Mutual Linear Feature Matching and Alignment Designed for Geometric Accuracy Enhancement of Graphical Cadastral Datasets

David Siriba and Monika Sester

Institute of Cartography and Geoinformatics (IKG), Leibniz Universität Hannover

Appelstraße 9a, 30167 Hannover, Germany

E-mail: (david.siriba, monika.sester)@ikg.uni-hannover.de

Abstract

In some parts of the world, legacy analogue geographic datasets continue to be used. This is because not only do they contain useful information, their replacement, for example, legacy graphical cadastral datasets, would involve length legal procedures and huge financial costs. Legacy and analogue graphical cadastral datasets that are not geo-referenced and contain large and uneven positional distortions provide a technical challenge when integrating them with other geographic datasets. In this paper, a linear feature matching and alignment approach towards the positional accuracy enhancement of a legacy cadastral dataset is presented. The approach entails a preliminary step in which the road network which is implicitly represented in the legacy graphical cadastral dataset is extracted. The extracted road network is then matched and aligned with a more accurate road network that is directly mapped using aerial photogrammetric methods. Matching of the datasets entails a strategy in which each of the road network dataset is used as a reference to identify the corresponding node features in the other dataset. The extracted road network is then preliminarily aligned with a more accurate dataset based on the matched node features by using a non-rigid transformation based on Thin Plate Splines. The preliminary alignment gives very promising results, particularly in the datasets locations that are near the determined corresponding node features. The next step which is not presented in this paper involves computing the transformation between the road network datasets and applying them to the original legacy cadastral dataset. Although the same approach can be adopted for similar legacy datasets, better results could be obtained by incorporating the knowledge about the process steps used to produce the legacy dataset.

Keywords: graphical cadastre, feature matching and alignment, Thin Plate Splines, accuracy enhancement

1. Background

Modern paradigm in land management (Williamson et al., 2010) requires that within a given jurisdiction there exists a seamless multipurpose cadastre that contains all the recognized land objects, such as Private property parcels and land use zones. This means that different cadastres containing different land objects can be integrated seamlessly even if they are maintained by different mapping agencies. A land object, as defined by Kaufmann and Steudler (1998), is a piece of land in which homogeneous conditions exist within its outlines, where these conditions are normally defined by law. Therefore, a piece of land where either a private or a public law imposes identical juridical parameters could be called a legal land object. The laws define the outlines of a right or a restriction within the land object. The legal land objects are normally described by boundaries, which demarcate where a right or a restriction ends and where the next begins as well as the contents of that right. Therefore, it can be concluded that private or public property parcels and land use zones are regarded as land objects.

A cadastre usually includes the geometry of land parcels linked to other textual records describing the nature of the interests, the ownership or control of those interests, and often the value of the parcel and its improvements (FIG, 1983). Cadastre maps are commonly categorized on the basis of their purpose, and this includes: juridical (or legal), fiscal and multipurpose. Another classification is based on the survey method adopted and the associated survey accuracy where the categories include numerical and graphical cadastres (Williamson, 1985). A numerical cadastre refers to a cadastral survey where the bearings and distances of the individual land parcels are determined from computations based on the actual land survey. On the other hand, a graphical cadastre refers to a cadastral survey where the bearings and distances of the individual land parcels are determined by scaling from photo maps or by other graphical methods. Graphical cadastres are typically represented graphically, although it is also common to find numerical cadastres represented graphically by maps, which may or may not contain measured data or dimensions of the parcel boundaries.

Since the introduction of automation in the cadastral domain aiming at enhancement productivity, many more developments have taken place. Chronologically, the developments include the introduction of the multipurpose cadastre, data modeling, and Internet-based cadastres (Williamson, 2010). However, in most developing countries, the developments have not gone beyond automation, let alone the non-existence of automation of the cadastre in some jurisdictions. Despite the existence of digital cadastres in developed countries, their basis is traceable to the legacy and analogue cadastres that preceded them. In jurisdictions that do not yet have digital cadastres, the dependence on legacy analogue cadastres is very common. A cadastre is considered to be legacy because it is created and used in the traditional way simply because it serves the basic purpose. The use of legacy cadastres persists notwithstanding the fact that newer and more efficient methods exist to create more accurate cadastres that could be used for multiple applications. The main reason why legacy cadastres continue to be used is the high costs and lengthy legal procedures associated with the creation of new cadastral datasets.

In jurisdictions where analogue graphical cadastres continue to be used, it will be necessary at one point in time to digitize them in line with the requirements of modern land administration. However, a number of technical limitation are presented by these analogue legacy cadastres that have to be taken into account. For instance:

- the cadastre in a jurisdiction may not necessarily contain all the land parcels. This is because the data contained in a cadastre are acquired over a long period of time,
- Use of both the numerical and the graphical cadastres.
- the cadastres have varying levels of positional accuracy. The difference in the levels of positional accuracy is as a result of the technical limitations of measuring instruments and techniques that were used at that time, .
- the possible use of different coordinate systems for the different parts of the cadastre.

Although it is expected that the numerical cadastres and the graphical cadastres can be combined seamlessly, this is however not possible. This is because graphical cadastres with different levels of geometric accuracy are used. For instance, in Kenya, one of the graphical cadastres was created basically to support land registration, thus was generated/ by tracing the outlines of land parcels from uncontrolled and unrectified aerial photographs. The cadastre, besides being graphically represented at an approximate scale, contains local geometric distortions; has no coordinate grid and therefore lacks geo-reference information; does not contain measurements for the land parcels let alone the absence of coordinates for the parcel boundaries.

To develop a jurisdiction-wide digital multipurpose cadastre by incorporating and combining the legacy graphical cadastres, two options for geometric accuracy enhancement can be considered. These include using either survey-accurate data or existing more accurate topographic datasets. The use of survey-accurate data relates to absolute positional accuracy, while the use of existing topographic data relates to relative accuracy. The difference between absolute and relative positional accuracy is presented by Stanislawski et al. (1996). Because of the high costs associated with capturing survey-accurate data, and the prevalence of recent and more accurate topographic datasets, the option of using existing topographical datasets appears less time consuming, and will yield accurate results and therefore equally reliable. The approach to use existing topographic datasets is conceived within the framework of data integration. The approach entails using existing topographic datasets to enhance the accuracy of the legacy cadastral datasets through data integration. This will then be followed by the edge matching of the individually enhanced cadastral datasets. The data integration approach requires that there are corresponding features in both the cadastral and topographic datasets. One challenge, however, likely to be encountered is the lack of directly corresponding features in both the cadastral and topographic datasets. This is particularly true, for example, with road features, which are the most common features in topographic datasets that are however not explicitly represented in most graphical cadastral datasets and – sometimes because of different times of data capture. After a discussion of approaches and techniques used today to enhance the accuracy of cadastral datasets, a three-step procedure for the positional accuracy enhancement of a legacy graphical dataset is presented.

