AGRICULTURAL INVENTORY CAPABILITIES OF MACHINE PROCESSED LANDSAT DIGITAL DATA

A-16

By David L. Dietrich, Ronald E. Fries and Dwight D. Egbert, General Electric Company, Space Systems Organization, Beltsville, Maryland

N76-17481

ABSTRACT

Agricultural crop identification and acreage determination analysis of LANDSAT digital data was performed for two study areas. A multispectral image processing and analysis system[1] was utilized to perform the man-machine interactive analysis. The developed techniques yielded crop acreage estimate results with accuracy greater than 90% and as high as 99%. These results are encouraging evidence of agricultural inventory capabilities of machine processed LANDSAT digital data.

INTRODUCTION

There is increasing interest in both the scientific and political arenas as to the feasibility of employing satellite data for surveying world agriculture. To be effective, satellite sensors must provide a repetitive synoptic view of agricultural areas, yet maintain the capabilities of accurate crop identification and area determination. Effective agricultural survey analysis of these remotely sensed data will probably require an interactive approach combining man's insights and the machine's "number-crunching" capabilities.

The ability to extract agricultural inventory parameters from LANDSAT digital data via an interactive processing system was investigated in this study. Two areas were selected for analysis: Williams County, North Dakota and Melfort, Saskatchewan. LANDSAT imagery and detailed ground truth for a 3 x 13km (2 x 8 mile) portion of each area were provided by the U. S. Department of Agriculture (USDA).

In this study, a man-machine interactive processing system performed the analysis of LANDSAT digital data. Specifically, the system, under the guidance of an analyst, performed multispectral agricultural crop identification and spatial area determination within the study areas. The results achieved from interactive analysis were then compared to detailed ground truth.

The results are encouraging and illustrate the agricultural inventory capabilities of using interactive processing to analyze LANDSAT data. The findings indicate the potential utility of this approach to carry out large area agricultural surveys.

BACKGROUND

LANDSAT digital data of Williams County, North Dakota from July 11, 1973[2] and of Melfort, Saskatchewan from August 1, 1973[3] were furnished by

[1]The General Electric IMAGE 100 System.
[2]LANDSAT ID No. 1353-17165.
[3]LANDSAT ID No. 1374-17324.

PRECEDING PAGE BLANK NOT FILMED

the USDA. Accompanying the digital data were detailed ground truth crop maps of a 3 x 16km (2 x 10 mile) area within each scene. Cloud cover obscured a portion of this selected area on the Williams County LANDSAT data, forcing the study area to be reduced to 3 x 13km (2 x 8 miles). To provide consistent results, the Melfort study area was also limited to 3 x 13km.

Agricultural practices in both Williams County and Melfort are similar. Both areas are influenced by severe winters and dry summers. As a result, spring wheat is planted rather than winter wheat, and the practice of "summer fallow" is commonly employed. A summer fallow field lies idle during the growing season, and is periodically tilled to control weeds and keep the soil surface loose. This practice reduces moisture loss by soil surface evaporation and weed evapotranspiration while allowing effective catchment of the infrequent rains. This fallow ground, viewed from LANDSAT, is in high contrast with surrounding vegetation.

USDA ground truth for Williams County identified 99% of the crops within the study area as spring wheat, barley, oats, sod, grass, and fallow. Based on findings by Erb (Ref. 1), it was suspected that wheat, barley and oats would be difficult to classify separately with LANDSAT data due to close and/or overlapping spectral responses. It was also suspected that sod and grass categories, consisting of poor to fair quality permanent pastures and hay harvesting areas, may also have similar spectral responses. Consequently, the crops within the study area were analyzed in three classes: small grains (spring wheat, barley, oats); sod (sod, grass); and summer fallow.

The major crop classes in the Melfort study area were spring wheat, barley, rape and summer fallow. Again, a composite small grain category (wheat, barley) was defined for analysis. The remaining crop area was classified as either rape or fallow.

