STUDY AND ANALYSIS OF MAP ALIGNMENT WITH MINIMUM ERROR

A DISSERTATION

Submitted in partial fulfillment of the requirements for the award of the degree

of

MASTER OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

By

AVNENDRA KUMAR

IIT Roorkee-ER&DCI, Noida
C-56/1 "Anusandhan Bhawan"
Sector 62, Noida-201 307

FEBRUARY, 2003

CANDIDATE'S DECLARATION

I hereby declare that the work presented in this dissertation titled **"STUDY AND ANALYSIS OF MAP ALIGNMENT WITH MINIMUM ERROR"**, in partial fulfillment of the requirements for the award of the degree of **Master of Technology** in **Information Technology**, submitted in **IIT, Roorkee – ER&DCI Campus, Noida**, is an authentic record of my own work carried out during the period from August 2002 to February, 2003 under the supervision of **Shri V. N. SHUKLA**, Director (Special Applications), Electronics Research and Development Centre of India, Noida.

The matter embodied in this dissertation has not been submitted by me for award of any other degree or diploma.

Date: 24/02/2003

Place: Noida

(Avnendra Kumar)

CERTIFICATE

This is to certify that the above statement made by the candidate is correct to the best of my knowledge and belief.

Date: 24/02/2003

Place: Noida

Guide:

(Mr. V.N. Shukla)

Director, Special Applications,

ER & DCI, Noida

ACKNOWLEDGEMENT

The work presented in this report would not have been completed without the guidance and support of many people. On this day, I feel deep sense of gratitude towards a visionary, **Prof. Prem Vrat**, Director, Indian Institute of Technology, Roorkee and **Shri R.K.Verma**, Executive Director ER& DCI-Noida, for creating the utmost congenial environment that sustains, enhances and incubates the thought.

I would like to take this opportunity to offer my deep sense of gratitude, **Prof. A. K. Awasthi**, Dean, PGS&R, Indian Institute of Technology, Roorkee for providing the best of course curriculum that has a lot of difference to the basic outlook of an individual.

I wish to express my deep sense of indebtedness and sincerest gratitude to my guide **Mr. V. N. Shukla**, Director Special Application and Course coordinator M.Tech. (IT) ER&DCI Noida, for his invaluable guidance & constructive criticism throughout this dissertation. He has displayed unique tolerance & understanding at every step of progress & encouraged me incessantly. I deem it my privilege to have carried out my dissertation work under his able guidance.

I am also grateful to **Prof. R.P. Agarwal**, course coordinator M.Tech (I.T) programme at Indian Institute of Technology-Roorkee, for providing necessary facilities for successful completion of this work.

I would also like thanks to **Mr. Satyajit Rath** and **Mr. Munish Kumar** (Project Engineer at ER& DCI, Noida.) for their valuable suggestions & making me to understand the problem in depth.

I am greatly indebted to all my friends who helped me directly and indirectly, and graciously applied themselves to the task of helping me with ample moral support and valuable suggestion.

Finally, I am heartily gratitude my Parents and all family members for their faith, support and continuously encouragement during this work.

(Avnendra Kumar)

Enrollment No.- 910310

CONTENTS

ABSTRACT

Since time immemorial, preparation of maps is taking place. In the process of map preparation size of the paper or drawing material plays a major role and defines the requirement of scale factor. When the size of presentation exceeds the limit, it is being splitted in to multiple sheets. When it comes to combine again, the geometrical data sêts are match from these splitted maps into a single map on the basis of common matched edge.

Edge matching is one of the most useful procedures in processing of spatial data. But several frequently used methods have showed their shortcomings. They are blind to erroneous data included in the original data, and most of them are sensitive to a threshold. An effective matching procedure is required to correct most of the errors included in the data. Geometric feature based edge matching procedure is one such method which is applied because every type of geographical element has innate geometric features represented by the geometric data (coordinates). These data are processed by comparing the geometric features. So the geometric features of elements may be regarded as credible data in the process of alignment of boundaries of a map.

A procedure has been implemented in which the used input vector data sets are considered to the scaled one. Then comparing the distance between two adjacent points of reference map edge to the candidate (common component) map edges. On the basis of the consecutive matching of the point's distance and application of geometric transformation, the resultant is a desired output map with minimum errors.

Chapter 1

INTRODUCTION

1.1 Objective

The objective of this dissertation work is to study the existing possible applied methodology used in map's boundary alignment and analysis them with minimal error.

1.2 Scope

The scope of the study and analysis of Maps alignment are included as:

- Experiencing Map alignment error propagation and analysis by understanding its background and terminology used.
- Exploring Map alignment error in composite boundaries using vectors data sets.
- Prediction of Mathematical Model and Computational outcome for Maps with minimum error by using proposed algorithm.

1.3 Introduction to Map Alignment

The Computer revolutions take over the stage from conventional means of information presentation to digital form with more accuracy and edits ability, mathematical computation made it more predictable and problem solving. This way computer has become the easiest, cheapest and most convenient means to reach the people in more understandable way even for the complex problems.

Before going into the detail it is necessary to define or know the meaning of Map alignment- it is a process of composition of two adjacent map on the basis of their common boundaries to present the right information about the piece of land without any confusion. Accurate alignment required where there are chances of getting distorted map with gap at boundaries which is known as no man's land or where there are chances of over lapping of land at the boundaries / edge which means the land having possessions of more people. Now a day every thing is going to be computerized and every concerned people or organization required the information on-line, every time in more accurate form in a quick and easy way so every government or private organization is interested to keep all the information in digitized form so that easy land management systems can be implemented. This way it is said that the geographical location based information systems requirement is increasing day by day.

This dissertation work presents the study and analysis of most existing technique and methodology in software industries and research works carried out in the fields of boundaries alignment of the maps because after digitization of maps they are required to be align and this process involved the human intervention. It means semi-automatic procedures are there to perform the alignment operation. The first approach gets into better shape from the visualization tools like AutoCAD Map 6.0 and AutoCAD 2000 from AutoDesk Corporation as well as Arc-VIEW, Arc-INFO from Environmental Science Research Institute.

The second approach is research oriented; this work comes under the Pattern Analysis and Machine Intelligence (PAMI) and System, Man and Computers (SMC) Journals. Still now computers can't carries out all the work but there is need of some human intervention, because 100% pattern recognition is not carried out so far in the map boundaries alignment. Human beings are the best recognizer.

The difficulty is that there are no satisfying matching rules to be obeyed when the original data include some error data. From the point of applications, however, a computer should mostly do matching work. That is to say, it should choose a proper method to work harder.

This work shows a progress dealing the map composition on the basis of similar edge. This is live problem in the field of boundary / edge matching and alignment or composition. The case implemented in this work is vector based in which the Geometrical Transformation plays a significant role. It required sets of input vector data sets, in which first input data vector set would be consider as the reference map and rest are the candidate's data sets of the maps which require the automatic boundaries alignment on the basis of similar edge using the proposed algorithm.

