Georeferencing of Satellite Imagery for Digital Soil Mapping

S. E. Smith, B. A. Dewitt, E. P. Gonzalez, and G. W. Hurt

ABSTRACT: As part of a statewide effort to produce digitized soil survey maps with a satellite imagery background, Landsat Thematic Mapper imagery was georectified to the map's projection system. Provided in this paper is a description of the manner in which soil surveys are conducted by the Soil Conservation Service in Florida and the procedure being used to convert traditional soil maps to a digital form. Also described is the process by which satellite imagery can be properly georeferenced to the digital soils maps and used as a background layer in the resulting geographic information system. Due to the large size of the project—the entire state of Florida—a study of the most efficient means of rectifying the imagery that still satisfied the spatial accuracy requirements of the Soil Conservation Service was made. An accuracy and time comparison of resampling techniques is presented in the paper.

Introduction

The United States Department of Agriculture's Soil Conservation Service (SCS), recently renamed the Natural Resources Conservation Service (NRCS), provides federal leadership for the National Cooperative Soil Survey (NCSS). The Soil Conservation Service has the federal responsibility for collecting, storing, maintaining, and distributing soil survey information for privately and non-federally owned lands in the United States according to Public Law 90-620, Public Law 74-46, and Public Law 89-560 (Soil Survey Staff 1983).

Soil Survey

The objective of the NCSS program in Florida is to provide an inventory of the state's soil resources. These inventories, or soil surveys, consists of three basic components. These are:

- soil maps (prepared on a photographic base) which show the geographic location of different kinds of soils,
- a narrative which describes the different kinds of soils and their use and management, and
- a set of tables which provides interpretations (attributes) for the soils for many uses.

A soil survey is an inventory of soils conducted on the ground. The survey results in a composite of maps, tabular data and written text that describes the soils of a particular area. Traditionally, soil surveys are conducted on a county level. The surveys are usually conducted in the field by soil scientists familiar with the local terrain. Guidelines for the procedures are provided by the *National Soil Survey Handbook* (1993); *Soil Taxonomy* (1975); *Procedures for Collecting Soil Samples and Methods of Analysis for Soil Survey* (1984); and *Keys to Soil Taxonomy* (1994).

The county-level soil survey manuals are published by the Soil Conservation Service. They are distributed through county extension agents, SCS field offices and government book stores. The survey contains a detailed, written description of the counties' soils and a description of the counties' climate, geomorphology, relief, drainage patterns, water resources, transportation network and general agricultural picture.

The bulk of the county soil survey is, however, a series of black-and-white rectified aerial photographs of the county at a scale of 1:20,000. The photography is taken by the SCS on a rotation of approximately five years. Superimposed on the photographs, which are reproduced as lithographs in the published surveys, are soil classes.

It normally requires several years for a team of soil scientists to produce a soil survey map of an average-sized county. Much of the time is devoted to field-work, such as soil profiling. Typically, between 40,000 and 60,000 holes (7.5-cm diameter by 2-m depth) are dug throughout the county for the survey. The number and type of soil horizons, textures, color, drainage and other important characteristics about each hole are recorded in a tabular form.

Holes are bored at specific locations that represent a particular slope, aspect or other landscape parameter. A skilled soil scientist then delineates the soil and landscape patterns observed in the field onto rectified vertical aerial photography using a zoom transfer scope. Each delineation or polygon

---

***S. E. Smith** and **B. A. Dewitt** are on the faculty, and **E. P. Gonzalez** is a graduate student, in the Surveying and Mapping Program, University of Florida, Gainesville, FL 32611. **G. W. Hurt** is a soil scientist with the U.S. Fish and Wildlife Service, St. Petersburg, FL 33702.*

then represents a homogeneous soil class. Soils are classified according to the SCS taxonomy system. In this system, soils are classified with increasing specificity into orders, suborders, great group, subgroup, family and series. Soil surveys are typically made at scales of 1:15,840, 1:20,000 and 1:24,000 in Florida.

Soil Geographic Data Bases

SCS has established three soil geographic data bases. They are the Soil Survey Geographic Data Base (SSURGO), the State Soil Geographic Data Base (STATSGO), and the National Soil Geographic Data Base (NATSGO) (Soil Survey Staff 1991). Each of these data bases is in some stage of development and quality control checks, both at the federal and state level. When completed, the SSURGO, STATSGO and NATSGO will comprise the SCS's geographic information system (GIS).

SSURGO

The most detailed level of information is SSURGO, which is the county soil survey geographic data base in Florida. Soil surveys have been produced on approximately 99% of the private and non-federal lands in Florida by using NCSS field mapping standards (Soil Conservation Service Staff 1981). Surveyors observe soils along delineation boundaries and determine map unit composition by field traverses and transect. Rectified aerial photographs are used as the field base maps at scales ranging from 1:15,840 to 1:24,000.

