Reprinted from

Symposium on

Machine Processing of

Remotely Sensed Data

June 29 - July 1, 1976

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
76CH1103-1 MPRSD

A LANDSAT DIGITAL IMAGE RECTIFICATION SYSTEM

Peter Van Wie and Maurice Stein

Goddard Space Flight Center
Greenbelt, Maryland

I. ABSTRACT

DIRS is a Digital Image Rectification System for the geometric correction of Landsat Multispectral Scanner digital image data. DIRS removes spatial distortions from the data and brings it into conformance with the Universal Transverse Mercator (UTM) map projection. Scene data in the form of landmarks or Ground Control Points (GCPs) are used to drive the geometric correction algorithms. The system offers extensive capabilities for "shade printing" to aid in the determination of GCPs. Affine, two dimensional least squares polynominal and spacecraft attitude modeling techniques for geometric mapping are provided. Entire scenes or selected quadrilaterals may be rectified. Resampling through nearest neighbor or cubic convolution at user designated intervals is available. The output products are in the form of digital tape in band interleaved, single band or CCT format in a rotated UTM projection. The system was designed and implemented on large scale IBM 360 computers with at least 300-500K bytes of memory for user application programs and five nine track tapes plus direct access storage.

II. INTRODUCTION

A. Overview

The need for geometrically corrected Landsat MSS digital data is strongly felt in a number of remote sensing application areas. A study of the 1973 and 1974 Mississippi River floods motivated the development of the Digital Image Rectification System at Goddard Space Flight Center.[1] This system has gone through several stages of development and is currently operational at Goddard. It is also offered for sale through COSMIC, the NASA software distribution facility at the University of Georgia. This paper will present the approach used in the development of DIRS, the distortion sources present in the data, the techniques which have been implemented, and the results achieved thus far.

B. Approach

The approach used in the development of DIRS was pragmatic. We were attempting to create a system which met our own needs within the context of available computing equipment. Our computer environment consists of two large scale IBM 360s: a model 91 and a model 75. Each has two megabytes of memory and a large collection of peripherals.

Image rectification, as opposed to registration, was desired since it allows image data to be combined with map data. Combinations such as photographic overlays were originally contemplated, but direct digital combination of map and image data was soon recognized as an important possibility. The rectification capability also allowed for multi-scene registration and thus was selected as the preferred approach to geometric correction of the MSS image data.

The first choice to be made was that of a map projection. The Universal Transverse Mercator (UTM) map projection was selected as the most appropriate. The UTM projection offers the advantages of a cartesian reference system with metric measure and is the primary projection for U.S.G.S. paper and digital topographic maps. See[2] for details on the UTM projection.

With these considerations in mind, we collected an assortment of existing techniques and implemented them. These techniques were derived from various sources including work done for NASA by IBM[3], TRW[4], CSC[5], as well as other routine photogrametric techniques and work done by the Defense Mapping Agency[6,7,8]. The development of new techniques was not our objective and was only employed where absolutely necessary.

C. Objectives

Our major objectives were:

1. To rectify full or partial Landsat MSS scenes with the maximum attainable accuracy and produce digital output products suitable for further machine processing and analysis.

2. To use available computing hardware, supporting software, and image processing techniques.

3. To produce efficient software which can process a full Landsat scene (30 megabytes of data) in an acceptable length of time.

4. To provide adequate flexibility to the user including the ability to trade off GCP location effort for accuracy to allow them to achieve the level of accuracy they require at the lowest cost.

III. DISTORTION SOURCES

Distortions exist in the MSS digital image data due to the combined effects of sensor operation, orbit and attitude anomalies, Earth rotation and from atmospheric and terrain effects. No attempt is made in DIRS to correct for atmospheric or terrain induced geometric distortions.[9,10]

A. Sensor Operation

1. Aspect Ratio. An aspect ratio of approximately 1.4:1 is introduced by the along scan oversampling of the MSS. The instantaneous field of view (IFOV) is 79 x 79 meters but along scan pixel separation (center to center) is only 57 meters. The line to line pixel separation is 79 meters giving no overlap or gapping in the cross scan direction.