1.4 Organization of Dissertation

Chapter 1 describes the objective, scope and introduction of the Map alignment, why and where it is used.

Chapter 2 provides the literature review and study of the applied methods in detail, which are required to design a procedure as in chapter 4.

Chapter 3 it describes the problem in detail and more emphasis on Map alignment error propagation and analysis.

Chapter 4 deals the numerical process life cycle and design and implementation of the procedure.

Chapter 5 discusses the results.

Chapter 6 concludes the dissertation work with the scope to the future work.

Chapter 2

LITERATURE SURVEY

Earlier Map alignment techniques were not efficient enough and introduced certain errors that have to be eliminated manually or maintained on the basis of familiarity of the region. The basic approaches have not been changed; they are still in use with some modifications. They are as discuss below-

2.1 Current Approaches to Map's Error Handling

Current approaches are directly related to geometric transformation [1] & [2] that alters the coordinate descriptions of the objects. These basic approaches are carried out with the help of the available tool such as AutoCAD Map 6.0 and ArcView / ArcInfo.

Main basic approaches are- Transformations and Rubber sheeting.

These two basic methods, which used in the maps boundaries alignment are-

(a) Transformation- it involves basic operations such as

1. Translation
2. Scaling
3. Rotation

(b) Rubber Sheeting- It involves Shearing. This approach is not commonly used because it deforms the shape of the object.

2.1.1 Transformation

(1) Translation

Translation is applied to an object by repositioning it along a straight line path from one co-ordinate location to another. The translation of a two dimensional point by adding translational distance x_t, y_t to the original position co-ordinate(x,y) to move the point to the new position (x', y')

$$x' = x + x_t \qquad y' = y + y_t \qquad (2.1)$$

The translation distance pair (x_t, y_t) is called a translation vector or shift vector. The above translational equation cab be present to a single matrix equation by using column vectors to represent co-ordinate positions and the translation vector as: -

$$P = \begin{bmatrix} x_1 \\ . \\ x_2 \end{bmatrix} \qquad P' = \begin{bmatrix} x'_1 \\ \\ x'_2 \end{bmatrix} \qquad T = \begin{bmatrix} x_t \\ \\ y_t \end{bmatrix} \qquad (2.2)$$

this allows us to write the two dimensional translation equations in the matrix form:

$$P' = P + T \qquad (2.3)$$

Translation is a rigid body transformation that moves object without deformation. That is every point on the object is translated by the same amount. A straight line segment is translated by applying the transformation equation to each of the line end points and redrawing the line between the new end point positions.

Polygons are translated by adding the translation vector to the co-ordinate position of each vertex and regenerating the polygon using the new set vertex co-ordinate and current attribute settings.

To perform the translation, the test data vector representing should be in normal form. [x y 1]. To this normal data vector, the translation vector is multiplied to achieve translation of the concerned body.

(2) Scaling

A scaling transformation alters the size of an object. This operation can be carried out for polygons by multiplying the co-ordinate values (x, y) of each vertex by scaling factors S_x and S_y to produce the transformed coordinates (x', y')

$$x' = x \,.\, S_x, \qquad y' = y.\, S_y \qquad (2.4)$$

Scaling factor S_x, scales object in the x direction, while scaling factor S_y, scale object in the y direction. The transformation equation (2.4) can also be written in the matrix form:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} S_x & 0 \\ 0 & S_y \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \qquad (2.5)$$

or

$$P' = S \cdot P \qquad (2.6)$$

Where S is 2 by 2 scaling matrix in equation (2.5).

Any positive numeric values can be assigned to the scaling factors S_x and S_y. Value less then one reduce the size of the objects, value greater then one produce an enlargement, Specifying a value of 1 for both S_x and S_y leaves the size of objects unchanged, when S_x and S_y assign the same value, a uniform scaling is produce that maintains relative object proportions. Unequal value for S_x and S_y result in a different scaling that is of ten used in design applications, where picture are constructed from a few basic shapes that can be adjusted by scaling and positioning transformations.

Object transformed with equation (2.5) are both scaled and repositioned. Scaling factor has value less then one, move object closer to the origin coordinate, and while value greater then one move coordinates position farther from the origin. The location of a scaled object by choosing a position, called the fixed point, which is to remain unchanged after the scaling transformation, can be controlled. Coordinate for the fixed point (x_f, y_f) can be chosen as one of the vertices, the centroid, or any other position. Scaling the distance from each vertex to the fixed point then scales a polygon relative to the fixed point. For a vertex with co-ordinate (x,y) the scaled co-ordinate(x', y') are calculated as

$$x' = x_f + (x - x_f) S_x , \quad y' = y_f + (y - y_f) S_y \qquad (2.7)$$

Thus it can be written that the scaling transformations to separate the multiplicative and additive terms are as such:

$$\left.\begin{aligned} x' &= x \cdot S_x + x_f(1 - S_x) \\ y' &= y \cdot S_y + y_f(1 - S_y) \end{aligned}\right\} \quad (2.8)$$

Where the additive terms $x_f(1 - S_x)$ and $y_f(1 - S_y)$ are constant for all points in the object.

Applying transformation equation (2.8) to each vertex and then regenerating the polygon using the transformed vertices scale polygons. Applying the scaling transformation equation to parameters defining the objects scales other objects.

(3) Rotation

A two dimensional rotation is applied to an object by repositioning it along a circular path in the x-y plane. To generate a rotation, specific rotation angle θ and the position (x_r, y_r) of the rotation point (or pivot point) about which the object is to be rotated is given. Positive values for the rotation angle define counter clock wise rotation about the pivot point, and the -ve value rotate the objects in the clockwise direction. This transformation can also be described as a rotation about the rotation axis that is perpendicular to the x-y plane and passes through the pivot point.

There is need to determine the transformation equations for rotation of a point position P, when the pivot point at the coordinate origin. If r is the constant distance of the point from the origin, angle Φ is the original angular position of the point from the horizontal, and angle θ is the rotational angle. Using standard trigonometric identities, it can express the transformed coordinate in term s of the angle θ and Φ as-

$$\left.\begin{aligned} x' &= r\cos(\Phi + \theta) = r\cos\Phi\cos\theta - r\sin\Phi\sin\theta \\ y' &= r\sin(\Phi + \theta) = r\cos\Phi\sin\theta + r\sin\Phi\cos\theta \end{aligned}\right\} \quad (2.9)$$

The original coordinates of the point in the polar coordinates are

$$x = r\cos\Phi \ , \ y = r\sin\Phi \quad (2.10)$$

Substituting equation 2.10 and 2.9, it obtain the transformation equations for rotating a point at position (x, y) through an angle θ about he origin as such-

$$\left.\begin{aligned} x' &= x \cos\theta - y \sin\theta \\ y' &= x \sin\theta + y \cos\theta \end{aligned}\right\} \qquad (2.11)$$

The equation can be written in matrix form:

$$P' = R \cdot P \qquad (2.12)$$

Where the rotation matrix is

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \qquad (2.13)$$

As with translations, rotation are rigid body transformation that move objects without deformation every point on an object is rotated through the same angle. Displacing each vertex through the specified rotation angle and regenerating the polygon using the new vertices rotate polygons. Curved lines are rotated by repositioning the defining points and redraw the curves.