The average field size in the Williams County study area was smaller than in Melfort. Williams County field sizes ranged from 0.4 to 97 hectares (1 to 240 acres), with an average of 10 hectares (25 acres). Melfort, Saskatchewan field sizes ranged from 4 to 130 hectares (10 to 320 acres), with an average size of 22 hectares (55 acres). In both study areas, field shapes varied from small narrow fields to irregular plots to large symmetrical fields.

These areas have similarities and differences that allow for an effective comparison of results, and present an interesting challenge for using interactive machine processing to analyze LANDSAT data.

APPROACH

In order to effectively utilize the vast amount of data being generated by LANDSAT and other remote sensing systems, special purpose electronic data processing systems have been developed. One such system, an interactive multispectral image processing and analysis system, was used in this study. The basic function of the system is to extract thematic information from multichannel digital image data. This is accomplished via statistical measurement of the radiometric properties of the multichannel data as guided by an operator's interactive commands. Interactive capability to modify computer parameters and decisions is always available; changes are quickly made and evaluated. In fact, in almost all modes of operation the system responds to man at least as quickly as he can decide what to do next. The advantage of an interactive system is that it makes the most efficient use of both man and machine. The pattern recognition capabilities and subjective judgment of the man are coupled to the "number-crunching" capabilities of the machine.

The system has four video channels, each of which is capable of storing eight-bit video image data, plus eight theme storage channels. The system uses a television compatible format of 512 x 512 picture elements (pixels) to put storage requirements at 10 million bits. For this, a solid-state memory (or rotating disc option) is used as a refresh and storage device. LANDSAT digital data can be entered directly from computer compatible tapes into the refresh memory.

In this agricultural study, the methods of interactive analysis utilized a variety of machine functions:

1. LANDSAT digital data for the Williams County study area were input at a magnification of 2X. At this scale each LANDSAT pixel (56 x 79m) is represented by a 2 x 2 pixel array on the interactive display. The scene was enlarged to facilitate visual examination of the study area. The entire screen display was 15 x 15km (Figure 1) and the study area was 3 x 13km (Figure 2).

2. The 3 x 13km study area was delineated on the color monitor with a polygon cursor (any user-defined polygon area whose position is recognized by the system during subsequent operations) so that numerical analysis results could be extracted for only the 3 x 13km study area.

3. Within the study area, small uniform portions of selected fields were designated as test and training sites. For example, portions of five known fallow fields, which were well distributed throughout the study area, were incorporated into a training site set. Portions of five other known fallow fields were selected to form the test site set. These training and test sets for all three crop categories are illustrated in Figures 3 and 4. Each set of fields represents approximately 2.5% of the total study area.

4. A first cut spectral signature including all four multispectral scanner (MSS) bands was obtained for each of the three crop classes from the defined training field sets. In training, multispectral brightness data (gray levels) within the training area are automatically measured, and their upper and lower spectral limits are used to define a single spectral cell. This spectral cell is the first cut signature of the crops within the training set. All screen pixels that lie between the bounds of this signature are then identified or alarmed on the color monitor image display.

5. The next step was to determine if each first cut crop signature truly alarmed only its own crop category. This accuracy of identification was determined by analyzing the pixels alarmed within each test field set.

   The errors of identification were presented as errors of omission and commission. For example, fallow omission and commission errors would be defined as follows:

$$\text{Omission Error (for fallow)} = 1 - \frac{\text{Number of pixels identified as fallow within the fallow test field set}}{\text{Total number of pixels defined as the fallow test field set}}$$

$$\text{\% Correct Classification} = (1 - \text{Omission Error}) \times 100$$

$$\text{Commission Error (for fallow)} = \frac{\text{Number of pixels identified as fallow within the grains and sod test field sets}}{\text{Total number of pixels defined as the fallow test field set}}$$

% Commission Error = Commission Error x 100

Omission and commission errors of these first cut signatures were thus determined for each crop category: small grains, fallow and sod.

6. First cut signatures were obtained and tested at three gray level resolutions as selected on the machine, 128 levels, 64 levels and 32 levels.

7. First cut signatures include all spectral values between defined gray level limits of each training set. Further refinement of this signature was attempted automatically through a multicell signature technique. That is, the all inclusive first cut signatures (large spectral cells) were reconstructed to include only the individual pixel gray level values contained within the training area. The results of this automatic refinement technique were not encouraging; therefore, a manual interactive refinement technique was attempted.

