A Multichannel Image Data Handling System For Agricultural Remote Sensing

D. A. Landgrebe, T. L. Phillips

D. A. Landgrebe, T. L. Phillips, "A Multichannel Image Data Handling System For Agricultural Remote Sensing," Proc. SPIE 0010, Computerized Imaging Techniques, (1 July 1967); doi: 10.1117/12.946705

Event: Computerized Imaging Techniques, 1967, Washington, D.C., United States

A MULTICHANNEL IMAGE DATA HANDLING SYSTEM FOR AGRICULTURAL REMOTE SENSING*

D. A. Landgrebe and T. L. Phillips
Laboratory for Agricultural Remote Sensing
Purdue University

Abstract

The development of an "image" processing system is presented as it relates to a research program to devise techniques to survey agricultural conditions from aerospace platforms. The motivation for the research and arguments for selected techniques are discussed. Finally, the current "image" handling methods are outlined, and a system is proposed which logically extends present capabilities.

*The work described here is supported by the NASA under contract NGR 15-005-028 and the U. S. Department of Agriculture contracts 12-14-100-8307(20) and 12-14-100-8926(20).

Introduction

NASA and the U. S. Department of Agriculture are currently co-sponsoring research to develop techniques to survey agricultural conditions from aerospace platforms. This program under the Laboratory for Agricultural Remote Sensing (LARS) is presently concentrating on the use of multispectral scanner imagery for this purpose. One of the goals of the program is to develop a means for making timely and accurate crop surveys (acreage planted, projected yield, etc.) using airborne or spaceborne equipment on a nation-wide basis. Due to the quantity of data necessary in systems of this type it becomes mandatory that, in addition to conventional methods, entirely automated methods of data analysis be developed as far as possible.

This paper is intended to point to significant problems in the area of data and image handling which arise in systems such as this. We will begin by indicating the need for agricultural surveys and some remote sensing technique considerations. This will show the motivation for selecting some of the features of the data gathering and the necessity for the multispectral approach. Though making possible an ultimate solution to the problem, the choice of a multispectral approach does tend to increase the quantity of data. As a result communication with the data by the researcher becomes cumbersome. Therefore, in the concluding sections of the paper techniques for handling this amount of data are discussed.

The Need for Agricultural Surveys

It is not the purpose of this paper to develop the need for agricultural surveys.[1] However, a sketch of some examples of this need are included here to indicate the motivation for the research.

There have been at least four levels of interest in agricultural surveys defined for remote sensing. These are local, regional, national, and international. The requirements for each of these interests are varied, but the analysis tools are similar.

Consider, for example, a local survey problem. The individual farmer could perhaps materially affect his personal income by the early detection of weed or disease conditions on his farm. On a regional scale, crop surveys and yield predictions could be used to minimize losses due to improperly allocated transportation facilities at harvest time. A little imagination, even on the part of any novice, will lead to numerous examples of this type for the use of agricultural remote sensing at all levels of interest.

Of course, the usefulness of an agricultural survey is highly dependent on the timeliness of the survey report. In the first example above, the knowledge of a weed or disease condition must be made available to the farmer in time to remedy the condition. Thus, the justification of any survey technique must include consideration of the required analysis time.

The techniques discussed in this paper provide high speed analysis possibilities. At present, automatic mapping of soil, water, and vegetation areas seem to be technically feasible for real time systems. The same is likely to become true for the mapping of terrain according to crop type.

Remote Sensing Technique Considerations

Consider now an operational system whose purpose is to survey an area according to crop type and local usage. In order to illustrate the amount of data generated by such a system, assume for example that we wish to cover an area of 20 miles by 20 miles. This is equivalent in area to a typical midwestern county. By conventional air photography techniques involving an human photointerpreter, it would take perhaps 100 man days[2] to give complete coverage of that area for a land use survey.

