LANDSAT Registration Methodology Used by U.S. Department of Agriculture's Statistical Reporting Service 1972 -1982 *

Paul W. Cook

Applications Section
Remote Sensing Branch
Research Division
Statistical Reporting Service
United States Department of Agriculture
Washington, D.C. 20250

May 1982

* Presented at the 1982 American Statistical Association's Meetings in Cincinnati, Ohio, August, 1982

TABLE OF CONTENTS

ABSTRACT

Utilization of the National Aeronautics and Space Administration's (NASA's) LANDSAT satellite digital data for crop acreage estimation by the U.S. Department of Agriculture's (USDA's) Statistical Reporting Service (SRS) has been made possible by the development of methodology to locate quite accurately (i.e., register) each LANDSAT scene to a map base. The purpose of this paper is to present a history of the development of this methodology and explain how linear regression analysis is used to obtain a relationship between the LANDSAT digital data and U.S. Geological Survey maps. Results from numerous registrations within the U.S. are used to compare three competing procedures and thereby choose the most accurate. This competition indicated that the use of a full term third-order linear polynomial would be best for SRS's location of ground information as well as to estimate county locations within the digital data.

KEYWORDS: Satellite, Digital Data, Linear Regression Analysis, Polynomial.

ACKNOWLEDGMENTS

As is always the case in any report which is generated by a research project of any magnitude, others in the Remote Sensing Branch have made various contributions to insuring that accurate and complete information is available or in doing much of the preparatory work involved in seeing that an analysis may proceed. For this reason I wish to extend my thanks to Adam Orita (formerly with SRS) who did many of the scene registrations mentioned in this report. Also, George Harrell and Sandy Stutson of our materials group contributed by digitizing the segment and strata data-thereby making further analysis possible. My thanks go to Richard Sigman, Bob Starbuck (formerly with SRS), Martin Ozga and Ray Luebbe as well for their addition of various computer software systems to allow more accurate and rapid registration to proceed.

Others outside our group have also aided. Walt Donovan formerly of the University of Illinois's Center for Advanced Computations has developed and refined our capabilities at BBN to do the registration work more rapidly and easily. Ed Schlosser of Lockheed, Johnson Space Center, has proved to be an invaluable resource of the last few years -- one who has always been willing to aid in every way possible. James Schoonmaker of USGS also taught control point location procedures.

The impact which a report has is also a function of its physical appearance and format. Consequently, I wish to thank Kathy Reiman and Bessie Johnson for their thorough and painstaking approach to preparing the typed form of this report.

I. INTRODUCTION

From the inception of NASA's LANDSAT sensor program in 1972, SRS's Remote Sensing Branch has been involved in the study of estimation of crop acreage by means of the data beamed to Earth from the MSS (Multi-Spectral Scanner) on board the spacecraft [20]. A regression estimator is used by combining SRS ground data and the LANDSAT data to produce acreage estimates [9]. Of central importance to that effort has been the use of SRS data from the JES (June Enumerative Survey) - a yearly area sampling frame survey designed to collect crop information from randomly selected areas (called segments) within the U.S. [10].

The JES is part of the on-going program within SRS to estimate crop production and other farm related data (cattle, hogs, and economic data) for the U.S. There are 44 state offices which are responsible for collecting, editing and analyzing farm data and disseminating reports detailing results of the surveys. The JES is just one of the methods used to obtain information, however, it is crucial to SRS's use of LANDSAT data because it provides the ground information needed to correlate the satellite observed data with crop and land cover types.

The JES provides information collected by a staff of trained enumerators sent out to each segment to obtain personal interview data from each farm operation within the randomly selected areas. Included in this information is the size, location (marked on an aerial photo), and crop type of each field within an outlined area of approximately 1 square mile for intensive agricultural areas.

The nature of the JES segments has spurred the development and search for a satisfactory method of geographic location of each LANDSAT scene (that is, registration). Because they had been randomly selected throughout each state, there were no fast and easy methods to locate each of the often 60 square mile areas within the 13,000 square miles of a LANDSAT scene. Simply printing out graphic representations of the LANDSAT data, called grey-scales, for each area within the scene did not provide rapid nor accurate enough location. Had fewer areas been required within each scene or had the segments been closely grouped together, the difficulties associated with segment location would have been reduced. The method of registration, which has been developed as described in Appendix B, has also provided a means of utilizing coordinate digitized locations of county and strata boundaries so that crop acreage estimates can be made at county and stratum levels.

SRS uses a regression estimator approach to produce acreage estimates for each crop type by developing a regression between ground gathered area information and the equivalent categorized LANDSAT area (see Glossary). A multivariate normal (Gaussian) classifier rule accomplishes the categorization of approximately each .8 acre according to the crop types found within the ground segments used for training the classification rule [11]. For each .8 acre, there are four types of energy readings. Thus for each crop type the sufficient statistics are the four dimensional mean vector and covariance matrix. This methodology is described in much greater detail in a previous paper [9]. The goal of this paper is to describe how SRS personnel have located ground features on LANDSAT data.