2.1.2 Rubber Sheeting

Rubber sheeting or conflation is a non-uniform adjustment of a data set based on the movement of known control points to new locations. For example, data collected by aerial survey may be inaccurate because of flight alignment and camera inaccuracies. By comparing this data with accurate ground survey data, the aerial data can be stretched or rubber sheeted over the accurate data using control points and monuments common to both data sets. Rubber sheeting is a process in which the shape of the object is deformed and it is very difficult to handle at this level.

Use rubber sheeting only:

(1) When absolutely necessary because it can severely compromise the relative accuracy of the data;

(2) As a last resort after exhausting other methods of object editing and coordinate adjustment;

(3) When attempting to get two or more different data sets from different sources to align geographically: for example, when stretching a new subdivision map into a pre-existing parcel map.

2.2 Geometric Featured Based Edge Matching

In the process of spatial data, Edge matching is one of the most useful procedure/ techniques. In which a lot of data can be organized in a proper form to meet different needs. This technique is often applied for building topological data, matching photogram metric models, and combining several large scale maps into a small scale map, Several other methods to implement edge matching procedure for different situations such as Node-matching method for building topological matching relation; Zipping method for line matching between adjacent map sheets and matching photogram metric models etc. An effective matching procedure should correct most of the errors included in the data, In generals matching results are good if the original data includes no error data. But if error data are included in the original data, this procedure will give unsatisfying results as new errors may be added or involved. In that condition, it will take a lot of interactive work to correct these data sets (both the new and the old). Geometric feature-based [3] method is especially good at detecting errors.

2.2.1 Edge Matching

Edge matching is the process to determine which edges/lines should be linked among other edge. For some cases, one edge will join with only other one and for some other cases more than two edges will be linked together. This depends on the features represented by data. The process of contour data matching was the previous one case and process of river, boundary, maps, and buildings matching are later one.

In organization of data, Edge-matching is a necessary process. It recovers the original states of the data. In capturing data, the same coordinate points may be digitized differently. One long line or curve may be divided into several pieces.

Besides, some errors such as redundant lines, missing lines may inevitably appear in the course of data capturing.

The mentioned data set must be improved in this matching procedure, so that available data are less prone to error. But, edge matching is an uneasy process that has been addressed by several authors. Although it is very simple for a human operator to do it because of human being's innate ability of recognition, it is particularly difficult to implement this procedure by a computer [4]. The difficulty is that there are no satisfying matching rules to be obeyed when the original data include some error data. From the point of applications, however, a computer should mostly do matching work. That is to say, it should choose a proper method to work harder [3].

2.2.2 Used Methods

The usual methods are generally grouped as-

(a) Distance-based Method
(b) Featured -based Method
(c) Zipped Method
(d) Geometric Feature-based Method

(a) Distance Based Method

There are no theoretical data to be referred in the matching procedure. There is need to find some matching rules (criterion) within the original data for determining matching lines. In common sense, distance between lines is the most possible criteria to be recommended. In most cases, the distance between two edges is simply defined as the minimum of distances between end points of different edges as Figure 2.1 (a. & b.) As shown in the Figure.2.1 (a), edges l1 and l2 have correspondingly end points N_{11}, N_{12}, and N_{21}, N_{22}. The distance D takes the following function:

$$D(l1,l2) = \min\{d(N_{11}, N_{12}), d(N_{11}, N_{22}), d(N_{12}, N_{21}), d(N_{12}, N_{22})\} \qquad (3.14)$$

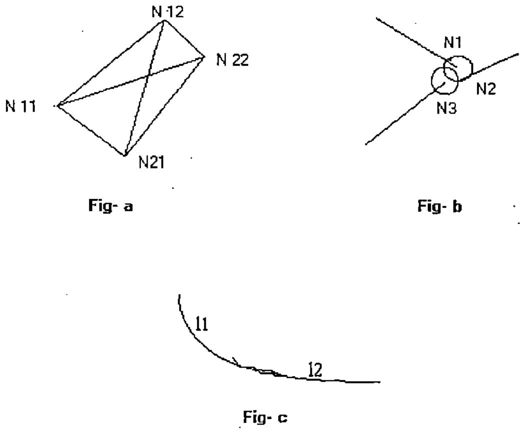

Figure: 2.1 Distance Point Matching (Fig-a: Minimum distance, Fig-b: Threshold consideration, Fig-c: Overlapping)

This method is also order-related. Figure 2.1(b) shows that case. For a given threshold e, if N_1 is regarded as a referred point, N_1 will only join with N_2 because of

$$d(N_1,N_2) <= e$$

and

$$d(N_1,N_3) > e.$$

But if N_2 is regarded as a referred point, (N_1, N_2, N_3) will take the same point that may be the original case. Besides, the above-defined distance may immediately show its disadvantages.

When these edges to be processed have overlapped parts illustrated in Figure 2.1 (c), Lines l_1 and l_2 have a rather long overlapped part. It is obvious that l_1 and l_2 are composed of one original line. Above method can only give right result when a threshold is assigned a rather large value which may cause a lot of confusing results can also be there, if the care is not taken any data sets are considered.

The method can be improved by redefining the distance between edges as following:

$$D\,(11,12) = \min\,\{d\,(p_1<i>,\, p_2<j>) \mid p_1<i> \in L_1,\, P_2<j> \in L_2\} \qquad (2.15)$$

Nevertheless, this method is still sensitive to a threshold, which will be determined by no means of ease.

Another disadvantage of this method is that it is blind to error data included in the original data. These error data may interfere with the procedure to produce bad results.

(b) Featured Based Method

In a lot of cases, the features/ attributes represented by these data sets are more important than their geometry. A road should be linked with a road; a river should be linked with a river. The matching criteria in this method are that, the lines of the same attribute are likely to be linked, no matter how far the distance might be. In fact, this method is often mixed with the above distance-based method [3].

(c) Zipping Method

This method is another useful method in matching adjacent map sheets as well in matching adjacent photogrammetric models. It works like a zipper. The procedure begins with a standard line. Two sides must have the same number of lines to be matched. The line on one side will be joined with the line on other side. But this procedure will get confused if the numbers of lines on one side are different from that on other side. Any procedure will be misled if it processes a set of data sets with wrong attributes/ features. However, in most actual situations, the original data do not include attribute errors but do include geometric errors, which is a reasonable pre-requisite [3].

