Tạp chí Tin học và Điều khiển học, T. 15, S. 2 (1999), 54-60

GEOMETRICAL CORRECTION AND GEO-REFERENCING FOR AUTOMATIC MAP DATA ENTRY

NGO QUOC TAO

Abstract. This paper presents the method of Geometrical Correction and Geo-referencing for Automatic Map Data Entry. We propose the method to find linear and polynomial transformations. We transform relative coordinate system maps vectorized into maps of projection coordinate systems. Beside, we use this method to merge two pages of an image (a map) into a page. The paper also present the selecting most appropriate projection-best fitting. This technique is implemented in the software packages MapScan and POPMAP. This technique is applied to maps such as topography, hydrography, transportation, satellite maps, etc.. Besides it is also used for geometrical correction of technical drawings, electronic circuits, fingerprints, etc..

1. INTRODUCTION

For automatic and semi-automatic map data entry system, the rubber sheeting correction and transforming relative coordinate map into real-world coordinate map with projection are very important. A map is the presentation of earth shape. There are many algorithms for rubber sheeting correction such as affine or polynomial transformations. In the software package MapScan, we use only the linear and bilinear transformations. These transformations are used in most of geographic information systems. In order to define these transformation, we can apply the root mean square method. Then we can compute effects of polynomial transformations. In order to transform a relative coordinate map to the real-world coordinate map, we choose special points in the relative coordinate map that we know the real coordinates of these points. We consider these points as control points. We use linear or polynomial transformations from the relative coordinate map to the real-world coordinate map.

After the vectorization process the map is in a relative coordinate system (plain x, y). The map is geo-referenced at this stage in order to convert the generated vector data into the real-world coordinate system using known ground control points and information about the map projection. For a given set of control points, the first component is a point in the relative coordinate source map and the second component is a point in the real-world coordinate systems. We must use the library GCTP [14] (General Cartographic Transformation Package: National Mapping Program Technical Instrudtions, U.S. Geological Survey, National Mapping Division) to convert the units of the second components of the set of control points into a necessary units of user. There are about thirty standard projections in this library such as UTM (Universal Transverse Mercator), Mercator, Albert, Lambert, Conic, etc..

We can transform a projection of a map. Each projection of a map needs less than 13 parameters. The parameters of the projection of the map can be *Radius of reference sphere (6370997 meters), Latitude of the standard parallel, Latitude of the first standard parallel, Latitude of the second standard parallel, CentMer Longitude of the central meridian, OriginLat Latitude of the projection origin, False easting in the same units as the semi-major axis, False northing in the same units as the semi-major axis, TrueScale, Latitude of true scale, Longitude down below pole of map, Scale factor at central meridian (Transverse Mercator) or center of projection (Hotine Oblique Mercator). CentLon Longitude of center of orojection. CenterLat Latitude of center of projection. Height of p[erspective point, Longitude of first point on center line (Hotine Oblique Mercator, format A) etc.. The parameters of each projection are supplied by users.*

Sometime we don't know exactly the projection of maps vectorized/digitized. Therefore we must recognize the projection of maps. We can do it by using RMS errors. We can choose some control points and extract parameters from these control points. Depending on the parameters, we can choose 10 projections having smallest root mean squares that are called the best projections (the best fitting projection).

In the Section 2 we discuss the detail of the rubber sheeting correction. In the Section 3, we introduce the combining of projection and rubber sheeting correction. Section 4 discusses the automatic defining of best appropriate projection. In the Section 5, we discuss the result of implementing the rubber sheeting map correction and map projections.

2. THE METHOD OF RUBBER SHEETING CORRECTION

Definition 2.1 (Control points). A control point is a pair (p, p') of points in two-dimension Euclidean space, where p is a point in a relative coordinate map and p' is a point in the earth surface with some projection. We call p and p' the first and second component of control point (p, p).

The polynomial method

Let $p = (x, y)$ be a point, we need define a point $p' = (x', y')$, where x', y' are functions of x, y. We have a control point $((x, y), (f_1(x, y), f_2(x, y))$. Now we need determine function $f_i : R^2 \to R$, $i = 1, 2$, where

$$f_k(x, y) = \sum_{i+j\leq n} a_{ij}^{(k)}\, x^i\, y^j\,. \tag{1}$$

In order to determine a polynomial with degree n, we need $(n + 1)(n + 2)/2$ parameters. Therefore we need $(n + 1)(n + 2)/2$ control points to define the function f_i.

