LARS Information Note 013167
Purdue University

LABORATORY FOR AGRICULTURAL REMOTE SENSING

Early Results in the Automatic Classification of Wheat

This note contains a brief report on early results in the automatic classification of crops from airborne multispectral data. This specific problem created here is the automatic detection of wheat in late June. It will be noted that at that time, wheat in the test site region (North-central Indiana) is near maturity. The results to be discussed here are contained in the four attached computer printouts which may be interpreted with the aid of the attached aerial photograph showing the flightline involved.

The key used on this photograph to designate the primary agricultural cover is as follows:

A - Alfalfa
C - Corn
O - Oats
R - Rye
S - Soybeans
W - Wheat
P - Pasture
RC - Red Clover
SC - Sweet Clover
Tim - Timothy
Sudan - Sudan Grass

Figure 1 shows a computer printout of the scanner imagery in pictorial form. The computer program which produces printouts of this type was written so that data editing may be done conveniently. Column numbers at the top of the printout, and line numbers at the left edge of the printout, are used to determine the address of any given data point. A comparison between the photograph and this

printout reveals a number of distinctive field patterns such that orientation from the photograph to the printout is easily possible. For example, consider field A at lines 1381 to 1543 and columns 1 to 115, which shows a wheat field with oats planted in the center of it. (Column numbers proceed from right to left due to characteristics of the airborne scanner.)

The results of a first, very unsophisticated, attempt at the automatic classification of wheat is shown on Figure 2. The method used in generating these results was as follows: In one channel of data the mean and variance of wheat samples were determined. The computer was then instructed to print a 'W' for each sample point whose response was sufficiently near this mean value. By comparing this printout with Figure 1 and the photograph, it is seen that the accuracy of classification is relatively high for a single channel of data.

Figure 3 shows the results of trying the same technique, but using four channels simultaneously. It is seen that the results have improved considerably.

It is reasonable to expect that even with this simple technique the accuracy can be improved as soon as more channels can be incorporated into the method. At present, due to (temporary) equipment limits, only four of the 18 channels of information are available to us.

The previous techniques may be best described as spectral matching techniques. Figure 4 shows the first attempt using a truly automatic pattern recognition technique. The method is still relatively unsophisticated and is as follows: We assume the four channels of data from wheat are four dimensionally Gaussianly distributed. For this first attempt we assumed that the remainder of the data was also governed by the Gaussian probability law. The two mean vectors and covariance functions were then determined. Thus, the probability

description of the two classes (wheat and "everything else") are completely known. At each sample point the ratio of the two probability functions was calculated and a classification was made depending upon whether this ratio was greater than or less than one.

In addition to making a classification at each point, the computer was instructed to keep a running total of the number of points classified into each class. The total number of samples classified in this example was 64,240, and of this total 5,469 were classified as wheat. From the altitude of the aircraft the area each sample point represents may be determined. Therefore, the above numbers can be converted into wheat acreage measurements.

From the above, the following observations and conclusions are noted:

- The value of the great flexibility of a digital approach for research purposes has been demonstrated. The results of Figures 1 through 4 were all obtained within days, and in some cases hours, of one another.
- In the problem above, the grossness of the differences in signatures of the two classes has permitted success with relatively simple procedures. Future problems will no doubt require more sophisticated techniques.
- The above problem has progressed to the point of being ready for new data collection of a semi-operational (large scale) nature.
- Effective data handling and analysis techniques have been demonstrated to be adequate for at least one task. Generally,

the data handling and analysis techniques either have been, or are being, improved to the point of requiring additional refinements in the data collection system.

- Work should proceed toward perfecting these analysis techniques for other classification tasks.