(d) Geometric Feature Based Method

Any set of data to be processed are likely to include inevitable some error data sets or unexpected data sets. This procedure is useful to detect existing errors. In this matching procedure no theoretical data sets are referred. However, every geographical element (such as river, road, boundary etc.) has its innate geometric feature that can be demonstrated by the coordinates of the elements. So the geometric features trends to be applied to this procedure in which relation between elements may also be included in the features. There may be a lot of geometric features for a type of element, some of which may not easily be implemented into a computer, and some of which may not be available the matching procedure. Choice of features for matching procedure depends on the type of element, namely on the data sets to be processed. This matching procedure determines linking edges by comparing the innate geometric features with ones shown by data sets. If the features shown by the data do not conflict with innate ones, these edges will be linked; otherwise they are not linking edges. Some of which are regarded as errors if they remain to conflict with innate features in matching [3].

2.3 Curve Matching

Curve matching algorithms have potential application in finding correspondence between maps and terrain image [5]. This approach has been motivated by an algorithms due to [6] solved the curve matching problem under the restrictive assumption that one curve to be matched is a proper sub curve of the other, namely as given two curves, such that one is proper sub curve of the other, find the translation and rotation of the sub curve giving its best least square fit to the longer curve.

However, this least square algorithm is applied in order to the problem of object recognition, for it one has to divide the boundary of the visible scene into sub-curves belonging to different objects. There are so many algorithms, which were developed for the two dimensional object recognition and location problems. In case of curve matching [7] had described two-dimensional shapes by set of critical points (such as discontinuity in the curvature) but this method is not applicable for curve because curves do not possess such critical points. This method [5] is more general and does not require a polygon approximation of the curves.

The problem that is tackled by [8] is more difficult, since they allow translation, rotation and scale change of the observed objects. This algorithm does not solve for the scale change. This algorithm is computationally more efficient than [6] introduced an object recognition technique, which is particularly efficient when a large database of models is involved. The deficiency of this technique is the need to use so called breakpoints.

Although it discusses the application of curve matching in 2-D object recognition, but the object is not to present than object recognition algorithms because objects recognition is one of its application. In The development of curve matching algorithms, authors were especially interested in general applicability and computational efficiency. The algorithm which are presented in this correspondence does not depend on special features of the curve such as the critical point of [8] or the break point of [9] the matching is based on information which is obtained from all sample points on the curve rather than on information obtained from some special points which may or may not exist, when a big number of sample points are included in the matching process, then the computational efficiency becomes a major concern.

"On Curve Matching", [5] presented two algorithms both are based on conversion of curves into numerical string. For the conversion of curves into numerical string, this algorithm exploits the most efficient string matching techniques presented by [9] and [10] as an exact integer substring comparison.

The above study work gives the detail description of the existing methodology but all these are used on the images of the maps, and some how these image maps are need the storage to perform the operation. After study and Analysis there is a

Chapter 3

PROBLEM DESCRIPTION AND ANALYSIS

Land being the general habitat of human being, its effective use forms a crucial aspect for the economic, social, and environmental advancement of all countries. Although it forms part of nature's bountiful resources, access to land is controlled by ownership patterns. In general Land is partitioned for administrative and economic purposes, and is used and transformed in myriad ways. The history of land records is as old as the Indian civilization. Maintenance of these records has undergone through a process of evolution as it passed through various administrative systems and socio-economic compulsions. In India land records are maintained by district administration for deciding ownership and boundaries of land or property. The process of defining and determining land in favour of an owner is called registry of land. Municipal Corporations, Development Authorities and other Govt. bodies also use the ownership information for various aspects such as revenue collection; yielding forecasting, land development etc.

Increasingly it has been recognized that policy makers, planners, land administrators, and individual citizens all have a need for information about the land and make significant use of spatial data on a day-to-day basis. Adequate knowledge of natural resources and accurate description and record of such knowledge forms the essential pre-requisite to their rational use and conservation. Land information is prime requisite for making decisions related to land investment, development and management.

Land, which is a scarce natural resource, has been regarded as a measure of wealth, status and power, from time immemorial. Any developmental activity is nearly impossible to conceive without taking land into consideration. And in this context since time immemorial, it has been a constant endeavor of human beings to pursue various aspects of such activities with ease. Therefore, though a printed map is useful, computerized systems offer improved ways of acquiring, storing, processing and retrieving such information. This endeavor of simplification is well achieved by an effective tool of Geographic Information System, when applied to, as complicated a process as land acquisition & development. Unlike the conventional procedure of handling various maps of different scale simultaneously, Geographical Information System helps to prepare and peruse the maps of multiple types at the same time. The concept of superimposition & overlaying is being used to store both non-spatial and spatial data in different thematic layers. This enhances data accessibility and also helps in analytical assessment of the information to be used prior to any land development process.

3.1 Problem definition

There was a time when any surveying department and other concerned authority used to keep hand made surveyed maps which are prepared by taking lots of reading of the similar area and keep them with very careful ways. Now every thing is going to be digitized, so there is requirement to get standardized maps in soft copies with minimal error.

Definition: It's a Process of edge composition on the basis of similar component edge matching, though improper alignment of the maps at the common boundaries exists.

This can be possible to get less error prone maps all over the concerned authority, with a minimum of effort, time & funds, which would have otherwise been required if maps were lost or miss placed due to any accident. Hence they can now be utilized for other purposes. It also enables the interested person to edit maps if some new plans are taking into action.

3.2 Problem Description

This problem deals with the methodology in map's boundary alignment. There is requirement to get the information about the land and analyze it effectively with minimal error, because as the map from different authorities of same or nearby piece of land is procured and aligned, there is always some overlapped area in the final map, usually called no-man's land (boundaries are not matched properly, away from each other) or some gaps at lands common boundaries (more possessions land), which is nothing but an error in the digitize maps. The propagation of these errors depends on so many factors discussed in detail and analyze at the problem analysis level latterly.

3.2.1 Estimating Error

It is noted that the map alignment accuracy is a relatively minor issue because users of maps are rarely aware of this problem, due to their familiarity with map notations and its underlying assumptions. For example, if a map shows that a drainage ditch runs parallel to a road, one assumes from world knowledge that the ditch is located close to the road. However, on the map, for purposes of better understanding, the map may appear to be offset laterally from the road by as much as the width of the road.

3.3 Problem Analysis

The majorities of map datasets are currently represented in vector format, and have an inherent representational error that arises from digitization or polygonalization processes. This is further discussed in this dissertation. Map alignment error minimization algorithms propagate this dataset error through various stages of computation, yielding a map product or results that often has unexpected errors. Such errors influence co-registration and tend to corrupt the derivation of range or elevation data. This further impacts the integration of map models (e.g., line models based on vector data) with map datasets.

All data have limited accuracy, which may be expressed in terms of positional error, abstraction or generalization error, measurement error, etc. Typical errors are presented next.