---

[1]This topic is discussed in several publications. See, for example, "Remote Sensing for Agricultural and Natural Resources from Space", R. B. MacDonald, David Landgrebe, published in Proceedings of 1967 National Symposium of American Astronautical Society.

[2]There are many ways which this figure could be estimated. Perhaps one conservative way is as follows: assume that the time necessary to interpret photography for such a survey is based upon the number of land use plots involved and that the average plot size is about 40 acres. Further assume that a photointerpreter can identify the use of any one plot and measure its area in 5 minutes and that he can work effectively at this rate for 5 hours a day. These assumptions lead to the figure of 100 man days for a 400 square mile area.

Then, extrapolating this figure to cover a whole state (approximately 100 such average size counties) it is seen that the interpretation time becomes 10,000 man days. It is clear that even using statistical sampling techniques such a survey could not be carried out on a nation-wide basis in this fashion in a timely manner.

What has been pointed out here is that the photography/photointerpreter approach by itself is not adequate for surveys of this type. However, it is clear that there are types of surveys (involving less data) which will require a photointerpreter because of his ability to make higher level decisions. Also, an amalgamation of automatic and photointerpretive techniques will surely be required in some instances.

After considering the total remote sensing survey system, data gathering, analysis and the dissemination of results, one comes to the conclusion that not only is data analysis a problem, it is the problem which presently limits feasibility. A possible solution then is to automate the analysis, and the problem then reduces to "what is the best way to do this". To answer this question, one must first consider what type of measurements can be made in a remotely sensed fashion. Modern remote sensing devices enable the measurement of spectral, spatial, and/or temporal variations of energy radiating from a source. A photograph, for example, indicates the energy radiated from an area in its proper spatial distribution. For the current work at LARS, the use of spectral information exclusively has been chosen for initial attempts, since this selection results in considerable simplification of algorithms necessary for the automatic classification of data points.[3]

To more clearly understand why the classification algorithms are simplified, consider how multispectral data is gathered. Figure 1 shows the diagram of a multichannel optical-mechanical scanner as might be mounted in an aircraft. The energy radiated by a specific ground resolution element at a given instant of time passes through the scanner optics and is divided according to its spectral wavelength and directed to an appropriate detector. The output of all such detectors are simultaneously recorded on a multiband instrumentation recorder. The transverse motion provided by the rotating mirror and the forward motion of the aircraft cause a continuous raster to be formed for each spectral band of the scanner output. An important feature of a system such as this is that by simultaneously sampling the output of all bands, one obtains a vector which contains all spectral information available about a given resolution element on the ground. Results from preliminary analysis of data gathered in this manner indicate that correct species identification will be possible using information from individual sample vectors only (no spatial information is required). It is this fact which results in simple algorithms for analysis.

Returning now to the operational survey system discussed at the beginning of this section, consider the proposed technique on the same 20 mile by 20 mile county. A multiband scanner with, for example, 50 foot by 50 foot resolution provides approximately 4 million resolution elements in this area. If one assume 12 bands of 8 bit data, this becomes 4×10^8 bits, and even after making allowances for formating, etc., this can easily be fitted on three reels of digital tape. If we now assume that a processor can make 1,000 decisions per second, we find the 20 mile by 20 mile area could be completely analyzed in about 1 hour and the average sized state (40,000 square miles) could be completed in just over four days.

In order to automate the analysis of the data, it was argued that spectral information is more convenient than spatial information. The question may be asked, then, "Why form an image (i.e. place the data points in their proper spatial orientation with one another)?" The key to the answer of this question is the data/human dialog. In the research situation, it is necessary for the researcher to be able to edit out specific data points. In the operational case at the very least, an human observer must be able to monitor data quality. These two functions can most easily be carried out by putting the data in image form. Examples of this data/human dialog will be shown below.

---

[3] There are at least two contexts within which data analysis must be carried out: research and operational. In the research situation, convenience of data handling is of prime importance, while in the operational case, speed of processing becomes more important. An additional constraint upon the researcher is that whatever method of analysis he uses in the research situation must still be applicable in an operational case.