2. Related Work

Various methods and techniques are developed and used to enhance the positional quality of digital cadastral datasets. These techniques are developed either as a step towards the integration of datasets or as a process in its own right. However, each method depends on the nature and characteristics of the cadastral dataset involved. To enhance the positional accuracy of cadastral datasets, the measurements of the existing land boundaries by modern and more accurate techniques is practically non-realistic solution (Arvanitis and Koukopouluo, 1999). In general, the schemes that could be used for accuracy enhancement of cadastral datasets can be categorized into three classes: numerical to numerical; graphical to numerical; and graphical to graphical.

Numerical to numerical cadastral data accuracy enhancement may be required because of the planar geometric discrepancies between the legacy datasets and modern GNSS positioning solutions (Hope et al., 2008) or because of the need simply to change from an old coordinate system to a new say GPS-based coordinate system (Felus, 2007). The sources of data for such a scheme are survey-accurate data from field measurements of parcel boundaries. Various transformation models are used to adjust the positions of the boundaries. Hope et al. (2008) use a global 4-parameter Helmert transformation, followed by a least squares adjustment of the cadastral vertices linked to the survey-accurate boundary data points that had been measured. In the final step, the position shifts, which are modeled through a Delaunay triangulation of the cadastral data points, are propagated to other points in the dataset. Felus (2007) presents a local rubber sheeting transformation with linear features followed by a boundary adjustment with area constraints.

Graphical to numerical data accuracy enhancement scheme would be required, for instance, when cadastral maps are digitized and non-homogeneities exist between the individual map sheets. The non-homogeneities may also exist between the graphically obtained coordinates and the more precise coordinates obtained through measurement and calculation. Accuracy enhancement can be carried out by processing existing survey-accurate data, or by carrying out field measurements of boundaries. Morgenstern et al. (1989) describe a procedure to remove such discrepancies while at the same time maintaining the positional conditions contained in the graphical structures of such maps. This scheme may also be used when there is need out of policy requirement to implement a legal coordinate based cadastre from a cadastral map that is simply graphical, i.e., no coordinates exist for the parcel boundaries. Steinberg (2000) and Klebanov and Forrai (2010) describe their experience and perspectives of the procedure used in Israel to modernize the national cadastre by implementing a coordinate based cadastre. The main idea behind the procedure is the creation of an optimal set of parcel boundary point coordinates within a cadastral block, and to create new legislation stating that authentic field marks will no more determine the position of cadastral boundaries but only the “legal coordinates” will. Klebanov and Forrai (2010), in particular, describe procedures used in three categories of street blocks defined according to their cadastral background. The categories are: blocks with solid cadastral background, blocks lacking solid cadastral background, and blocks with mixed cadastral background. The specifics of each procedure depend on the background of the cadastre. Considering the varying position accuracy across the original cadastral map, the adjustment of the boundary is realized through a rubber sheet transformation that employs a triangulation to subdivide the area into small units and the transformation is calculated from constraints.

Graphical to graphical data accuracy enhancement scheme would entail adjusting one graphical cadastral dataset with reference to another graphical cadastral dataset of a higher positional accuracy. Usually, alternative graphical cadastral datasets rarely exist unless created specifically for the purpose of improving the accuracy of the existing datasets, in which case they would simply be replacements of the existing datasets. In this case a topographical dataset of a higher accuracy can be considered as the reference graphic dataset. The use of one geographic dataset to enhance the positional accuracy of another geographic dataset is usually carried out in the context of spatial data integration. Other applications that take advantage of the concepts of spatial data integration include the propagation of updates from one dataset to another and fusion of datasets to create new datasets that have more potential for more applications (Mustiere, 2006; Gösseln et al., 2006). Various aspects of spatial data integration are discussed in (Saalfeld, 1988; Devogele et al., 1998; Olteanu et al., 2006; Savopolo and Armenakis, 2002; Sherren et al., 2004; and Yuan and Tao, 1999). In the existing literature, a significant number of spatial data integration problems involve transportation datasets such as roads. There is however, little progress reported with regard to the integration of boundary and cadastres. This is because it is more difficult to integrate these datasets since neither of them commonly has ground signatures in images as transportation features do (Jensen et al., 2004). In the next section, an approach for the graphical to graphical data accuracy enhancement is presented.

3. Graphical to graphical positional accuracy enhancement scheme

In this section, a methodology for enhancing the positional accuracy of a graphical dataset is presented. The methodology is based on the graphical to graphical data accuracy enhancement scheme, and it consists of three main steps: extraction of a road centerline dataset from the graphical cadastral dataset; feature matching; and feature alignment and adjustment. Figure 1 illustrates a flow diagram of these steps.

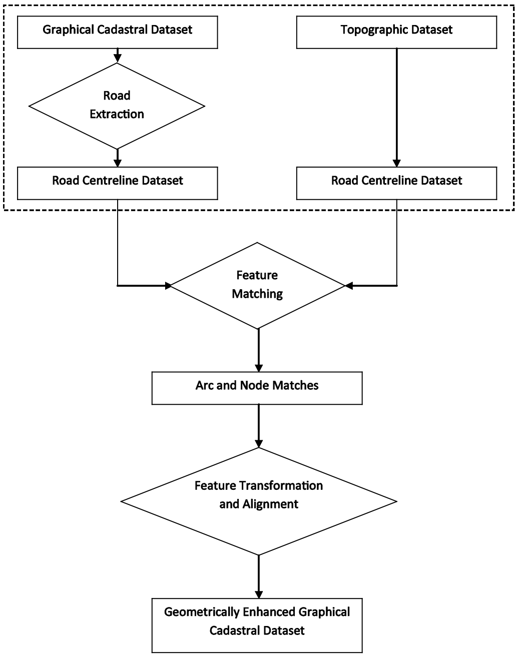

Figure 1: Flow diagram showing the steps in the methodology.

