A DUAL INTERVAL APPROACH FOR THE ALIGNMENT OF GEOMETRIES IN VECTOR DATA SETS

G. v. Gösseln, M. Sester

[a] Institute of Cartography and Geoinformatics, University of Hannover, Appelstr. 9a, 30167 Hannover, Germany – {vongoesseln,sester}@ikg.uni-hannover.de

KEYWORDS: Cartography, GIS, Adjustment, Correction, Integration

ABSTRACT:

Today, a growing availability of geo datasets can be observed, together with an increasing accessibility of the data via web-services. In spite of this favourable situation, there are still some problems left on the way to true interoperability, which are due to the heterogeneity of the data sets: even when data relate to the same phenomena in reality, there are still geometric discrepancies left. The reasons for this are manifold: different underlying data model, different acquisition time, different acquisition type and finally different human operators who performed the measurements. In order to allow for a true interoperability a geometric adaptation of corresponding objects is necessary. Only then, overlay and buffer operations can be executed, as well as geometric calculations.

In the paper, an approach for matching and adaptation of geometric objects is presented, which leads to one consolidated data set.The adaptation process starts with finding corresponding objects in both data sets using matching techniques. These corresponding objects serve as "anchor objects" according to which the adaptation of the rest of the data can be performed using rubber sheeting. The adaptation consists of two different adaptation operations, namely the well-known ICP-algorithm and the Dual Interval Adaptation (DIA), that allow for a global adaptation of whole objects, and local adaptation of parts of objects, respectively.

1. BACKGROUND

Due to different underlying data model, different acquisition time, different acquisition type and finally different human operators, the content of different data sets differ in geometry, accuracy and actuality, even when they relate to the same area and the same objects in reality. Integrating different data sets is usually a fundamental prerequisite for spatial data analysis, furthermore, it allows for a consistent representation and for the propagation of updates from one data set to the other. To achieve these integration benefits, an approach will be presented which allows not only for the alignment of one data set (candidate) to a reference data set, but by using a weighting factor also data conflation can be performed which enables combining the information of two data sets to derive a new one.

The workflow which will be presented in this paper can be executed on any kind of vector data set with polygon object geometries.

2. RELATED WORK

Prior to any geometric data integration and alignment, a semantic matching has to be performed (Kuhn 2003, Volz, 2005). A conceptual framework for the integration of geographic data sets, based on a domain ontology and surveying rules, was developed for update propagation between topographic data sets (Uitermark, 2001). There are concepts which also take relation and data attributes into account (Cobb et al., 1998, Walter & Fritsch, 1999). Data integration can also take different cardinalities of the matching partners into account (Sester et al., 1998b).

The integration of vector data and raster data is being investigated with the aim of enriching a 2D-vector data set with 3D-information (Butenuth & Heipke, 2003).

Integration can be used for data registration, when one data set is spatially referenced and the other has to be aligned to it (Sester et al., 1998a). The alignment of different geometries can be performed using the iterative closest point algorithm (ICP), which originally has been developed by (Besl & McKay, 1992) for aligning 3D objects. In Kohonen (1997) the learning vector quantization (LVQ) approach has been presented which associates an entire feature vector sequence, instead of a single feature vector. The usage of geometric harmonization for cadastre data sets has been presented by Hettwer & Benning (2000). Doytsher and Gelbman (1995) presented a rubber sheeting algorithm for cadastral maps. This algorithm is used to ensure the correct alignment of the complete data set.

3. GEOMETRIC ALIGNMENT

The goal of the alignment is to derive one consolidated data set that includes the properties of both data sets it is based on. The prerequisite is to first identify corresponding objects using matching. In our case we focus on areal objects that relate to the same area, therefore, a simple geometry overlay matching procedure was used. After possible matching candidates are identified, in a first step it will be decided, if there is a global fit between the two geometries. This is done using the ICP algorithm, leading to a global adaptation of the whole object. There still might be some local discrepancies left, which can be adapted, as long as they are in the range of allowed differences between the objects. The result of the local adaptation algorithm DIA are displacement vectors of vertices from the original geometry of the objects to the new, consolidated geometry. Based on these displacement vectors for all the reference objects, the whole data set can be harmonized using rubber sheeting. In the following, we will concentrate on the adaptation process.