8. The analysis objective was to obtain spectral crop signatures with characteristically low omission and commission errors. The first cut signatures were thus manipulated manually to obtain the lowest errors possible. This interactive procedure, called histogram trimming, allows the machine operator to adjust the spectral range (large cell gray level limits) of any one or all of the LANDSAT spectral bands that comprise the four channel signature. Initial trimming was made on a quantitative evaluation of the machine-displayed histograms in terms of variance, skew, etc. The new alarm created by the adjusted signature was then quantitatively tested through omission and commission error determinations.

9. Signature refinement through histogram trimming was repeated until "best results" were obtained. In this study, best results were defined as a time efficient analysis technique that produced a signature which minimized both omission and commission errors when considered simultaneously. That is, some slight trade-off between accuracy and speed of analysis was allowed.

10. The best signature for each crop class alarmed or mapped certain portions of the 3 x 13km study area. The best result classification theme maps for Williams County are shown in Figures 5 and 6. The exact number of pixels alarmed in each crop class were converted to acres for crop area comparisons with known ground truth acreage.

11. Throughout the study, MSS bands 5 and 7 appeared most sensitive in differentiating between crop classes. Therefore, classification using only these two bands of data was also attempted. The reduction in the number of spectral bands to be analyzed not only speeds analysis, but also requires less memory in the automated data processing system. These time and space savings could facilitate simultaneous multitemporal and multispectral analysis.

12. The techniques and parameters that produced the best results for Williams County were then applied to obtain signatures for the

Melfort, Saskatchewan study area. Repeating the developed classification techniques on the Melfort study area provided a check of classification and area accuracies found in Williams County. It also provided a measure of the time required to produce results rather than to investigate techniques. The Melfort study area is shown in Figure 7, and the best results classification are shown as Figure 8.

RESULTS

The results of interactive analysis of both study areas were encouraging. The developed techniques yielded acreage accuracies greater than 90% for each crop category in both the Williams County and Melfort study areas.

Williams County, North Dakota

Initial training and test statistics were the decision criteria for determining the number of spectral gray levels (32, 64 or 128 levels) that could be used most effectively on the interactive system. A gray level resolution of 32 levels was not specific enough to adequately differentiate between vegetative classes. Though initial test field percent correct classification was excellent, the commission errors were also quite high. A resolution of 128 levels introduced too many gray levels, which tended to be too restrictive and thus decreased the percent correct classification within some of the test fields. Also, visual examination of the histograms was difficult at 128 levels. At a 64 gray level resolution, initial training and testing yielded the most equitable trade-off between percent correct classification, commission error, and ease of interactive manipulation.

The primary criteria in selecting the most effective classification mode was user-interaction-efficiency coupled with accurate classification results. A first cut signature (single large cell classification) was accepted as the mode capable of providing the most rapid response to interactive commands. In addition, data operations and statistical results were more readily performed, effectively displayed and efficiently interpreted in the single cell mode.

Crop classification, based on the first cut signatures of the training fields, initially yielded high test field classification accuracies, but also yielded high commission errors. Manual refinement of the signatures, using histogram trimming, was then accomplished. Final interactive results of training and testing are represented in the first four columns of Table I. For the three categories (small grains, fallow and sod) the average percent correct classification of test fields was 93.3% with an average commission error of 3.6%.

Some tests were performed to determine if barley or oats could be accurately distinguished from wheat, or if a grass could be determined spectrally exclusive of sod. The similarity of the spectral responses within these categories, however, made more detailed stratification impractical. Therefore, all classification and acreage determinations were made using the small grains, sod or fallow categories.

The refined signatures were then applied to individual crop acreages over the entire 3 x 13km area. The number of LANDSAT pixels classified in each category was converted to hectares (1 LANDSAT pixel = 0.45 hectares) and then compared to USDA ground truth. The best results show acreage accuracies of 99.3%, 98.2% and 91.7% for small grains, fallow and sod, respectively (see Table I). Unclassified pixels amounted to 4.5% of the study area or 181 hectares (447 acres). Overlap or conflict pixels (pixels classified as two crops) amounted to 1.8% of the study area. Most of the conflict was between the vegetative classes of small grains and sod. Fallow fields were of high enough contrast with vegetated areas to minimize conflict pixels.