3.1 Road Extraction

Preprocessing is necessary to ensure that the graphical cadastral dataset is approximately geo-referenced; digitized and in the appropriate data structure required in the subsequent steps. An exploration of the sample cadastral maps presented in the cadastral template project, www.cadastraltemplate.org (Steudler et al., 2004), shows that in most countries road corridors are not explicitly represented in the cadastral maps. Instead, only the outlines of land parcels are drawn, leaving out gaps between some of them to indicate the likely location of existing or planned road reserve hereafter referred to as road corridor. This situation can be attributed to the traditional emphasis by land registry and cadastral authorities to register and map only privately owned land at the expense of public owned lands. Moreover, road as public lands are topographically mapped independently by different (government/mapping) agencies.

In case where road corridors are included in the cadastral map as land parcels, i.e., polygon parcels, then part of data pre-processing would involve the extraction of these road centerlines by identifying the road corridors and generating their centerlines. Otherwise, a technique to derive the roads from the cadastral dataset is required. In case roads in the corresponding digital topographical dataset are represented by dual lines, pre-processing would involve collapsing the dual lines to obtain single lines representing the centerline for simplified feature matching.

3.2 Feature Matching

After preprocessing, the extracted rod centerline, hereafter referred to as the source dataset is approximately aligned with the mapped road network hereafter referred to as the target dataset. The road centrelines from the two datasets are then matched to obtain corresponding point and linear features. It is important that an approach used in feature matching considers the relative positional accuracies of the datasets, in particular the inherent local distortions in the legacy graphical cadastral map. Matching is based on the relative positional accuracies of the datasets by considering one of the datasets to have higher positional accuracy.

Feature matching methods that use various strategies are considered in (Kieler, et., al, 2009; Walter and Fristch, 1999). Irrespective of the method used, in this work, the primary objective in feature matching is to obtain high matching certainty as opposed to high matching rate and processing speed (Zhang and Meng, 2007); corresponding linear features; and even distribution of the matches. High matching certainty ensures that only the most likely matches are obtained and minimal visual validation incase it is required. Linear feature correspondence is required as the basis for data geometric accuracy enhancement because many topographic maps are linear based – as opposed to point based - and therefore the majority of counterpart elements are linear (Doytsher et al., 2001). In addition, linear features contain more semantic information unlike points, which are dimensionless. Finally, the even distribution of the matches will ensure that a few feature locations do not have a disproportionately large influence on the transformation, with potentially detrimental results. In this work, matching of the nodes is obtained by adopting the procedure in Kieler et al. (2009). The procedure is divided into two main steps: preliminary matching of both node and arc features, and final matching of node features. The approach is specifically used it yields high matching certainty. The approach requires a threshold distance between the nodes when carrying out feature matching. This is empirically determined based on the relative positional discrepancies between the source and target datasets.

Feature matching involves matching node features extracted from both the source and target datasets. The matching process used here is accomplished in two steps: preliminary matching of corresponding nodes and arcs features in the two datasets followed by a final matching of node features. The preliminary matching of nodes is based on a threshold distance, which is derived from the relative positional differences of the datasets. The relative positional accuracy of the target road dataset is obtained from the mapping specifications, and that of the graphical source dataset is determined based on the knowledge of the lineage of the dataset. In this work, the positional error modeling described in Siriba (2009) is used as the basis for determining the positional accuracy of the graphic cadastral dataset. These steps are elaborated below.

i. For each node in the source dataset, nodes in the target dataset that lie within the specified threshold distance are identified. This process is carried out for the arcs as well, i.e., for each arc in the source dataset, the arcs in the target dataset that are within the threshold distance are identified. During the preliminary matching of nodes, even the most unlikely corresponding nodes are identified. The erroneous node correspondences are filtered-out during the second step.

ii. The final node matches are identified by considering the corresponding incident arcs as well. For each preliminary corresponding node pair, the incident arcs for each of the nodes are identified as well for the corresponding node in the target dataset. If the node matches have incident arcs that were also matched during the preliminary matching, the node pair is considered to be the most likely match. The node match is further ascertained if the angles the corresponding incident arcs create with the respective node are within a specified tolerance.

Although only 1:1 matches are expected for the nodes, 1:m, m:1 and n:m matches are also possible during the preliminary matching. A strategy that exploits a two-way matching is considered in order to receive matches that may be erroneously omitted. The strategy implemented here involves identifying for each node in the source dataset, the most likely nodes in the target dataset (as explained in earlier). Then, the process is carried out again, only this time from the target dataset to the source one, i.e., vice-versa. This is carried out to recover erroneously omitted matches and to ensure higher statistical certainty.

3.3 Feature Transformation and Alignment

Preliminary feature transformation is carried out after the final matching of nodes. It involves adjusting the road centrelines using a non-rigid transformation, which takes into account the local positional distortions. During the transformation the matched points are brought into exact alignment while the unmatched points are transformed locally according to the positional displacement of the matched points.

Non-rigid transformations utilized here include rubber sheeting and models that exploit a radial basis function. In this study, the Thin Plate Splines (TPS), which is a radial basis function, is used as the basis for the coordinate transformation during feature alignment and adjustment by taking advantage of the matched nodes and line segment vertices. In contrast to rubber sheeting transformation, TPS aims at obtaining a parametric solution.

TPS is commonly used for representing coordinate mappings from R^2 to R^2. It is commonly used to represent shape deformations, for example, image warping or shape detection, and has been applied in biological forms, which exhibit non-rigid transformation. The graphical cadastral dataset in this study is considered to exhibit uneven and non-rigid positional distortions. Therefore, adopting TPS in geometric transformation should yield more reliable results. In general, the aim of using TPS is to minimize the amount of change in landmark positions and therefore minimal shape deformations; and avoid disruption between objects by ensuring that topologic and geometric structures between elements are maintained.

