Relative Curve Orientation in the Alignment of Inconsistent Linear Datasets

David N. Siriba, Daniel Eggert, Monika Sester

Institute of Cartography and Geoinformatics (IKG), Leibniz University of Hannover, Germany
Appelstraße 9a, 30167 Hannover, Germany.
E-mail: {david.siriba, daniel.eggert, monika.sester}@ikg.uni-hannover.de

ABSTRACT

An approach for the alignment of a linear dataset of inconsistent positional accuracy with a more accurate one is presented. The approach consists of two main steps: path-matching and alignment based on relative curve orientations of the corresponding feature followed by a non-rigid transformation based on Thin plate Splines (TPS). The results obtained with this approach on two experimental cases show significant improvement in alignment and geometric accuracy of the lower accuracy datasets.

1. INTRODUCTION

Geometric alignment of spatial features is one of the typical problems of data integration and usually carried out so that spatial features from two datasets that describe the same object coincide. It is achieved through point-based transformation that utilizes corresponding point features, which represent either some identified point landmarks or nodes mainly in a road network. Point-based transformation is adequate only for simple cases where positional discrepancies between the datasets are low and almost evenly distributed. In case the discrepancies are high and unevenly distributed, point-based approaches are not sufficient for aligning linear and polygonal features as they only guarantee that the control points will coincide. This problem is often encountered when integrating legacy datasets that contain large positional distortions with more geometrically accurate and newer datasets. Therefore corresponding linear features could be used to determine a more suitable geometric transformation between the datasets, in addition to point features. This is because linear features contain more semantic information unlike points, which are dimensionless.

This paper presents a vector based approach for the alignment of two road dataset in which one of them is of inconsistent geometric accuracy. The approach exploits the relative curve orientations of corresponding linear features in the datasets. The approach consists of three steps: point feature matching and preliminary alignment using Thin Plate Splines (TPS), a non-rigid transformation; path matching; and final linear feature vertex correspondence and alignment. The approach is described after an overview of the related work on curve alignment.

2. OVERVIEW OF CURVE ALIGNMENT

Curve alignment is considered to be good if: human-perceptually similar curves are matched; translation, rotation and scaling are invariant to the alignment process; general and significant features should be considered to have higher priorities over local and minor ones and distortions are tolerable to a certain extent (Li et al, 2006).

Curve alignment methods can be classified into two categories: methods based on rigid and non-rigid transformation (Sebastian et al., 2003). Methods based on rigid transformation rely on matching feature points by finding the optimal rotation, translation and scaling parameters that minimize the distance between the two curves. These methods are however sensitive to articulations, deformations of parts and other variations in the object form.

Non-rigid transformation methods model articulations and other deformations by finding the mapping from one curve to another that minimizes a performance function consisting of stretching and bending. The minimization problem in the discrete domain is transformed into matching of shapes based on curvature, bending angle or absolute orientation as a measure of similarity. The approach for non-rigid curve matching was pioneered by Cohen et al. (1992). The basic premise of the approach was to match high curvature points along the curves, while maintaining a smooth displacement curve. The problem was cast in terms of minimizing an energy functional penalizing “bending” and “stretching” in a physical analogy similar to the one used in formulating active contours (Kass et al., 1988). This approach is sufficient for the alignment of parametric curves with features that exhibit a lot of flexibility however not appropriate for polygonal curves which represent spatial features.

Doytsher et al. (2001) describe a technique specifically for the alignment of topographic linear features. This technique entails projecting the vertices from both polylines towards one another. The points in the source curve that have local distortions will be mapped to wrong vertex positions in the reference polyline. Figure 1 (left) illustrates a situation where the geometric accuracy of the corresponding curves (reference in black and source in grey) is almost the same. While Figure 1 (right) shows a situation where the source curve is distorted and therefore its vertices do not have possible matches in the reference feature.