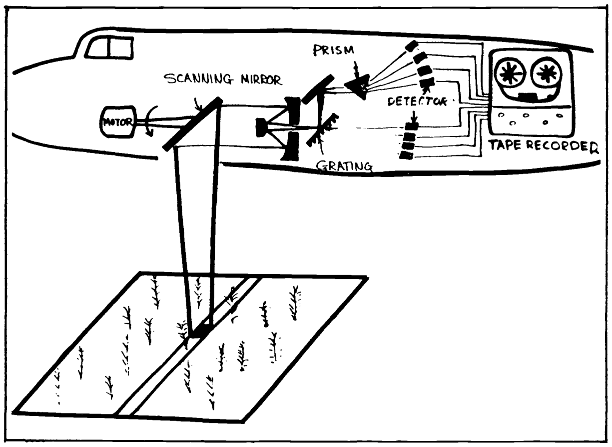


Figure 1. A Multi-channel Optical-mechanical Scanner.

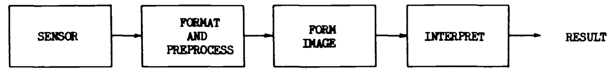


FIGURE 2a. IMAGE INTERPRETATION TECHNIQUE

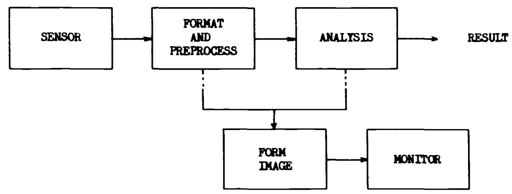


FIGURE 2b. AUTOMATIC ANALYSIS TECHNIQUE

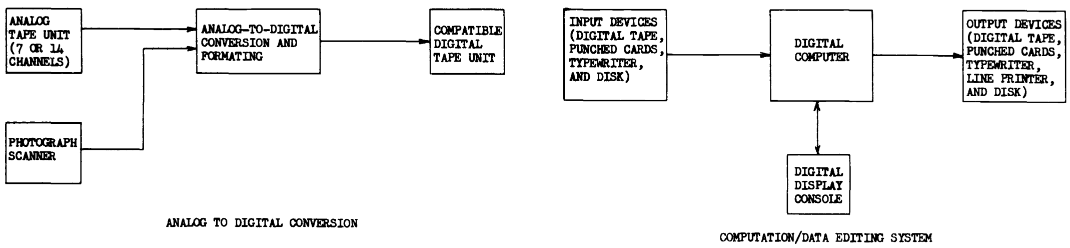


FIGURE 3.

