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

AUTOMATED PIXEL SCREENING AND SELECTION TECHNIQUE

C.L. DAILEY, G.M. CHAPMAN

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

ABSTRACT

The automated pixel screening and selection technique was developed to improve the efficiency of estimating crop proportion within segments by use of Landsat multi-temporal spectral information. By replacing the efforts of an analyst in identifying pixels appropriate for labeling, the automated procedure reduced the time for this task. It also reduced the turnaround time from 3 days to 1 day so that subsequent tasks in the estimation process could be initiated sooner. The purpose of the automated pixel screening and selection was to select pixels to be sent through an automated labeling procedure. This was necessary because the various random sampling methods will often include in their selection mixed pixels, which are difficult to label correctly. Mixed pixels are those located along field boundaries that are spectrally or temporally mixed. For the mixed pixels sampled, an alternate pixel associated with it is labeled in its stead. To accomplish this, the automated pixel screening and selection technique designates whether pixels are pure or mixed and selects alternates for mixed pixels. This technique produced comparable proportion estimates to those obtained when analysts selected pixels. In addition, the labeling accuracies for automatically selected pixels, in general, were comparable to those obtained for analyst-selected pixels. In fact, the labeling accuracy for automatically-selected alternate pixels showed an improvement from that obtained for mixed pixels.

I. INTRODUCTION

To estimate crop acreage using Landsat data, one approach involves sampling portions of Landsat imagery that represent 5x6-nautical mile areas on the ground and obtaining acreage estimates for each of these areas (segments). Segment acreage estimates for the crop of interest are obtained from estimates of the proportion of the crop in the segment. One general procedure for obtaining crop proportion estimates for segments from Landsat scenes requires that a sample of the 22,932 picture elements (pixels) of the imagery be selected for labeling. The crop proportion estimate is made on the basis of the labels. Since errors in labeling directly affect crop proportion estimates, this step is extremely important. Some pixels, however, are difficult to label correctly; specifically, mixed pixels. Mixed pixels occur when pixels that are located on the boundaries of fields have a mixed spectral signature, or they occur when pixels are mixed temporally by appearing to be within one field in one acquisition of Landsat imagery and within a different field in another acquisition of Landsat imagery. Part of the mixed pixel problem is limited resolving power of the Landsat multispectral scanner (MSS). Each pixel in the MSS data represents a 57x79-meter area of ground, large enough so that the spectral measurements may easily represent more than one crop. Several attempts have been made to deal with the problem that mixed pixels present in estimating crop proportions.

One proposed method of resolving the problem was to ignore the mixed pixels by neither labeling them nor using them in estimating crop proportions. However, the results showed that this procedure introduced a bias into the proportion estimates because estimates were no longer being made from representative samples.[1]

As a result, analysts were required to associate each mixed dot (a dot is a pixel that is to be sent through a labeling procedure) with a particular field and to select a pixel from the interior of that field to be labeled in the place of the mixed dot. The label for this "alternate" dot then could be associated with the mixed pixel. This method of handling the mixed dot provided representative samples and resolved the labeling problem presented by the mixed pixel.[1] However, designating dots as mixed or pure and selecting alternate dots required 2 to 4 hours of

---

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

the analyst's time per segment for 209 dots. It was hoped that this task could be automated.

With this goal, technology was borrowed from a previously designed automated procedure that used field delineations in producing estimates. This technology was modified so that the designation of dots as mixed or pure and the selection of alternate dots could be performed automatically. By doing this, proportion estimates could be made more efficiently.

II. TEST OBJECTIVES

There were two major objectives for this study. The first was to specifically define settings within the technology from among the various options. The second was to determine the effect of the automated technique on proportion estimation and labeling accuracy and to examine the relationship between the purity designations of the automated technique, the analyst's procedure, and data from ground observations.

III. AUTOMATED PIXEL SCREENING AND SELECTION TECHNIQUE DESCRIPTION

The technology was designed to approximate the approach used by analysts through interpretation. The automated pixel screening algorithm consists of two functions.

1. Designation of dots as pure or mixed. The spatial/color field extraction program developed by T. B. Dennis groups pixels into fields (pure) and boundaries (mixed).[2] A maximum of seven acquisitions can be used to form the field/boundary map, thus accounting for both spatially and spectrally mixed pixels.

2. Selection of alternate pure pixels to represent mixed dots in the automatic labeler.

Before the procedure was tested, the selection of alternate dots had to be specifically defined. Options for three decisions had to be specified before the technology was defined to select alternate pixels.

1. The area that would be searched to locate a pure pixel.

2. The selection of an alternate from two or more equally distant candidates.

3. The manner of handling the mixed pixel if a pure dot could not be found within the search area.

The search area is comprised of groups of pixels equidistant from the center of the mixed dot. Each group of dots has an associated group number which indicates the spatial proximity of its pixels to the mixed dot; lower numbers are assigned to the groups nearer the mixed dot. For example, group 1 is comprised of those dots nearest the center of the mixed dot (30 meters); group 2 are those dots 45 meters away, etc.