Line segmentation (vectors) are digitized in accordance with SCS standards and specifications using United States Geological Survey's (USGS) 7.5 minute topographic quadrangles series (Cartographic and Geographic Information System Staff 1987). The digitizing is performed by the SCS or by cooperating water management districts and county governments. Approximately 75% of county soil surveys have been digitized.

The SSURGO data set was designed to be used for farm, ranch, development, city, or county resource planning and management. More than two hundred different attributes are recorded. Farmers, foresters, and agronomists use SSURGO to maximize food and fiber production. Planners, community officials, engineers, developers, builders, and home buyers use SSURGO to select sites for development and to select practices in order to insure proper performance. Environmentalists use SSURGO to understand, protect, and enhance the environment.

STATSGO

The next less-detailed level of information is STATSGO. STATSGO is the state soil survey geographic data base. It covers the entire state of Florida. Soil maps were made by combining the SSURGO data based on similar topography, geology, vegetation, and climate. Map unit compositions were determined by transecting the SSURGO data. These line segments were also digitized following SCS standards and specifications. Digitizing was performed by SCS in cooperation with the Florida Department of Environmental Protection. The 1:250,000 scale maps of the USGS 1-degree by 2-degree topographic quadrangle series were used as the digitizing base.

The STATSGO data base was designed to be used for regional, multistate, river basin, state, and multicounty resource planning, management and monitoring. STATSGO data can be used to compare the suitability of large areas for a particular land use. Areas of suitable soils, as well as areas where the soils are not suited, can be identified.

NATSGO

The lowest level of detail is NATSGO, which is the national soil survey geographic data base. The boundaries of federally-designated lands such as the Major Land Resource Areas (MLRA) and Land Resource Regions (LRR) were used (Soil Conservation Service Staff 1981) to form the NATSGO data base. The NATSGO data set was designed to be used for national-level resource planning and management, and modeling of national-level climate changes. Map unit compositions for NATSGO were determined by sampling data of the 1982 National Resources Inventory (Soil Conservation Service Staff and Iowa State University Statistical Laboratory 1987).

USGS's U.S. base map at a scale of 1:5,000,000 was the base map for NATSGO. The soil type boundaries were digitized in accordance with SCS standards and specifications.

Satellite Image Background For A Digital Soils Map

Soils mapping is an iterative process that involves intensive ground work and aerial photographic interpretation. Satellite imagery has, thus far, played a very limited role in soils mapping at any scale. Reasons for this include limited experience of soil scientists with satellite imagery, poor spatial resolution of imagery compared with aerial photography and image processing requirements unique to satellite imagery.

Over the past several years, however, satellite imagery has been found useful in a variety of agricultural applications, including soil mapping. Long-term projects such as AGRISTARS, co-sponsored by the National Aeronautics and Space Administration (NASA) and the Department of Agriculture (DA), provided ample evidence that satellite imagery was useful for soil mapping. In the AGRISTARS program, soils experts received training in satellite imagery processing and interpretation. Over the same period, the quality of the imagery greatly improved and facilities became more "user friendly."

The advent of GIS has moved the SCS toward development of a digital soil map. Digital satellite imagery as a layer is a logical next step. The imagery's format is automatically utilizable in a GIS, making data conversion no problem. Additionally, updated imagery can easily be input to the GIS, and storage of satellite data is vastly more efficient than with aerial photography.

For commercial satellite data, the primary choices are either Landsat Thematic Mapper (TM) or SPOT Multispectral (XS). Landsat TM has seven spectral channels, six of which have a 30-m resolution with the seventh, a thermal channel, at 120 m. SPOT XS imagery, on the other hand, has a 20-m resolution but has only three spectral channels. In addition, a single TM image covers an area 185 km by 185 km, whereas a single SPOT image covers 60 km by 60 km. For the requirements of a soil survey, it was determined that Landsat's spectral advantages outweighed SPOT's better spatial resolution. The large size of the project (the entire State of Florida) also required using the most efficient imagery in terms of minimizing the number of scenes required to cover the state. Fifteen Landsat scenes are needed to cover Florida compared with over 100 SPOT scenes.

Georeferencing of Imagery

To make Landsat imagery useful in the SCS's soils GIS, the imagery must be adjusted to a geographic map projection. This process is often called georeferencing. It converts the image from an arbitrary coordinate system into that of the GIS.