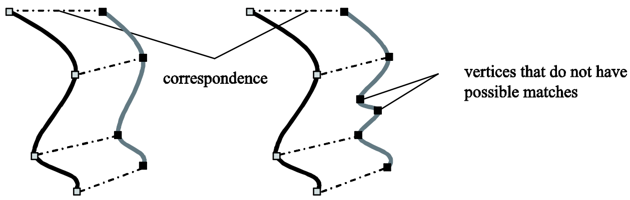

Figure 1: Correspondence between the characteristic points (vertices) of the source (grey) and reference (black) polylines

The uneven positional differences due to local distortions require a non-rigid transformation that takes curve characteristics into account. Song and others (2006) developed a data conflation approach based on the active contour models (snake) to improve the positional accuracy of TIGER roads to high-accuracy roads. The resulting accuracy obtainable is limited by the size of the cell specified during the rasterization of the reference dataset. We propose a vector-based approach, avoiding that problem, while exploiting the corresponding characteristic curve points and a non-rigid transformation based on Thin Plate Splines to achieve a better alignment.

3. LINEAR FEATURE ALIGNMENT BASED ON RELATIVE CURVE ORIENTATIONS

The approach consists of two steps analogous to the conventional second and third steps of the classical data conflation. We assume in our approach that pre-integration has already been carried out, therefore only the remaining two steps as developed in this study are described. The two steps are path matching and feature alignment based on relative curve orientation.

3.1 Path Matching

The bulk of feature matching algorithms in literature involve road datasets. The matching algorithm developed by Walter and Fritsch (1999) was based only on the geometric structure of the features in the datasets without regard to the inherent network structure. A network-based matching algorithm developed by (Mustiére, 2006) considered road networks at different scales. This was extended by Mustiére and Devogele (2008) by considering networks with different scales. This study adopts the generic network matching approach and considers road network datasets that have significant differences in the geometric accuracies.

A network matching algorithm generally and specifically as developed in this study consists of three main steps: pre-matching, final node-matching and path-matching.

- *Pre-matching*: nodes and arcs in the source dataset that lie within a specified threshold distance from node and arcs respectively in the reference dataset are identified. The threshold distance can be determined empirically from the relative accuracies of the datasets.
- *Final node-matching*: the final matching of the nodes is carried out using geometric and semantic criteria to filter out the unlikely matches from the pre-matching operation. 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 confirmed as a valid match if

the angles of the corresponding incident arcs make with the respective node are within a specified angular tolerance.

- *Path-matching*: geometric, semantic and topologic criteria to identify corresponding paths in the networks are used instead of individual incident arcs. Path matching is used instead of matching of the individual incident arcs so as ensures that only one-to-one correspondences are obtained, which are required for the alignment operation.

In this study path-matching algorithm involves finding the corresponding paths between a given pair of corresponding nodes obtained after the final matching process. In order to identify matching polylines the shortest paths between nodes from one dataset are determined. In case a (shortest) path between the corresponding nodes from the other dataset exists a potential match is found. Determining the shortest path employs the Dijkstra shortest path algorithm, DSPA (Dijkstra, 1959). Since the DSPA is based on weighted graph edges, the weight of an edge is set to the length of the underlying polyline. The identified match between the two shortest paths still needs to be validated. For this purpose, various validation methods can be applied. The lengths of the paths considered are compared. In case the length ratio is within a given threshold (e.g. 1:2) the path match is then considered valid.

Some optimizations are necessary in order to increase the matching speed and quality. First, the overall shortest path within the graph is referenced. In case a match is found, the paths are removed from the graphs and a new overall shortest path needs to be determined. Depending on the given graph, applying the DSPA many times may become a time consuming task. Since the first shortest paths are likely to consist of a single edge, the path matching step is split into two. The first step incorporates matching only single edges from dataset to the corresponding shortest path in the other dataset and vice versa. This includes only one DSPA between two points in one graph instead of DSPA between all points in both graphs. After that step all one-to-many and many-to-one path matches are removed from both graphs, which are now likely to be split into various unconnected sub-graphs. This improves the DSPA search, because not all nodes are connected and therefore fewer paths exist in the graph. Finally, apart from the paths that have a one-to-one polyline correspondence, all the other path that consist of one-to-many and many-to-many polyline correspondences are combined to form intermediate one-to-one polylines for the purpose of alignment.