2. Mirror Velocity Nonlinearity. The MSS oscillating scaning mirror used to sweep out the image lines does not maintain a precisely constant angular velocity. Therefore, along scan distortions are introduced causing pixel compression or pixel stretching at various places long the scan line. This is perhaps the most difficult distortion to correct since it is poorly measured and changes gradually with time. Two standard mirror models are selectable in DIRS and provisions are made for user defined mirror models.

3. Band to Band Misregistration. The MSS senses six scan lines in each of four spectral bands on each mirror sweep. The light reflected off the scanning mirror falls on the end of a 4 x 6 array of fiber optic tubes which conduct it to the spectral filters and detectors. Displacement along the four element axis of the fiber optic array induces an along scan offset in the observed ground position. This spectral misregistration is corrected (to within a pixel) by the introduction of pad pixels by the NASA Data Processing Facility (NDPF) where CCTs are generated from the video data tapes. A subpixel misregistration remains which is not corrected by DIRS.

4. Sensor Delay. The 24 MSS detectors are strobed sequentially by a precisely timed clock to initiate the A/D conversion of the radiance data. The small time delays between the readout of each of the six sensors in a given band allow the scanning mirror to move forward a small amount. This gives rise to a small along scan displacement from the first through the sixth scan line in each swath.

B. Orbit and Attitude Anomalies

Significant distortions are caused by inconsistencies in spacecraft orbit and attitude. The major component of these distortions is linear and can be corrected with a single affine transformation of the entire image using three ground control points (GCPs). However, small continuous variations in altitude, velocity, yaw, pitch and roll add nonlinear distortions on the order of 200 to 300 meters over the scene.

C. Earth Rotation

Image Skew. As the Landsat spacecraft travels southward it scans the Earth from West to East. The rotation of the Earth causes each successive mirror sweep to begin a bit farther to the West. The overall geometric effect is to skew the image. The magnitude of this distortion is proportional to the cosine of the latitude and is greatest at the equator.

Rotational Delay. The scanning of six lines with each mirror sweep causes the Earth rotational skew of an image to be a step function. The Earth rotation correction in DIRS is accomplished in two steps. The major portion is corrected by treating the distortion as a linear (continuous) function over the entire scene area. This skews the image as described previously. This linear correction leaves a residual saw tooth distortion which is corrected (along with sensor delay distortion) when the image resampling is done.

D. Miscellaneous

Other geometric adjustments are necessary to compensate for the effects of Earth curvature, MSS versus map perspective and projection.

E. Distortion Categories

The distortions present in digital MSS data can be grouped according to the way in which they effect image geometry. Three general categories can be defined.

Global Continuous. This category includes all those sources which operate over the full extent of a scene and are mathematically continuous. Attitude, orbit and aspect ratio are examples of distortion sources which fall in this category.

Swath Continuous. These are distortions which operate in a consistent and continuous manner on each set of six scan lines in a mirror swath. Examples are mirror velocity nonlinearity and Earth curvature.

Swath Discontinuous. These are distortions which are discontinuous in nature and operate in a regular pattern within the lines of a swath or between two adjacent swaths. Sensor delay and the residual Earth rotation distortion (rotational delay) are in this category.

IV. TECHNIQUE

The rectification process used in DIRS consists of six major steps and a number of special techniques.

A. Location of Ground Control Points

Ground control points are features or landmarks which are visible in the image and whose map coordinates can be determined. They are used to relate image geometry to the desired map projection. The image coordinates of a landmark are the sample and line number (i.e., column and row) in the digital image array where the landmark is located. Manual and automatic techniques for determining GCP image coordinates are provided in DIRS.

The manual method involves the generation of shade prints of the GCP area with an image coordinate border. These shade prints are made on the high speed line printer and use a combination of normal print characters to produce 16 levels of gray. Shade prints can be generated at full resolution (one print position per pixel) or at expanded resolution (3x3, 5x5, 10x10 print positions per pixel) using cubic convolution interpolation. The edges of the shade print are annotated with the correct sample and line coordinates so that if a feature can be seen on the shade print its coordinates can be measured.

