates at end of season. This has been accomplished with a high rate of data processability at a small cost in variance. The obvious extension from spring small grains to summer crops is now being tested, and extensions to large unit proportion estimates are under development.

Table 3. Results of Procedure SSG4 Pilot Experiment for Demonstration of Four-State Aggregation.

[Area estimates are expressed as a percentage of the USDA standard ±1 CV]





AUTHOR BIOGRAPHICAL DATA

Thomas B. Dennis. Currently, he is responsible for the design and development of techniques to solve Earth observation problems in the areas of proportion estimation and technology development. His specific areas of skill include computer software development and data processing. Dr. Dennis has accumulated experience in Landsat image analysis and has been involved in several Landsat labeling/proportion estimation tasks. After receiving a B.S. in mathematics (1968) and an M.S. in mathematics (1973) at North Texas State University, he received a Ph. D. in mathematics from the University of Houston in 1977.

Robert B. Cate. Currently, he is responsible for providing technical leadership in developing mathematical methodology for the analysis of remotely sensed satellite data for agricultural inventory. Specific areas of skill include regression and nonparametric models of crop development and yield, error models, single and multicrop area estimation procedures, and quantified color theory for preprocessing of remotely sensed radiometric data. Dr. Cate has accumulated over 30 years experience in collection and analysis of data on natural resources and their utilization, 20 years of this experience being in foreign countries, primarily Latin America and South Asia. Awarded a B.S. in botany by Dartmouth College in 1946, he received an M.S. in soil science (1960) and a Ph. D. in soil science (1971) at North Carolina State University.

Charles V. Nazare. At Lockheed, he was responsible for assessment of accuracy of crop identification techniques in the AgRISTARS program. Currently with Intergraph Corporation in Houston, he is involved in project planning and consulting with oil exploration clients. In previous appointments he has been responsible for the collection and interpretation of visible and near infrared reflectance spectra of rock types for identification of lithology, and he has participated in geotechnical investigations for site selection power plants and EHV transmission line corridors for environmental impact analysis. He received a B.Sc. in geology (1969) at the University of Bombay and an M.S. in geologic remote sensing (1973) from the University of Michigan.

Margaret M. Smyrski. Currently, her responsibilites are in the performance analysis and evaluations of ongoing experiments. She conducts statistical testing of experiment hypotheses of remote-sensing-based agricultural information systems with the use of the statistical computer package, SAS. Additional responsibilities include the experiment design of available and specially designed technology to meet project objectives. Previous experience focused primarily on statistical consulting for statistical computer packages and aiding in the design of experiments. She received a B.A. in mathematics from the University of Chicago in 1978 and an M.S. in statistics from Purdue University in 1980.

Thomas C. Baker, Jr. Currently, he is active in the design, development, and integration of statistical sampling techniques in support of remote-sensing-based technology for the estimation of the amount of change in the crop area of a target region over time. Other recent activity includes the development of extensions to a multiyear aggregation technique. Areas of specialization include statistical computing, stochastic mathematical programming, and simulation. Awarded a B.S. in mathematics at Abilene Christian University in 1974, he received an M.S. in statistics (1975) and a Ph. D. in statistics (1978) at Texas A&M University.