3.2 Feature Alignment based on relative curve orientation

Curve orientation along polyline, captures the geometric characteristics of the features they represent and employed for identifying the corresponding points (vertices). The approach is based on the paradigm that high curvature points usually possess an anatomical meaning, and are therefore good landmarks to guide the matching process (Cohen et al., 1992). The approach has three main steps: i) computation of the direction change ii) the identification of the matching vertices and iii) final alignment of the linear features.

3.2.1 Direction Change

To compare the direction changes between the vertices of the curves, normalization is required. Based on the relative normalized position, a direction change is calculated at each vertex. The first and the last vertices are assigned a direction change of zero, since no direction change takes place there. Figure 2 shows an example of a pair of corresponding polylines. The source polyline is represented in grey, while the reference polyline is represented in black.

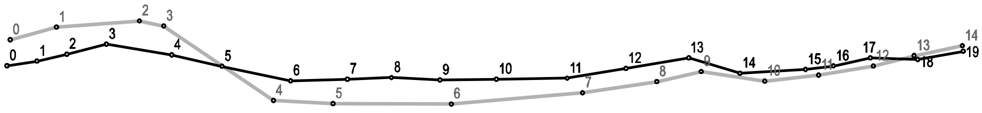

Figure 2: Corresponding source and reference polylines

A corresponding graph of the direction change at each vertex plotted against its relative normalized position is shown in Figure 3. The similarity of the example polylines based on the direction change of their points is evident. How the actual matches are identified is described in the next step.

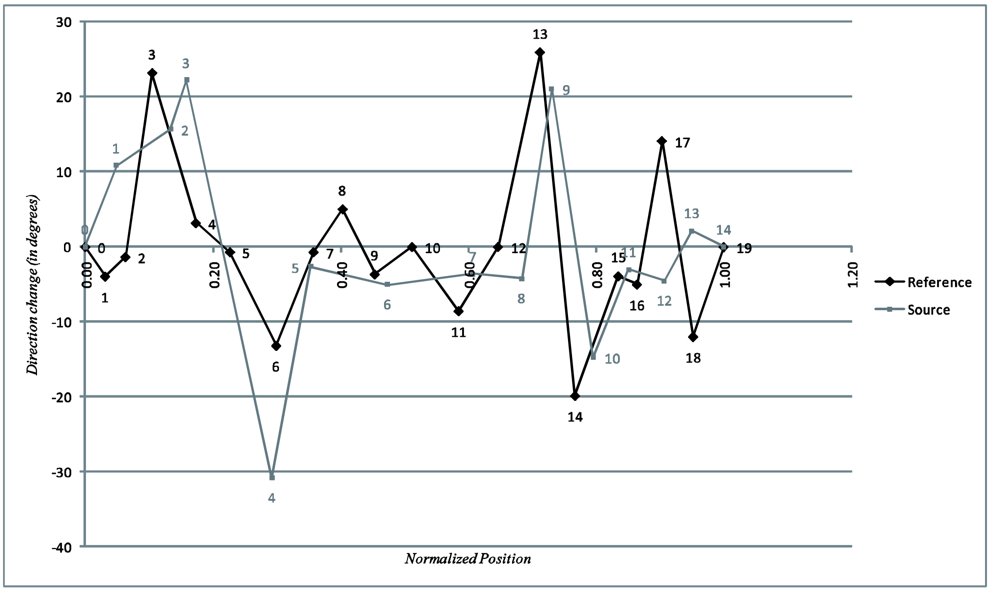

Figure 3: Direction change at the vertices for the source and reference polylines shown in Figure 2

3.2.2 Matching vertices

Matching of the vertices depends on the distance between the relative normalized positions of the vertices. Matches with the least distance are considered to be the most probable. A further condition is to ensure that the vertices have comparable amplitude, i.e., the perpendicular distance of the vertices and the straight line connecting the start and end points of the respective polylines. The vertex matches are ascertained further if the curvature (i.e., the derivative of the curve's tangent angle with respect to position on the curve at that point) at the vertices is nearly the same. Table 1 contains the final matches between the vertices.