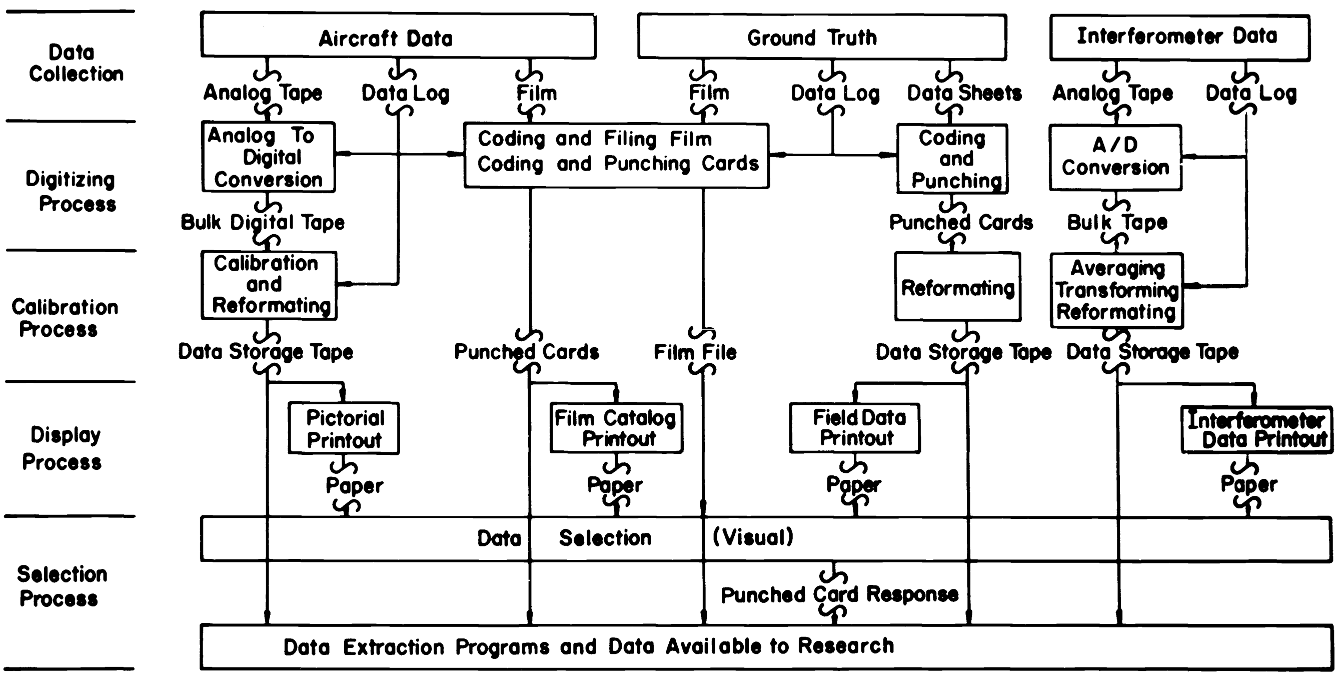


Figure 4. Phase I Data System.

Summarizing then, the use of spectral information for data analysis allows us to relegate use of an image to a secondary role in the analysis procedures. Images are best interpreted by human observers and because of the data load the use of man in this loop must be minimized. This is illustrated in figure 2. Figure 2a shows the way image information is involved in a typical computer aided image analysis technique. On the other hand, with the spectral approach the concept illustrated in figure 2b becomes feasible. The key here is that in the former case man is directly involved in the analysis, whereas in the latter case he is only monitoring the operation. Notice also in the latter case, should the spectral approach not provide the desired answer, subsequent human image analysis is still possible.

The Current "Image" Handling System

In order to prevent misunderstanding in discussing this work, we wish to review the definition of two terms: data handling and data analysis. The term data handling refers to any type of data manipulation (e.g. analog-to-digital conversion, calibration, formating, editing, etc.) which is necessary with data of this sort. The term data analysis refers to the algorithm by which the data will be reduced to useful information. Examples would be automatic pattern recognition, image enhancement, statistical analysis, and the like. Within the context of these two definitions, the primary concern of this paper is data handling rather than data analysis.

As indicated above, the current status of automatic remote sensing techniques is still in the research phase of existence. Also, early results have shown that large quantities of data must be analyzed to develop meaningful reduction methods for efficient operational systems. Therefore, the primary consideration used in selecting data processing hardware for LARS was speed and convenience (to the researcher) of data acquisition as well as flexibility for changing handling techniques and reduction algorithms. Since a digital approach can best meet these requirements of the research situation[4] the hardware shown in figure 3 was conceived and is now being placed into operation as it becomes available. The first portion of this hardware converts the data to digital form and stores it on magnetic tape. Thereafter, all data handling and analysis are carried out on a general purpose computer as shown at the bottom of figure 3.