3.1 Derivation of object correspondences using matching

An area-based matching process is used for the creation of links between object candidates. Because of the fact that all used datasets refer to the same area, objects representing the same real world object must share a certain amount of space. If they do not intersect, the objects must not be handled as corresponding partners. Objects which have been identified as corresponding objects in the matching process, are stored in a so called relation set, which consists of two sets of objects; these sets are forwarded to the alignment process. The correspondences between the objects in the two data sets can be of cardinality 1:1 to n:m.

3.2 Adaptation using a four-parameter transformation (ICP)

The iterative closest point algorithm (ICP) developed by (Besl & McKay, 1992) has been implemented to achieve the best fitting between the objects from one data set to another using a rigid transformation. The original ICP algorithm was targeting at the alignment of three-dimensional objects using a 7 parameter transformation. In our case the problem is reduced to a 2D problem which requires four parameters (position, scale, orientation). At the end of the process the best fit between the objects using the given transformation is achieved, and a link between corresponding objects in the different data set is established. Evaluating the transformation parameters allows to classify and characterize the quality of the matching: in the ideal case, the scale parameter should be close to 1; also rotation and translation should be close to 0. Assuming, that the registration of the data sets is good, a greater scale factor can be an indicator, that the difference between two objects is due to a change at the real world object, which occurred between the data acquisition of different data sets (Goesseln & Sester, 2004).

The application of the iterative adaptation using the ICP approach showed very good results in the global adaptation of the objects and in the decision, whether the objects really match and can be adapted. However, after the global adaptation, there are still some local discrepancies, which lead to sliver polygons in the overlay of the two geometries. This is especially the case for long and elongated natural objects like rivers, for which typically no single transformation can be found that leads to a perfect fit.

3.3 Adaptation using a vertex-based approach (DIA)

In order to compensate for local discrepancies in the object boundaries a different approach, called DUAL INTERVAL ALIGNMENT (DIA) has been developed and implemented, aligning the geometry of matched features calculating the transformation of single vertices. For every point in one object a suitable neighbor in the corresponding partner object is determined using the criterion of proximity. The alignment approach evaluates the distance between these coordinates. The adaptation is applied based on an interval that determines, in which distance objects are considered as neighbors. These values depend on the type of object and on the map scale. Then, the new position of the vertex is determined as a weighted mean of both original vertices. The basic workflow for the DIA strategy is presented in Goesseln (2005).

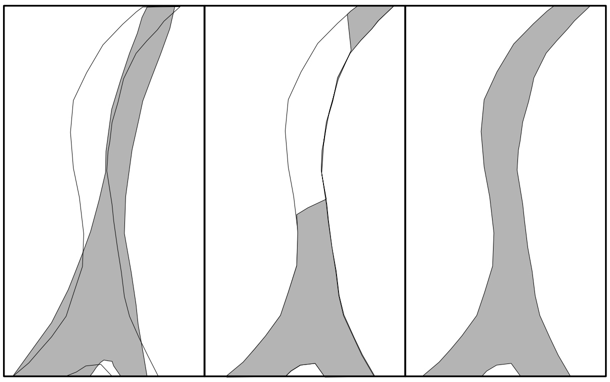

Fig. 1: Left – Aligning a river from a reference (dark border) and a candidate data set (filled), Middle – candidate object misaligned, Right – candidate object aligned with estimating segment orientation

The integration of a weight to the alignment process does not only take different accuracies of geometries into account, but opens this approach to a much wider range of conflation tasks, E.g. one data set is handled as a reference data set and must not be changed: This can be achieved by setting its weight to 1 and the weight of the corresponding partner to 0. In other cases, when two data sets have to be aligned and no single data set can be selected as reference the alignment is performed using the general idea of conflation by aligning two data sets to a new improved data set using equal weights.