(1) Surveying error

Surveying is the first step (Except for early planning) in all about the smallest engineering or architectural projects and is often a least step before the owner accepts the finished construction. It is stated that no surveying measurement is exact. The surveyor continuously deals with errors. If the work is to be well done then it must understood thoroughly the nature of sources and behavior of the errors that affected the results. So, the errors are inaccuracies in measurements caused by the type of equipment used or by the way in which the equipment is used. Errors in surveying measurements are some times classified as to their source, i.e., instrumental, personal, or manual [11].

(i) Instrumental Error

Instrumental Errors are those due to imperfections in the instruments because of deficiencies in their manufacture or in account of improper relationships between the different parts. The incorrect length of the steel tape or the faulty adjustments of the plate bubble tubes of a transit are examples. It should be understood that the instruments are never perfect; therefore, proper corrections and field methods are applied to bring the measurement with in the desired limit of precision. It may be added that some of the principal advances in the art of surveying within recent years have been affected by improvements in the design and manufacture of the instrument.

(ii) Personal error

It arises from the fallibility of the senses of sight and touch on the part of the observer. He must frequently estimate fractional parts of scale divisions, and he must manipulate the instrument with dexterity. These operations are ever done exactly and the magnitude of the resultant errors depends largely on the co-ordination on the sense of sight and touch, and on the skill of the observer. Reading the division on a graduated circle or fixing the line of sight of a transit on a given object are examples.

(iii) Natural Error

Natural errors have their sources in the phenomena of nature, such as changes in the temperature, differential refraction of the atmosphere, wind and curvature of the earth. Such source of error are quite beyond the control of the observer, but he can take proper precautions and adopt his methods to field condition so as to keep the resulting errors within proper limits. In surveying measurements two kinds of errors are present, namely systematic and accidental errors. These can be defining in terms of their behavior as to sign and magnitude both when the field conditions are constant and when the field conditions are changing.

(iv) Systematic Errors

Systematic error are those which (i) for constant field conditions are constant in both in sign and magnitude; and which (ii) for changing field conditions are usually constant in sign but variable in magnitude, in other words errors that occur in same direction there by tending to accumulate so that the total error increases proportionally as the number of measurements increases. Systematic Error are also caused by imperfection in the manufacturing of the equipment – not mistakes in the manufacturing process, but an inability to achieve absolute perfection, Equipment may also cause systematic errors because it is damaged or out of adjustment.

(v) Accidental Error/ Random Error

Random error are that occur randomly in either direction there by tending to cancel one another or can say are those error whose signs are just as likely to be plus as minus under both constant and changing field conditions. Although the total errors dose increase as the number of measurements increase, the total error becomes proportionally less and the accuracy becomes greater as the number of measure increase.

Errors in locating a point

The locations are not mathematically exact, either on paper or in the field. Surveyor should appreciate the relative accuracy of the combination of distance and angle that can be used [12].

Method 1: - A point can be located in relation to one another point by a direction and distance.

Method 2: - A point can be located in relation to two other points by a direction from each of the two points or by distance from each of the two points.

Method 3: - A point can be located in relation to two other points by a direction from one point and a direction from the other point.

If the systematic and random errors are eliminated and no mistakes are made, accuracy will increase as the length of the survey increase. It is always not possible to identify each error as systematic or accidental although most fall under one type or the other.

(2) Decision Error

Vague geographic relations such as west-of are used in practices. In conventional map analysis, precision is usually adapted to accuracy. It is necessary to distinguish between the meanings of the terms precision and accuracy as they are used in describing physical measurements and the subsequent computations [11].

Precision

It refers to the care and refinement with which any physical measurement is capabilities made. It relates to the expertness of manipulation on the part of the observer or to the of the instrument used.

Accuracy

It refers to the difference between the final measured value of a quantity and its absolute, or true, value.

- Precision limit of paper is approximately 0.5mm or one line width, and map error increases with humidity and use, due to paper war page.

- Area accuracy is thus approximately m^2.
- Map processes (planimetry, dot counting, transparency overlay) have approximately the same error (one line width).

The ability to change scale and combine data at different scales implies that map precision is not necessarily adapted to accuracy.

Example: Maps storing systems generally do not warn users if datasets of different scale (e.g., 1:24,000 vs. 1:1,000,000) are combined and the result is displayed at 1:50,000 scales. This causes loss of information in changeover from the high- (low-) resolution dataset.

Observation:

Most vector-based systems perform vector operations (line intersection, overlay, area computation) at full computational precision, without regard for dataset accuracy. This way it is found there is decision-making error when plotting any surveyed record.

(3) Ground Truth

Users are often surprised by errors when maps are validated against ground truth. The goal of Map alignment with minimum error analysis is a measure of uncertainty associated with every map product (e.g., datasets, maps, and analyses).

Ideal: A set of confidence limits with each map.

Problem

Cartographic map and digitization errors, other errors listed previously can produce a diversity of error modes and magnitudes in maps datasets. So there is propagation of error when map is going to be convert in to vectorize form by scanner software or by any geographical information system software but they needed extensive editing. This conversion also depends upon scan resolution and type of instrument used such as drum or flatbed electromechanical scanner. Digitized maps also formed with the help of sensors. The sensors collect data directly in digital form (e.g. digital cameras) but in that case one third photos of any piece of land, use to be repeated to get the right information. Still, scanning play major roll because most of the maps are paper maps. As it is found that the presentation of right information is more accurate when using vector data but automatic creation of vector data via scanning is more problematic, such as:

- docs must be clean
- lines at least .1mm
- complex line work adds error
- lines shouldn't be broken with text.
- text may be interpreted as lines
- automatic feature detection (road versus railroad) difficult

Some more common errors are such as:

- Dangling arc (node missing at one end), dangling node, node arc intersection, overshoot, undershoot, open polygon, sliver, gap. So at every step extra care is required.

Reality

Error measures are likely on points and lines, maybe some for area, since supporting theory can be derived rigorously from map coordinate information. Hence, the key issues in this section are measurement, estimation, and prediction of maps error. The following observations pertain:

- Maps attributes are often non-numeric and cannot (in some cases) be indexed in a physically faithful way by a subset of the real numbers.

The root problem in this case can be stated as follows:

- All physical science is defined in terms of mathematics, which is based on measurement theory and the formal logic.

In contrast, language and humanistic pursuits (from which the process of classification arises) have as yet no rigorous supporting or descriptive mathematics.

Thus, attributes or labels, which are linguistic entities, cannot in many Instances be rigorously assigned numerical values or ranked in a physically significant way.

(4) Total Error

It is the sum of the in accuracies in a completed job. Since accuracies are either positive or negative, it is an algebraic sum. Surveying accuracy is determined by completing a circuit to the beginning (or to another known point). The error of closure is the difference between the actual positions of the finishing point (either vertically or horizontally) and the location of the same point determined mathematically as result of the circuit. At the latter stage when the maps are captured in the papers or cloths form for digitization, unavoidable errors used to be present due to tension in the paper or cloth or the error at the data-capturing machine (scanner types) and then to get the sets of vector data from these scanned maps.