The automatic method of determining GCP image coordinates is accomplished through edge correlation. The GCP feature must first be identified using the manual technique and then extracted and entered on the GCP library tape. A second scene over the same ground location can then be processed using the automatic GCP location capability.

The edge correlation technique was developed for NASA by Computer Sciences Corporation for the Large Area Crop Inventory Experiment (LACIE) and is a very reliable technique.[4] It is based on the generally valid assumption that image edges are invariant with time. It is well known that other image qualities such as color, intensity, or texture can change dramatically with time and season. These variations introduce much of the difficulty in using other correlation techniques. Since edges represent the shape of ground features they are relatively invariant with time.

The edge extraction process begins with the computation of a reflectance gradient value at each pixel location. This is done over the target subimage area (from the manual GCP location) and the search area in the second scene. A gradient histogram is then constructed for each of the subareas. The user specifies a threshold to determine the percentage of edge pixels. Generally, about twenty percent of the pixels are used as edge pixels. The gradient histogram is then analyzed to find the nearest break point to split the gradient range into the desired percent of edge pixels and non-edge pixels. This break point will be different in the target area histogram than it is in the search area histogram but will produce approximately the same percentage of edge pixels in both areas. The pixels whose gradient value exceeds the break point value are then coded as binary ones and all other pixels are binary zeros. The resulting binary edge images are then cross correlated and the point of maximum correlation identified. The GCP coordinates in the target subimage are then mapped into the search area and this point represents the image coordinates of the GCP in the second scene. Accuracies of plus and minus one sample and line are achieved with edge correlation. The reliability of the method is in excess of eighty percent.

The final operation in the location of GCPs is the determination of map coordinates. Generally seven and a half minute U.S.G.S. topographic maps are used for this purpose since they are very accurate and provide UTM coordinate tick marks. For maps without UTM coordinates the latitude and longitude values may be used. DIRS converts these geodetic coordinates to UTM automatically. Care must be taken to insure that the same part of the GCP feature is used for determining both map and image coordinates. For example, if a road intersection is used, the centroid of the intersection is usually selected as the precise GCP location and the coordinates of this centroid are determined in both the map and image.

B. Create Global Mapping Functions

The ground control point data developed in step A is used to define functions which make transformations between the map and image spaces. Three types of global mapping functions are available in DIRS, affine transformation, two dimensional least squares polynomials and attitude model functions. The choice of the most appropriate method depends on the number of GCPs and their distribution over the image area.

Affine Transformation. The affine transformation is one which maps a triangle from one two dimensional space into a triangle in a second two dimensional space. The mapping is constructed such that the verticies of the first triangle map into the verticies of the second triangle and all other points within, on and outside the first triangle map into proportionate locations relative to the second triangle. An affine transformation accounts for the following distortions:

Translation
scale change
rotation
aspect ratio
skew

The mapping is linear in the sense that straight lines map into straight lines. Mathematically, it is not a true linear function since the translation term violates the linearity condition

$$f(a+b) = f(a) + f(b) \qquad (1)$$

Affine transformations are expressed as functions of the form

$$X = X_o + a_1U + b_1V \quad (2)$$

$$Y = Y_o + a_2U + b_2V \quad (3)$$

where (X,Y) represent coordinates in one space and (U,V) coordinates in the other space. A set of affine triangles formed by the ground control points produces a piecewise-linear global mapping function over most of the image area. Borders outside the network of affine triangles are mapped using a large "master" triangle formed by three widely spaced GCPs.

Some limitations of the affine transformation are that it is a piecewise linear approximation to the "true" nonlinear distortion function. Also, the ground control points required to form an adequate set of affine triangles can be difficult to obtain. Some large image areas may be sparce in features suitable for GCP usage. The affine method is preferred where either small image areas are needed from the rectified scene or where high accuracy is not mandatory.