The relationship calculated from the sample segments and classifièd LANDSAT pixels for each crop type is used to expand crop area to stratum, substate and state levels by the following formula.

$$Y_r = \sum_{n=1}^{L} N_h \left(\bar{y}_h + \hat{b}_h \left(\bar{X}_h - \bar{x}_h\right)\right)$$

where $\hat{b}_h$ is the estimated regression coefficient between classified pixels and sampled ground acreage,

$\bar{y}_h$ is the average of the ground gathered crop acreage for the h^{th} stratum,

$\bar{X}_h$ is the average number of classified pixels of the given crop per sample unit for all units in the h^{th} stratum,

$\bar{x}_h$ is the average number of classified crop pixels for the sampled units in the h^{th} stratum,

and N_h is the total number of units within stratum h [9, 21].

After experimentation with various techniques of locating specific crop areas within the LANDSAT data, SRS has found that the best procedure is to register the LANDSAT data to a map base by means of a bivariate polynomial regression analysis. The objectives achieved by this registration include: (1) accurate location of training and test data, (2) determination of county, land use strata and count-unit (sub-strata) boundaries, and (3) the ability to aggregate estimates by whatever area (or subarea) as may be deemed appropriate [5]. Since only a polynomial equation is required to relate the satellite and map data, it is not necessary to resample the LANDSAT data in any way; instead it is left in its original format as obtained from the EROS data center.

Another consideration for the use of registration methods was the need for accurate location of each pixel to be input into the classification algorithm. The inclusion of pixels from either boundary areas or other crops into the training signature of each crop will cause greater crop classification errors and reduce the overall effectiveness of the LANDSAT data. Consequently, a modification of the registration procedure was introduced to allow for shifting the location of each segment as needed for proper location of the fields' data. The aerial photo (see Glossary) and field level crop data of each segment, map data where necessary, and a grey-scale print-out (see Glossary) of the area about each segment are used in adjusting the position of each segment by the lightness - darkness values associated with fields of different crops.

SRS's present method of registration has evolved from work done by researchers both within and outside other government agencies [1, 2, 5, 13, 18]. Initially, only a linear first order polynomial was used as was done by others during the 1974 and 1975 period. However, this first order linear polynomial was found to be effective over only a 1000 row by 2000 column (see Glossary) portion of the LANDSAT scene. This made it more difficult to consider areas covered by multiple scenes and led to consideration of higher order polynomials. Experimentation with such polynomials led to the eventual adoption of the third order polynomial during the projects done in 1977 using 1976 collected data. Extensive testing of the third-order polynomial was done by SRS to compare it with a modeling method which had been implemented at Johnson Space Center [7]. The results of this test are presented in the next section.

II. LANDSAT Scene-to-Map Registration

A. Original X-Format LANDSAT Data

Every method used by SRS researchers to register a given LANDSAT scene to a map base has made use of ground control or registration points because accuracy of segment location has been of paramount concern. Selection of ground control points is consequently an important component in registration and can help determine how well the resulting calculated coefficients can be used.

Various types of areas have been shown to be of value as control points since they can be easily selected both within the LANDSAT data and within U.S. Geological Survey (USGS) maps. Of most value in use as registration points are road intersections, streams with a sharp bend, centers of small lakes, and centers of small forested areas. Determination of an accurate location for such points is necessary both within the LANDSAT digital data (grey-scales) and the USGS maps.

Primarily, the improvements in registration accuracy have been made by selection of better ways to analyze the relationship between the selected locations of the registration points within the LANDSAT data and the USGS map base. Actual selection of the points has remained very much dependent upon the individual who chooses them. No method of improved analysis capability will make up for improper or imprecise choice of the control points. Major constraints to the registration accuracy are made by the LANDSAT data itself and the USGS maps [Appendix A].

Not all LANDSAT data can be registered to map base with the same degree of accuracy. Sometimes, internal problems within the sensor or in the LANDSAT ground processing systems have caused breaks within the geometric integrity of the data, thereby causing troubles in registration. This was particularly evidenced in early 1980 data, but it was corrected for the 1981 data. In some areas USGS maps do not meet the National Map Accuracy Standards [22], often because of age. In some instances, the maps may simply not have useful features that are needed for registration. Such occurrences frequently contribute to rather wide variations in the accuracy of the registration within a given scene.

When a linear first order polynomial was used to register small areas within a LANDSAT scene in Texas for the 1974 Texas project [3], these limitations in registration accuracy were first encountered by SRS. Extension to larger areas by means of the linear first order polynomial proved to be too inaccurate, since there were not enough terms to model the sensor characteristics across the entire LANDSAT scene.

Full scene registration errors using the first order linear polynomial were usually in excess of 300 meters RMS (Root Mean Square Error, see Glossary) and exhibited excessive maximum errors in row and column. Usually, 1000 rows by 2000 columns (see Glossary) was the largest area for which RMS errors of less than 100 meters was possible. This size area proved to be too small for practical use. Segments are normally scattered throughout the entire scene and require that the full area be accurately registered.