To control errors so that total error not excessive, source of error must be understood and methods and equipments must be chosen which will reduce the total errors to allowable levels without wasting time. This way it is find that the size of total error depends on the precision of the equipment and the way in which the equipment is used.

Chapter 4

DESIGN AND IMPLEMENTATION

4.1 Numerical Process Life Cycle

Numerical computing involves formulation of mathematical models of physical problems that can be solved by using basic arithmetic operation [13]. The process of numerical computing can be roughly divided into the following four phases that are shown in Figure 4.1:

1. Formulation of the mathematical model
2. Construction of an appropriate numerical method
3. Implementation of the method to obtain a solution
4. Validation of the solution

The formulation of a suitable mathematical model is critical to the solution of the problem. A mathematical model can be broadly defined as a formulation of certain mathematical equation that expresses the essential features of a physical system or process.

The formulation of a mathematical model begins with the statement of the problem and the associated factors to be considered. Real life problem have many uncertainties and unknowns. It might, therefore, be necessary to make certain assumptions for approximating and to include only those features of the problem that are consider critical to the final solution. An over simplified model have only limited usefulness. The model may be enhanced later, if necessary. The model refinement may make the solution procedure more difficult.

Once a mathematical model is available then the first step would be to try to obtain an explicit analytical solution. In most cases the mathematical model may not be amenable to analytical solution or they may not be solved efficiently using analytical techniques. As mentioned earlier, a numerical method is a computational technique, which involves only a finite number of basic arithmetic operations.

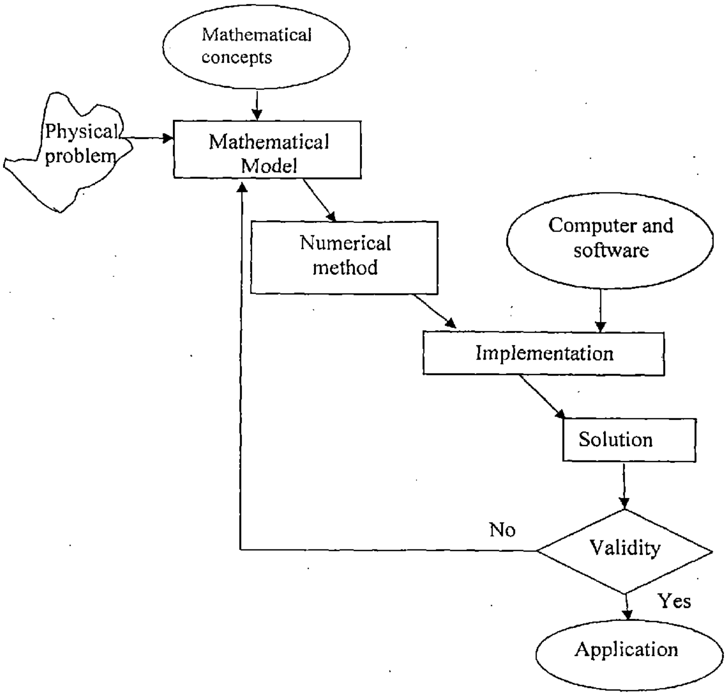

Figure: 4.1 Numerical Computing Processes

For given problem, there may be several alternative arithmetic numerical methods [13]. So there is need to consider various factors or trade-off before selecting a particular method. These are such as type of equation, type of computer available, accuracy, speed of execution and programming and maintenance effort required. Modeling is a process of translating a physical problem into a mathematical problem. The process involves-

1. Making a number of simplifying assumptions.
2. Identification of important variables.
3. Postulation of relationship between the variables.

The third phase is the numerical computing process in the implementation of the method selected. This phase is concerned with the following three tasks-

1. Design of an algorithm.
2. Writing of a program.
3. Execution it on a computer to obtain the result.

Once, if able to get the results, the next step is the validation of the process. Validation means the verification of the results to see that it is within the desired limits of accuracy. If it is not , then we must go back and check each of the following:

1. Mathematical model itself.
2. Numerical method selected.
3. Computational algorithm used to implement the method.

This may mean modification of the model, selection of an alternate numerical method or improving the algorithm (or a combination of them). Once, a modification is introduced, the cycle begins again. Figure: 4.2 show how the numerical cycle moves from the real world to mathematical world and back.

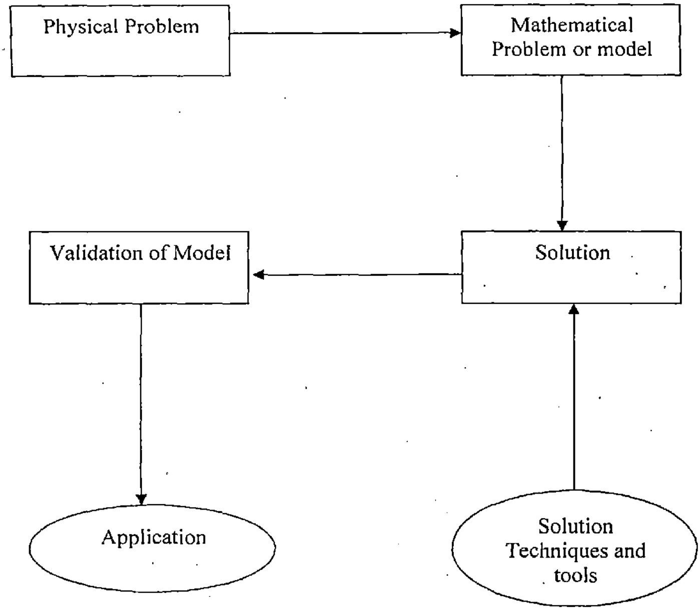

Figure: 4.2 Different View of Numerical Computing Process

4.2 Design Structure

The design procedure required certain pre-assumption which are discuss below and shown in the Figures.

(1) Data Structure used for maps data operations are Link List and Arrays.