The linear method

Replacing $n = 1$ to equation (1), we obtain the following linear method

$$\begin{cases} y_1 = f_1(x, y) = a_{10}^{(1)}\, x + a_{01}^{(1)}\, y + a_{00}^{(1)} \\ y_2 = f_2(x, y) = a_{10}^{(2)}\, x + a_{01}^{(2)}\, y + a_{00}^{(2)} \end{cases}$$

The bilinear method

Replacing $n = 2$ to equation (1) we obtain following bilinear method

$$\begin{cases} y_1 = f_1(x, y) = a_{20}^{(1)}\, x^2 + a_{11}^{(1)}\, xy + a_{02}^{(1)}\, y^2 + a_{10}^{(1)}\, x + a_{01}^{(1)}\, y + a_{00}^{(1)} \\ y_2 = f_1(x, y) = a_{20}^{(2)}\, x^2 + a_{11}^{(2)}\, xy + a_{02}^{(2)}\, y^2 + a_{10}^{(2)}\, x + a_{01}^{(2)}\, y + a_{00}^{(2)} \end{cases}$$

Calculating Error for each and for all the control points

Definition 2.2. Error RMS (Root Mean Square) of a transformation for a set of control points $C_p = \{((x_i, y_i), (x_i', y_i'))\}$. In the case of the theoretical model $x' = f_1(x, y)$, $y' = f_2(x, y)$, then the errors for each and all control points are written as follow:

$$\mathrm{RMS}\,(C_p) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left[(f_1(x_i, y_i) - x_i')^2 + (f_2(x_i, y_i) - y_i')^2\right]}\,. \tag{2}$$

The basic idea of rubber sheeting correction is to define the parameters a_{ij} of the transformation from equation (1) so that RMS (C_p) is minimum. We show that parameters which are deduced [9] for linear transformation are not correct. Therefore, we determine the general formulation of $a_{pq}^{(k)}$. In order to determine $a_{pq}^{(k)}$ $p, q = \overline{1, n}$ we use the smallest root square. By replacing (1) into (2) we have the following equation:

$$\mathrm{Error}\,(p_k) = \sqrt{\Big(\sum_{i+j\leq n} a_{ij}^{(1)}\, x_k^i\, y_k^j - x_k'\Big)^2 + \Big(\sum_{i+j\leq n} a_{ij}^{(2)}\, x_k^i\, y_k^j - y_k'\Big)^2}$$

and

$$\mathrm{RMS}\,(C_p) = \sqrt{\frac{1}{m}\sum_{k=1}^{m}\Big[\Big(\sum_{i+j\le n} a_{ij}^{(1)} x_k^i y_k^j - x_k'\Big)^2 + \Big(\sum_{i+j\le n} a_{ij}^{(2)} x_k^i y_k^j - y_k'\Big)^2\Big]}\,. \tag{3}$$

The partial derivative of $\mathrm{RMS}^2(C_p)$ with respect to $a_{pq}^{(l)}$ is equal to 0, where $l = 1, 2$, $p+q \le n$. We have the following linear system:

$$\begin{cases} \sum_{k=1}^{m}\Big(\sum_{i+j\le n} a_{ij}^{(1)} x_k^{i+p} y_k^{j+q} - x_k' x_k^p y_k^q\Big) = 0, \; p+q\le n, \; 0\le p,q\le n \\ \sum_{k=1}^{m}\Big(\sum_{i+j\le n} a_{ij}^{(2)} x_k^{i+p} y_k^{j+q} - y_k' x_k^p y_k^q\Big) = 0, \; p+q\le n, \; 0\le p,q\le n \end{cases} \tag{4}$$

The equation (4) can be written into two linear equation systems as follows

$$\Big\{ \sum_{i+j\le n} a_{ij}^{(1)} \sum_{k=1}^{m} x_k^{i+p} y_k^{j+q} = \sum_{k=1}^{m} x_k' x_k^p y_k^q, \; p+q\le n, \; i,j = \overline{1,n} \tag{5a}$$

and

$$\Big\{ \sum_{i+j\le n} a_{ij}^{(2)} \sum_{k=1}^{m} x_k^{i+p} y_k^{j+q} = \sum_{k=1}^{m} x_k' x_k^p y_k^q, \; p+q\le n, \; i,j = \overline{1,n}\,. \tag{5b}$$

Let us denote

$$\begin{aligned} u_{\theta(i,j)}^{(k)} &= a_{ij}^{(k)}, \quad A_{\theta(i,j),\theta(p,q)}^{(k)} = \sum_{k=1}^{m} x_k^{i+p} y_k^{j+q}, \\ b_{\theta(p,q)}^{(1)} &= \sum_{k=1}^{n} x_k' x_k^p y_k^q, \quad b_{\theta(p,q)}^{(2)} = \sum_{k=1}^{n} y_k' x_k^p y_k^q, \end{aligned} \tag{6}$$

where $\theta_{(i,j)} = (i+j)^*(i+j+1)/2 + j$. We can prove that $\theta(i,j)$ is uniquely defined for each pair (i,j).