The non-linear sinusoidal nature of the motion of the MSS's rotating mirror quite naturally imparts higher order irregularities to the array of pixels (picture elements) obtained from the satellite. Indeed, the pixel size varies across every line since the amount of overlap between pixels is determined by the speed of the mirror's movement [Appendix A]. Consequently, it is intuitively evident that higher order linear polynomials should be more appropriate in modeling such irregularities since they would be able to account more fully for the effects of the non-uniform motion of the mirror assembly. Baker in [1] presents a cogent justification for the use of polynomial equations for aircraft sensor arrays that is also applicable to LANDSAT MSS data.

Following this line of reasoning, the second order linear polynomial was used during the 1975 Illinois project on the advice of researchers at the University of Illinois' Center for Advanced Computation (CAC). Also, as part of the registration process, the LANDSAT data was deskewed, i.e. made to resemble more a north-south orientation so that it more closely overlaid maps of the area.

The CAC deskewing algorithm required the location of control points on both the LANDSAT photo (1:500,000 scale) and USGS maps of at least 1:250,000 scale. After selection of at least 10 such points, coefficients for a first order linear polynomial were calculated.

Two deskewing coefficients were calculated concurrently to determine the angle at which the scene would be resampled [4]. An additional program was next run on an IBM 370 to accomplish the final scene transformation based upon the calculated coefficient. The resampled data array of LANDSAT data output from this program was used to print out the grey-scales within which ground control points would be selected.

The deskewing algorithm caused some problems with creating a final registration file and its use was abandoned in the 1978

Iowa study after it was learned how to select parameters to prevent the data from being resampled onto a new grid by a nearest neighbor resampling method. Also, since the map and segment data could be mathematically warped to correspond to the satellite's coordinate system, this step had been unnecessary anyway -- especially since SRS had no need to provide map-like data output products at that time.

Continued efforts to improve the method of registration resulted in two major changes during 1976 and 1977. First, the third order linear polynomial was tried on scenes previously registered by the second order polynomial and was found to give lower root mean square errors (RMS error, see Glossary). It also was observed that segments needed less movement when the third order polynomial was used rather than the second order. The only drawback to the use of the third order polynomial was the need to provide more degrees of freedom (that is, registration points) for its 2 sets of 10 coefficients to be reliable.

The second change in registration capability during this time period was the partial implementation of the DAM-COEFF procedure from the DAM program written by Ed Schlosser at NASA/JSC [6, 17]. Although it was not made capable of estimating the segment locations, the DAM-COEFF procedure was set up to analyze registration points used and select outliers to be deleted. Its calculations required the scene nadir (see Glossary), scene center, and heading of the satellite for each scene. This information was used to model the satellite's characteristics (roll, pitch, yaw, and altitude) and correct the LANDSAT pixel sizes (mathematically) to conform to a pseudo-UTM coordinate system. In this way, higher order components of the satellite's characteristics were removed and a linear first order polynomial fit made to the map base. Theoretically, this method was thought to require fewer points (in the range of 16) to effect a successful registration.

Since three competing methods were available by the end of 1977, a test of all three was devised to select the best procedure for use during SRS's 1978 Iowa Project [9]. Numerous control points had been selected in each scene during recent projects and the decision was made to use randomly chosen points located throughout the scene as check points to determine the accuracy of each method.

Two methods were used in testing the three procedures. The first compared the second order, third order polynomials, and DAM-COEFF using a reduced number of control points. The second compared the third order polynomial and DAM-COEFF when a larger number of control points were available.

First, all the available LANDSAT scenes that had been registered by late 1977 were reexamined and all available precision control points were used in the testing. To keep the test as fair as possible and to exclude any outliers, all three methods were used to select the worst 10% of the control points available. These worst 10% of points for all methods were deleted from use and the remaining points used to provide both control and check points.

In Table I the results of this first test are listed. RMS (root mean square error, see Glossary) values in meters are ordered according to the best to worst results for the third order polynomial. This test also was meant to determine registration accuracy when using only a minimal grid of 18-25 points. This number of points was chosen since it was presumed that DAM-COEFF would give accurate results with as few as 16 points. It was hoped that 25 points would be sufficient for the third order linear polynomial as well.

The RMS accuracies are given only for the check points (not for control) because the check points should give a better assessment of the actual accuracies achieved. As can be noted, the number of check points ranged from 27 to 58 with most scenes having in excess of 30 check points. The accuracies for the second order linear polynomial are in every case the worst values for the three methods. Consequently, it was concluded that the second order linear polynomial is the least desirable method, so it was no longer used.