Two dimensional Least Squares. In this method four functions are computed using a least squares fit to the GCP data. These functions are

$$S = f_1 (E,N) \quad (4)$$

$$L = f_2 (E,N) \quad (5)$$

$$E = f_3 (S,L) \quad (6)$$

$$N = f_4 (S,L) \quad (7)$$

where S is the sample coordinate, L the line coordinate, N the Northing coordinate and E the Easting coordinate. These functions are polynomials in two variables and may be defined as 1st, 2nd, 3rd, 4th or 5th degree polynomials. The number of GCPs available determines the maximum degree of the polynomials. The polynomials have the general form:

$$Z = C_0 + C_1X + C_2Y + C_3X^2 + C_4XY + C_5Y^2 + C_6X^3 + C_7X^2Y + C_8XY^2 + C_9Y^3 + \ldots \quad (8)$$

These functions are continuous and global and therefore offer advantages over the affine transformation if an entire scene or a large portion of a scene is to be rectified. However, they have inherent limitations due to the need for a large number of ground control points and the requirement that the edges and corners of an image be well covered by GCPs. Some of the edge problems can be minimized by using "auxiliary" control points. ACPs are pseudo ground control points created at image locations where GCPs are needed. The image coordinates for an ACP and a suitable affine triangle (formed by existing GCPs) are specified by the user and the UTM coordinates at the ACP are computed using the affine transformation. ACPs are somewhat inaccurate but with judicious placement they can constrain the least squares functions and introduce very little error. Typical locations for ACPs are directly out from the image corners at a distance of 500 samples and 500 lines. Some experimentation may be needed in a particular situation to achieve the best results. The remaining problem with the least squares approach is the number of GCPs required to obtain the best results. In excess of thirty GCPs are needed to get good results with fifth degree polynomials. This problem is reduced through the use of automatic GCP location using edge correlation where possible.

Attitude Model. The final option for a global mapping function is the use of an attitude model for the Landsat spacecraft. This involves the determination of functions relating time to spacecraft position, velocity, altitude, yaw, pitch and roll. DIRS has the capability to use these functions to geometrically correct the image data, but it does not have the ability to compute the functions from the GCP data.[11] The Goddard Trajectory Determination System (GTDS)[12] is used to compute these functions from the GCP data. Approximately nineteen GCPs are needed to obtain the best fit for all of the attitude and orbit functions. This method offers the advantages of requiring fewer GCPs and having a somewhat lower sensitivity to the GCP dispersion pattern.

C. Create an Interpolation Grid

The global mapping functions computed in step B could be used directly to transform the image data into the desired map projection. This is not practical, however, due to the enormous amount of computation required to evaluate these functions over millions of points. The solution to this problem is the creation of an interpolation grid. The functions can be evaluated at each mesh point in a grid covering the image area and then an efficient bilinear interpolation technique can be used to map points within each grid cell.

The interpolation grid is composed of approximately twenty horizontal and twenty vertical grid lines. These lines are defined in the UTM map space by functions of the form

$$Y = MX + B \quad (7)$$

and

$$AX + BY + C = 0 \quad (8)$$

where Y is Northing and X is Easting. The UTM intersection coordinates for each pair of horizontal and vertical grid lines is then computed. These intersections (or mesh points) are then transformed through global mapping functions to compute the corresponding image coordinates. When each mesh point is known in both UTM and image coordinates we are ready to map the image data in an efficient manner. The mapping is actually an inverse mapping, i.e. it proceeds from the map space (output coordinate system) back into the image space (input coordinate system). This arrangement has the advantage of allowing interpolation to occur in the input (image) space instead of the output (map) space. This proves to be a much more convenient approach than forward mapping.

D. Step through Interpolation Grid

The desired end product of a digital image rectification is a uniformly spaced matrix of pixels which produces an image in conformance with a (UTM) map projection. It is impossible to maintain a uniform lattice in both the image and map space since the mapping between the two is nonlinear. The regular lattice of points in the map space form a nonuniform lattice in the image space. The objective of stepping through the interpolation grid is to identify the map coordinates of the uniformly spaced points in the map space. These are the locations at which image intensity values are desired for the rectified output image.

E. Compute Image Coordinates

The map location selected in step D is now transformed into image coordinates using bilinear interpolation within the local grid cell. Pad pixels are introduced at the beginning and end of each resample line to form a rectangular output pixel array. The image coordinate computation is done by breaking the resample line into segments. One segment corresponds to the portion of the resample line between two vertical grid lines. The image coordinates at the intersection of the resample line and the two vertical grid lines are computed and an equation of the line segment in the image space is derived. Actually, two equations are derived, one for the sample coordinate and one for the line coordinate. These equations are evaluated at each resample location to produce the image coordinates at which resampling is to be done.