These groups were searched sequentially until a pure alternate was found. The size of the search area was determined by the number of groups searched. There were two options for selecting an alternate from two or more pixels equidistant from the mixed dot.

1. Arbitrarily select the first one encountered.

2. Select the one spectrally closest (by Euclidean distance over four channels of spectral data) to the mixed pixel.

There were three options for handling the mixed pixel when a pure alternate was not found in the search area.

1. Leave the dot unlabeled.

2. Give it an arbitrary label.

3. Label it as other dots are labeled with the automatic labeler.

IV. TEST APPROACH AND METHODOLOGY

Thirty Landsat segments dispersed throughout the U.S. Northern Great Plains from crop years 1976-79 (table 1) were used first for establishing option setting and then in testing. The option settings were determined by processing the segments through the automated pixel screening technique, allowing it to search the first 19 groups. By noting the group numbers at which a pure alternate was found, the search area could be defined. Ground-truth crop codes obtained from tapes storing digitized information of ground observations allowed the tie option to be defined as the option that more frequently selected pixels of the same crop code as the mixed dot. Based on these ground-truth crop codes of mixed dots for which an alternate could not be found, a decision was made to use automated labels for these dots to avoid potential software problems caused by leaving dots unlabeled and to avoid the introduction of a bias by the use of an arbitrary label.

Table 1. Development Data Set.

	YEAR				
	76	77	78	79	
MN		1	2		3
MT		4	1		5
ND	4	2	7	3	16
SD		4	2		6
	4	11	12	3	30

After these settings were specified, the segments were processed using the SSG2 procedure.[3] This procedure requires analysts to designate purity of dots and to select alternate dots to replace mixed dots in the automated labeling step. Based on the automated labels, proportion estimates were made for the crop of interest. These segments were reprocessed using the same procedure with the automated technique substituting for efforts of the analysts in designating purity and selecting alternate dots. All other processing remained the same, so a comparison of results would indicate the effect of the automated technique. Finally, the segments were reprocessed once more without the selection of alternate dots to provide another comparison.

V. TEST RESULTS AND CONCLUSIONS

A. DETERMINED SETTINGS

The options for selecting an alternate dot were defined to include a search area comprised of the first nine pixel groups, the spectral tie option, and using automated labels for the mixed dot when a pure alternate was not found within the first nine groups. A pure alternate was found within the first nine groups over 99 percent of the time. The spectral tie option selected alternates that had the same ground-truth crop code as the mixed pixel more frequently (70 percent) than did the arbitrary option.

To reiterate, using automated labels for mixed dots when an alternate could not be found in the search area avoided potential software problems and bias introduction.

B. PROPORTION ESTIMATION, LABELING ACCURACY, PURITY DESIGNATIONS

Initially, random-sample, spring small grains proportion estimates for the 30 segments were derived using ground-truth data to label the 209 grid dots (every tenth pixel of every tenth line), the automatically selected grid dots with alternates, and the grid dots designated as pure (a subset of the 209). Table 2 shows the resulting mean errors and standard deviations using each of the three sets of labeling targets.

The automated pixel screening technique produced unbiased results while using only pure grid dots, significantly underestimated the ground-truth derived proportions, and introduced bias into the estimates. These results confirmed the trend identified with the use of analyst-selected alternate dots.[1]

Testing was continued with the selected options to examine the interaction between the automated pixel screening and selection with the SSG2. The results compiled in table 3 show that no significant difference was found between proportion estimates based on analyst-selected dots and those based on dots selected by the automated procedure.

Table 2. Random Sample Spring Small Grain Proportion Estimates Based on Ground-Truth Labels.

LABELING TARGET	209 GRID DOTS	AUTO GRID & ALTERNATE DOTS	PURE GRID DOTS ONLY
MEAN ERROR %	-0.8	-0.6	-7.5
STANDARD DEVIATION OF ERRORS	2.6	2.6	5.4

Table 3. SSG2 Proportion Estimation Accuracy Results.

LABELING TARGETS	AI	AUTO	ORIGINAL GRID
MEAN ERROR*	2.0	3.0	1.3
STD. DEV. OF THE ERRORS	7.35	7.61	8.41
MEAN ABS. ERROR	5.9	6.7	(-)
RELATIVE MEAN ERROR	7.5	11.3	5.1
MEAN GT PROPORTION	26.6	26.6	25.7
SAMPLE SIZE	30	30	29

However, when the original grid dots were used, proportion estimates showed better accuracy than when alternate dots were used.

To understand why the original grid dots obtained better proportion estimates, labeling accuracy results as shown in table 4 were examined.