The remaining two methods, the third order linear polynomial and DAM-COEFF, were then compared non-parametrically by means of the sign test. There were four cases in which the DAM-COEFF algorithm had proved to be superior. It can be concluded at the $\alpha = .0923$ level ($\alpha_1 = .0730$, $\alpha_2 = .0193$) for the 12 nontied observations that the two methods are equally effective since 4 falls between 3 and 10. This result is based on the assumption that neither method had been presumed to be better than the other initially.

Another test was devised in an attempt to examine the differences between the third order linear polynomial and DAM-COEFF. In this test, a minimal set of check points consisting of 20% of all the available ground control points was chosen after the deletion of the 10% worst points as determined by both methods concurrently. Three different systematic samples of 20% of the control points were selected from each LANDSAT scene as a means of averaging out the effects of choosing different control sets. These results are presented in Table II.

As was done for Table I, the results for the two methods are presented in order of the best to worst RMS values for the third order polynomial. Again making use of the sign test it is observed that there are four of the observations for which DAM-COEFF is superior and one tie. Therefore, we accept the two methods as being equal at the $\alpha = .0768$ level (i.e. $\alpha_1 = .0592$ and α_2 .0176) since 4 lies within the interval of 4 and 11.

Because neither test indicated the third order polynomial was inferior to the use of DAM-COEFF, the decision was made to continue the use of the third order linear polynomial because it was easier to implement. The poor showing of the second order linear polynomial during the first test, however, was sufficient to exclude it from future use.

An examination of Tables I and II allows the comparison of results using the third order linear polynomial when the number of check points used is increased. These results are listed in Table III for the scenes done using a different number of control points.

Use of the sign test, under the assumption that more points should give greater accuracy, indicates that it can be concluded at $\alpha = .1334$ that the larger control point set is no better than that of the 18-25 control point set. Consequently, the number of control points used in future studies has been maintained at the level of 25 - 36 points since this number appears to be adequate to maintain accurate registration when utilizing the third order linear polynomial.

These results had not been anticipated, since it had been assumed that finding more control points would lead to much greater accuracy in the predictions of the locations of all areas within the scene. Such was not the case for the scenes that were examined. Therefore, it was concluded that a smaller number of well-chosen points should be adequate for full scene registration. Generally, it is best to select extra points which may be easily deleted if needed to insure that the final number of control points will equal or exceed twenty-five points.

TABLE I
Check Point RMS Results for 25 or Less
Control points on 13 LANDSAT Scenes

Scene	Third Order Linear Polynomial	DAM COEFF	Second Order Linear Polynomial	No. of Control No. of Check (No. of Deletes)
2537-17480	45	104	184	25 (0 delete) 58 check
2228-15524	53	81	169	25 (0 delete) 35 check
2897-17332	54	63	189	25 (0 delete) 58 check
2211-15580	55	112	150	25 (0 delete) 35 check
2175-15592	60	64	182	25 (0 delete) 33 check
2450-16232	73	80	171	25 (0 delete) 56 check
2450-16235	114	114	248	24 (1 delete) 27 check
2435-16404	125	122	256	20 (0 delete) 32 check
2435-16410	132	108	235	20 (0 delete) 30 check
2470-16344	209	215	320	19 (1 delete) 35 check
2435-16413	255	217	426	22 (3 delete) 61 check
2470-16342	294	274	443	19 (1 delete) 35 check
2470-16335	464	478	734	18 (2 delete) 35 check

TABLE II
Registration Accuracy Meters RMS Error of Check Points
Averaged from 3 check sets of 20% of all Control Points

Scene	Third Order Polynomial	DAM-COEFF
2537-17480	43	116
2897-17332	47	67
2211-15580	53	131
2175-15592	58	70
2467-16714	60	70
2228-15524	62	84
2228-15531	69	173
2450-16232	72	87
2435-16404	106	97
2435-16410	117	117
2450-16235	122	127
2470-16342	204	203
2435-16413	216	236
2470-16344	257	261
2467-16165	366	360
2470-16335	549	502

TABLE III
-Third Order Polynomial-
Minimal and Maximum
Control Point Sets

Results from TABLES I and II

Scene	18-25 Control Points	35-64 Control Points
2537-17480	43	45
2228-15524	62	53
2897-17332	47	54
2211-15580	53	55
2175-15592	58	60
2450-16232	72	73
2450-16235	122	114
2435-16404	106	125
2435-16410	117	132
2470-16344	257	209
2435-16413	216	255
2470-16342	204	294
2470-16335	549	464

B. New P-Format LANDSAT Data

During April 1979, an updated data format for the LANDSAT scenes became available at the U.S. Department of Interior's EROS Data Center. This format, the P-Format, replaced the original X-Format which had been used while SRS developed its registration procedures. NASA's Goddard Space Flight Center established the methods to implement this data format.

The new data format is changed considerably from that of the previous X-Format [20]. All the data have been resampled to provide a 56.9 meters by 56.9 meters pixel size (the previous size was 56.9 meters by 79.7 meters). Also, various systematic variations within the data are now removed as part of the routine processing. Major causes of distortion from the sensor (optics, scan mechanism, detector array geometry), the space craft (attitude and altitude variations), and terrestrial effects (for example, Earth rotation) are computed. Also, ground control points are used for some scenes to further correct the imagery and map it to the Hotine Oblique Mercator (HOM) projection using NASA's Master Data Processing System.