Given two sets of 2D corresponding points, say X(source) and V(target), consisting of points $\{X_a, a = 1, 2\ldots k\}$ and $\{V_a, a = 1, 2\ldots k\}$, TPS fits a mapping function $f(v_a)$ between corresponding point sets X and V by minimizing the bending energy, I_f,given as:

$$I_f = \iint_{R^2} (f_{xx}^2 + 2 f_{xy}^2 + f_{yy}^2) dxdy \qquad (1)$$

And, has the form:

$$f(x,y) = a_1 + a_x x + a_y y + \sum_{i=1}^{k} w_i U(\|(x_i, y_i) - (x, y)\|) \quad (2)$$

Where:

- First 3 terms (a_1 , a_x:, a_y) are the coefficients of the global affine transformation
- Remainder of the terms : describe (non-global) non-linear transformation
  - $U(r) = r^2 \log r$: with r being the Euclidean distance between any two points
  - w_i is the TPS coefficient at point *i.*

In equation (2), TPS is a 2D generalization of the cubic spline and includes the affine transformation as a special case in its regularized form.

In order for the function $f(x, y)$ to have continuous second derivatives, the following conditions have to be satisfied:

$$\sum_{i=1}^{k} w_i = 0 \quad (3) \qquad \sum_{i=1}^{k} w_i x_i = 0 \quad (4) \qquad \sum_{i=1}^{k} w_i y_i = 0 \quad (5)$$

This, together with the interpolation condition $f(x_i, y_i) = v_i$, yields a linear system for the coefficients, which can be written in matrix form as:

$$\begin{bmatrix} K & P \\ P^T & O \end{bmatrix} \begin{bmatrix} w \\ a \end{bmatrix} = \begin{bmatrix} v \\ o \end{bmatrix} \quad (6)$$

Where

- $K_{i,j} = U(\|(x_i, y_i) - (x_j, y_j)\|)$
- The i^{th} row of *P* is (1, x_i, y_i)
- *O* is a 3 x 3 matrix of zeros
- *o* is a 3 x 1 column vector of zeros
- *w* and *v* are column vectors formed from w_i and v_i respectively
- *a* is the column vector with elements a_1, a_x, a_y

An assessment of approximation techniques for determining the solution of TPS transformation is presented in Donato et al. (2002). This is particularly useful when the number of corresponding source and target values is large. The methods include simple sub-sampling, basis function subset and matrix approximation. Simple sampling exploits a randomly selected subset of corresponding points. Basis function subset involves using a subset of the basis functions with all of the target values. Matrix approximation approach involves the use of a full set of basis functions and a full set of target values. In this study since the number of the corresponding node points is relatively small, in the range of less than 500 points, all the matched points are used to determine the solution.

Up to this point, the preliminary transformation and alignment of features was based on the matching of nodes. No attempt was made to carry out the final matching of the arcs. This is because of the required 1:1 correspondences between the arcs. To exploit the correspondence between the linear features as the basis for the positional accuracy enhancement and alignment of the cadastral dataset, a technique that exploits paths instead of the individual arcs is suggested. The technique uses a shortest path algorithm to establish corresponding paths between the derived and mapped road datasets. The corresponding paths are then considered as the corresponding linear features that have to be aligned.

Although these corresponding linear features are considered to be similar, their vertices are not necessarily conjugate to each other. To align the corresponding linear features, a functional relationship between the features has to be established. The main objective during the alignment is to determine the displacements of the vertices of the derived road dataset features. A method based on the active contour (snake) concept (Siriba and Sester,

2010) is considered among other methods. The displacement values of the line segment vertices together with those of the nodes are the basis for transforming the graphic cadastral dataset.

4. Experimental Results and Analysis

4.1 Description of Experiment Data

The source dataset used in this study was derived from a graphical cadastral dataset created by directly tracing the outline of parcel boundaries from un-rectified, mosaicked and enlarged aerial photographs. Therefore, the dataset is fundamentally graphical, has geometric distortions due to the disregard for the relief distortions, besides having no coordinate grid. Only the outlines of land parcels boundaries are included and the roads are not explicitly represented. The topographical dataset used as the basis for accuracy enhancement was mapped from aerial photographs using stereo photogrammetry for a target cartographic scale of 1:2,500 and contour interval of 1m.

The geometric accuracy of the original graphical cadastral dataset was evaluated based on the approach described in Siriba(2009) and was found to range between 0 to 40m, while that of the target topographic dataset was evaluated based on the mapping specifications used. At a cartographic scale of 1:2,500, the nominal positional accuracy is then reasonably assumed to be about 0.5m. Figure 2, shows the extent of the original graphical dataset after preliminary geo-referencing, and covers an area of approx. 3.5Km by 2.5Km.

4.2 Derivation of Road centreline from the Cadastral Dataset

The extraction of the source dataset from the approximately geo-referenced graphical cadastral dataset was carried out in two steps. First, road corridors are extracted from the cadastral dataset yielding a polygon road dataset. In the second step, road centerlines are created based on the road polygons. It is important that the topology of the input dataset is validated before any further process is carried. This is because if slivers and gaps exist between the polygons representing the land parcels, some false road corridors will be created.

a) Derivation of Road Corridors from the Cadastral Dataset

The process described here assumes that the original graphical cadastral dataset does not explicitly include road corridors as land parcels. The desired output in this step is a polygon dataset of road corridors that has the same spatial extent as the input cadastral dataset. This is achieved by buffering the input dataset with a distance sufficient enough to cover the entire extent of the dataset and all the distances between non adjacent polygons representing land parcels. This is then followed by combining the buffers using a polygon union procedure. The result is one large polygon, which is once again buffered with a negative distance but equal in magnitude to the distance used in the earlier buffer procedure.

The road corridor dataset is then obtained by the intersection of the input dataset whose adjacent polygons had been combined with the large polygon generated earlier in this step. This procedure can generally be described as buffer union. It entails a combination of the positional processes of buffering, union and intersection of the resultant polygons. Figure 2 shows the outline of the land parcels (in grey) as digitized from a map and the road corridors (parcels) in gray fill as derived from the land parcels. This dataset is then used as the basis for the derivation of the road centerlines (in black).

b) Derivation of road Centrelines from road corridor dataset

To generate road centrelines from a road corridor dataset two approaches were considered. The first approach, which is vector based, is presented in Haunert and Sester (2008). This is an area collapse method based on straight skeletons. The second approach, which is raster based, converts the road corridor polygons to raster and then generates road centerlines in raster domain. The main operation is line thinning followed by the extraction of the road centerlines. The raster approach was first carried out in ArcScan, an extension of the ArcGIS software system (ESRI, 2008). The operation in ArcScan involves adjusting some setting depending on the required result. The settings include, for example, the required level of line smoothing and how the junctions are to be treated. The derived road centrelines are shown in black in figure 2. Because of the potential distortions introduced by the thinning and smoothing procedures , the vector approach was instead considered.