Labeling accuracy of dots selected by the automated technique was comparable to the labeling accuracy of the analyst-selected dots. The alternate dots were hypothesized to have better labeling accuracy than the mixed grid dots. Indeed, when alternate dots were selected, labeling accuracy of small grains improved an average of 18.3 to 25.6 percentage points over labeling accuracy obtained from corresponding mixed grid dots. Nonspring small grains accuracies were comparable when the different types of dots were used.

The significantly better labeling accuracy of alternate dots over mixed grid dots seems to

Table 4. SSG2 Labeling Accuracy Results Shown as Percent Correctly Labeled.

		SMALL GRAINS	NON SMALL GRAINS	OVERALL
PURE GRID DOTS	AUTO AI	75.4 (73.5)* 71.5 (69.4)	85.1 (86.6) 87.6 (87.0)	82.7 (83.4) 83.5 (82.6)
ALTERNATE DOTS	AUTO AI	66.9 (48.6) 82.4 (56.8)	82.1 (85.0) 82.6 (82.0)	78.3 (75.9) 82.6 (75.0)
TOTAL DOTS	AUTO AI	73.3 (67.5) 73.4 (67.5)	84.4 (86.4) 86.9 (86.4)	81.6 (81.6) 83.4 (81.6)

*Original grid dot results over 29 segments.

contradict the proportion estimate accuracy results. Relative-count proportion estimates based on alternate dots, pure grids, and mixed grids were calculated. The alternate dots, pure grids, and mixed grids were calculated. The alternate dots estimates (1.8 percent mean error) were closer to those of the pure grid dots (3.7 percent mean error).

The smaller proportion estimation mean error for the original grid dots appears to be due to a combination of overestimation of pure grid dots and an underestimation of the boundary grid dots.

Finally, the purity of the pixels chosen by the analyst and by the automated technique was compared with ground-truth purity. Ground-truth purity was defined as all six subpixels* which make up a single Landsat pixel having the same crop code. The automated technique was more conservative in designating a dot as pure (75.5 percent pure grid dots) than the analyst (85.5 percent). Although there was a discrepancy in the percents of grid dots designated pure by the two methods, both agreed with the ground-truth definition at a comparable rate. Agreement of purity designation between ground-truth definition and the analyst selection method was 73 percent and between ground-truth definition and the automated method was 70 percent. Even between the two methods, purity designation agreed 74 percent of the time, again showing the automated method as a viable substitute for the analyst selection method.

The times required to do the automated pixel screening and selection method and the analyst interpretation method were recorded. The automated method used approximately 5 minutes of terminal time and 4 minutes of central processing unit (CPU) time. The analyst method averaged 217 minutes of the analyst's time and approximately 20 minutes of terminal time. Turnaround time is also improved from 3 days to 1 day by using the automated method.

In summary, automated pixel screening and selection provided representative samples of the scene. Proportion estimation accuracy and labeling accuracy were comparable to those based on analyst-selected dots. Both analyst and automated methods showed comparable rates of agreement with the purity designation from ground-truth definition. The amount of time spent on pure pixel selection was significantly reduced by using the automated method.

C. FURTHER RESEARCH

Perhaps labeling accuracy for automatically selected dots could be improved by selecting alternate pixels that are further into the fields rather than those near the edge, as it was tested in this study. Hopefully, this possibility will be studied.

Another potential study would apply the automated pixel screening and technique on corn/soybean segments. Presently, it is being tested with another spring small grains proportion estimation procedure.

REFERENCES

1. Chapman, G. M.; and Carnes, J. G.: Evaluation of the Procedure 1A Component of the 1980 U.S./Canada Wheat and Barley Exploratory Experiment. To be published.

2. Dennis, T. B.: Spatial Color Sequence Proportion Estimation Technique. SR-L0-04028, JSC-16843, LEMSCO-15641, NASA/JSC (Houston), Dec. 1980.

3. Waggoner, J. T.; and Phinney, D. E.: Project Procedures Designation and Description Document. Vol. 1, FC-L1-00715, JSC-17154, NASA/JSC (Houston), June 1981.

---

*Note: Each pixel on the tapes storing digitized data from ground observations is made up of six subpixels, each having its own crop code. The crop code for a pixel is determined by the crop code represented most frequently in the six subpixels. If two crop codes are equally represented, the first crop code observed is used.

AUTHOR BIOGRAPHICAL DATA

Christine L. Dailey. In the AgRISTARS program, she is currently involved in development and testing of crop identification technology using Landsat data. Earlier, she worked as an image analyst for foreign crops during the LACIE program. She did agricultural remote sensing applications research while earning an M.S. in geography from the University of Illinois in 1976.

Gary M. Chapman. For the AgRISTARS program, he has been involved in experiment design and performance assessment technology. He has been engaged in this capacity since receiving an M.S. in statistics from Kansas State University in 1979.