Because so many corrections have been made in processing the P-Format data, it had been hoped that registration of the LANDSAT data to map base would proceed much more rapidly and require use of only a linear first-order polynomial. This supposition was tested on a number of scenes, but it was soon evident that the third-order polynomial still performed significantly better. Consequently, the third-order linear polynomial has continued to be used in registration of the LANDSAT data obtained in the P-Format as well.

The changes made in processing the P-Format data tape have also resulted in the development of additional registration methods. Since the data are resampled before the photo product is produced, the LANDSAT photo product represents the true pixel locations quite accurately. Also, the tape records give coefficients for relating each pixel to its appropriate Hotine map coordinates. Both of these developments can be valuable in registration work.

The photo product may be used in registering the LANDSAT digital data to map base. A 1:250,000 scale photo product is used as a replacement for the grey-scale maps formerly used in registration. Many scenes have been registered using this method and a report presenting the results is in progress.

Use of the Master Data Processor produced registration is also being researched. If successful, this would eliminate SRS's need to locate control points and provide much faster registration. Research to implement the algorithms required in the operational system is now underway.

III. Segment Location on LANDSAT

A. Background

The segment location problem has been made much more manageable by the use of registration techniques. Several associated procedures were introduced by University of Illinois personnel to make the overall process proceed more effectively and efficiently. Further modifications of the procedures originally introduced in 1975 have primarily been made in the software to provide easier use whereas the overall plan of action is still intact.

During data collection for the JES, field locations for every segment within the sample drawn are obtained by field enumerators who draw the outlines onto aerial photographs. Although these photos are not always current, new aerial photography is flown where needed and careful editing is done to confirm that field boundaries are properly drawn.

Preparations for digitizing the photo segment data consist of drawing the field boundaries onto acetate overlays and then marking the points at which the digitizer should take readings. The Altek coordinate digitizer used in this operation is accurate to $\pm$0.005 of an inch and takes each reading in 0.001 of an inch increments. Each segment data file consists of these vertices coordinates to define the polygons which approximate the field boundaries.

To establish the relationship between each aerial photo and its matching USGS quadrangle map, corresponding points are chosen both on the photo and map. These points are digitized and a least squares linear fit is made to determine the best coefficients to be used in calculating the (latitude, longitude) coordinates for each digitized point. The scene registration file will then be used to predict the location of the digitized segment and field boundaries within the LANDSAT digital data.

B. Accuracy of Segment Location

Because the LANDSAT scene registration and the segment calibration contain some inaccuracies, provision has been made for each segment to be shifted to the area which most closely matches its field patterns. This is done by the use of segment plots which are used to overlay the computer printout of the area within which the segment is predicted to be contained. Printed on each plot is the location of a corner point given in LANDSAT row, column coordinate values. Field boundaries and other segment information is also printed (see figure 1).

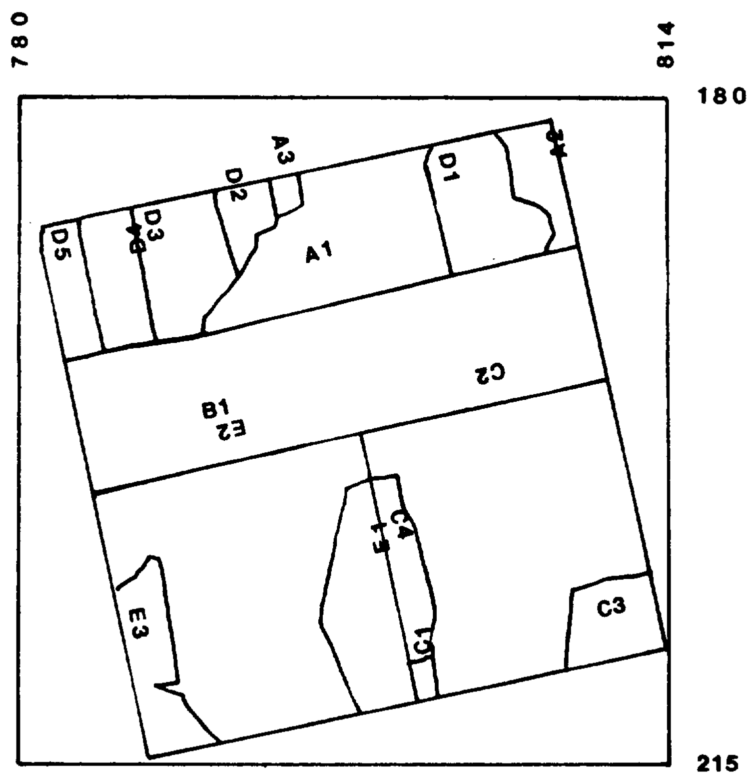


Figure 1. Plot of Digitized USDA/SRS Kansas Segment 305

When checking a segment LANDSAT plot, the analyst will use the segment photo, ground truth listing, segment plot drawn to LANDSAT scale, and the USGS quadrangle map of the area to aid in determining the exact position that is the segment's perceived location. When the scene registration and segment calibration are both accurate, it is usually no more than a ±2 pixel movement in either row or column that gives an exact match. Movements larger than this are generally caused by poor calibration of the photo to the map. Should this occur, the segments are usually recalibrated or both redigitized and recalibrated as needed.

