An Algorithm for Automated Map Mosaicing Prior to Georegistration

Guofan Shao*, Huazhong Zhu§ and Walter L., Mills, Jr.*

* Department of Forestry and Natural Resources, Purdue University,West Lafayette, IN 49707-1159, USA
§ Chinese Ecosystem Research Network, CAS P.O. Box 9717, Beijing 100101, P.R. China

Abstract
An automated process of mosaicing non-registered map sheets or images is proposed and examined with a case study. When there are not any or enough control points available from individual map sheets or images, it is sometimes necessary to assemble all the adjacent map sheets prior to the rectification of an entire map mosaic. Conventional manual methods can easily result in human errors in map mosaicing. The proposed technique involves the shifting and rotation of adjacent map sheets with the condition of minimizing the overall errors. A weighting factor that reflects the analyst's confidence in various connection points and map sheets can be incorporated in the mosaic process. An initial application of this mosaic technique indicated its great potential for broad applications. A visual basic computer program was written and thus this technique can be applied to a wide range of applications.

I. INTRODUCTION

The original definition of rectification or transformation is the process of producing, from a tilted or oblique photograph, a photograph from which displacement due to tilt has been removed (ASCE, ACSM, and ASPRS, 1994). A photograph may be rectified optically, graphically, or mathematically. With the availability of computers, mathematical rectification can correct non-verticality of photograph and, in suitable cases, relief-caused displacement. This digital technology has also been used broadly in registering satellite imagery and other images (Campbell, 1996; Jensen, 1995). Several different mathematical models, such as polynomials or Fourier series, may be used to accomplish the transformation processes (McGlone, 1996). All transformation process requires at least three or four control points that are evenly distributed and clearly identifiable on an image or map sheet.

In most cases, if not all, it is necessary to assemble map sheets for adjacent areas into a single map. If adjacent map sheets are pre-registered, the processes of map mosaicing can be accomplished with straight forward techniques provided by many GIS (Geographic Information Systems) or image processing software systems. For example, the USGS topographic maps have been scanned and mosaiced by MAPTECH® software (www.maptech.com). However, it is not always possible to easily or accurately rectify every map sheet. There are situations when there are not sufficient control points available on each individual map sheets, such as old survey maps or aerial photographs for deserts or water bodies, to register each map sheet that is to be mosaiced. In these cases, it is necessary to assemble all the adjacent map sheets and then to rectify the entire map mosaic. Mosaicing of non-registered map sheets can be done manually by placing two adjacent map sheets on a digitizing tablet and digitizing the mosaiced sheets. This is often impractical because the resulting composite sheet could be too large for a digitizer and the maps could be damaged through folding or cutting them (Star and Estes, 1989). Since this line-up procedure is a manual operation, subjective errors can not easily be avoided (Demers, 1997). This is particularly important when there are a large number of map sheets that need to be mosaiced because potential errors could accumulate through the connecting of a chain of map sheets.

This paper proposes a precise and computer-automated method to perform map assembly prior to rectification. This technique consists of three steps: (1) scan all the map sheets and locate connection points along the edges of map sheets on screen; (2) choose a reference map sheet from the mosaic map area and append all other map sheets to the reference map sheet; and (3) transform individual map sheets to the entire mosaic and then register the mosaic. While step 1 and 3 are standard GIS techniques, step 2 will be explained in detail in this paper.

II. SHIFTING AND ROTATING CONNECTION POINTS

When appending one map sheet to its adjacent map sheet(s), connection points or line-up points on the edge of the map sheet are shifted and then rotated. It is assumed that there are n ($n \geq 2$) connection points

1082-4006/00/0601-97$5.00

between each pair of the map sheets, the rotation angle is a degrees, and the shifts of x, y coordinates are b_x and b_y. Due to pre-existing errors in the original drawing, tracing, and/or scanning, it is impossible to have every pair of connection points merged into a single position after the shifting and rotation. Assuming that the differences in x and y values for the ith pair of connection points between one map sheet and another are Δx_i and Δy_i, let

$$\Delta x_i = (x_i' \cos\alpha - y_i' \sin\alpha + b_x) - x_i \quad (i = 1, 2, \ldots, n) \quad (1)$$

$$\Delta y_i = (x_i' \sin\alpha + y_i' \cos\alpha + b_y) - y_i \quad (i = 1, 2, \ldots, n) \quad (2)$$

where, x_i' and y_i' are x, y coordinate values for the ith connection point from the map sheet that needs to be shifted and rotated; x_i and y_i are x, y coordinate values for the ith connection point from a reference map sheet.