Figure 2: Extent of the experimental data with an outline of the derived roads extraction.

4.3 Feature Matching

After determining the positional accuracy of the cadastral dataset based on positional error modeling, double the value of the maximum error was taken as the planar distance threshold during node and arc feature matching. The maximum positional error was multiplied by 2 (or any other value if required) to ensure that the correct matching candidates are included in the initial preliminary matches. In this study, the maximum positional error for the cadastral dataset was determined to be 40m; therefore, a threshold distance of 80m was used in the matching process. For the final matching of nodes, angle of the incident arcs was specified not to exceed 45° (decimal degrees).

Table 1 gives a summary of the preliminary and final matching of nodes. The total number of preliminary node matches was 402 which included both 1:1 and 1: m matches. Since for node matching only 1: 1 matches are expected, the uncertain matches should be filtered out during the final matching. This is achieved by considering the angle of the incidence arcs on the matching nodes.

Table 1: summary of node matching

	Number of node features	Preliminary node matches	Final node matches	Final node matches	Valid node matches	% Certainty
Source-to-target						
Source	328	219	122	69	72	95
Target	231	138	102	69	72	95
Target-to-source						
Source	328	216	123	65	72	90
Target	231	140	102	65	72	90

The final matches could still contain matches that are uncertain. The reference and target road datasets contained 328 and 231 nodes, respectively. In the preliminary matching, when the source dataset was specified as the reference, a total of 219 nodes in the source dataset were matched to 138 nodes in the target dataset. This result contained both 1:1, 1: m and m:1 matching. In the final matching, total of 122 nodes in the source dataset were matched to 102 nodes in the target dataset. This result contains 1:m and the unwanted m:1 node matches. During a visual validation, out of the 122 final matches only 67 were validated, i.e., 1:1 matches. When the target dataset was used as the reference, the results are as shown in the last two row of Table 1. This means that in the first case 3 matches were omitted, while 7 matches were omitted in the second case. This strategy therefore helped to recover these matches, with a certainty of 95% and 90% respectively.

Figure 3 (i) and (ii) show sample cases of node matches, which were missed when the source dataset was taken as the reference. In the example, node matches (98-111) and (272-136) were missed in 3(i) and 3(ii) respectively. Figure 3 (iii) and (iv) show sample cases where node matches were missed when the target dataset was taken as the reference dataset. The omitted nodes matches are respectively, (233-87) and (285-97). The main reason for the missed matches is the difference in the valency of the nodes involved. For example, in (i), node 98 in the source dataset has a valency of 3, while the corresponding node 111 in the target dataset has a degree of 1. According to the algorithm used, it is then possible to establish the correct match (98 -111) when the source dataset is taken as the reference than when the target dataset is taken as the reference.

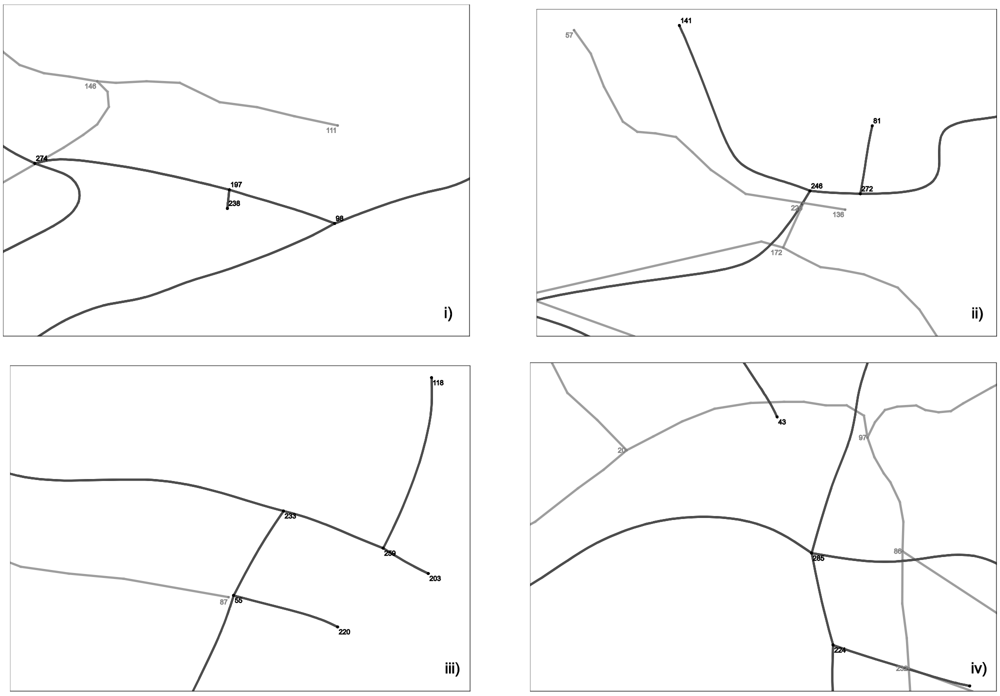

Figure 3: Examples of omitted node matches recovered via the two-way matching.

The results of the two-way matching strategy are illustrated in Figure 4. A total of 72 matches were obtained using this strategy. The objective of uniform distribution of the matching nodes was obtained, thus guaranteeing a uniform effect from the nodes during transformation (see figure 5).