In order to illustrate in more detail the way this hardware is utilized, figure 4 shows the data handling system (called the Phase I Data System) now in use. The current system has the capability of dealing with three different types of data now required in the research program: interferometer data, ground truth data, and the previously described aircraft scanner data. Schematically, all three types of data proceed through similar steps in the system, and we will concentrate our discussion upon the more challenging problem presented by the aircraft data. After analog-to-digital conversion, the aircraft data is calibrated and reformated to a packed form with ancillary information added. It is in this form that the data is archived.

As mentioned above, one of the chief functions of the handling system is that of data editing. The researcher must, with as much speed and convenience as possible, be able to edit from large quantities of data available to him any smaller portion which he needs for analysis purposes. And he must be able to extract from the great body of data that which corresponds to any point (or block of points) on the ground in any spectral band he desires.

In order to accomplish this in the Phase I Data System, a computer program has been written which recreates the image in raster form from a single arbitrarily selected channel. Figure 5 shows the results of applying this program for three different bands. The bands shown are from the blue, red-orange, and near-infrared portions of the spectrum, respectively. Figure 6 shows a conventional panchromatic air photo of the area overflown for comparison. This line printer technique is being used as a substitute for the Digital Display shown in figure 5.

The format of the data storage tape is such that by obtaining the row and column number of the desired data point from the printout, one obtains the address on the data storage tape of a complete

[4]Note that this choice does not preclude the use of analog methods for higher speed in an operational system.