F. Resample

Resampling refers to the determination of an image intensity value at a given location. Typically, this location falls between and not on exact pixel centers and some form of interpolation is required. Two interpolation techniques are available in DIRS. The simplest and most efficient method is Nearest Neighbor. In this method the pixel whose center is nearest to the resample location is used to supply the intensity value at the resample location. This method introduces up to one half sample and line of geometric error. For a large subimage area or full scene this is acceptable and produces adequate results in a short period of computation time.

The second resampling technique provided in DIRS is cubic convolution. This method was developed by TRW and is an efficient approximation to the theoretically optimum interpolation using $\sin(x)/x$. While cubic convolution is much more efficient than $six(x)/x$ it is a great deal slower than nearest neighbor.

Whether resampling is done using nearest neighbor or cubic convolution, DIRS allows the option of setting the spacing between pixels and lines. Since Landsat pixels are spaced 57 x 79 meters and since even spacing is desired in the output array, it is necessary to introduce some redundancy to preserve all the information in the original image. A sample and line spacing of 50 meters is frequently selected for output pixels. Fifth meter spacing gives a scale of 1;1,000,000 when the image data is displayed on a film recorder with a 50 micron spot size. Spacing as small as 10 meters has been used successfully with DIRS for small subimage areas.

6. Special Techniques

Steps A through F above describe the general approach used in DIRS. This covers the correction of all distortions in the category "global continuous" as discussed in Section III-E. Special methods are implemented to remove swath continuous and swath discontinuous distortions.

Since swath continuous errors could also be viewed as global continuous errors they could be corrected through the global mapping functions. However, because these distortions are highly consistent from swath to swath, there is no need to "load" the global mapping functions to correct them. Whenever distortions can be removed without resorting to scene data (GCPs) it is advantageous to do so. This reduces the total number of GCPs needed to achieve the same accuracy. The method used for swath continuous distortion corrections is:

1. Adjust the sample coordinate of each GCP by subtracting the swath continuous errors.

2. Generate the interpolation grid using the adjusted GCP data. The image coordinates at mesh points in this grid now correspond to an image which does not have swath continuous distortions.

3. Reintroduce the swath distortion by adding it to the sample coordinate at each mesh point in the interpolation grid. This forces the grid to map back into the correct (raw) image coordinates in the real image space.

The effect of these operations is to superimpose the swath continuous distortion functions on the global mapping functions.

The third category described in Section III-E is swath discontinuous distortions. These distortions must be corrected in the resampling process. Both nearest neighbor and cubic convolution resampling algorithms in DIRS correct for sensor delay and Earth rotational delay distortions. The technique used involves repositioning the input image pixels prior to resampling. The repositioning is the inverse of the original distortion and removes the distortion, but it also creates an irregular array for the resampling algorithm. See Figures 1 and 2 to compare the original and repositioned pixel arrays used for cubic convolution resampling.

An additional technique provided in DIRS is the removal of sensor delay and Earth rotational delay distortions from the expanded shade prints. This produces more realistic appearing features on the shade prints and helps in determining the image coordinates of the GCP features. This procedure introduces a small error in the image coordinate associated with the given feature. This error is eliminated when the GCP table is preprocessed prior to developing the global mapping functions. The sample coordinates of the GCPs are adjusted to put back the error due to sensor delay and Earth rotational delay which had been taken out of the shade prints.

V. RESULTS

DIRS has been used to rectify eight scenes thus far. Both Landsat 1 and 2 scenes have been processed. The accuracy of the results depend primarily on three factors:

(1) The number of GCPs
(2) The distribution of GCPs
(3) The mirror model used.