By replacing the equation (6) into (5a) and (5b) we obtain the following linear system:

$$\Big\{ \sum_{i=1}^{\frac{(n+1)(n+2)}{2}} u_i^{(k)} A_{ij}^{(k)} = b_j^{(k)}, \; 0\le j\le \frac{(n+1)(n+2)}{2}, \; 1\le k\le 2\,. \tag{7}$$

The equation (7) can be solved by Gauss method. For $n = 1, 2$ we can obtain linear transformation or degree two polynomial respectively.

3. COMBINING PROJECTION WITH RUBBER SHEETING CORRECTION

In this section we discuss the transform of a set of points of plane-coordinates into a set of points with real-world coordinates. The problem of geometrical correction is based on a set of control points C_p so that $(p, p') \in C_p$, p' is a point of a map in a specified projection and coordinate system.

Let C_p be a set control points. Let us denote $P_f(C_p)$ be a set of points that consist of the first components of control points of C_p and $P_t(C_p)$ a set of points that consist of the second components of control points of C_p.

We must find a transformation that image of $P_f(C_p)$ is approximate to $P_t(C_p)$ so that RMS (C_p) is minimum. The difficult problem is to compute RMS (C_p) by length unit, but unit of $P_t(C_p)$ can be angular measures such as degree or radian units.

Therefore, we must transform a control point (p, p') in to control (p, p'') by using library GCTP, where p'' is calculated by length unit (met, kilometer, mile). Using this transformation C_p is replaced by $C'_p = \{(p, p'') \mid (p, p') \in C_p$ and p'' is the image of $p,\}$ and geometrical correction is based on the set of control points C'_p.

4. THE METHOD FINDING SOME BEST PROJECTIONS

4.1. Definitions

Denote Pro_set be a set of map projections in the projection library. Denote **Pro_user** be a set of map projection supported by users so that **Pro_user** $\subseteq$ **Pro_set.** From the set of projections **Pro_user**, k best projections can be automatically detected.

Let $E = \mathrm{Rec}(P_t(C_p))$.

Depending on E, a data base and a set of projections Pro_user proposed by users, a software can automatically select some best projections.

The important problem is to select appropriate parameters Par_i *based on a projection* $P_i \in$ *Pro_user.*

Definition 4.1.

1. Denote $C(P)$ be the center of a set of points P which is a point so that sum of squares of distances from $C(P)$ to each point in P is minimum.
2. Denote $C_f(C_p)$ and $C_t(C_p)$ be the center of the first components $P_f(C_p)$ and the second components $P_t(C_p)$ respectively.

Definition 4.2. Given a set of points $P = \{(x, y)\}$. Let (x_0, y_0) be the center of P. Denote $\mathrm{Rec}(P)$ be a rectangle depending on P on that its center is (x_0, y_0) and it satisfies:

1. There a two points $p_1(x_1, y_1)$ and $p_2(x_2, y_2)$ of P so that
$$x_0 - x_1 = \max_{p(x,y)\in P}(x_0 - x) \text{ and } x_2 - x_0 = \max_{p(x,y)\in P}(x - x_0)\,.$$
2. There are two points (x_3, y_3) and (x_4, y_4) of P so that
$$y_0 - y_3 = \max_{p(x,y)\in P}(y_0 - y) \text{ and } y_4 - y_0 = \max_{p(x,y)\in P}(y - y_0).$$
3. The width $\mathrm{Rec}(P)$ is $2^* \min(x_0 - x_1, x_2 - x_0)$ and the length of $\mathrm{Rec}(P)$ is $2^* \min(y_0 - y_3, y_4 - y_0)$.

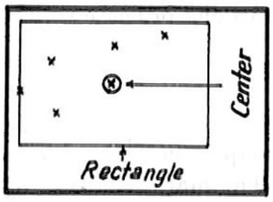

Figure 1. The center of a set of P and the rectangle $\mathrm{Rec}(P)$ surrounding P

4.2. Selecting parameters for best fitting projections

In this subsection we discuss the method selecting parameters based on projections.

Proposition 4.1. *For azimuthal projections, the closer points are to the center of projections, the smaller errors are.*

- Depending on the Proposition 4.1, we can consider $C_t(C_p)$ as the center of projections.