Similar results were achieved when either two (MSS bands 5 and 7) or four LANDSAT spectral bands were used for classification. This result supports the findings of LANDSAT-1 investigators Landgrebe (Ref. 2) and Wiegand (Ref. 3) and others who have indicated that MSS bands 5 and 7 alone are sufficient for most agricultural applications. Band 5 (0.6 - 0.7um) is in a spectral region that is selectively absorbed by chlorophyll. Almost all energy in the spectral region of band 7 (0.8 - 1.1um) is reflected by vegetation. These two bands are therefore quite sensitive to changes in vegetation and vegetation cover, and little if any advantage is gained by using all four bands. Band 4 may actually introduce an obscuring effect due to the low contrast characteristic of this spectral range that results from atmospheric scattering. The important advantage of utilizing only bands 5 and 7 is in the reduction of interactive manipulations and the number of digital operations used throughout the classification procedure.

In summary, the techniques and parameters that were found to produce the best results were: 64 gray level resolution; and MSS bands 5 and 7, with manual histogram trimming of the first cut signature. The achieved acreage accuracies for the 3 x 13km study area all exceeded 90%. The developed techniques present an efficient, accurate way of extracting crop information from LANDSAT data using interactive processing capabilities.

Melfort, Saskatchewan

To provide a check of the procedures developed for Williams County, similar classification techniques were applied to the Melfort, Saskatchewan study area. The Melfort scene was expected to provide a reliable test of the techniques since it contained larger fields with some different crops; the crops were farther into the growing season; and the area is 565km north of Williams County, North Dakota.

Interactive classification yielded crop acreage accuracies of 98.9%, 96.0% and 98.9% for small grains, fallow and rape, respectively (see Table II). The unclassified area was 11.8% or 498 hectares (1231 acres) of the total 4229 hectares (10,460 acres) within the polygon cursor defining the 3 x 13km study area. The unclassified areas were primarily composed of roads, spaces between fields and other nonhomogeneous areas. The overlap pixels represented 4.7% or 199 hectares (492 acres) of the study area. The majority of overlap or conflict existed between the small grain and rape classes. The ground truth indicates that a few of the rape fields contained portions of wild oats and weeds. The spectral variations within these fields influenced both crop identification and crop acreage results. Similar to the Williams County site, the fallow category conflicted less than the vegetative classes.

During the training and classification of the small grain crop category within the Melfort study area, unsatisfactory results were obtained when 64 gray level resolution was used. Consequently, the system parameters were changed back to a 128 gray level resolution. The spectral properties of small grains and rape were so nearly the same that the difference of only one gray level on the 128 level scale made a significant difference in the commission errors. This points out the need to exercise caution when generalizing from the previous results obtained with only 64 gray level resolution. It also demonstrates the advantages of the man-machine interactive approach, which allows periodic human intervention in the automated classification. The final results for the small grain category in the Melfort study area shown in Table II are for a gray level resolution of 128. All other figures in both Table I and Table II are for a 64 gray level resolution.

CONCLUSIONS

Man-machine interactive processing was used to perform agricultural crop identification and acreage determination analysis of LANDSAT digital data for two 3 x 13km study areas: Williams County, North Dakota and Melfort, Saskatchewan. Crop acreage accuracies as high as 99% were achieved by applying the techniques developed in the study. These techniques involved analysis of LANDSAT digital data MSS bands 5 and 7, at 64 gray level resolution (in some cases 128 level resolution), using first cut signatures refined through histogram trimming.

The crop identification and acreage accuracies obtained were similar in both study areas. These techniques may not be successful in all areas, but results for these particular study areas provide encouraging evidence of the utility of a man-machine interactive processing system for agricultural inventories. Unfortunately, implementation of these techniques to a larger area inventory was beyond the scope of this pilot project.