An image layer in a GIS can be visualized as a two-dimensional arrangement of grid cells, each containing a numerical quantity corresponding to the amplitude of the reflected energy from the ground at the corresponding location. This numerical quantity can be called the digital number (DN) or gray level. There will be a separate layer, each with its corresponding set of digital numbers, for every spectral channel (color) in the image. The grid is defined on the basis of the chosen map projection, with the coordinates of the cell centers corresponding to nominal values. This is illustrated in Figure 1.

The general process of georeferencing starts with an empty GIS image layer. Then, using the coordinates for each cell, the correct digital number from the satellite image that corresponds to the cell's ground position is found and placed in the cell. When all cells have been populated in this fashion, the GIS layer can be displayed as an image with the pixel grid defined in terms of (x, y) UTM coordinates instead of some arbitrary (row, column) image system.

Two main subprocesses within georeferencing warrant further explanation. First, given a set of GIS cell coordinates (x, y) how is the correct location in the image determined?

Coordinate Transformation

The theoretically correct approach involves a complicated relationship between the map projection, the three-dimensional earth surface, and parameters that describe the imaging satellite's orbital characteristics and data acquisition. Software that would allow this approach is highly complex and not generally available. A compromise was reached where a two-dimensional affine transformation was used to relate the satellite image coordinates directly to the map coordinates. Because of the pixel size (30 m for Landsat TM) and the nature of Florida's topography (minimal relief) the error due to this compromise was assumed to be negligible.

In the 2-D affine transformation, parameters that relate UTM (x, y) coordinates to image (row, column) coordinates must be determined. These parameters are shown in Equations (1) and (2).

$$R = a_0 + a_1 x + a_2 y \qquad (1)$$
$$C = b_0 + b_1 x + b_2 y \qquad (2)$$

where

x, y	are UTM coordinates
R, C	are row, column image coordinates
$a_0, a_1, a_2, b_0, b_1, b_2$	are transformation parameters.

As a preliminary step, a set of at least three control points (having known UTM coordinates) are located in the satellite image by their row and column coordinates. Using these control values, the affine transformation is solved (by least squares) giving the most probable parameters (a and b terms) that relate the two coordinate systems. Because of this preliminary step, any cell in the GIS image layer can be readily transformed to obtain its location in the satellite image.

Control Point Selection

Control points for the affine transformations were obtained from 1:100,000 scale USGS topographic maps. These points consist primarily of road and airport runway intersections. Control point locations