Segment orientation

Calculating the distance based on the nearest neighbor enables very good alignment result, but can result in topologic errors. As it can be seen in Fig. 1 a relational

error occurs between two river representations from different data sets.
By using the conformation approach and evaluating the distance with the described interval, the distance criteria fails by aligning the right side of the stream to the left side of its corresponding partner. This requires the integration of an additional criteria. In a first step the orientation of two corresponding polygons is compared. In the following step, the orientation of the polygon segments are calculated for all corresponding objects.
If a point and its corresponding partner are selected using the distance criterion, the direction to each corresponding successor is calculated. If the difference between these directions exceeds a given threshold, the points must not be aligned due to the assumption that they do not represent the same "side" of the real world object.

3.4 Results

The combination of both adaptation approaches delivers very good results, offering the possibility to assess the geometric discrepancies by evaluating the resulting ICP-parameters, and aligning large object groups or partially matching objects using DIA.
The resulting parameters of ICP can be used for the investigation and the evaluation of the influences which were responsible for the geometric discrepancies:

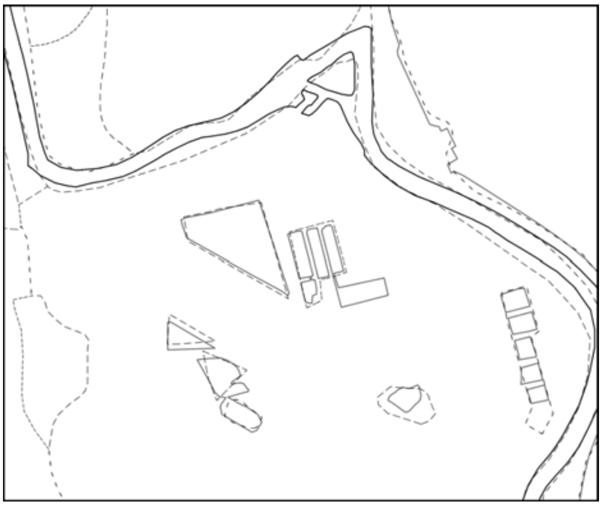

Fig. 2a : Superimposition of candidate (thin, dotted line) and reference data set (dark lines).

An object which can be aligned by just using translations with a small change in scale can be judged as minor error based on digitization or different acquisition methods. A larger scale factor is an indication for changes on the real world objects.
In Fig. 2b the results of the different alignment methods can be seen. The ICP algorithm using an iterative four-parameter transformation is very suitable for the alignment of objects which already have a similar geometry. The alignment parameters which are the results of the ICP algorithm can give a first hint whether the geometric discrepancies are due to different acquisition methods (a., d.) or to changes which occurred to the real world object (c.).

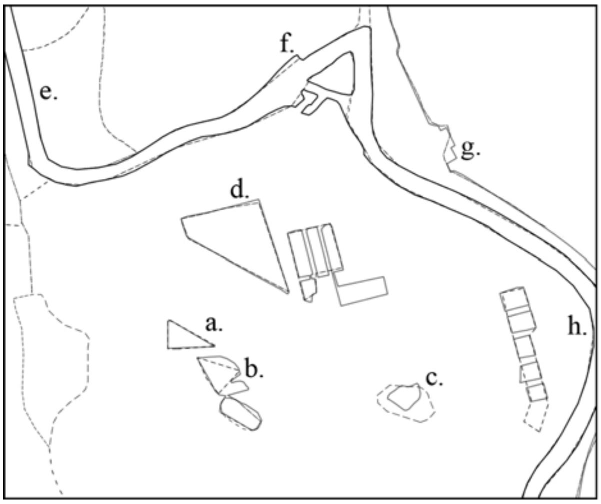

Fig. 2b : Result of the approach, candidate (thin, dotted line) aligned on the reference data set (dark lines).

The resulting scale factor which was calculated for the alignment of object c. was rated as too large and therefore no alignment was performed. Of course changes of the topography can not be discovered by simple evaluation of these parameters. For object b. the algorithm achieved a best fit with four parameters below certain thresholds, but the remaining differences between the geometries still have to be corrected using DIA.

