LARS INFORMATION NOTE 21667

Purdue University

LABORATORY FOR AGRICULTURAL REMOTE SENSING

Early Results in the Automatic Classification of Wheat

Abstract

The primary purpose of this information note is to give a brief report on early results in the automatic classification of crops from airborne multispectral data. Preceding the presentation of these results is a brief description of the remote sensing program at Purdue University.

LARS INFORMATION NOTE 21667

Purdue University

LABORATORY FOR AGRICULTURAL REMOTE SENSING

Early Results in the Automatic Classification of Wheat

Program Objectives

Research efforts at Purdue University to develop and apply remote sensing techniques in agriculture are nearing the end of their first year. Research on remote multispectral sensing techniques are directed at establishing methods whereby various soil and agricultural crop parameters may be determined remotely from aircraft and spacecraft systems through a program of comparative multispectral sensing. Crop and soil parameters to be studied include species identification, state of maturity, disease conditions, soil types, and soil moisture conditions.

The objectives of this research program are as follows:

- To determine the degree to which selected major crops of the corn belt region, such as corn and soybeans, can be differentiated on the basis of their multispectral response at various times during the growing season.
- To determine the amount of variation and to identify the major sources of variation in the multispectral response of selected agricultural features such as soil conditions and major crop species of the corn belt region at various times during the growing season.

- To determine and prescribe methods for gathering information from the ground that will allow prediction of multispectral response characteristics obtained by remote multispectral sensing techniques.

To achieve these objectives, the research at Purdue University involves three interrelated study areas:

1. Biophysical studies in the laboratory and field, designed to:
   - Learn more about natural variations in plant and soil reflectance spectra and the factors which influence variations.
   - Determine optimum portions of spectrum in which to work with remote sensors.
   - Through the use of ground-based instrument systems, determine the optimum time during the growing season to obtain remote multispectral data with airborne systems.
2. Remote multispectral sensing studies, which involve aircraft flights over selected agricultural areas in the vicinity of Lafayette, Indiana. Such aircraft systems have capabilities to obtain simultaneous data on reflectance and emission characteristics of areas flown over in many spectral bands between 0.32 and 15 microns wavelength. Such data is currently recorded electronically on analog tapes and later converted to digital form.

   Data from this phase of the program is being used to determine the feasibility for use of optical-mechanical scanner systems to conduct surveys which would supplement the current U. S. Department of Agriculture, Statistical Reporting Service crop reporting survey system.

3. The third major study area at Purdue University is that of data handling and pattern recognition techniques. If remote sensing survey systems are to be thoroughly investigated even as to the feasibility of becoming an economically practical procedure, the amount of data to be precessed are so enormous and the analysis techniques are so specialized, that data handling and pattern recognition problems become a crucial phase of the current program.

   Initial efforts in this phase of the research, primarily using multispectral scanner data obtained at three different times during the growing season, are being directed toward the following classification tasks, using various pattern recognition techniques:

   - Wheat vs everything else
   - Oats vs everything else
   - Bare soil vs vegetation vs water
   - Percentage ground cover in a given bare soil/vegetated area
   - Corn vs soybeans
   - Soybeans vs everything else
   - Alfalfa vs everything else

   In conjunction with these efforts, studies will be carried out to determine which spectral bands are the most useful for each task.

LARS Information Note 1 describes in tutorial fashion the pattern recognition approach as applied to this work.

Early Results in the Automatic Classification of Wheat

The primary purpose of this information note is to provide a brief report on early results in the automatic classification of crops from airborne multispectral data. The specific problem treated here is the automatic detection of wheat in late June. It will be noted that at that time wheat in the test site region (Northcentral Indiana) is near maturity. The results to be discussed here are contained on the four attached computer printouts which may be interpreted with the aid of the attached aerial photograph showing the flight line 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 realtively 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 measurement.

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

- o 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.
- o 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.
- o The above problem has progressed to the point of being ready for new data collection of a semi-operational (large scale) nature.
- o 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.