These results were obtained using only multispectral analysis. Many investigators, both at General Electric and elsewhere, concur that the accuracy of an agricultural survey will improve by combining multispectral and multitemporal input data. Crop calendar parameters could provide the inputs necessary to differentiate between crops with similar spectral characteristics, such as wheat and barley.

To summarize, the results illustrate the capability to rapidly extract accurate agricultural survey information from LANDSAT digital data via an interactive data processing system. Man-machine interactive data processing systems provide rapid and accurate crop classification through the utilization of the most efficient analysis capabilities of both man and machine.

REFERENCES

1. Erb, R. B. ERTS-1 Agricultural Analysis. Type III Final Report, LANDSAT-1 Investigation No. 1600A. August, 1974.

2. Landgrebe, D. A Study of the Utilization of ERTS-1 Data From the Wabash River Basin. Type III Final Report, LANDSAT-1 Investigation No. 1049A. October, 1974.

3. Wiegand, C. Reflectance of Vegetation, Soil and Water. Type III Final Report, LANDSAT-1 Investigation No. 1039A. November, 1974.

TABLE I. WILLIAMS COUNTY, NORTH DAKOTA TEST FIELD

AND STUDY AREA RESULTS

CROP CATEGORY	TEST FIELD ANALYSES				3 X 13KM STUDY AREA		
	OMISSION		COMMISSION		COMPUTER CLASSIFIED AREA IN HECTARES (ACRES)	GROUND TRUTH AREA IN HECTARES (ACRES)	% CORRECTLY CLASSIFIED
	PIXEL NO. CORRECT	% CORRECT	PIXEL NO. INCORRECT	% ERROR			
SMALL GRAINS	212	99.5	8	3.8	1567 (3871)	1578 (3899)	99.3
FALLOW	188	87.0	9	4.2	1406 (3475)	1432 (3538)	98.2
SOD	196	93.3	6	2.9	975 (2409)	1064 (2628)	91.7

TABLE II. MELFORT, SASKATCHEWAN TEST FIELD AND

STUDY AREA RESULTS

CROP CATEGORY	TEST FIELD ANALYSES				3 X 13KM STUDY AREA		
	OMISSION		COMMISSION		COMPUTER CLASSIFIED AREA IN HECTARES (ACRES)	GROUND TRUTH AREA IN HECTARES (ACRES)	% CORRECTLY CLASSIFIED
	PIXEL NO. CORRECT	% CORRECT	PIXEL NO. INCORRECT	% ERROR			
SMALL GRAINS	189	87.5	6	2.8	1656 (4094)	1675 (4140)	98.9
FALLOW	217	100.0	5	2.3	1632 (4033)	1696 (4190)	96.0
RAPE	216	99.5	6	2.8	644 (1592)	652 (1610)	98.9

Figure 1. MSS band 5 monitor image display (15km x 15km) containing Williams County study area.

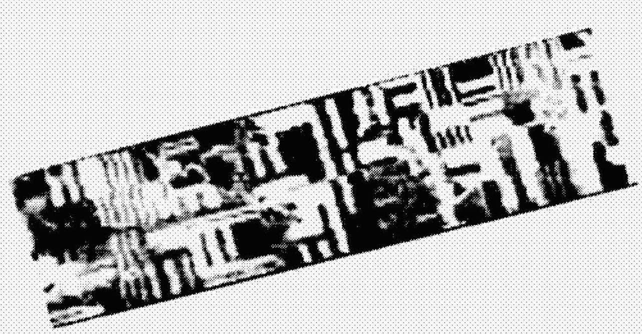
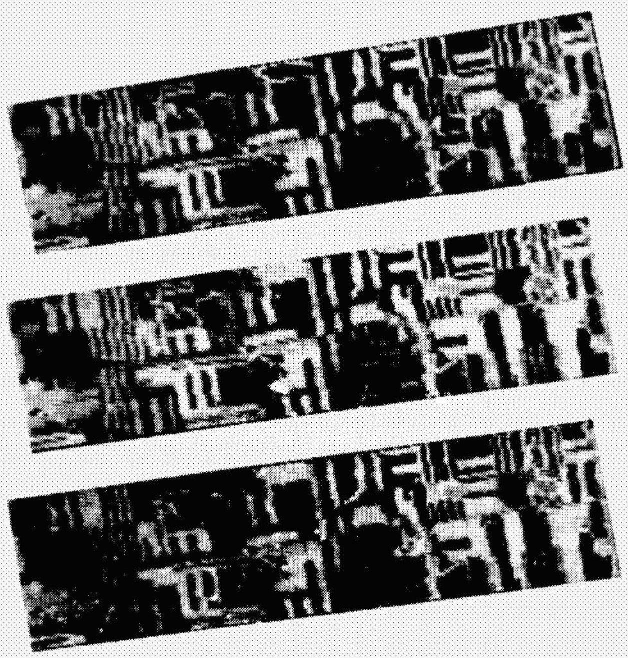