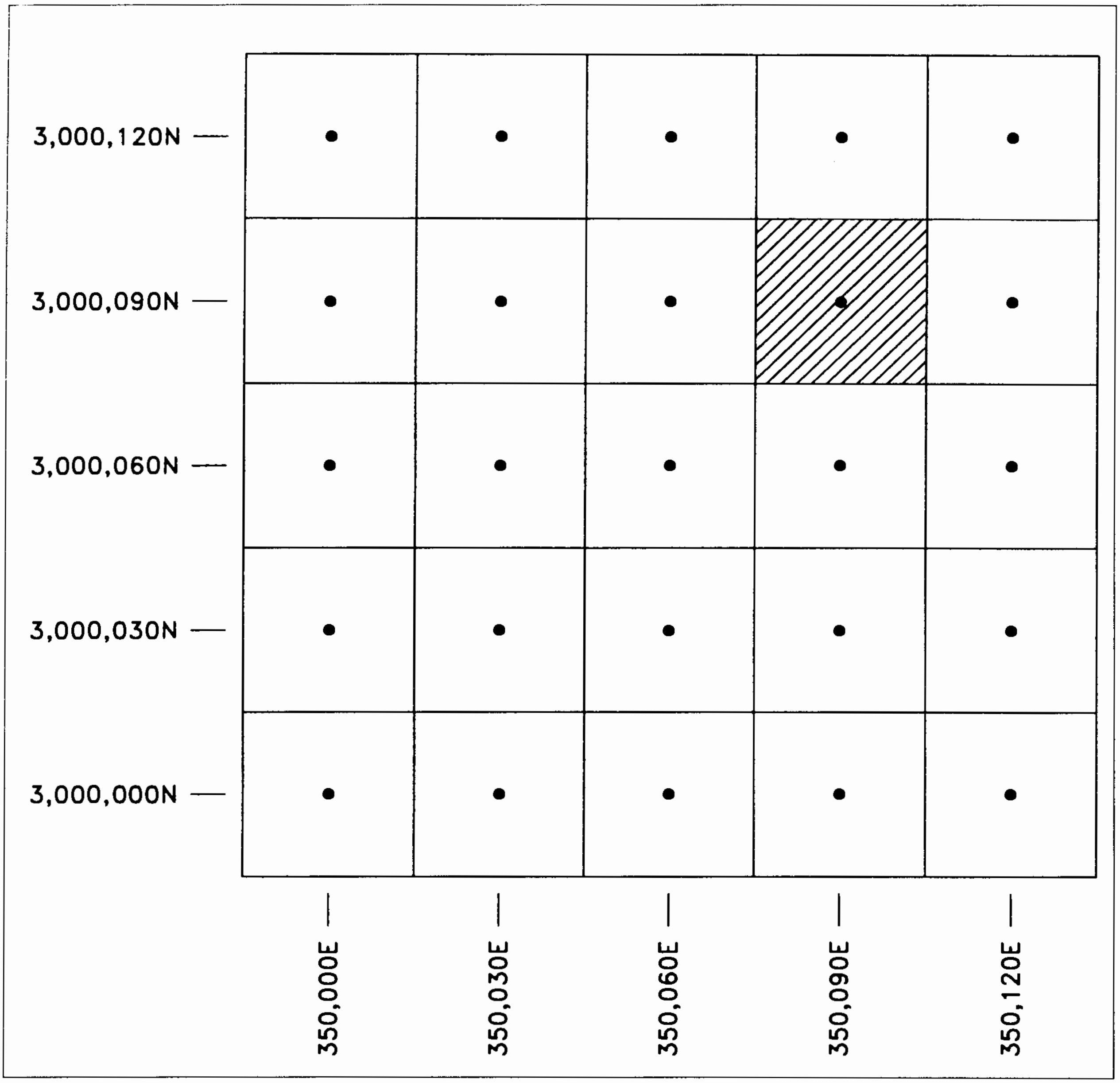


Figure 1. Two-dimensional grid cells with cell centers corresponding to nominal values.

from the maps were determined by scaling from grid ticks. The corresponding image point coordinates were obtained by pointing with a cursor on a computer screen to extract the row, column values.

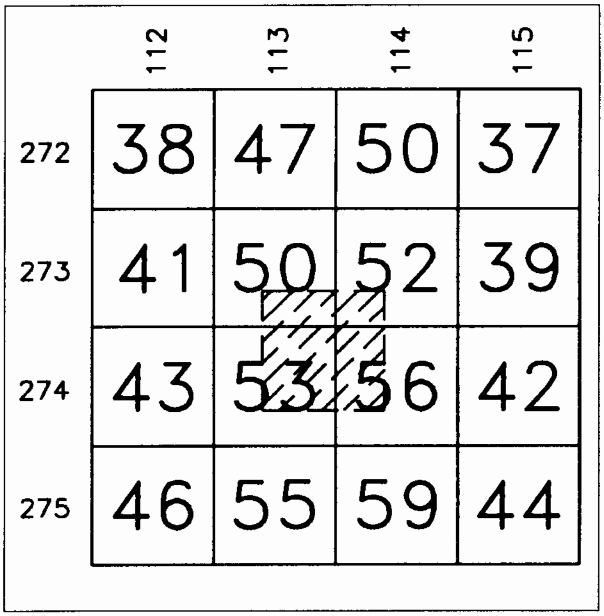
Selection of control points largely depends on what is available on both the maps and imagery. Ground control should be spread evenly throughout an image to minimize collinearity and strengthen the geometry of the transformation. An affine transformation requires at least three points and a noncollinear configuration in order for a solution to be computed. Ideally, points should be taken from the corners and sides of the image, besides those spread throughout the interior. Practical considerations of image-identifiable points made this difficult, however. In the end, 36 control points were selected for each full Landsat scene in this project.

Affine transformation

Once the lists of both image and UTM coordinates for the control points had been compiled, an affine transformation was computed. Residuals were checked and, if larger than one pixel (30 m) in magnitude, the point was thrown out. Once the remaining residuals passed the test, the transformation was judged acceptable. In no case did the actual number of points used for the transformation fall below 16 per full scene.

Resampling

The second main subprocess is a consequence of the fact that, in general, the row and column coordinates from the affine transformation are not integer values. Since the satellite image has digital numbers only for integral locations, a procedure is required to



	112	113	114	115
272	38	47	50	37
273	41	50	52	39
274	43	53	56	42
275	46	55	59	44

Figure 2. A 4 × 4 subarea of image pixels with superimposed grid cell at a fractional location.

obtain the proper value from the satellite image. This procedure is commonly called resampling.

Acquisition of a digital image involves discrete sampling of a continuous analog signal, i.e., the reflected energy from the ground. These digital samples are made in a (distorted) grid pattern, with each grid cell or pixel containing a DN representing the lightness or darkness at its corresponding ground location. When a digital satellite image is acquired, no attempt is made to have the pixels line up with any particular map projection coordinates. It is therefore necessary to perform resampling to obtain a digital sample at an intermediate (i.e., fractional) row, column location. Resampling involves interpolation between existing pixels (DNs) to synthesize a pixel which corresponds to the fractional location.