The task is to minimize the differences between all the paired connection points. Let $q = \cos\alpha$, $k = \sin\alpha$, and

$$f(q,k,b_x,b_y) = \sum_{i=1}^{n}(\Delta x_i^2 + \Delta y_i^2) =$$

$$\sum_{i=1}^{n}\left((qx_i' - ky' + b_x - x_i)^2 + (kx_i' + qy_i' + b_y - y_i)^2\right) \quad (3)$$

With a condition of $q^2 + k^2 = 1$, a system of five equations are composed as follows:

$$\lambda\frac{\partial Q}{\partial q} + \frac{\partial f}{\partial q} = 0 \quad (4)$$

$$\lambda\frac{\partial Q}{\partial k} + \frac{\partial f}{\partial k} = 0 \quad (5)$$

$$\frac{\partial f}{\partial b_x} = 0 \quad (6)$$

$$\frac{\partial f}{\partial b_y} = 0 \quad (7)$$

$$q^2 + k^2 = 1 \quad (8)$$

where, $Q = q^2 + k^2 - 1 = 0$ and λ is a multiplier that needs to be determined for holding (4) and (5).

This system can then be expressed as

$$2q\lambda + \frac{\sum_{i=1}^{n}\begin{pmatrix}\partial(x_i'q - y_i'k + b_x - x_i)^2 + \\ \partial(x_i'k + y_i'q + b_y - y_i)^2\end{pmatrix}}{\partial q} = 0 \quad (9)$$

$$2k\lambda + \frac{\sum_{i=1}^{n}\begin{pmatrix}\partial(x_i'q - y_i'k + b_x - x_i)^2 + \\ \partial(x_i'k + y_i'q + b_y - y_i)^2\end{pmatrix}}{\partial k} = 0 \quad (10)$$

$$\frac{\sum_{i=1}^{n}\begin{pmatrix}\partial(x_i'q - y_i'k + b_x - x_i)^2 + \\ \partial(x_i'k + y_i'q + b_y - y_i)^2\end{pmatrix}}{\partial b_x} = 0 \quad (11)$$

$$\frac{\sum_{i=1}^{n}\begin{pmatrix}\partial(x_i'q - y_i'k + b_x - x_i)^2 + \\ \partial(x_i'k + y_i'q + b_y - y_i)^2\end{pmatrix}}{\partial b_y} = 0 \quad (12)$$

$$q^2 + k^2 = 1 \quad (13)$$

Thus, a system of five equations with six variables including λ is derived as

$$q\lambda + q\sum_{i=1}^{n}(x_i'^2 + y_i'^2) +$$

$$b_x\sum_{i=1}^{n}x_i' + b_y\sum_{i=1}^{n}y_i' - \sum_{i=1}^{n}x_ix_i' - \sum_{i=1}^{n}y_iy_i' = 0 \quad (14)$$

$$k\lambda + k\sum_{i=1}^{n}(x_i'^2 + y_i'^2) -$$

$$b_x\sum_{i=1}^{n}y_i' + b_y\sum_{i=1}^{n}x_i' + \sum_{i=1}^{n}x_iy_i' - \sum_{i=1}^{n}x_i'y_i = 0 \quad (15)$$

$$q\sum_{i=1}^{n}x_i' - k\sum_{i=1}^{n}y_i' + nb_x - \sum_{i=1}^{n}x_i = 0 \quad (16)$$

$$q\sum_{i=1}^{n}y_i' + k\sum_{i=1}^{n}x_i' + nb_y - \sum_{i=1}^{n}y_i = 0 \quad (17)$$

$$q^2 + k^2 = 1 \quad (18)$$

To eliminate λ, multiply Eq. 14 by k and Eq. 15 by q. Hence,

$$k\left(-b_x\sum_{i=1}^{n}x_i' - b_y\sum_{i=1}^{n}y_i' + \sum_{i=1}^{n}x_ix_i' + \sum_{i=1}^{n}y_iy_i'\right) =$$

$$q\left(b_x\sum_{i=1}^{n}y_i' - b_y\sum_{i=1}^{n}x_i' - \sum_{i=1}^{n}x_iy_i' + \sum_{i=1}^{n}x_i'y_i\right) \quad (19)$$