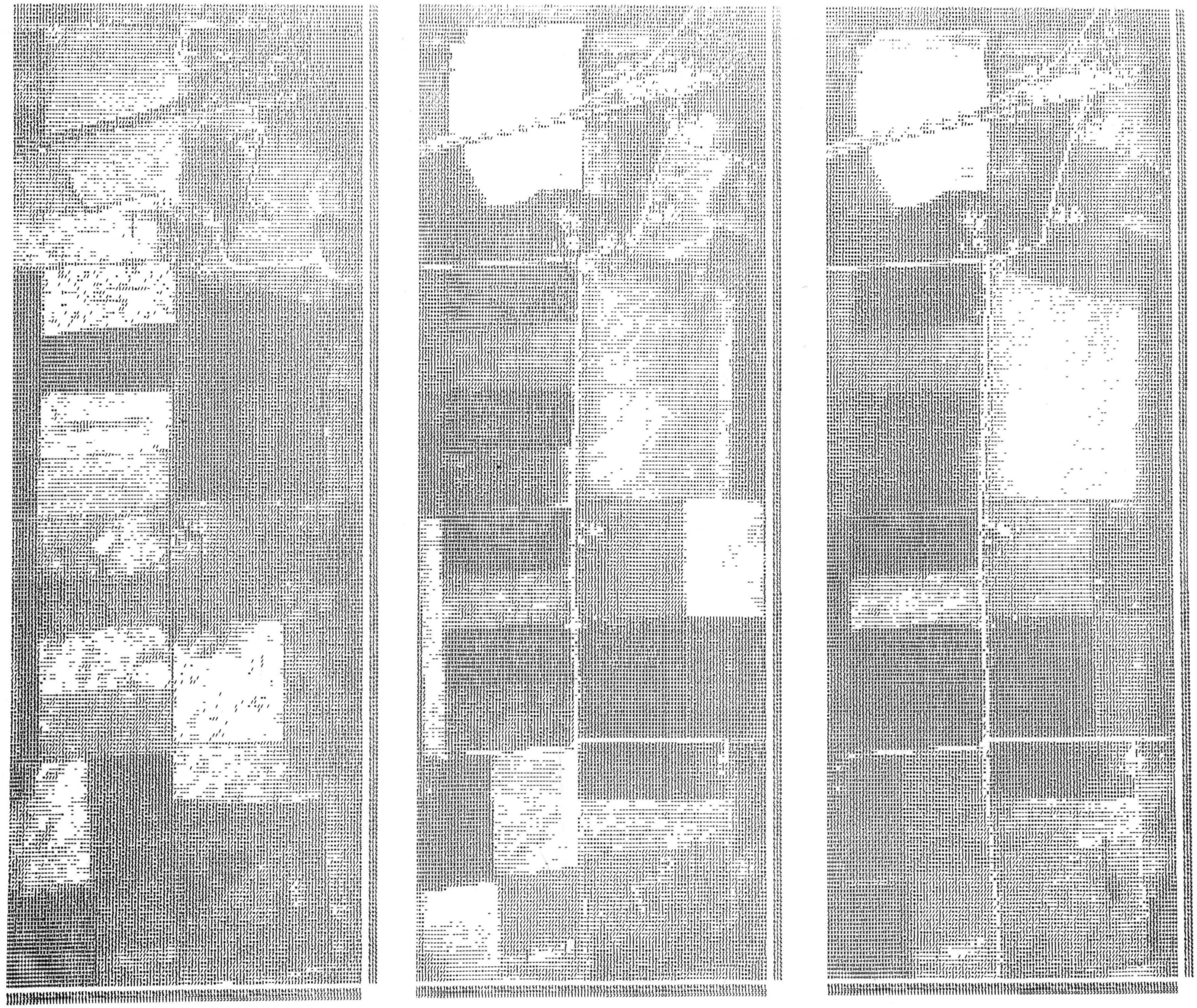
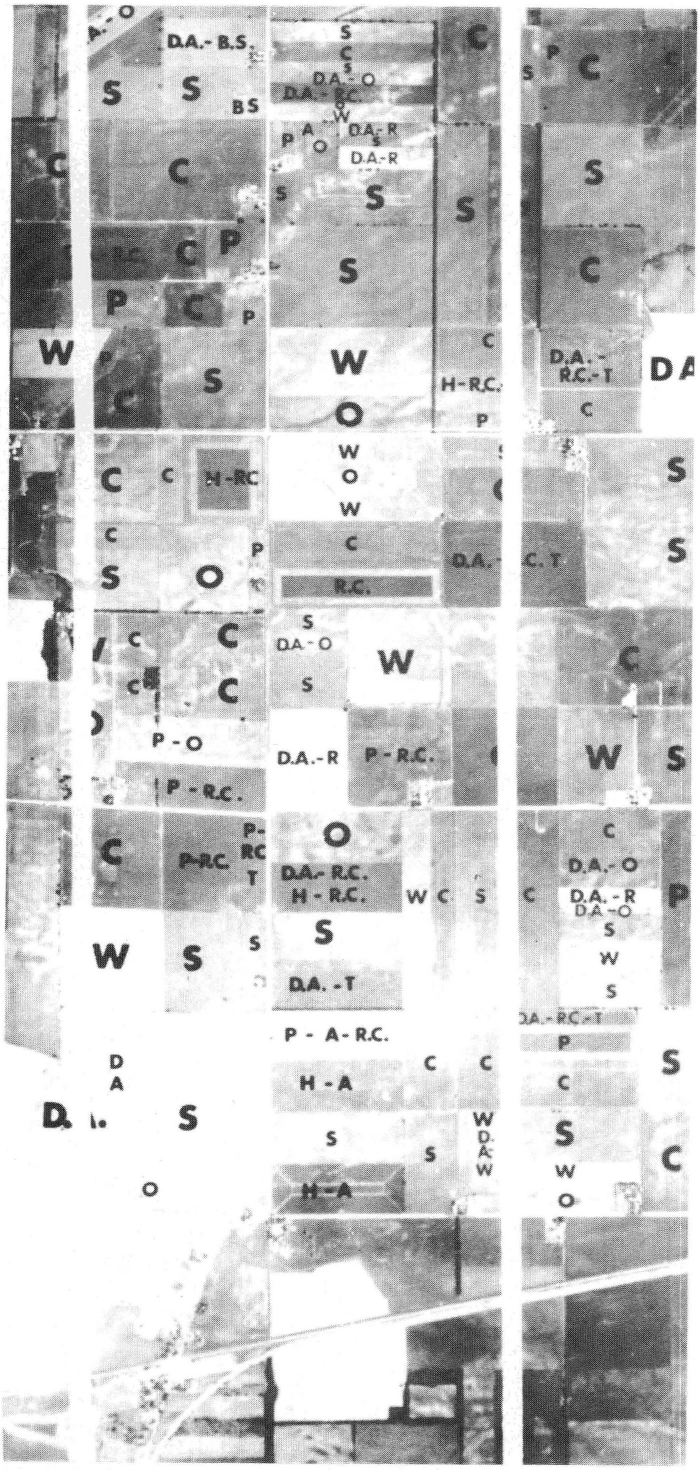