Proposition 4.2. *When conical and cylindrical projections have an intersection circle, the closer points are to the intersection, the smaller RMS errors are.*

- Depending on this Proposition 4.2, we can consider $C_t(C_p)$ as the center latitude and the center longitude of projections.

Proposition 4.3. *When conical and cylindrical projections have two intersection circles, the closer points are to the center latitude and the center longitude of projections, the smaller RMS errors are.*

- Depending on the Proposition 4.3, we can consider edges of rectangle Rec(C_p) as the standard latitude and longitude parallels.

4.3. Best fitting algorithm

Depending on a set of control points C_p and a list of projections **Pro_user**. MapScan can analyse C_p and compute parameters such as STDPR1 (Latitude of the first standard parallel), STDPR2 (Latitude of the second standard parallel), CentMer (Longitude of the central meridian), OriginLat (Latitude of the projection origin). These parameters are enough to determine errors for all projections of **Pro_user**. Depending on C_p, MapScan can calculate the RMS errors for all projections of **Prouser** and select k first best projections. Therefore, the best fit algorithm can be written as follows:

Input: Control points C_p.
A list of projections **Pro_user** in the GCTP library.
Output: 10 best projections.
Step 1: Depending on control points C_p and Rect(C_p), compute the parameters:
STDPR1, STDPR2, CentMer, OriginLat.
Step 2: For each proj $\in$ **Pro_user**,
(RMS(proj), project) $\rightarrow$ **list** /* computing the RMS error, store it to a queue */
Step 3: Sorting the list of projections **list** by advancing,
Printing k first projections and their RMS errors.

5. EXPERIMENT

We implemented the linear and second order transformations with map projections in the MapScan-the software package for automatic data entry. MapScan can concert relative maps into the real-world maps. Relative maps are obtained from vectorisation or imported from vector map such ad (Atlas Gis, AutoCad, MapInfor, POMAP, for DOS or POPMAP for Windows). MapScan can support about thirty map projections, Among projections, the none-earth projection is a normal geometrical correction.

5.1. Selecting control points

After the vectorization process the map is in a relative coordinate system (plain, x, y). The map is georeferenced at this stage in order to convert the generated vector data into the real-world coordinate system using known ground control points and information about the map projection.

If the units of the real-world coordinate system is measured by degrees, users can be inserted by the formats *dddmmmsss* or *decimal*. For example, if user want to input the angular value 45'30" (45 degrees and thirty minutes), they can input the numbers 049030030 or 45.51.

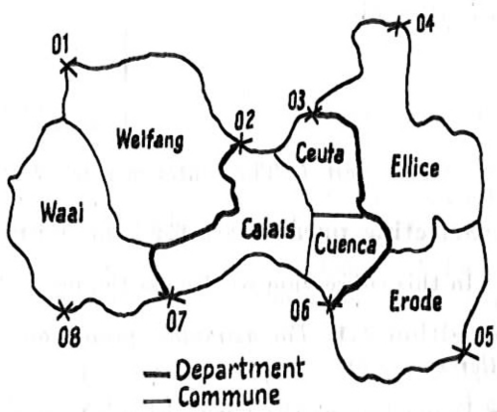

Figure 2. The set of control points

Metaland is a mall fictitious country. We will geo-reference the map based on eight ground control points identified on the map. The real world coordinates of the control points are:

Control point	Longitude	Latitude
01	0.248 degrees East	0.231 degrees South
02	0.838 degrees East	0.505 degrees South
03	1.079 degrees East	0.390 degrees South
04	1.374 degrees East	0.070 degrees South
05	1.606 degrees East	1.250 degrees South
06	1.156 degrees East	1.079 degrees South
07	0.610 degrees East	1.065 degrees South
08	0.251 degrees East	1.125 degrees South

5.2. Selecting a set of projections

Specify the control points by digitizing them on the map and entering their coordinates. Note that East longitudes and North latitudes are entered as positive numbers while West longitude and South latitudes entered as negative numbers. From the **Projections** list box (this means no projection is applied), specify the unit and format of the coordinates. When completed, click on the **OK** button. Digitize the control points on the map and enter the coordinates in the Map X and Map Y boxes respectively. When the coordinates in the Map X and Map Y boxes are completed, from the **Projections** list box click on the **Best fit** button. From **Select projections** list box, select a list of projections. When the selecting were completed, click mouse on the **OK** button to accept or **CANCEL** to escape the best fit processing. If **OK** button is clicked, MapScan will print ten best projections and their RMS errors respectively.