Residual distortions on the order of 50 to 100 meters have been observed in the most recent rectifications. These results were achieved using 19 ground control points and the attitude model capability for generation of the global mapping functions. A single county (Madison County, Iowa) was rectified and extracted from four separate Landsat 2 scenes. Madison County is shown in Figure 3. It covers an area of 24 x 24 miles. Figure 4 is a print of the extracted image data of Madison County from one of the four rectified scenes. The borders of the extracted image area do not coincide exactly with the map borders because the extraction run used approximations to the geodetic coordinates of the corners of the county. The resampling was done with cubic convolution at 50 meter steps. The four extracted subimages of Madison County have been used for a multitemporal classification study of Corn growth and phenology. The four subimages have been input to the General Electric Image 100 and show scene to scene registration accuracy of approximately 50 meters (one sample) for three of the scenes and 100 meters for the fourth scene. The slope of the resampled lines was forced to be identical in the four DIRS recitification runs to permit accurate registration on the Image 100.

VI. CONCLUSIONS

DIRS is operational and performing up to expectations. The objectives described in Section II-C have been achieved. Further work will be required to determine the limits of accuracy possible with DIRS. Additional work could also be done to improve efficiency but the existing run times are within acceptable limits. A full Landsat scene can be rectified using nearest neighbor resampling at 50 meter steps in less than 5 minutes of CPU and 1.6 minutes of I/O time on a 360/91 computer. This does not include the run time for reformatting the four CCTs and generating shade prints.

REFERENCES

1. Van Wie, Peter, et al., "Digital Image Rectification System Preliminary Documentation," NASA/GSFC, Nov., 1975.

2. Department of the Army, "Universal Transverse Mercator Grid," Department of the Army Technical Manual, TM5-241-8, July, 1958.

3. Bernstein, Ralph, "All Digital Precision Processing of ERTS Images," IBM Final Report to GSFC under Contract No. NAS5-21716, April, 1975.

4. Rifman, S.S., et al., "Experimental Study of Digital Image Processing Techniques for Landsat Data," TRW Final Report to GSFC under Contract No. NAS5-20085, January, 1976.

5. Nack, M.L., "Interim Report on Temporal Registration of Multispectral Digital Satellite Images Using Their Edge Images," CSC Report to GSFC under Contract No. NAS5-11999, April, 1975.

6. Quinones, J., "UTM Forward." U.S. Army Topographic Command, Report No. 945, Department of Computer Services, Computer Memorandum, July, 1970.

7. Quinones, J. "UTM Inverse," U.S. Army Topographic Command, Report No. 946, Department of Computer Services, Computer Memorandum, July, 1970.

8. Fisher, David A., "Zone to Zone Transformation," U.S. Army Topographic Command Report No. 963, Department of Computer Services, Computer Memorandum, February, 1971.

9. National Aeronautics and Space Administration, "ERTS Data Users Handbook," NASA/GSFC Document 71SD4249, September, 1971.

10. Thomas, Valerie L., "Generation and Physical Characteristics of the ERTS MSS System Corrected Computer Compatible Tapes," NASA/GSFC X-563-73-206, July, 1973.

11. Puccinelli, Edward F., "Ground Location of Satellite Scanner Data," Photogrammetric Engineering and Remote Sensing, April, 1976.

12. "Goddard Trajectory Determination System, Users' Guide," by Computer Sciences Corporation and Goddard Space Flight Center, July, 1975.

1,1	1,2	1,3	1,4
2,1	2,2	2,3	2,4
3,1	3,2	3,3	3,4
4,1	4,2	4,3	4,4

Figure 1

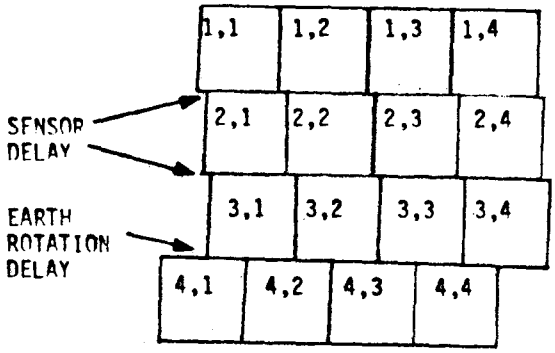

Figure 2.