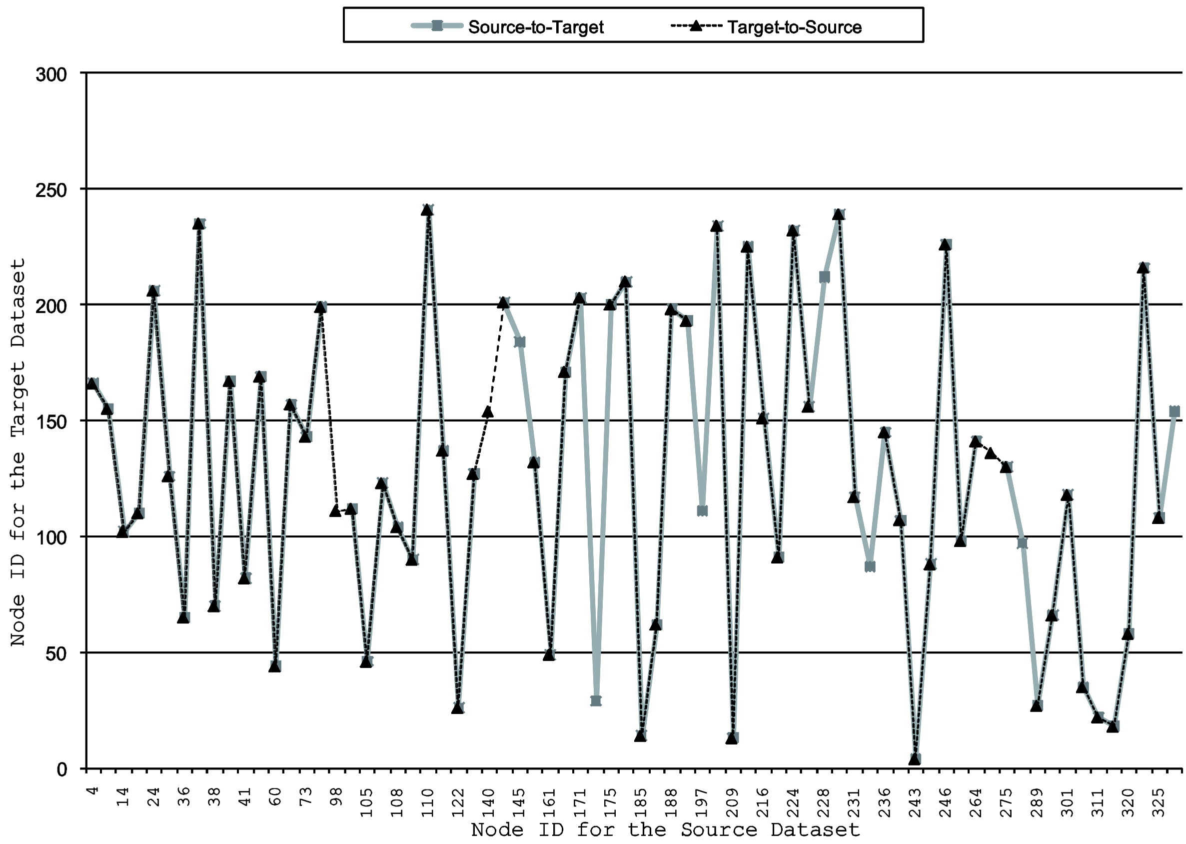

Figure 4: Illustration of mutual two-way node matching.

4.3 Transformation and Linear Feature Alignment

Figure 5 illustrates the source road dataset before (in black) and after transformation (dotted black), while the target road dataset is represented in grey. The source road dataset was transformed based on the transformation value derived from matched nodes, which are also shown in the figure. The matched nodes are uniformly distributed, although it can be seen that some possible node matches were missed

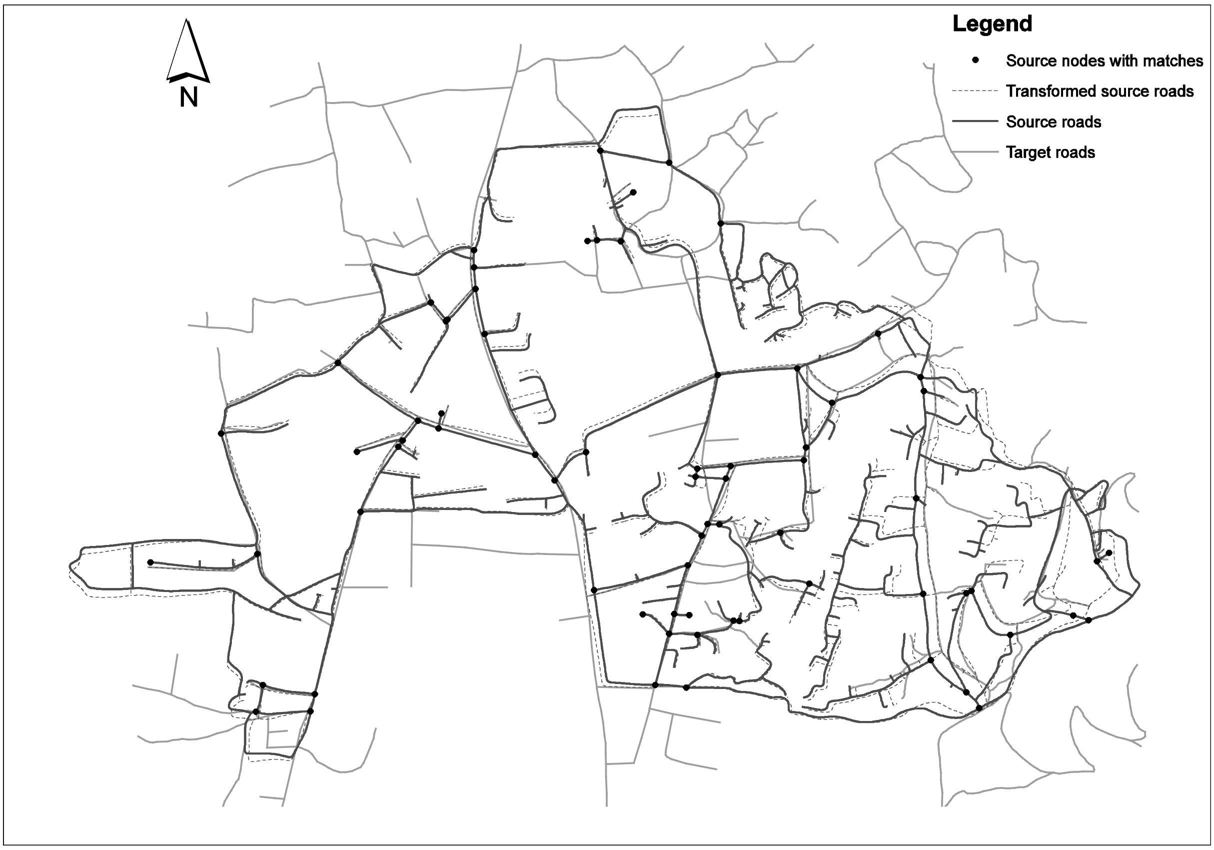

Figure 5: Transformed source road dataset

On the basis of the validated node matches, the entire source dataset is transformed using the non-rigid TPS transformation. Figure 6 a) illustrates in more detail part of the target dataset (in grey), the source dataset (in black), while 4b, the target dataset (in gray), the source dataset (in black) after feature matching and transformation are illustrated.