From Eq. 16 and 17, b_x and b_y are derived as

$$b_x = -\frac{q\sum_{i=1}^{n}x_i' - k\sum_{i=1}^{n}y_i' - \sum_{i=1}^{n}x_i}{n} \quad (20)$$

$$b_y = -\frac{q\sum_{i=1}^{n}y_i' + k\sum_{i=1}^{n}x_i' - \sum_{i=1}^{n}y_i}{n} \quad (21)$$

Substituting b_x and b_y in Eq. 19 with Eq. 20 and 21, it is derived that

$$k\left(-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_i' - \sum_{i=1}^{n}y_i\sum_{i=1}^{n}y_i' + n\sum_{i=1}^{n}x_ix_i' + n\sum_{i=1}^{n}y_iy_i'\right) =$$

$$q\left(\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i' - \sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i' - n\sum_{i=1}^{n}x_iy_i' + n\sum_{i=1}^{n}x_i'y_i\right) \quad (22)$$

Eq. 22 and 18 form a system that is composed of two equations with two variables: q and k. Hence,

$$q = \pm \sqrt{\frac{1}{1+\left(\dfrac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i' - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i' - n\sum_{i=1}^{n} x_i y_i' + n\sum_{i=1}^{n} x_i' y_i}{-\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i' - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} y_i' + n\sum_{i=1}^{n} x_i x_i' + n\sum_{i=1}^{n} y_i y_i'}\right)^2}} \quad (23)$$

$$k = \pm\sqrt{1-q^2} \quad (24)$$

$$b_x = -\frac{q\sum_{i=1}^{n} x_i' - k\sum_{i=1}^{n} y_i' - \sum_{i=1}^{n} x_i}{n} \quad (25)$$

$$b_y = -\frac{q\sum_{i=1}^{n} y_i' + k\sum_{i=1}^{n} x_i' - \sum_{i=1}^{n} y_i}{n} \quad (26)$$

The above solution was derived with equal importance for all connection points. If a weighting factor w_i of ith connection point is introduced for adjusting the importance or confidence of connection point i, the f function is then expressed as

$$f(q,k,b_x,b_y) = \sum_{i=1}^{n} w_i\left(\Delta x_i^2 + \Delta y_i^2\right) \quad (28)$$

where, w_i is ranges from 0 to 1. If $w_i = 0$, the ith connection point is not involved in the computation. In other words, the ith point is not important; if $w_i = 1$, the ith connection point has 100% of weighting effect; if $0 < w_i < 1$, the ith connection point is partially weighted.

The solutions for a weighted scenario are derived as

$$q = \pm \sqrt{\frac{1}{1+\left(\dfrac{\sum_{i=1}^{n} w_i x_i \sum_{i=1}^{n} w_i y_i' - \sum_{i=1}^{n} w_i y_i \sum_{i=1}^{n} w_i x_i' - \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i^2 x_i y_i' + \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i^2 x_i' y_i}{-\sum_{i=1}^{n} w_i x_i \sum_{i=1}^{n} w_i x_i' - \sum_{i=1}^{n} w_i y_i \sum_{i=1}^{n} w_i y_i' + \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i^2 x_i x_i' + \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i^2 y_i y_i'}\right)^2}} \quad (29)$$

$$k = \pm\sqrt{1-q^2} \quad (30)$$

$$b_x = -\frac{q\sum_{i=1}^{n} w_i x_i' - k\sum_{i=1}^{n} w_i y_i' - \sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad (31)$$

$$b_y = -\frac{q\sum_{i=1}^{n} w_i y_i' + k\sum_{i=1}^{n} w_i x_i' - \sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \quad (32)$$

III. SNAP CONNECTION POINTS

After the shifting and rotation, connection points at a single location are still not merged with each other because their Δx and Δy are not equal to zero. There are at least two points (one point from each map) at a joint position depending on the number of adjacent maps at an intersection point. Let m be the number of connection points that merge at a joint location and their coordinate values be (x_1, y_1), (x_2, y_2), …, (x_m, y_m). The merging location (x, y) of these connection points can be defined with a condition of Minimize