The grey-scales provide a means of checking the lightness and darkness patterns within the LANDSAT data. Usually data from Band 5 is used in this part of the procedure. The amount of movement needed is then either noted by a count of the number of pixels moved in row or column (nearest 0.5 pixel) or digitized. A file of shifts is then made and individual segment calibration files are generated for each segment which has been shifted.

IV. CURRENT RESEARCH

Personnel with NASA's Earth Resource's Laboratory (ERL) are currently working with SRS to improve SRS's registration methodology. They have examined both the scene-to-map procedures and the scene-to-scene or multitemporal registration as well. Preliminary results of their work are available now.

The ERL researchers have developed two methods of improving the LANDSAT scene-to-map registration. They have also examined the accuracies of currently available scene-to-scene registration algorithms.

An examination of the MSS P-format data registration has shown what accuracies can be expected from the NASA supplied tick mark registration [8]. Since this registration is available with every scene, the effort required in locating registration points could be eliminated in those cases for which the registration is sufficiently accurate. For the 12 scenes that were examined, it was found that 8 would provide sufficient accuracy to locate SRS segments within a search window of 10 LANDSAT columns by 10 LANDSAT rows. Should further tests show this ratio of usable registrations to hold for those scenes used in SRS studies, then the P-format registration will be used to provide the starting point for a segment matching algorithm.

An Automated Segment Matching Algorithm (ASMA) has been developed by ERL to establish a match between the digitized segment boundaries and the LANDSAT digital data [7]. Development of this procedure required that an algorithm be determined that provided a means of edge detection within the LANDSAT pixel data. Instead of manually moving the digitized segment outline of a segment over a printed grey-scale, the computer shifts the segment outline over a 10 pixel row by 10 pixel column window to match gradient values calculated from the LANDSAT data. Preliminary testing of this method has indicated that appropriate criteria for matching can be determined. Further testing is needed to determine the reliability of correct shifting as compared with that done by hand.

Use of the P-format data registration in conjunction with the ASMA should dramatically reduce registration times. For those cases for which the P-format registration is too inaccurate, scene-to-map registration can still be performed using a 1:250,000 scale LANDSAT photo and 1:250,000 scale USGS maps. Although somewhat less accurate than registration done with grey-scales, scene registration can be done much more quickly. Typically, registration accuracy will be within 6 pixels in either row or column - quite sufficient for use with the ASMA.

Of course, the ASMA has criteria for determining when segments are sufficiently well matched with the LANDSAT data. Those segments incorrectly matched will be checked by hand. When the P-format registration is shown to be incorrect, the photo-to-map method will be used.

Scene-to-scene registration has also been extensively researched in cooperation with ERL [19]. Preliminary results indicate that present algorithms provide approximately 33-40 meter accuracy within the LANDSAT scene. Plans to improve the scene-to-scene registration methods will not be formulated until a full review has been conducted of research done to this time.

Glossary

1. Aerial Photography - a photo taken by low flying aircraft so that the fields are readily visible. In some cases color infra-red photos are obtained in this manner, however, SRS normally uses black and white photos for regular surveys obtained from ASCS (Agricultural Stabilization and Conservation Service, USDA).

2. Band (channel) - A designation for a portion of the light spectrum sensed by the MSS (multi-spectral scanner) for which digital data is provided. The sensors provide data from 0.5 - 0.6 μm , 0.6 - 0.7 μm , 0.7 - 0.8μ m , and 0.8 - 1.1μm which are called bands 4, 5, 6, and 7 (or alternatively channels 1, 2, 3, and 4, respectively).

3. Digitizer - an electronic device used to measure a position within coordinate axes quite accurately (often to within 0.001 inch).

4. Grey-scales - Assignment of pixel values to computer print-out symbols of varying levels of lightness or darkness makes possible the presentation of the CCT (Computer Compatible Tape) as a computer generated picture.

5. LANDSAT Scene - The specific area of the Earth's surface covered by a given LANDSAT CCT.

6. Pixel (picture element) - A four-tuple of data imaged by the MSS containing the 4 bands of data on the earth's surface. The P-format data resamples this to 56.9 meters by 56.9 meters. The X-Format data had a 79.1 meter by 56.9 meter sized pixel.

7. Registration (control) points - accurately locatable points within both the LANDSAT digital data and the USGS (U.S. Geological Survey) maps which cover the chosen area.