After the preliminary feature alignment, the matched nodes are brought into exact alignment, while the line segments only partially (see figure 6 b. The road centrelines derived from a cadastral map have to be aligned somehow with the road centrelines from the topographical dataset. This is necessary in order to determine the movement of the vertices of the linear segments.

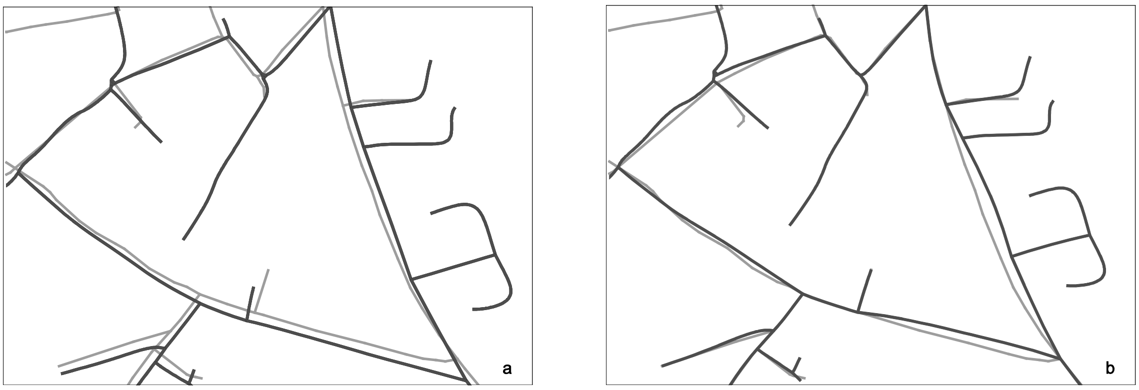

Figure 6: Target road dataset in black, source road dataset in grey (a) and the source road dataset after transformation in dotted grey (b)

A preliminary feature transformation was carried out based on the matched nodes. Two cases have to be distinguished here: if there is only one arc connecting two matched nodes in each data set, then there is a 1:1

correspondence between the arcs. If, however, there are several arcs connecting the two matched nodes in each data set, we propose to calculate the shortest path between the matched nodes using the Dijkstra shortest path algorithm (Dijkstra, 1959). The vertices along the matching paths are then used to create some intermediate linear features for transformation and alignment. Based on the corresponding linear features, a non-rigid transformation was carried out and applied to the original graphical cadastral dataset. Figure 7 shows the target road centerline (in black) and the original graphic cadastral dataset (grey) before and figure 8 shows the same graphical cadastral dataset after the transformation and alignment.

Figure 7: Graphical cadastral dataset before transformation and alignment

Figure 8: Graphical cadastral dataset after transformation and alignment

Visually, the graphical cadastral dataset is perfectly aligned in the sections whether there were point and linear feature correspondences. For sections that are enclosed by corresponding road segments, the positional accuracy is as good as that of the target road dataset. However for the rest, the extent to which the positional accuracy has been enhanced is part of the on-going further work. One possibility involves obtaining more accurate measurement of the boundaries of some of the land parcels to be used as a reference.

Discussion and Outlook

In this paper, a methodology has been presented for feature matching and alignment. The methodology is considered as an intermediate process in the overall process of positional accuracy enhancement of a graphical cadastre. The methodology consists of three steps: extraction of roads that are implicitly represented in a graphic cadastre; feature matching, which involves matching of both nodes and arcs that constitute the datasets; and finally linear feature alignment.

The vector approach that was used for the extraction of roads from a graphic cadastral dataset yielded sufficient results. This is because it avoids the potential distortions that are introduced by thinning in the raster approach. The assumption that all the gaps between land parcels are basically roads is not necessarily true, because some of these gaps could be representing riparian reserves. Again, not all the extracted roads actually exist and may not be contained in the target road dataset. The main weakness of the adopted approach is the creation of roads at the outer extents of the dataset, which is not always so. An alternative technique should be established that avoids this relatively inferior problem.

A two-way matching strategy was adopted, in which the source and target datasets were taken as the reference dataset alternately. This made it possible to recover matches that otherwise would have been omitted because of some significant differences in some parts of the datasets. The differences were mainly due to the differences in the degree of the nodes being matched. Whereas it was possible to identify the nodes corresponding to a node with a degree greater than 3 in the reference dataset, the corresponding node in the other dataset with a degree of 1, the reverse was however not possible.

Feature alignment, the final process in this methodology, was carried out in order to align the line segments between the matched nodes. The alignment was first carried by using the vector potential field technique and then by use of the relative orientations of the corresponding linear features obtained through the Dijkstra shortest path algorithm. Although the use of vector potential field yields promising results, the use of different iterations and the possible creation of kinks demand that an approach that considers the relative orientations of the features being aligned.

The matching and alignment results from this process are used to transform the original graphic cadastre from which the source road dataset was extracted. Two aspects that have to be considered include the differential accuracy levels and the legal validity of the new transformed cadastre. After the final transformation of cadastral dataset, the final quality levels may not necessarily be uniform; therefore the quality information need to be reported at the local rather than the global level. The legal validity of the new coordinates of the parcel boundaries, i.e., how to deal with the problem of changing registered areas of the parcels is one of the main issues that has to be addressed. This depends on the hierarchy of boundary evidence as used in the respective jurisdiction and on existing regulations.

In conclusion, the extent to which the positional accuracy is improved depends on the positional accuracy of the road dataset used as the target, and on the certainty of matching and alignment of the road datasets. The results presented here are based on the matching results of nodes only. Further work entails the alignment of the linear features, and the transformation of the entire data exists in the cadastral dataset.