$$F = \sum_{k=1}^{m}\left((x-x_k)^2 + (y-y_k)^2\right) \quad (33)$$

Differentiating Equation 33 with respect to x and y yields

$$\frac{\partial F}{\partial x} = nx - \sum_{k=1}^{m} x_k = 0 \quad (34)$$

$$\frac{\partial F}{\partial y} = ny - \sum_{k=1}^{m} y_k = 0 \quad (35)$$

Hence,

$$x = \frac{\sum_{k=1}^{m} x_k}{m} \quad (36)$$

$$y = \frac{\sum_{k=1}^{m} y_k}{m} \quad (37)$$

The importance or confidence assigned a map can be incorporated in this analysis by changing the function F (Eq. 33) to

$$\text{Minimize} \quad F = \sum_{k=1}^{m} W_k\left((x-x_k)^2 + (y-y_k)^2\right) \quad (38)$$

where, W_k is a weighting factor for kth map. W_k ranges from 0 to 1. If $W_k = 0$, it means that the kth map will be moved to snap to other map(s). Only one map can have a weighting factor of zero for a merging location.

When this step is accomplished, a frame or web of connection points is established. All the map sheets can be joined together by performing standard transformations available in most GIS and image processing software systems (Campbell, 1996; Jensen, 1995; McGlone, 1996).

IV. DATA STRUCTURE FOR COMPUTER PROGRAM

The automated map mosaicing procedure described above can be programmed with most computer languages including those associated with GIS packages. A Visual Basic computer language (http://www.microsoft.com/) program written with was written and used for this project. This program assumes input data with a data structure as shown in Table 1.

The order of map sheet number (in the first column) is the order for computation sequence. The No. 1 map sheet is used as a reference or starting point, which is normally the most reliable map sheet found in a corner or the center of the desired mosaic. The map sheets with higher numbers are appended to the map sheets with lower numbers. The connection points (in the third column) are numbered in such a way that the connection points that share the same location between adjacent map sheets should have the same number (Figure. 1). No duplicate connection point numbers can be used on different locations across the entire mosaic. However, every connection point may have unique x, y coordinates. The map weighting factor used in Eq. 38 (column 2, Table 1) can be set as 1 for the initial computation.

Figure 1 is an example of four map sheets. In this case, map sheet 1 contains connection points 1, 2, 3, 4, 5, and 6 with $n_1 = 6$; map sheet 2 contains connection points 2, 3, 4, 7, 8, and 9, with $n_2 = 6$; map sheet 3 contains connection points 4, 5, 6, 10, 11, and 12 with $n_2 = 6$; and map sheet 4 contains connection points 4, 9, 10, and 13 with $n_4 = 4$. Among the 13 connection points, 2, 3, 5, 6, 9, and 10 are shared by two maps and 4 is shared by four maps. The x, y coordinates of these shared connection points are different between adjunct map sheets.

The output table contains four columns (Table 2). The first column shows connection point number ($N = 13$ in Figure 1); second and third columns contain corrected x, y coordinate values for each connection point; the fourth column lists error information (standard deviation) for each connection point. The connection

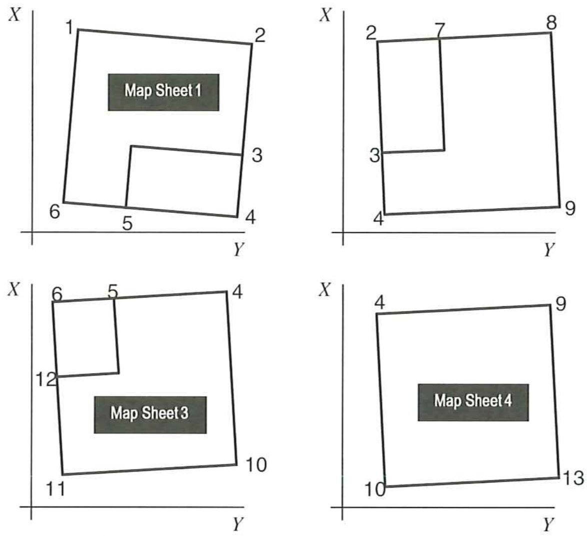

Figure 1. An illustration of numbering connection points. Although the shared connection points have the same numbers, their x, y coordinate values are different between adjunct map sheets.