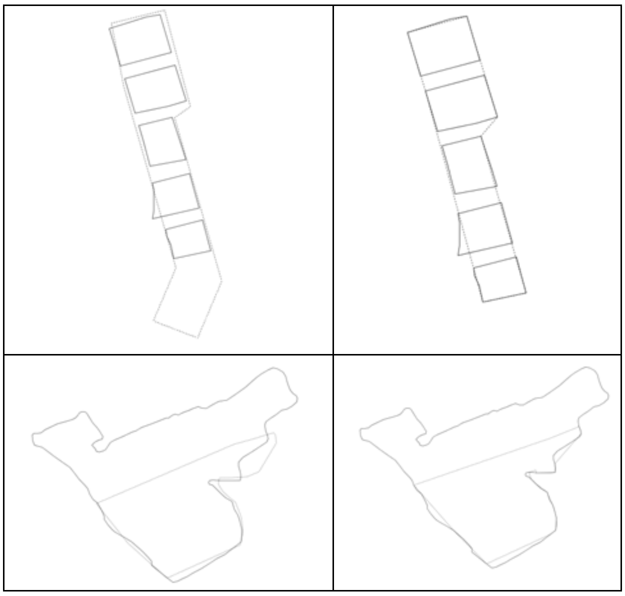

Fig. 3a-d : Alignment of n:m relations (first row), Partial alignment (second row)

The results d to h are enhanced using DIA, but they are still not the expected results, as there are still some differences in the geometries left. As it was said before, the DIA strategy is capable of aligning n:m relations. A sufficient solution for object h (Fig. 2b) can be found by adjusting the applied thresholds to fulfill the special requirements and to derive a satisfying solution (Fig. 6c &d). The importance of circumspectly chosen thresholds can be seen in Fig. 3c & d, the partial alignment produced appealing results and the alignment could be limited to the correct regions.

4. BALANCED VERTEX DISTRIBUTION

4.1 Addition of new vertices

As it was described in chapter 3, the dual interval approach is working directly on the vertices of the object which have to be aligned. Therefore a well adapted vertex-distance is required on the candidate object, which should be integrated. An too large distance between the vertices of the candidate object would result in aligned segments which do not consider higher detailed segments in the reference objects geometry (see Figure 5b, chapter 5). As our ultimate goal is to be able to determine a mapping from one data set to the other and vice versa, this implies that the number of vertices in both objects is the same. Therefore the candidate objects must be preprocessed in advance to ensure a sufficient and similar vertex distance, that is short enough to cover every detail in the reference geometry, but which is not to short for the segment orientation.

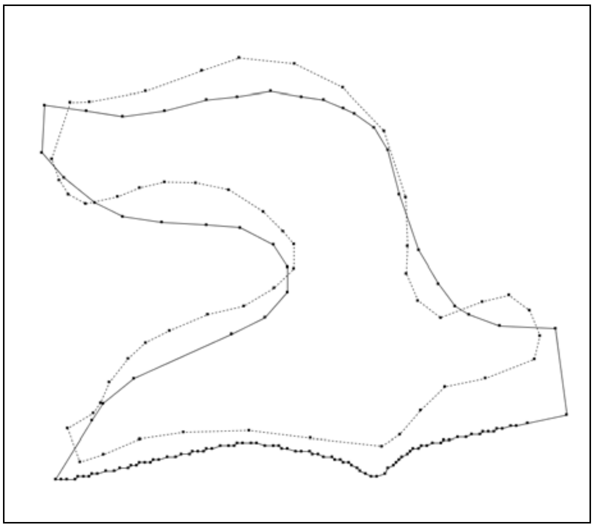

Fig. 4a: Object with very irregular distribution of vertices (dark line), and target object (dotted line).