There are several techniques available for resampling of digital images, though three particular ones are by far the most prevalent. These three are known as

- nearest neighbor interpolation
- bilinear interpolation, and
- bicubic interpolation

(Moik 1980). Other, more computationally intensive techniques are generally not employed since they tend to be sensitive to noise which exists in satellite imagery.

Nearest Neighbor Interpolation

The nearest neighbor interpolation is simplest of the three. As its name implies, the DN chosen will be that of the image pixel whose center is closest to the center of the grid cell. From a computational standpoint, all that is required is to round off the fractional row and column values to the nearest integral value. Figure 2 shows a 4 × 4 subarea from a digital image with a superimposed pixel at a fractional location (R=273.68, C=113.37). Rounding these values to the nearest integer yields 274 and 113 for the row and column indices, respectively. Thus, the resampled value is 53.

Bilinear Interpolation

A second resampling method, which involves greater computational complexity, is bilinear interpolation. In this approach, the four surrounding pixels are selected and linear interpolation is performed in two directions. This will be illustrated by example using the values from Figure 2. First, interpolated values (N_1 and N_2) are computed along rows 273 and 274. Equations (3) and (4) illustrate the calculation.

$$N_1 = 0.37 \bullet (52 - 50) + 50 = 50.74 \qquad (3)$$

$$N_2 = 0.37 \bullet (56 - 53) + 53 = 54.11 \qquad (4)$$

Next, an interpolation is performed in the column direction, yielding the result (DN) given in Equation (5):

$$DN = 0.68 \bullet (54.11 - 50.74) + 50.74 = 53.03 \quad (5)$$

Finally, since digital numbers are generally integers themselves, the value from Equation (5) is rounded to 53.

Bicubic Interpolation

Bicubic interpolation, also known as cubic convolution, is the third resampling technique commonly used. Explanation of this technique requires a little background in sampling theory. First, an assumption is made that the original signal has been sampled above the Nyquist rate, which is generally satisfied for imaging sensors (ASP 1983). The Nyquist rate is, in essence, the sampling frequency required to faithfully record the high (spatial) frequency content of the scene. Given this assumption, the so-called "sinc" function allows an (almost) exact reconstruction of the original scene. The form of the sinc function is shown in Figure 3. If the images had an infinite number of rows and columns, and all pixels were used for the interpolation, the sinc function would yield a perfect reconstruction. Practicality, however, dictates that interpolations are carried out using only small neighborhoods surrounding the interpolated pixel. A cubic spline approximation to the sinc function is the form generally used for bicubic interpolation. The shape of the spline is given in Figure 4 while Equations (6)–(8) express the functional relationship (Billingsley 1985). For comparison, Figure 4 also shows the shape of nearest neighbor and bilinear interpolation expressed in the form of weighting functions. Note that the cubic spline most nearly approximates the sinc function of

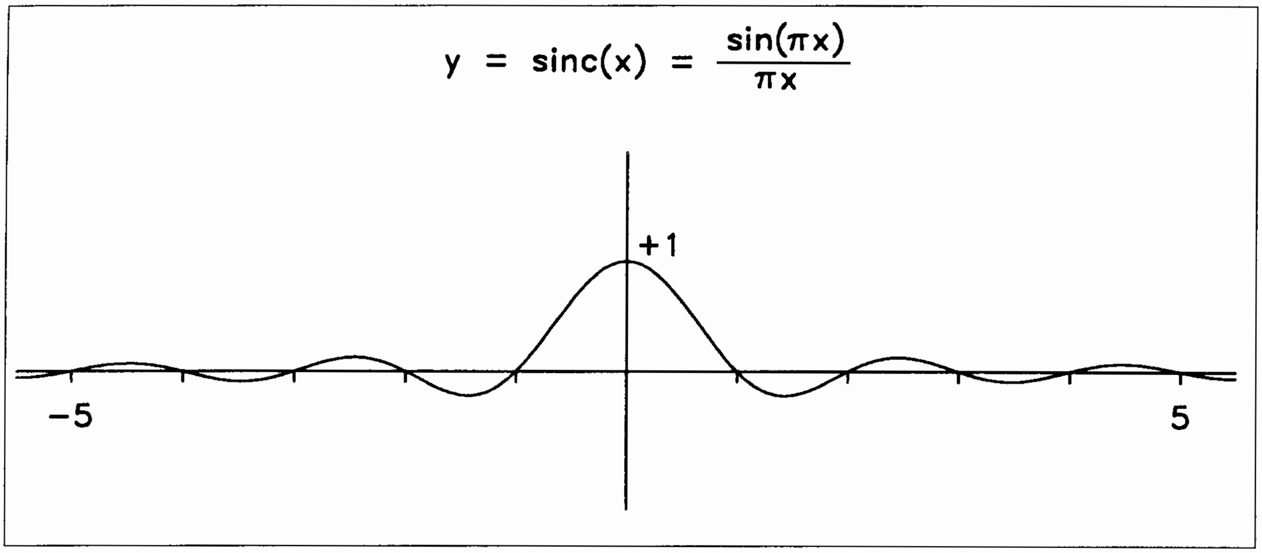


Figure 3. Form of the sinc function.