(2) Consider the maps as sets of polygon for performing the operations, as shown in Figure: 4.3 below-

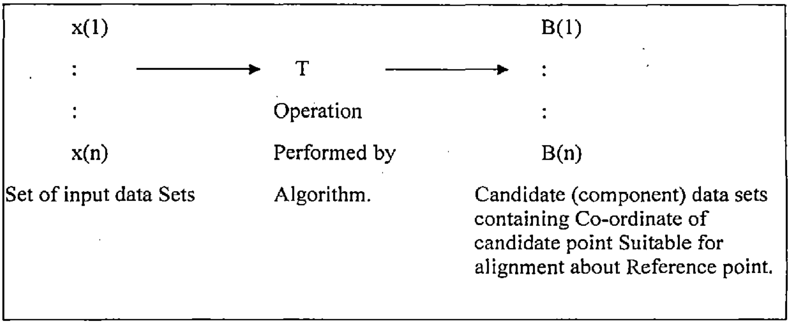

Figure: 4.3 Operational Structures

(3) Now there is requirement to know how these data sets would be store in memory, these would be store in array as the pre define structure as shown in the Figure: 4.4 as below-

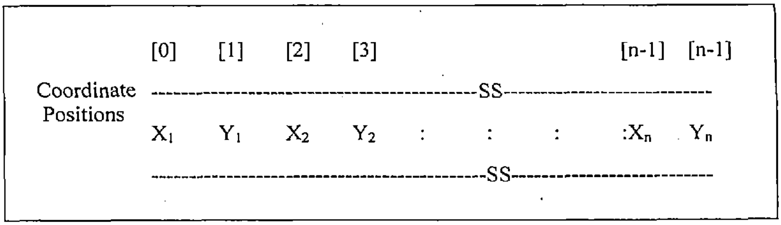

Figure: 4.4 Data Set Organized in Memory

(4) Now the next is fetching the points from memory and it would be by traversing of the points from reference data sets to candidate data sets in the anticlockwise direction. This shows the algorithm approach on reference and candidate data sets. This is shown in the Figure. 4.5 on the next page-

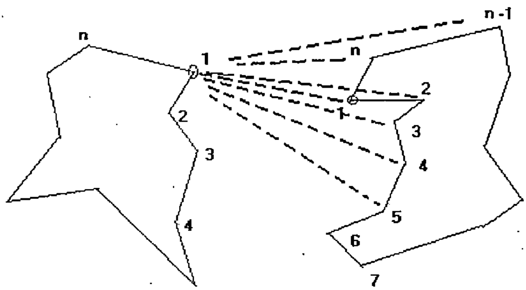

Figure: 4.5 Algorithm Approach on Reference and Component Data Set.

4.2.1 Basic Algorithm Structure

Variables used - i& j for counting, D is for calculated distance, Poly is for map Polygon, sqr for square of variable value, N is number of sets of candidate polygon map. It is considered that candidate polygon holds the common component of reference polygon. The generalized steps are given below:

Step No. 1 ⟶ Get polygon data from file for a number of polygons (maps).

Save the data as:

Poly 1[i][j]

Poly 2[i][j]

Poly 3[i][j]

:

Poly n [i][j]

Step No.2 ⟶ Normalize all polygons on (0, 1) scale.

Step No.3 ⟶ Calculate inter-point distance for each polygon and save it, as following:

Distance_Poly 1[i][j] ← D_1 ($D_1 = \sqrt{}(sqr(X_k-X_i) + (sqr(Y_k-Y_i))$)

: :

: :

Distance_Poly n [i][j] ← D_n ($D_n = \sqrt{}(sqr(X_i-X_n) + (sqr(Y_i-Y_n))$)

Step No.4 ⟶ Choose any reference polygon about which other polygons can be moved.

Step No.5 ⟶ for each distance in

$D_i = \{D_1\ D_2\ D_3 \ldots\ldots\ldots D_n\}$, where i = 1 to n, Compare

D_i with $N_j = \{D_{1 \text{ to } n}$ of $-N_1\ N_2\ N_3 \ldots\ldots\ldots\ldots\ldots N_n\}$,

Where j = 1 to n.

Step No. 6 ⟶ if no points matches, Print ⟶ No matching possible.

Else say matching found.

Step No. 7 ⟶ Choose next polygon to move about new combined polygon.

Step No. 8 ⟶ Stop.

4.3 Implementation

Chapter two discusses the Geometric Transformation, The Transformation such as Translation, Rotation, Scaling and Reflection have been formulated and implemented for limited number of inputs, now the attempts are going on to implement for "N" number of data sets as input. These data sets would be the vector data set and data structure to store these vector data set are arrays and chained memory link list, which read data sets from flat files.

4.3.1 Algorithm for Vector Data Map Alignment

Step No.1 ⟶ Read reference polygon map data $X_1(n)$ from the file.

Read candidate polygon map data $X_2(n)$ from the file.

Step No.2 ⟶ Transfer polygon data from link list to actual buffer for both reference and candidate polygon.

Step No.3 ⟶ Find all the distance D_i $\forall$ i = 1….no of points of reference and candidate polygon.

Step No.4 ⟶ Initialize and find the optimal candidate side that can be placed with reference polygon. Set the index at which at least 3 point from both candidate and reference polygon match.

Step No.5 ⟶ Extract the candidate points (X_1, Y_1) and extract reference point (X_2, Y_2).

Step No. 6 ⟶ Compute translation factor for the problem domain. . i.e. for both the candidate and reference polygon.

$$\text{Translation factor } T_x = X_2 - X_1$$

$$\text{Translation factor } T_y = Y_2 - Y_1$$

Step No. 7 ——→ Perform rotation and translation about the reference point for candidate point.

Step No. 8 ——→ Draw the final polygon.

Step No. 9 ——→ Stop.

The data flow diagram is shown on the next page.