References

Arvanitis, A. and Koukopoulou, S. (1999). Managing data during the update of the Hellenic Cadastre, in: FIG Commission 3 Annual Meeting and Seminar, FIG Commission 3, Budapest, Hungary.

Butenuth, M., von G¨osseln, G., Tiedge, M., Heipke, C., Lipeck, U. and Sester, M. (2007). Integration of Heterogeneous Geospatial Data in a Federated Database, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 62(2007), 328 346.

Cowen, D., Saalfeld, A., Price, K., Jensen, J., Ramsey, D., Lapine, L. and Broome, F. (2004). A Research Agenda for Geographic Information Science, CRC Press.

Devogele, T., Parent, C. and Spaccapietra, S. (1998). On Spatial Database Integration, International Journal of Geographical Information Science, vol. 12(4), 335 – 352.

Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs, Numerische Mathematik, vol. 1, 269 – 271.

Donato, G. and Belongie, S. (2002). Approximate Methods for Thin Plate Spline Mappings and Principal Warps, in: 7th European Conference on Computer Vision Part III Lecture Notes in Computer Science, vol. 2352, Copenhagen, Denmark, pp. 21 – 31.

Doytsher, Y., Filin, S. and Ezra, E. (2001). Transformation of Datasets in Linear-based Map Conflation Framework, Surveying and Land Information Systems, vol. 61(3), 159 – 169.

Envrionmental Systems Research Institute (ESRI) (2008). ArcScan for ArcGIS Tutorial, http://webhelp.esri.com/arcgisdesktop/9.3/pdf/ArcScan for ArcGIS Tutorial.pdf (accessed on 29.06.2010).

Felus, Y. A. (2007). On the Positional Enhancement of Digital Cadastral Maps, Survey Review,, vol. 39(306), 268 – 281.

Haunert, J. H. and Sester, M. (2008). Area collapse and road centerlines based on straight skeletons, GeoInformatica, vol. 12(2), 169 – 191.

Hope, S., Gordini, C. and Kealy, A. (2008). Positional Accuracy Improvement: Lessons Learnt from Regional Victoria, Australia, Survey Review, vol. 40(307), 29 – 42.

International Federation of Surveyors (FIG) (1995). Statement on the Cadastre, FIG Publication No. 11.

Jarroush, J. and Even-Tzur, G. (2006). Re-Instating Parcel Boundaries by Real Time Kinematik GPS, Survey Review, vol. 38(302), 643 – 653.

Kaufmann, J. and Steudler, D. (1998). Cadastre 2014: A Vision for a Future Cadastral System, in: XXI FIG-Congress, Brighton, UK.

Kieler, B., Huang, W., Haunert, J. H. and Jiang, J. (2009). Matching of River Datasets of Different Scales, in: M. Sester, L. Benard and V. Paelke (Editors), Advances in GI-Science: Proceedings of the 12th AGILE Conference -Lecture Notes in Geoinformation and Cartography, 2009, Springer-Verlag Berlin Heidelberg, pp. 135 – 154.

Klebanov, M. and Forrai, J. (2010). Implementing a Coordinate Based Cadastre in Israel: Experience and Perspectives, in: FIG congress: Facing the Challenges -Building Capacity, Sydney, Australia, p. 17.

Morgenstern, D., Prell, K. M. and Riemer, H. G. (1989). Digitisation and Geometrical Improvement of Inhomogeneous Cadastral Maps, Survey Review, vol. 30(234), 149 – 159.

Musti´ere, S. (2006). Results of Experiments on Automated Matching of Networks at Different Scales, in: ISPRS WG II/3, II/6 Workshop, Hanover, Germany.

Olteanu, A., Musti´ere, S. and Ruas, A. (2006). Matching Imperfect Spatial Data, in: M. Caetano and M. Painho (Editors), Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.

Saalfeld, A. (1988). Conflation: Automated Map Compilation, International Journal of Geographical Information Systems, vol. 2(3), 217 228.

Savopol, F. and Armenakis, C. (2002). Merging of Heterogeneous Data for Emergency Mapping: Data Integration or Data Fusion?, in: Sysmposium on Geospatial Theory, Processing and Applications, Ottawa, Canada.

Sheeren, D., Musti´ere, S. and Zucker, J.-D. (2004). How to Integrate Heterogeneous Spatial Databases in a Consistent Way?, Lecture Notes in Computer Science -Advances in Databases and Information Systems, vol. 3255(2004), 364 – 378.

Siriba, D. (2009). Positional Accuracy Assessment of a Cadastral Dataset based on the Knowledge of the Process Steps used, in: 12th AGILE Conference on GIScience, Hannover, Germany.

Siriba, D. and Sester, M. (2010). Linear Feature Alignment based on Vector Potential Field, in: Joint International Conference on Theory, Data Handling and Modelling in Geospatial Information Science, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 38 Part 2, Hong Kong, China, pp. 400 – 405.

Stanislawski, L. V., Dewltt, B. A. and Shrestha, R. L. (1996). Estimating Positional Accuracy of Data Layers within a GIS through Error Propagation, Photogrammetric Engineering & Remote Sensing, vol. 62(4), 429 – 433.

Steinberg, G. (2000). Implementation of a Legal digital Cadastre in Israel, in: FIG Working Week, Seoul, Korea.

Steudler, D., Williamson, I. P., Rajabifard, A. and Enemark, S. (2004). The Cadastral Template Project, in: FIG-Working Week 2004, Athens, Greece, http://www.cadastraltemplate.org/ (accessed on 16.07.2010).

Walter, V. and Fritsch, D. (1999). Matching spatial datasets: a statistical approach, International Journal of Geographical Information Science, vol. 13(5), 445 – 473.

Williamson, I., Enemark, S., Wallace, J. and Rajabifard, A. (2010). Land Administration for Sustainable Development, ESRI Press Academic, Redlands, California.

Williamson, I. P. (1985). Cadastres and Land Information Systems in Common Law Jurisdictions, Survey Review, vol. 28(217), 114 – 129.

Yuan, S. and Tao, C. (1999). Development Of Conflation Components, in: Proceedings of Geoinformatics'99 Conference.

Zhang, M. and Meng, L. (2007). An iterative road-matching approach for the integration of postal data, Computers Environment and Urban Systems, vol. 31(2007), 597 – 615.