Table 1. The structure of input data set for automated map mosaicing

Map Sheet Number &	*Weighting Factor*	*Connection Point Number*	*Weighting Factor*	*x Coordinate*	*y Coordinate*
1	W_1	1, 2, ⋮ } n_1	$w(1)$, $w(2)$, ⋮	$x(1)$, $x(2)$, ⋮	$y(1)$, $y(2)$, ⋮
2	W_2	a, b, ⋮ } n_2	$w(1+n_1)$, $w(2+n_1)$, ⋮	$x(1+n_1)$, $x(2+n_1)$, ⋮	$y(1+n_1)$, $y(2+n_1)$, ⋮
⋮	⋮	⋮	⋮	⋮	⋮
M	W_M	α, β, ⋮ } n_M	$w(1+n_{M-1})$, $w(2+n_{M-1})$, ⋮	$x(1+n_{M-1})$, $x(2+n_{M-1})$, ⋮	$y(1+n_{M-1})$, $y(2+n_{M-1})$, ⋮

Table 2. The output form after shifting, rotating and snapping all connection points.

Connection Point Number	X Coordinate	Y Coordinate	Standard Deviation
1			
2 } N			
⋮			

points with large errors could be reduced through iterative calculations. The easiest way to eliminate the ith connection point is to set the weighting factor (w_i in Eq. 28) equal to 0. The corrected x, y coordinate values are used to transform or rectify every map sheet. The first order or higher order transformation may have to be applied in order to minimize RMS errors. A mosaic as a whole can be registered with some known geo-referenced control points that are clearly locatable throughout the mosaic.

V. APPLICATION ASPECT

The method was applied initially for assembling forest inventory map sheets for a biosphere reserve, Denezhkin Kamen, in Russia. Located in the North-Ural Province of the Ural mountain region, the reserve covers a forested area of 78,192 ha (Sekerina, 1998). In 1984, a Russian Federal Forest Inventory Expedition prepared a forest inventory map that consists of 54 sheets, scaled at 1:10,000. These map sheets contained polygon lines for quarter nets, sized 1 – 3 km, and forest stands but did not contain geo-reference markers. Each map sheet contained 3 to 12 quarters and the corners of the quarters along the edges of map sheets were used as connection points. All the map sheets were scanned and connection points were interactively located using image processing software Erdas Imagine (Erdas, 1997).

The initial transformations resulted in acceptable RMS errors for a majority of map sheets. By eliminating some questionable connection points, all the map sheets were assembled again resulting in higher accuracy. The overall RMS error for the final transformation was less than 20 m (Sekerina, 1998).

This experiment showed that the automated mosaic technique is particularly useful when there are not enough reliable control points that can be located on at least some of map sheets. Its has the following five advantages:

(1) mosaic procedure is repeatable by different analysts;

(2) connection points can be precisely located on screen by zooming to each map sheet;

(3) map shifting and rotation represent the conventional manual map-placing procedure but result in higher accuracy;

(4) large errors can be avoid by changing the weighting factors and eliminating the worse connection points;

(5) it is faster and simpler (as long as a computer program is available) than the conventional manual method.

Further tests are needed in several aspects including connection points selection, weighting factor control, computation order sorting, involving more reference maps, and merging with GIS software.

ACKNOWLEDGMENTS

Authors thank the Department of Forestry and Natural Resources, Purdue University, for providing convenient working conditions and computer facilities. The Chinese Academy of Sciences for supporting the second author's short-term visit at Purdue University. Thanks to Ms. Sekerina who provided data for the experiment, Ms. Zhuang who helped digitizing all the maps and Drs. Rhodes and Le Master whom provided financial support.

REFERENCES

[1] ASCE, ACSM, and ASPRS, 1994, *The glossary of the mapping* sciences. American Society of Civil Engineering, American Congress on Surveying and Mapping, and American Society for Photogrammetry and Remote Sensing.

[2] Campbell, J. B., 1996, *Introduction to Remote Sensing*, Guilford Press, New York.

[3] Demers, M. N., 1997, *Fundamentals of Geographic Information Systems,* New York: John Wiley & Sons, Inc.

[4] Erdas, 1997, *Erdas® Field Guide*. Printed in the United States of America by Erdas, Inc.

[5] Jensen, J. R., 1995, *Introductory Digital Image Processing*. Prentice-Hall, New Jersey.

[6] McGlone, C., 1996, Sensor modeling in image registration. In: *Digital Photogrammetry*, Cliff Greve (ed.), 115-123. American Society for Photogrammetry and Remote Sensing.

[7] Sekerina, A.Y., 1998, Development of a GIS Database for Application in Russian Biosphere Reserves. Master Degree Thesis, Department of Forestry and Natural Resources, Purdue University.

[8] Star, J and Estes, J., 1989, *Geographic Information Systems: An Introduction*. Englewood Cliffs: Prentice-Hall.