Table 1: Correspondence of vertices between two corresponding curves**.**

Source curve	0	3	4	9	10	11	12	14
Reference curve	0	3	6	13	14	15	16	19

3.2.3 Curve alignment

Once the points in the curves have been matched, the unmatched points from the source curve are scaled accordingly and delimited by the matched points. The alignment of the curves is thus limited by the flexibility of the source curve, depending on the number of points. For each point in the corresponding curves in the source dataset, a displacement vector is determined, which then is used to determine a parameterized non-rigid transformation for the rest of the polylines in the dataset.

The non-rigid transformation adopted here exploits the Thin Plate Splines (TPS). TPS is adopted in this approach because transformation parameters are required that will also be used to transform polylines in the source dataset that do not have corresponding polylines in the reference dataset. Moreover, since the source dataset contains uneven local positional distortions, TPS provides the best non-rigid transformation for representing such shape deformations, which exhibit nonlinear and local character. TPS are used as interpolating functions and consequently ensure an exact correspondence between the points. Some theory and assessment of approximation techniques for determining the solution of TPS transformation is presented in Donato and Belongie (2002).

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experimental Data

The approach was implemented as a computer application and applied to two example datasets as illustrated in Figure 4. In Case 1 (Figure 4-left), the dataset-pair consists of a road network dataset mapped at a scale of 1:2500 and the corresponding road network obtained by unevenly distorting the mapped road dataset. Case 2 (Figure 4right) consists of a road dataset mapped at a scale of 1:2500, while the corresponding road dataset was derived from a geometrically inaccurate cadastral dataset mapped at 1:2500. In both cases, the relative positional discrepancy between the datasets was generally in the range of 0 to 40 m. In Case 1, the source dataset is simulated and a possibility to obtain a perfect alignment of the features, since there are only a one-to-one correspondences. In Case 2, from a real situation, there locations with feature corresponding at all. This is mainly because of the difference in the time of capture between the two datasets.

4.2 Path Matching

The threshold distance was determined based on the knowledge of the maximum positional error in the source dataset. The certainty of node matching in the two cases was respectively 98% and 95%. The validation-criterion of potential matching paths is based on the relative lengths and a value of 1:2 was empirically set. It is however possible that those paths follow different routes, yet they could have their relative lengths well within this value. In this case, an alternative criterion is required.

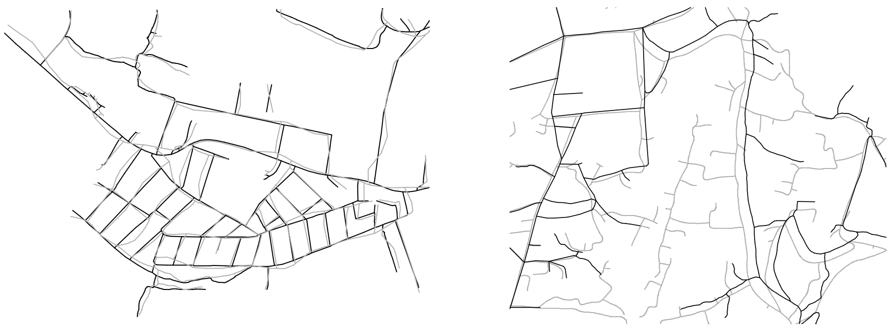

Figure 4: Inconsistent linear datasets with reference in black and the source in gray: Case 1(*left*) and Case 2 (*right*)

4.3 Feature Alignment based on relative curve orientation

Figure 5 shows the results of alignment after path-matching for Case 1 (left) and Case 2 (right) respectively. In both cases there are sections in the dataset where the alignment is poor. This is because of either missed point correspondences during point matching or complete lack of corresponding features.