Figure 3, whereas bilinear and nearest neighbor are less consistent approximations.

$$f_1(x) = (a+2)x^3 - (a+3)^2 + 1 \quad \text{for } 0 \le |x| \le 1 \qquad (6)$$

$$f_2(x) = ax^3 - 5ax^2 + 8ax - 4a \quad \text{for } 1 \le |x| \le 2 \qquad (7)$$

$$f_3(x) = 0 \quad \text{for } |x| \ge 2 \qquad (8)$$

where

a = free parameter equal to the slope of the function at x = 1. Generally, a = -0.5 for best results.

The computational process is analogous to that of bilinear interpolation in that it is performed first along the rows and then down the single, fractional column. The computations are conveniently expressed in matrix form, as shown in Equation (9). In this equation, the **R** and **C** matrices consist of coefficients derived from Equations (6) and (7) and the **D** matrix contains the digital numbers from the 4 × 4 neighborhood surrounding the interpolated pixel.

$$\mathbf{R} \cdot \mathbf{D} \cdot \mathbf{C} = [r_1 r_2 r_3\, r_4] \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & d_{22} & d_{23} & d_{24} \\ d_{31} & d_{32} & d_{33} & d_{34} \\ d_{41} & d_{42} & d_{43} & d_{44} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} \qquad (9)$$

Interpolating across the rows (based on the fractional column position) is done by forming the product **D** · **C**. Subsequently, **R** is multiplied by the product to obtain the final interpolated value. Actually, the sequence of multiplications does not matter.

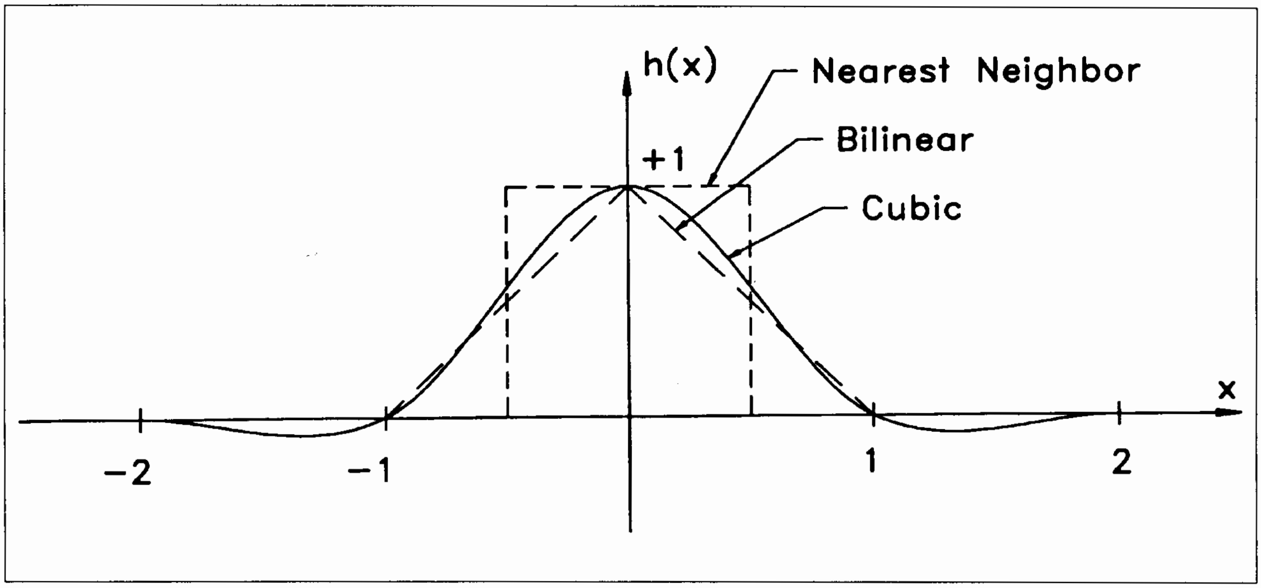


Figure 4. Shape of the cubic spline, bilinear interpolation, and nearest neighbor, in the form of weighting functions.