8. Row and Column - These are the names for the along track (row) and across track (column) pixel coordinate grid. P-Tapes contain 2983 rows and 3548 columns whereas the X-Format had 2400 rows and 3240 columns.

9. RMS (root mean square error) - A measurement of the accuracy of a least squares fit to the data under analysis. It is calculated from the formula:

$$RMS = \sum_{i=1}^{n} \frac{(\text{Actual-Predicted value})_i^2}{n}$$

where n = the number of observations.

10. Scene Nadir - The point on the earth's surface at which a tangent line to the earth would be perpendicular to the direct line from the satellite.

REFERENCES

1. Baker, J.R., E.M. Mikhail, Geometric Analysis and Restitution of Digital Multispectral Scanner Data Arrays, Purdue University, Laboratory for Applications of Remote Sensing, West Lafayette Indiana, LARS Information Note 052875, 1975.

2. Bernstein, R., "Digital Image Processing of Earth Observation Sensor Data," IBM Journal of Research and Development, 20 (1), 1976, pp 40-57.

3. Cook, Paul W., Texas '74 Remote Sensing Project, USDA Statistical Reporting Service, August, 1977.

4. Donovan, Walt, Oblique Transformation of ERTS Images to Approximate North-South Orientation, Center for Advanced Computation, University of Illinois at Urbana-Champaign, Urbana, Illinois, November 1974, CAC Technical Memorandum, No. 38.

5. Donovan, Walt, and Martin Ozga, Retrieval of LANDSAT Image Samples by Digitized Polygonal Windows and Associated Ground Data Information, CAC, University of Illinois at Urbana-Champaign, August 1975, CAC Technical Memorandum No. 57.

6. Foreman, Grant, "PROC DAMCOE," USDA, Statistical Reporting Service, May 1976.

7. Graham, M.H., An Algorithm for Automating the Registration of USDA Segment Ground Data to LANDSAT MSS Data, National Aeronautics and Space Administration, National Space Technology Laboratories, Earth Resources Laboratory, December, 1981.

8. Graham, M.H., and R. Luebbe, An Evaluation of MSS P-format Data Registration, National Aeronautics and Space Administration, National Space Technology Laboratories, Earth Resources Laboratory, and U.S. Department of Agriculture, Economics and Statistics Service, Research Division, April, 1981.

9. Hanuscak, George, and others., Obtaining Timely Crop Area Estimates Using Ground Gathered and LANDSAT Data, USDA, Economics, Statistics, and Cooperatives Service, August, 1979.

10. Houseman, Earl E., Area Sampling Frame in Agriculture, USDA, Statistical Reporting Service, November, 1975.

11. Lachenbruch, Peter A., Discriminant Analysis, Hofner Press, 1975.

12. Malila, William A., Ross H. Hieber, and Arthur P. McCleer, Correlation of ERTS/MSS Data and Earth Coordinate Systems, Environmental Research Institute of Michigan (ERIM), August 1973.

13. Malila, W.A., J.M. Gleason, and R.C. Cicone, Investigation of Spatial Misregistration Effects in Multispectral Scanner Data, Environmental Research Institute of Michigan (ERIM) May 1976.

14. Malila, W.A., R.H. Hieber, R.C. Cicone, Studies of Recognition With Multitemporal Remote Sensor Data, ERIM, May 1975.

15. Ozga, Martin, Walter E. Donovan, and Chapman Gleason, An Interactive System for Agricultural Acreage Estimates Using LANDSAT Data, cooperative paper between CAC, University of Illinois (Ozga, Donovan) and USDA, Statistical Reporting Service, Washington, D.C., March 1977.

16. Ozga, Martin, Selection, Sampling, and Tabulation of Masked Files in EDITOR, CAC Technical Memorandum No. 79, Center for Advanced Computation, University of Illinois at Urbana-Champaign, Urbana, Illinois, December 1976.

17. Schlosser, Edward H., and M.L. Brown, Detection and Mapping (DAM) Package, NASA, Johnson Space Center, Houston, Texas, June 1976 (revised).

18. Schoonmaker, Jr., James W., "Geometric Evaluation of MSS Images from ERTS-1".

19. Seyfarth, B.R., and P.W. Cook, An Analysis of LANDSAT MSS Scene-To-Scene Registration Accuracy, National Aeronautics and Space Administration, National space Technology Laboratories, Earth Resources Laboratory, and U.S. Department of Agriculture, Statistical Reporting Service, Research Division, December, 1981.

20. Thomas, Valerie, L., Generation and Physical Characteristics of the ERTS MSS System Corrected Computer Compatible Tapes, Goddard Space Flight Center, July 1973.

21. Von Steen, Donald H., and William H. Wigton, Crop Identification and Acreage Measurement Utilizing LANDSAT Imagery, USDA-SRS, March 1976.