Figure 2. MSS band 7 monitor image display of 3 x 13km Williams County study area.

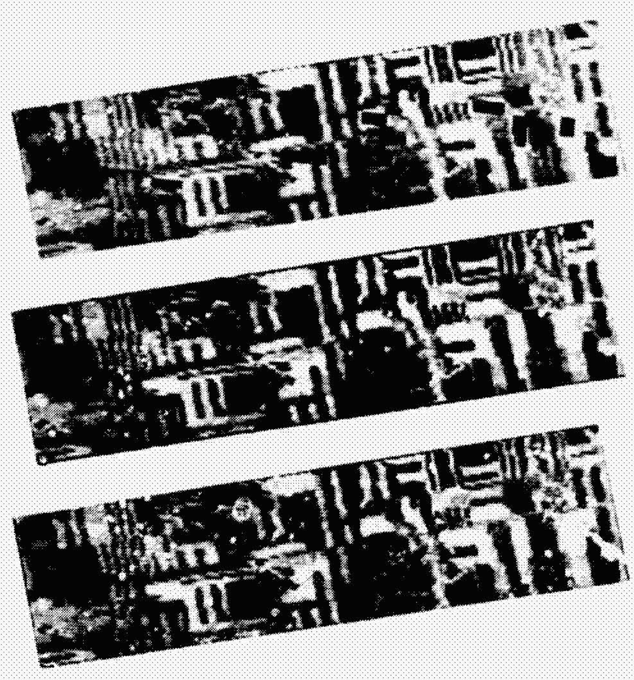


(a)

(b)

(c)

Figure 3. Williams County training fields (black) for (a) small grains, (b) fallow, and (c) sod.



(a)

(b)

(c)

Figure 4. Williams County test fields (black) for (a) small grains, (b) fallow, and (c) sod.

(a)

(b)

(c)

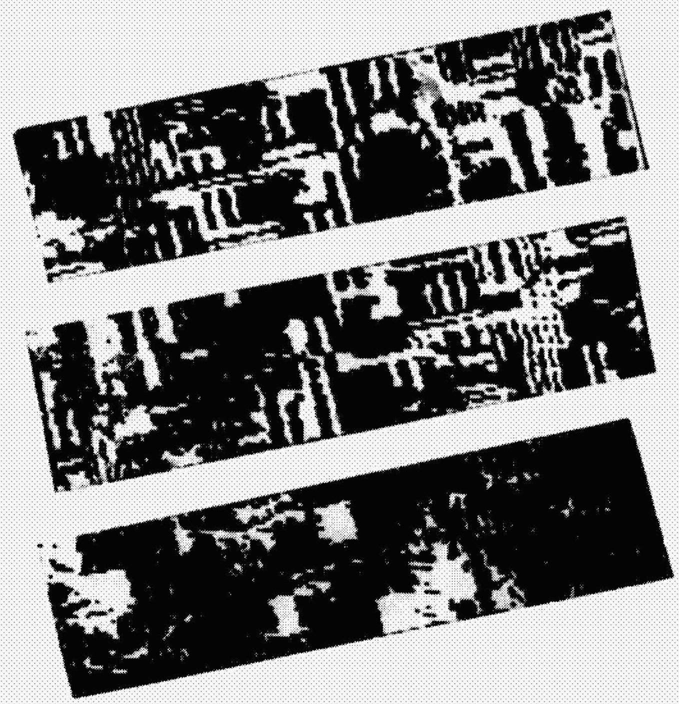


Figure 5. Classification results (white) for Williams County study area (a) small grains, (b) fallow, and (c) sod.