Figure 5. Pictorial Printout for Three Channels



The categories represented by the symbols are as follows:

W - Wheat

O - Oats

S - Soybeans

C - Corn

A - Alfalfa

T - Timothy

R.C. - Red Clover

B.S. - Bare Soil

H - Hay

P - Pasture

D.A. - Diverted Acres

Figure 6. Aerial Photo of Flight Path Area.

sample vector (all bands) for that point. An extraction program is then available for extracting this desired point from the data storage tape.

Of course, there is a great deal of detail involved in the organization and design of a system of this sort. And it is not possible nor desirable to cover all of it here. However, as an illustration of this detail we would like to briefly describe the method used to obtain the desired contrast scale. The symbols representing the various gray levels were selected arbitrarily based upon their symmetry and relative gray level as determined by observer comparison. Various blocks of symbols were considered until an uniformly changing set of 16 symbols representing gray levels had been selected.

The raw data is in 8 bit form; thus there are 256 possible gray levels per waveband. Before making a pictorial printout of any particular data set it is necessary to decide what subset of these gray levels will be assigned to each symbol. That is, within the 256 possible gray levels the boundaries of 16 bins must be defined.

For these particular printouts, the bin edges were selected by first computing a histogram of the entire block of data to be printed out, then selecting bin edges from the histogram in such a way that the use of each bin is equally likely. This method has been found to be much faster than attempting to set bin edges arbitrarily by hand. Actually, it has also been found that, for this application, 10 instead of 16 bins are quite adequate.

There are other uses for the printout technique. For example, they provide an ability for qualitatively judging the quality of the data on the data storage tape. Another application is in the evaluation of classification results. To illustrate this point a typical analysis result is shown in figure 7. This figure shows the result of attempting to classify all sample points into "wheat" or "not wheat" categories. Clearly, whenever a researcher attempts a classification task he must be able to evaluate his results. It is apparent from this figure that a result printout, as shown in the figure beside the pictorial printout, provides a convenient mechanism for evaluating the result.

It may be noted from figure 5 or 7 that only every other data point and every other row have been printed out. If greater resolution in the imagery is required, it is, of course, possible to print out every data point and every row.

A careful and detailed look at the design of the Phase I data System will show that an important factor in the organization of this system is researcher convenience. When each of these programs was designed and written, and each time a data format was chosen, the primary consideration was that of convenience to the researcher. It is the feeling of the authors that the rate of progress achieved on this or any other research program is as much due to this simple consideration as it is to the proper choice of the sophisticated analysis algorithms necessary to actually carry out the analysis.

The Digital Display

We should now like to describe the next step in an image editing system such as this one. As the research in this program proceeds from early feasibility studies to a more operational nature, the quantity of data which must be handled increases steadily. While the pictorial printout method has proven itself for editing and qualitative analysis of results so far, it is clear that it cannot continue to do so due to the large quantities of data which must be dealt with in the future. Faster methods of man-machine and man-data communications must become available. To help minimize this problem the digital display system of figure 3 was conceived and is now under construction.