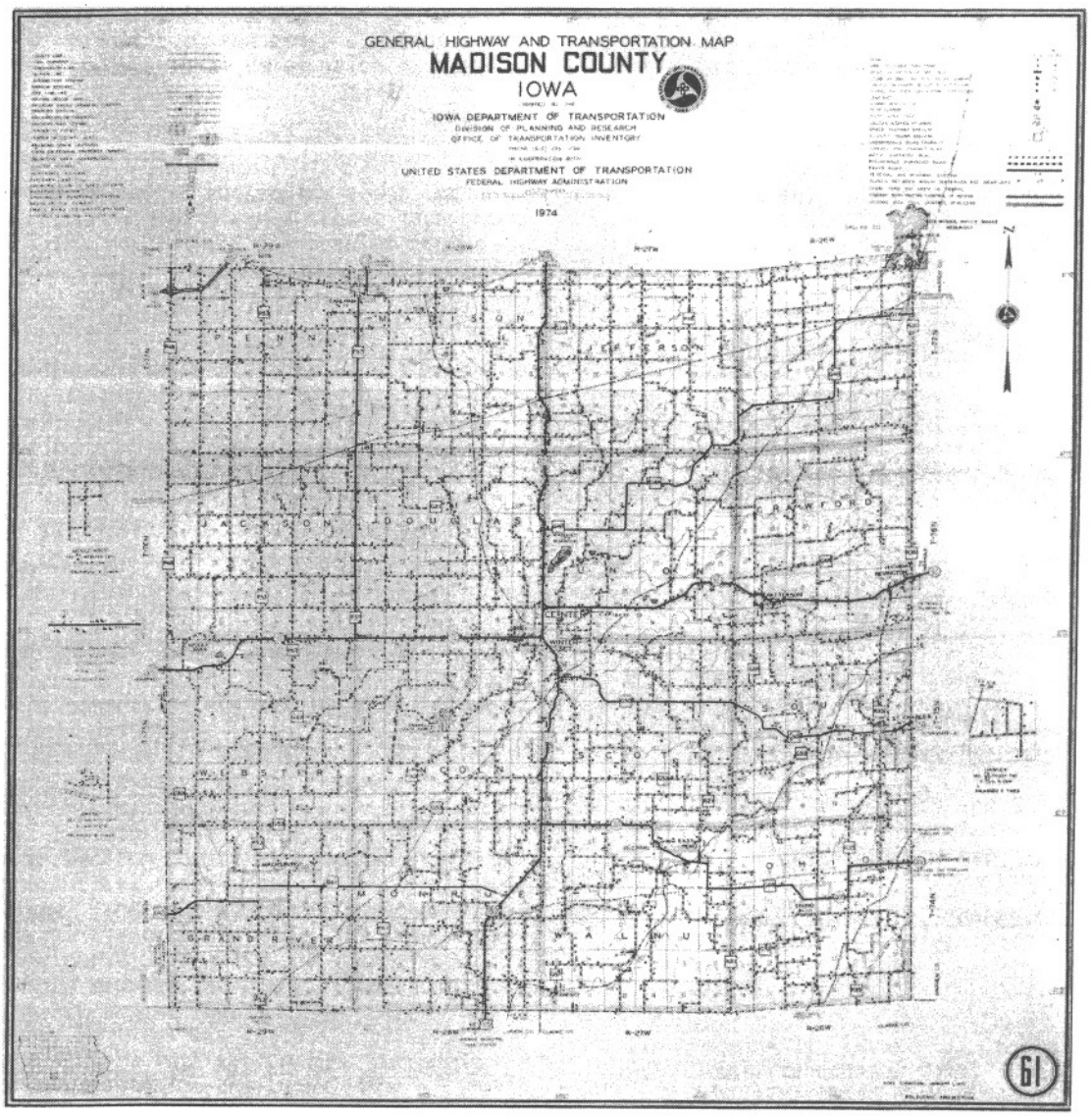
GENERAL HIGHWAY AND TRANSPORTATION MAP
MADISON COUNTY
IOWA
IOWA DEPARTMENT OF TRANSPORTATION
DIVISION OF PLANNING AND RESEARCH
OFFICE OF TRANSPORTATION INVENTORY
UNITED STATES DEPARTMENT OF TRANSPORTATION
FEDERAL HIGHWAY ADMINISTRATION
1974
61

Figure 3.