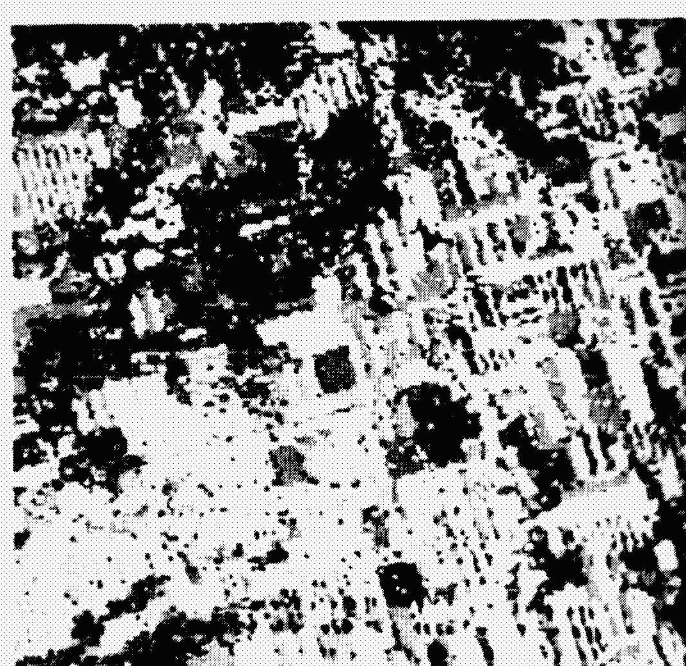


Figure 6. Classification results for 15km x 15km image display of Williamd County containing study area: small grains (white), fallow (medium gray), sod (dark gray), and unclassified (black).

(a)

(b)

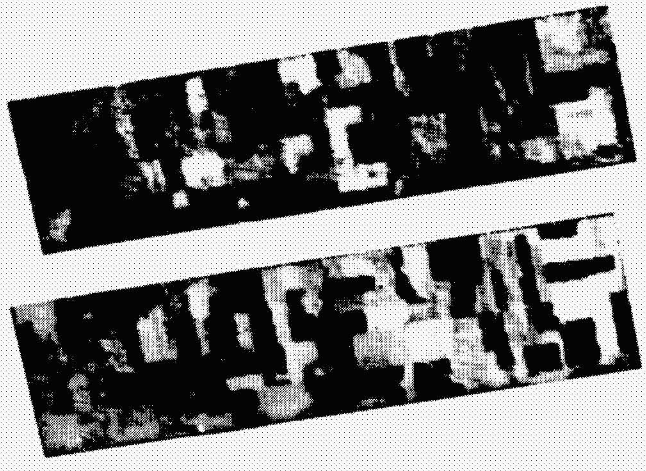


Figure 7. Monitor image display of Melfort study area: (a) MSS band 5 and (b) MSS band 7.

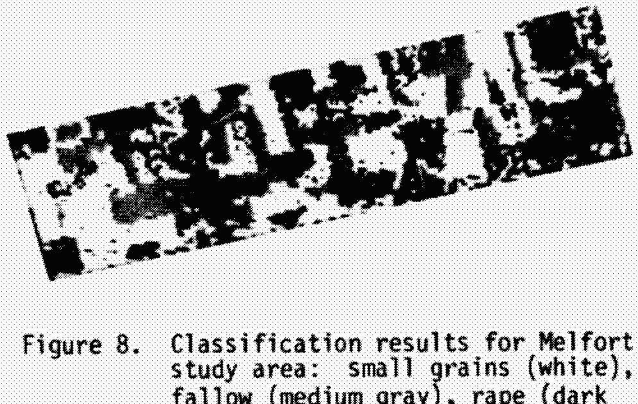


Figure 8. Classification results for Melfort study area: small grains (white), fallow (medium gray), rape (dark gray), and unclassified (black).