Referring to the example of Figure 2, the bicubic interpolation computation procedure begins by computing the elements of the **R** matrix. Given the fractional row location (273.68), an interpolation weighting function will be computed for the two rows above (272 and 273) and the two rows below (274 and 275). The distance (x) from the fractional row is determined for each of the four surrounding rows and the corresponding function (Equation (6) or (7)) is selected. The result of this computation is listed in Table 1.

Table 1. Row interpolation weight matrix, **R**, computed for example of Figure 2. Slope value of a =-0.5 was used.

Element	Row	Distance x	Weighting function	Value
r_1	272	1.68	$f_2(x)$	-0.03482
r_2	273	0.68	$f_1(x)$	0.31565
r_3	274	0.32	$f_1(x)$	0.79315
r_4	275	1.32	$f_2(x)$	-0.07398

In a similar fashion, the elements of matrix **C** are computed as shown in Table 2.

Following these preliminary calculations, the matrix product from Equation (9) is formed. This product, using example values, is shown in Equation (10). The resultant value, 54.3, is then rounded to 54, the nearest integer.

$$[-0.03482 \quad 0.31565 \quad 0.79315 \quad -0.07398] \times \begin{bmatrix} 38 & 47 & 50 & 37 \\ 41 & 50 & 52 & 39 \\ 43 & 53 & 56 & 42 \\ 43 & 55 & 59 & 44 \end{bmatrix} \begin{bmatrix} -0.07343 \\ 0.73373 \\ 0.38282 \\ -0.04312 \end{bmatrix} = 54.3 \qquad (10)$$

Resampling Pros and Cons

Figure 5a is an example of an unrectified "raw" image. This figure defines the basis for comparing the three resampling methods. The nearest neighbor method has the distinct advantage of being the quickest of the three techniques in terms of computational time. The nearest neighbor method also has the advantage of not modifying the sensor data which is important if image classification will be performed (not so in this project). However, since a "smooth" interpolation is not being performed, the resultant appearance can be somewhat "blocky," as illustrated in Figure 5b.

The primary advantage of bilinear interpolation is the smoother appearance of the result, as shown in Figure 5c. This appearance is slightly compromised by the fact that some high frequency detail is filtered out. In other words, edges in the scene are slightly less distinct. In terms of computational time, bilinear interpolation is slower than nearest neighbor but faster than bicubic.

Table 2. Column interpolation weight matrix, **C**, computed for example of Figure 2. Slope value of a =-0.5 was used.

Element	Row	Distance x	Weighting function	Value
c_1	112	1.37	$f_2(x)$	-0.07343
c_2	113	0.37	$f_1(x)$	0.73373
c_3	114	0.63	$f_1(x)$	0.38282
c_4	115	1.63	$f_2(x)$	-0.04312

Bicubic interpolation is the most correct resampling technique of the three on the basis of signal processing theory. It achieves the "smooth" appearance without sacrificing as much high frequency (edge) detail, as illustrated in Figure 5d. However, this enhanced appearance comes at a penalty in terms of computational time.

Resampling Time Comparison

To test the relative computational time requirement and appearance of the three resampling approaches, a preliminary test was performed where three bands of a scene were resampled using each method. Using the time for nearest neighbor as a baseline, the bilinear interpolation took 1.5 times longer and bicubic interpolation took 3.2 times longer. Since the task before us would require many hours of computational time on a PC, the bilinear interpolation method was chosen due to its reasonable tradeoff between computational efficiency and appearance of the image. Since the image is being used primarily as a visual backdrop for the soil boundaries, this method was deemed to be suitable.

Results and Conclusions

For many reasons, the Florida SCS is interested in creating a digital soil map base. The procedure for transforming the existing paper maps to a digital format is tedious and must be performed to exacting standards, but the end product is worthy of the effort. Satellite imagery as a backdrop to the digital soil map is sensible because the imagery is already in a digital form and can be easily updated. The imagery, however, must be carefully rectified to the soil maps so that, when layered together, all elements will spatially match. Figure 6 shows an example of the final appearance of the soil boundaries superimposed on the resampled image layer.

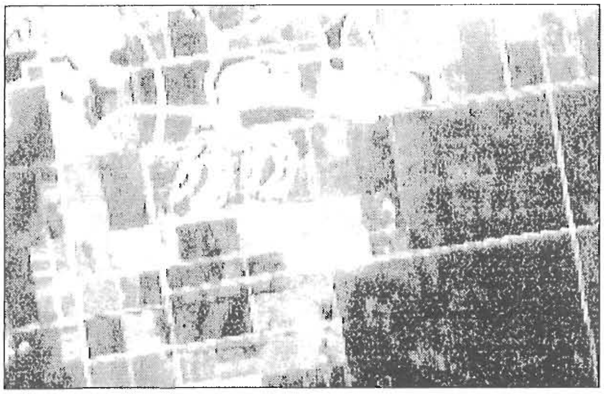


Figure 5a: Raw image.