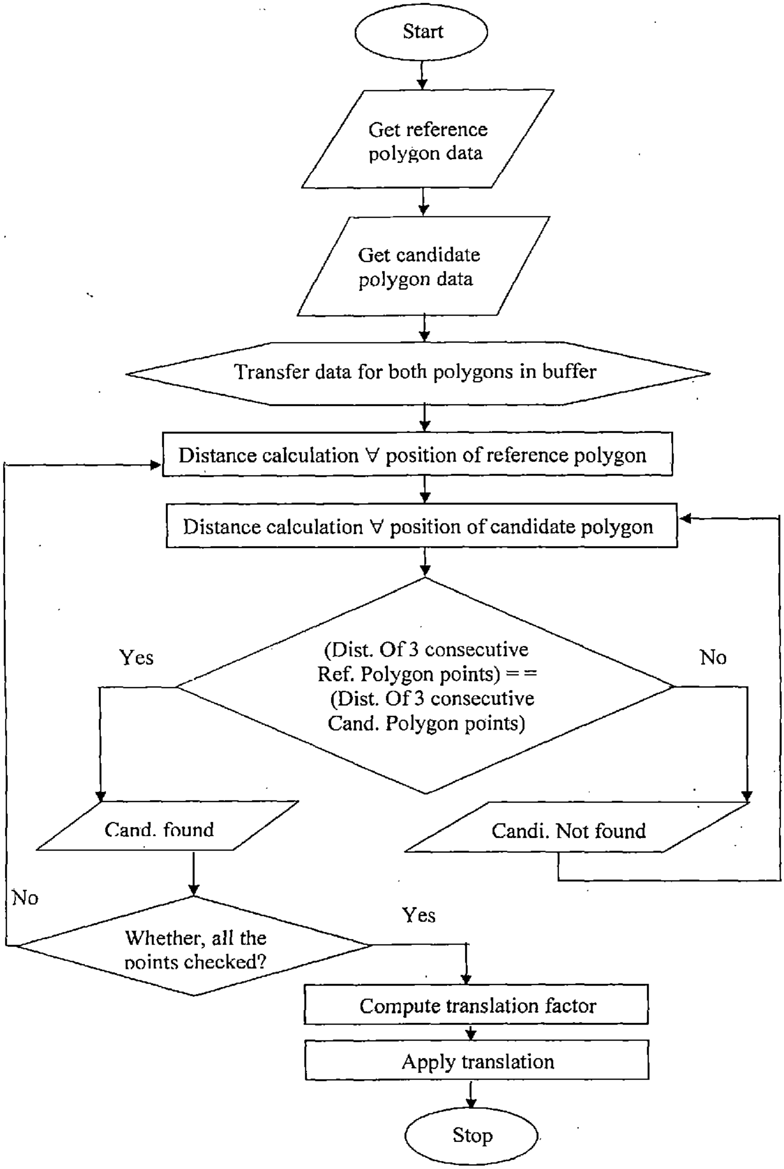

Figure: 4.6 Flow Chart for Vector data Map alignment

RESULTS AND DISCUSSION

Geometric feature based method is reliable because geometric features of element is constant. The method can only give right result when pre-assumptions are well satisfied. The First step is to read the input files in proper way. A case has been shown in Figure. 5.1 below- in which data from file 1 used as data.1 and data from file 2 as data.2

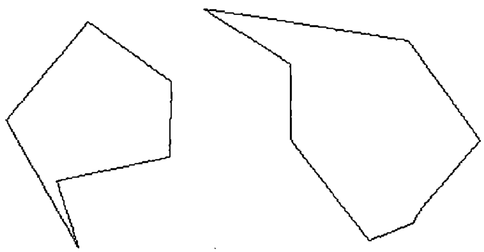

Figure: 5.1 Construction of Polygon from Input Data Files

and construct the polygon map If it is not under the mentioned criteria then the out put of the program may cause a bit of confusing results, so if the care is not taken in any input vector data sets, it will take very long time to finish. The results obtained by implementing the algorithm are experimental rather then by simulation.

On the next page Figure. 5.2 presents one case in which boundaries of two input vectors data sets are shown after implementing the translation. This translation is based on the basis of similar edge recognition at the consecutive points in reference as well as in candidate (component) input vector data sets.

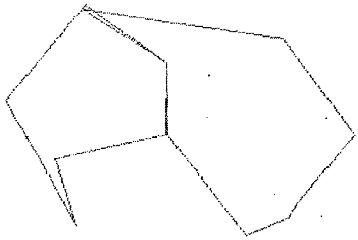

Figure: 5.2 Aligned Polygon Map after Translation

The main advantage of this method implemented in this work is to make the system more robust and minimum man handling. In other words at the input level there is only need to handle the vector data files with care. Then rest of the work of getting right candidate (component) for combining would be perform by the program, as it is not in the current use of any software dealing the map alignment. People of civil engineering use Mat Lab for mosaic two images by selecting the edges of the maps image not the vector data sets.

The procedure implemented shows that how the errors in Map alignment can be reduced using geometric knowledge of the maps, this could be achieved only by finding out the right component of the candidate polygon map. Along with the advantage, there is some disadvantage of this method. Nevertheless, this method is still sensitive, which has determined by no means of ease.

- The procedure implemented here is sensitive to the distance error as shown above in Figure. 5.2, which has some mismatching at the one matched point.

- Another disadvantage of this method is that it can be blind to error included in the original input vector data sets. These input data sets error might interfere with the procedure to produce good results.

The implemented approach also produced the result with error, this could be only because the approach used for the translation factor only and less number of consecutive edges for deciding the matching between the reference and candidate polygon.

Chapter 6

CONCLUSION

6.1 Conclusion

The system has tried to produce the robust result by taking some samples as input datasets (co-ordinate of the maps). The procedure presented in this work is a step in the field of pattern or object recognition for vector data sets to produce the automatic combined map by handling the input datasets. This procedure has gone under some assumptions, which were required at the primary stages. To make the system more efficient for handling large amount of input vector datasets, a proper memory management is required. This effort required to bring the work in more realistic ways or can say the minimum error prone maps for better understanding and minimum human intervention.

6.2 Suggestion for Future Work

Now the next step to implement here would be towards the more error minimization using more accurate input vector data sets, for it an obvious improvement of the present method can be achieved by using proper angle matching and scaling factor. As it is found, that the edge matching, on the basis of well-known features, matching was improper or erroneous, this error is due to un-identification of the error in input vector data sets. So, more precaution is required at digitization level to avoid distortion in maps. This process of error minimization can be carried out to enhance the method of Map alignment

REFERENCES

1. Donld Hearn & M. Pauline Baker "computer graphics", 2nd Edition, Prontice hall of India Private Limited, pp. 183 -215, 2002.

2. James D Folley, A.V.Dam, S.K.Feiner and John F.Hughes, "Computer Graphics- Principal and practices " 2nd Edition, Pearson Education Asia, pp. 201-227, 2002.

3. Xin Chang Zhang "Geometric Feature-based Edge-Matching" City and Resource Planning Department, Zhongshan University, 135 Xingang Xi Road, Guangzhou, 510275.P.R.China. http://www.divcom.otago.ac.nz /SIRC/GeoComp/GeoComp98/99/gc_99.htm

4. Raymond J. Hintz and Mike Z. Zhao "Demonstration of Ideals in Fully Automatic Line Matching of Overlapping Map Data", Auto-Carto 9 Proceedings. pp. 118, 1990.

5. Haim J. Wolfson "On Curve Matching", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.-12, No.-5, pp. 483-489, May 1990.

6. A. Kelvin, E Schonberg, J.T. Schwartz, and M. Sharir, " Two Dimentional Model Based Boundary Matching using Foot Prints," Int. J. Robetics Res.,vol 5, no. 4, pp 38-55,1986.

7. H. Freeman, " Shape Description via the use of Critical Points," pattern Recogn.,vol 10,pp 159 - 188, 1978.

8. N. Ayache and O. D Faugeras, " Hyper: A New Approached for the Recognition and Positioning of Two Dimensional Objects," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI -8, pp.44-54,Jan 1986.

9. P. Weiner," Linear Pattern Matching algorithms," in Proc. 14th Annu. Symp. Switching and Automata theory, IEEE Computer. Soc., pp. 1-11, 1973.

10. M. McCreight, "A Space Economic Suffix Tree Construction Algorithm, " J. ACM, Vol.23, no. 2, pp 262-272, Apr.1976.

11. Kreisle, William E, "History of Engineering Survey", Journal of Surveying Engineering ASCE, Vol. 114, No.3, August 1998.

12. Berkholder, Earl F," Calculation of horizontal / level distance", Journal of Surveying Engineering ASCE, Vol. 117, No.3, August 1991.

13. E. Balagurusamy, "Numerical Methods", TMH Publishing Company Limited, pp 3-10,2002.