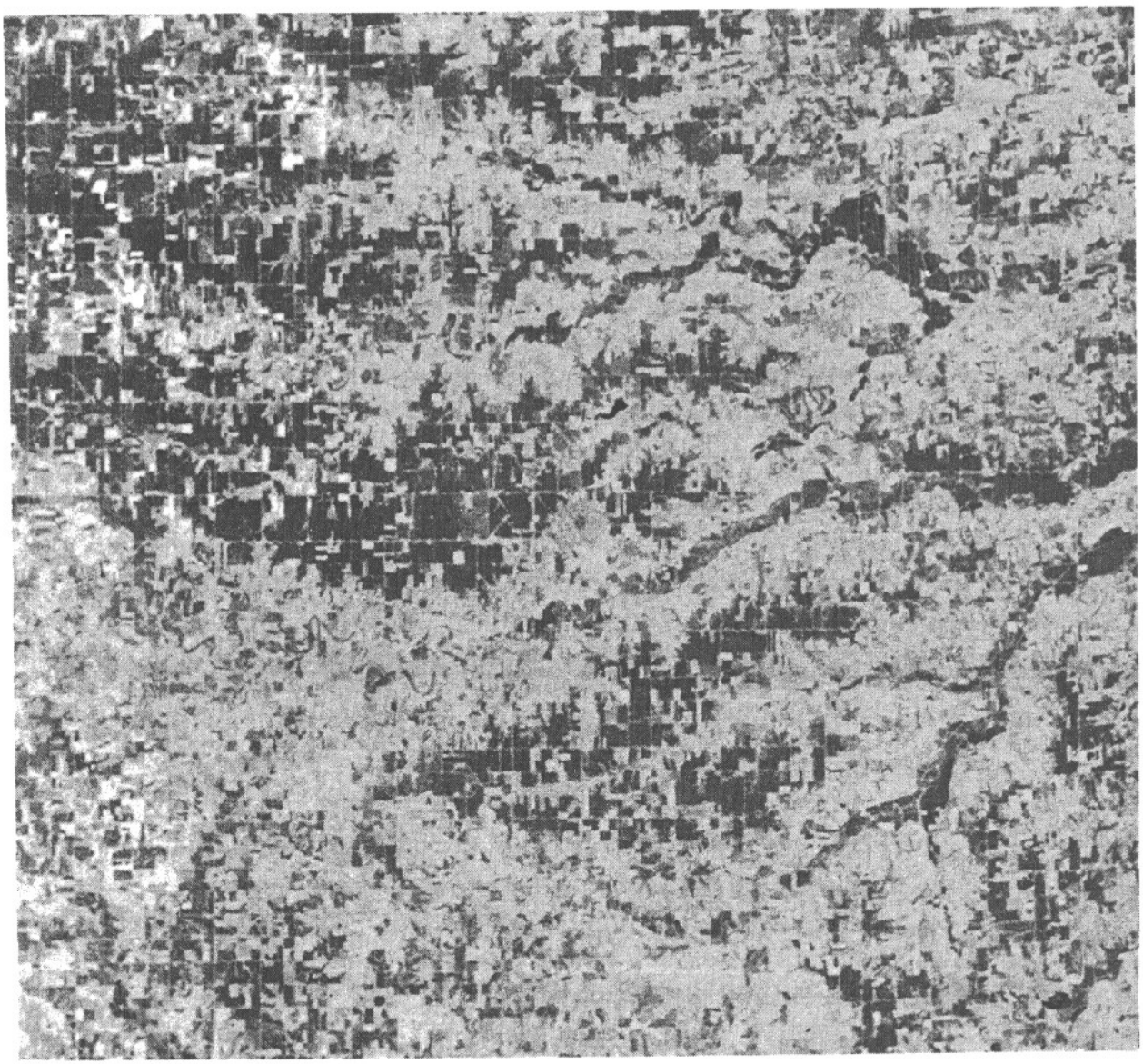

Figure 4.