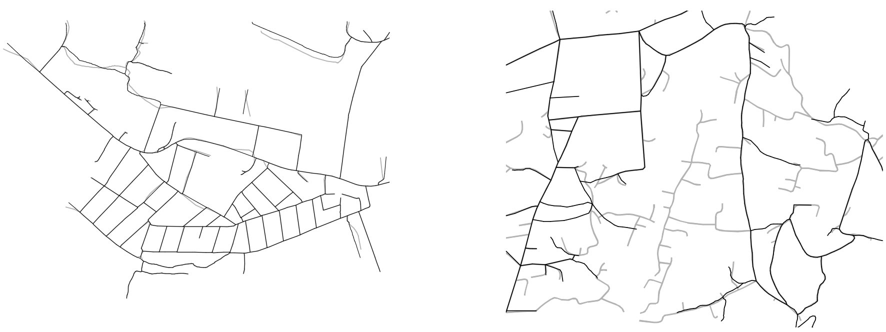

Figure 5: Final alignment of the linear datasets: Case 1(*left*) and Case 2 (*right*)

The results presented in Figure 5 satisfy one of the qualities of good alignment, i.e., matching the human intuition. In addition, there is a significant improvement in positional accuracy of the source dataset as a result of the alignment. The improvement of the positional accuracy was evaluated based on the Haudorff distance of the corresponding linear features in both datasets. In Case 1, the positional accuracy of features in source with correspondences in the reference datasets was in the range of 0.63 m to 51.8m (with an average of 16.8m). This was improved to an average positional accuracy of 2.8m. In Case 2, the positional accuracy of the features in source dataset ranging from 0m to 47.8m (an average of 10m) was improved to an average of 1.7m. Following this approach, there is clearly a significant improvement in the alignment of features in the source dataset.

5. CONCLUSION AND OUTLOOK

This paper has outlined with illustrations a vector-based approach for the alignment of inconsistent linear datasets. It was demonstrated that there is a significant improvement in the alignment by ensuring that the relative characteristics of the features are maintained and thus matching the human intuition of good alignment. In addition, the approach yielded significant improvement in the positional accuracy of road network dataset of low positional accuracy. The approach which exploits point and path matching together with the TPS transformation corrected the positions of the source road network significantly and produced roads with positional accuracy equivalent to those of the reference road network.

In order to guarantee the positional accuracy of polyline in the source road network without correspondences in the target road network, further work is required. This would entail establishing whether the terrain has any influence on the positional inaccuracy of the source dataset and if so, how the terrain information could be incorporated in the approach to improve the accuracy.

BIBLIOGRAPHY

Cohen, I., Ayache, N. and Sulger, P., 1992 Tracking Points on Deformable Objects Using Curvature Information, in: Proceedings of the 2nd ECCV, pp. 458-466.

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, 21-31.

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

Kass, M., Witkin, A. and Terzopoulos, D., 1988 Snakes: Active Contour Models, International Journal of Computer Vision, vol. 1(4), 321-331.

Li, Z., Luo, X. and Gao, C., 2006 Multi-Resolution Curve Alignment Based on Salient Features, in: 18th ICPR, vol. 2, Hong Kong, China, 357-360.

Mustière, S. 2006 Results of Experiments on Automated matching of networks at Different Scales. In: ISPRS WG II/3, II/6 Workshop, Hannover, Germany

Mustière, S and Devogele, T, 2008 Matching Networks with different Levels of detail. GeoInformatica, vol. 12(4), 435-453.

Sebastian, T. B., Klein, P. N. and Kimia, B. B., 2003 On Aligning Curves, IEEE Transactions on Pattern Analysis, vol. 25(1), 116-124.

Song, W., Haithcoat, T. L. and Keller, J. M., 2006 A Snake-based Approach for TIGER Road Data Conflation, Cartography and Geographic Information Science, vol. 33(4), 287-298.

Walter, V. and Fritsch, D., 1999 Matching Spatial Datasets: A Statistical Approach, International Journal of Geographical Information Science, vol. 13(5), 445-473.