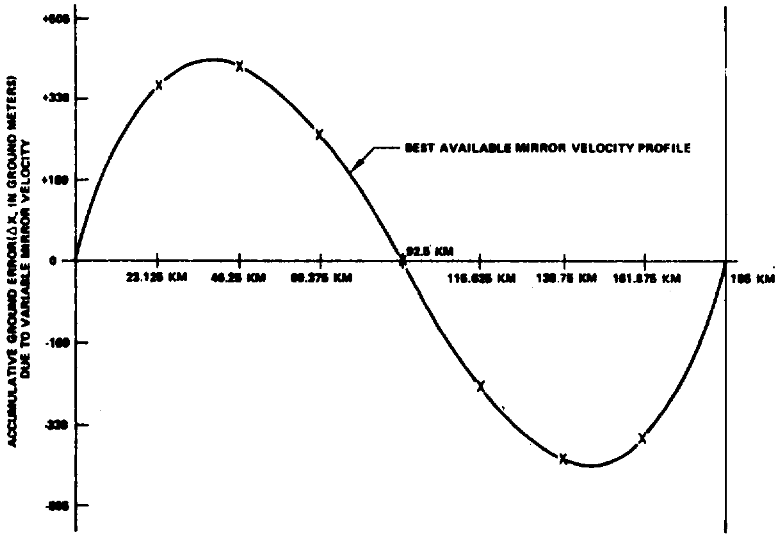
22. Topographic Maps, U.S. Geological Survey, pamphlet, 1975.

Appendix A

Much of the detailed operating characteristics of the MSS scanner are clearly presented in the ERTS Users Handbook, available from NASA. Of interest to us, however, are the characteristics which affect geometric accuracy and digital classification techniques. For example, sampling is timed so that sensors from all 4 bands observe the same area of ground within 0.1 pixel.

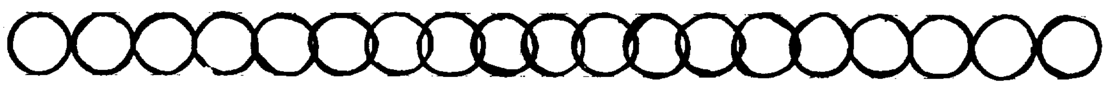
Also of importance in this respect is the operation of the oscillating flat mirror which directs incoming light onto the square optical fiber ends of the twenty-four detectors. Only a $\pm$ 2.89 degree movement is necessary to scan the entire 185 km cross-width swath. Because its motion is not constant, however, samples are not all equally divided up along the swath. Figure 1 clearly indicates the error in distance across-track caused by this kind of motion whereas figure 2 indicates the physical overlapping which occurs.

Figure 1. Across Track Errors Caused by the Variable Mirror Velocity Profile



Mirror Velocity Profile for the Active Mirror Scan

Figure 2. Overlay of Pixels, Corresponding to a Variable Mirror Velocity



NOTE: 1. Although circles are used to represent the pixels, they are actually squares.

2. Not drawn to scale.

Appendix B

This section is meant to present a step-by-step procedure to be followed in doing the registration of a LANDSAT scene to USGS map base. This should further clarify the discussion of the registration methodology as presented in the body of the paper.

When greatest accuracy is desired, the scene-to-map method is done utilizing grey scales obtained from the LANDSAT digital data. When speed is of greater concern, then a 1:250,000 scale photo of the LANDSAT scene and its corresponding 1:250,000 scale USGS map are used instead.

Suppose we consider only the case of the utilization of grey-scales generated by assigning darkness values to print characters to represent each pixel location. The following steps would be followed:

(1) Using a black and white 1:1,000,000 scale transparency of either band 5 or band 7, initial control points would be chosen from among road intersections, road-river intersections, rivers, and other distinguishable features also on a 1:1,000,000 scale USGS index map. These points should be located on an approximate 6 x 6 or 7 x 7 grid throughout the scene.

(2) A Cartesian coordinate digitizer with 0.001 inch resolution would be used in determining the location of each potential registration point by digitizing both the photo and map used in step (1). This procedure will provide a row, column and latitude, longitude pair for each chosen point. These values are derived from a special program found in the system on which SRS does data analysis.

(3) The area about each point is expanded to fill a computer sized sheet (11 x 15 inches) and printed out again using SRS's analysis package.

(4) Also, output from step (2) is a map index for each point so that selection of the proper 7-1/2 minute or 15 minute USGS map can be made for the grey-scales print to be overlaid. Selection of an appropriate matching point on both the grey scales and map is done at this time.

(5) The coordinate digitizer and appropriate computer program are next used to determine the map coordinates for each point. The appropriate LANDSAT pixel location is determined by selection of the center of a LANDSAT pixel. The resulting output from this step is a file of corresponding points containing row, column, and latitude, longitude coordinates for each point.

(6) Analysis of the selected points is then accomplished by producing both a first order and a full term third order linear fit to the coordinates obtained in step (6). The first order polynomial is used to detect major errors and eliminate the possibility of the third order erroneously accepting bad points. The coefficients of the third order polynomial are saved for use in predicting the location of SRS segments and county boundary data.