This display system will, under program control, reconstruct an image in raster form using one or a combination of channels and present it in 500 by 500 element frames using 16 adjustable intensity levels at 30 frames per second.[5] A keyboard for calling up software will facilitate the man-machine dialog by controlling picture motion, light pen operation, the overlay of data from various channels, and other similar functions. Thus, in addition to making possible high speed data editing and monitoring of data

---

[5] Note that this calls for a writing speed of 30 megabits/sec.

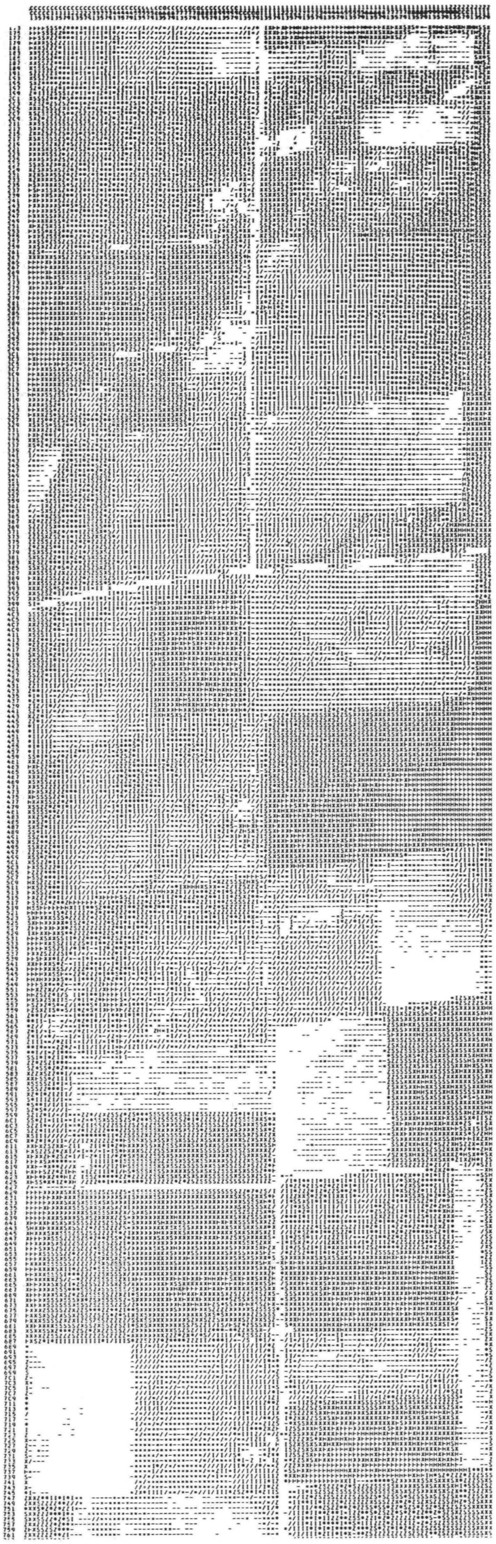


Figure 7a. Pictorial Printout of One Channel of Scanner Data.

Figure 7b. Classification Result of "Wheat versus Everything Else" Based on 12 Channels of Data.

and results, such other functions as contrast enhancement and color reconstitution becomes not only possible, but highly feasible.

Closure

In summary, the necessity for automatic methods of data analysis has led to the multispectral approach. The prime reason for this is the relatively simple availability of a meaningful measurement vector. This approach appears, at least in some cases, to be leading to an analysis system not requiring active human participation. However, man must always be able to communicate with his data. Therefore, hardware such as the digital display and software techniques to permit this must be the subject of active research.

The major purpose of this paper, then, is the proper consideration of data handling and analysis methods for this application. It is easy to fall into the trap of generating methods which give good results in the research situation. The real problem then is to devise research techniques which can also be applied in the operational case.

Acknowledgements

The authors wish to acknowledge the contribution to this work made by the entire LARS staff, but in particular the following: Paul Anuta, R. B. MacDonald, and Philip Swain.