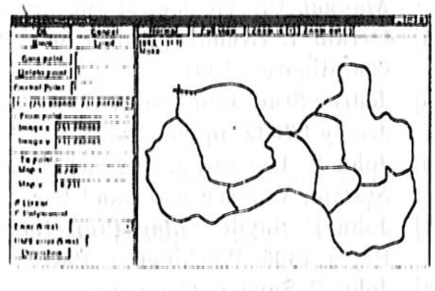

Figure 3. Select control point

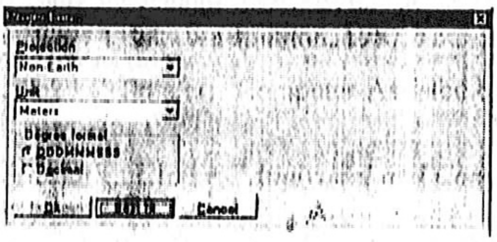

Figure 4. Select the projection, unit and format

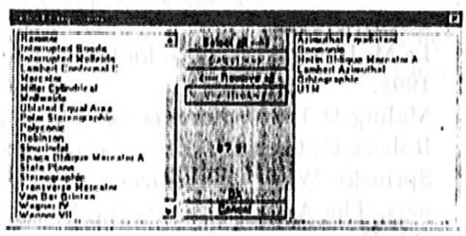

Figure 5. Select a list of projections

5.3. Changing parameters and processing

Users can select appropriate projections by inputting parameters into dialog box of projections, MapScan will compute RMS(C_p) error for each projection and display ten projections having smallest RMS(C_p) errors. Therefore users can choose one suitable projection and change parameters for this projection. When users accept these parameters and the projection, they can click on to the **OK** button to convert map into the real-world system map.

6. CONCLUSION

In this paper we have dealted with geometrical correction based on linear, second order transformations and showed some error coefficients of linear transformations [9]. Therefore we have proposed the general method for computing the coefficients of polynomials used for geometrical correction. The formulations of polynomial coefficients are easily computed. Linear transformations and second order transformations have been implemented in the software POPMAP and MAPSCAN.

Acknowledgment

The author would like to thank Professor Bach Hung Khang for his supports, suggestions and encouragement in studying the problem of rubber sheeting correction and rectification. The author would also like to thank Professor Vu Duy Man and Dr. Dang Van Duc for their helpful comments so that the presentation of the paper could be improved. Thanks are also given to Dr. Luong Chi Mai for her help and suggestions.

REFERENCES

[1] ESRI, Inc. Understanding Geographic Information System (GIS) software, ESRI South Asia OTE LTD, Singapore 238868, August 1995.

[2] Gerald I. Evenden, *Cartographic Projection Procedures*, Release 4, Interim Report, September 17, 1993.

[3] Gerald I. Evenden, *Cartographic Projection Procedures for the UNIX Environment - A user's Manual*, US. Geological Survey Open-File Report 90-284, May 13th, 1990.

[4] Gerald I. Evenden, Notes on a Method to Transform Digitized Coordinates to Geographic Coordinates, 1991.

[5] Jeffrey Star, John Ester, *Geographic Information Systems*, Prentice Hall, Englewood cliffs, New Jersey 07632, pp. 96-125.

[6] John E. Lee and J. Michael Walsh, *Map Projection for User with the Geographic Information System*, Western and Land Use Team, Washington, DC, July 1984.

[7] John P. Snyder, *Map Projections - A Working Manual*, US. Geological Survey Professional Paper 1395, Washington, 1987.

[8] John P. Snyder, *Computer-Assisted Map Projection Research*, Geological Survey Bulletin 1629, May 8th, 1995.

[9] Keith C. Clarke, *Analytical and Computer Cartography*, Prentice Hall, Englewood cliffs, New Jersey 07632.

[10] Le M. Bussgayevskiy, John P. Snyder, *Map Projections - A Reference Manual*, Taylor & Francis, 1995.

[11] Maling D. H., *Coordinate Systems and Map Projections*, Pergamm Press, 1992.

[12] Robert G. Cromley, *Digital Cartograpy*, Prentice Hall, Englewood cliffs, New Jersey 07632.

[13] Sprinsky W. H., *Transformation of Positional Geographic. Data from Paper-Based Map Products*, The American Cartographer, 1987, pp. 359-366.

[14] US. Department of the Interior, *GCTP General Cartographic Transformation Package*, Software, September 1990.

Received July 15, 1998
Revised June 1, 1999

Department of Pattern Recognition and Image Processing
Institute of Information Technology.