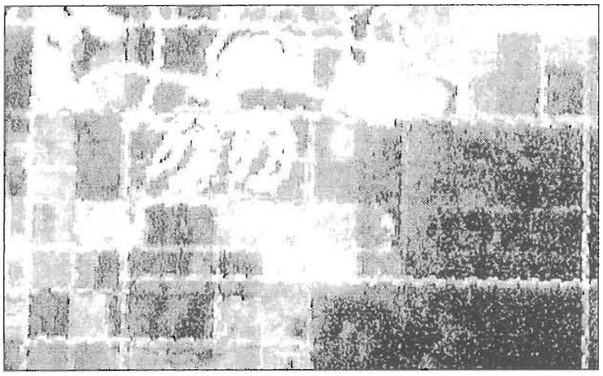


Figure 5b. Nearest neighbor interpolation.

Figure 5c. Bilinear interpolation.

Figure 5d. Bicubic interpolation.

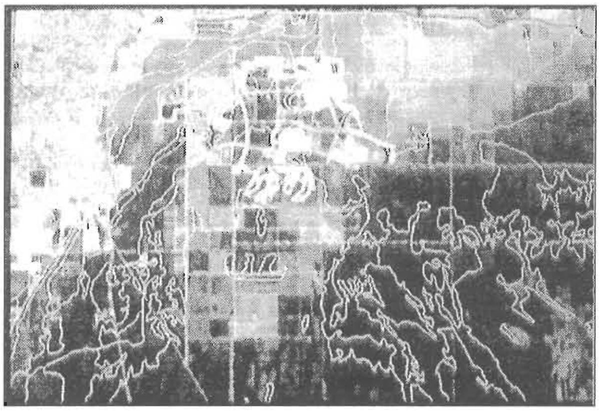


Figure 6. Resampled image with superimposed soil boundaries.

REFERENCES

American Society of Photogrammetry. 1983. *Manual of Remote Sensing*. 2nd ed. Falls Church, VA: ASP.

Billingsley, F.C. 1985. "Review of Image Processing Fundamentals." *Digital Image Processing*. SPIE, vol. 528.

Cartographic and Geographic Information Systems Staff. 1987 *Technical Specifications for Line-Segment Digitizing of Detailed Soil Survey Maps*. Soil Conservation Service, U.S. Department of Agriculture. Washington, D.C.: Government Printing Office.

Moik, J.G. 1980. *Digital Processing of Remotely Sensed Images*. NASA SP-431. Washington, D.C.: Government Printing Office.

Soil Conservation Service Staff. 1981 *Land Resource Regions and Major Land Resource Areas of the United States*. Soil Conservation Service, U.S. Department of Agriculture. Agricultural Handbook no. 296. Washington, D.C.: Government Printing Office.

Soil Conservation Service Staff. 1984. *Procedures for Collecting Soil Samples and Methods of Analysis for Soil Survey*. Soil Conservation Service, U.S. Department of Agriculture. Washington, D.C.. Government Printing Office.

Soil Conservation Service Staff and Iowa State University Statistical Laboratory. 1987 *Basic Statistics: 1982 National Resources Inventory*. Soil Conservation Service, U.S. Department of Agriculture, Statistical Bulletin no. 75 Washington, D.C.: Government Printing Office.

Soil Conservation Service Staff. 1993. *National Soil Survey Handbook*. Soil Conservation Service, U.S. Department of Agriculture. Washington, D.C.. Government Printing Office.

Soil Survey Staff. 1975. *Soil Taxonomy. A Basic System of Soil Classification for Making and Interpreting Soil Surveys*. Soil Conservation Service, U.S. Department of Agriculture. Washington, D.C.: Government Printing Office.

Soil Survey Staff. 1983. *National Soils Handbook*. Soil Conservation Service, U.S. Department of Agriculture. Washington, D.C.: Government Printing Office.

Soil Survey Staff. 1991 *State Soil Geographic Data Base (STATSGO): Data Users Guide*. Soil Conservation Service, U.S. Department of Agriculture, Miscellaneous Publications no. 1492 Washington, D.C.: Government Printing Office.

Soil Survey Staff. 1994. *Keys to Soil Taxonomy*. Agency for International Development and U.S. Dept. of Agriculture—Soil Management Support Services, Technical Monograph no. 6. Ithaca, NY: Cornell University. ■