4.2 Segment based simplification

Fig. 4a shows an example where the distribution of the vertices at the object boundary is very irregular. The segment orientation estimated for every vertex to be aligned - based on the angle between the adjacent line segments – fails if the vertex distance is to small. In Fig. 4a the bottom line of the object was derived by automatic vectorization from aerial images, which ends up in line segments with a length about 40 cm and even less. The line segments, following the boundaries of the pixels from their raster source, lead to angles which do not correspond to the boundary of the reference objects.

To avoid these irregular and unbalanced vertex distributions, a simplification strategy has been developed which takes the line segments of the corresponding reference object into account. This approach is working in two steps.

In the first step the median-length of the segments from the reference object is calculated, by addition of new vertices on the candidate object it is ensured that there will be no segment which is longer than half the length of the median.

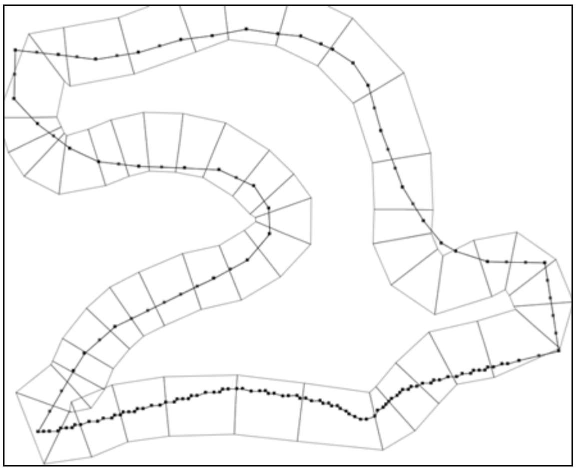

Fig. 4b: Search-spaces and candidate object (superimposed)

In the following step (see Fig. 4b) a buffer area is defined for every segment on the reference object. These buffers are build up using the thresholds from the DIA approach; the usage of the DIA thresholds ensures, that all vertices which will be transformed during the alignment process will be evaluated for line-simplification in the preprocessing. Vertices outside of these buffers will exceed the DIA thresholds and therefore will not be transformed.

Using these buffer areas the corresponding vertices from the candidate object are selected and evaluated to adjust the distribution of the vertices according to the length of the reference segment and the level of details from the candidate object: if the distribution of the vertices is higher and would deliver segments which are to short, they are simplified. If the vertex spacing is lower, then new vertices are inserted. Using this strategy allows to align a candidate to a reference data set – but not necessarily vice versa. Furthermore, the derivation of weighted data set with an averaged geometry requires an intelligent strategy to ensure that both data sets have a well-balanced distribution of vertices.

5. MUTUAL INSERTION OF VERTICES

Following the preprocessing to ensure a well balanced distribution of points, the relations between the vertices from one data set to its corresponding partners are determined. These links are created using the criterion of proximity as it is described in chapter 3, and the segment orientation enhances the topological correctness of the point correlation.

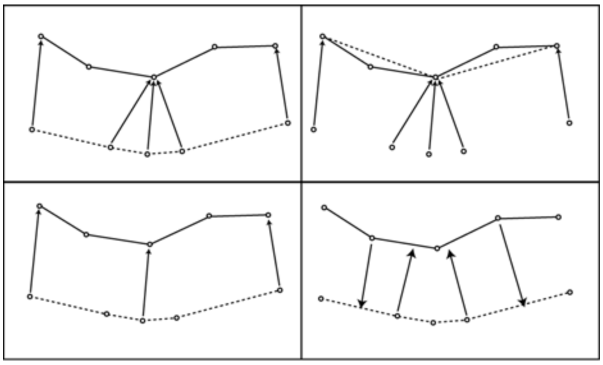

Fig. 5a-d: (First row, left) Relation based on distance evaluation, (right) resulting in an invalid geometry and unrelated vertices. (Second row, left) Selection of shortest neighbor, (right) allocation of new targets.

However, even if a similar object spacing of both data sets is given, there will still be some wrong candidates related using only the proximity criterion. Therefore, more information has to be put into the selection of possible partner vertices to ensure sufficient results. Figure 5 shows a direct/indirect relation strategy that enhances the process of creating relations between the vertices.

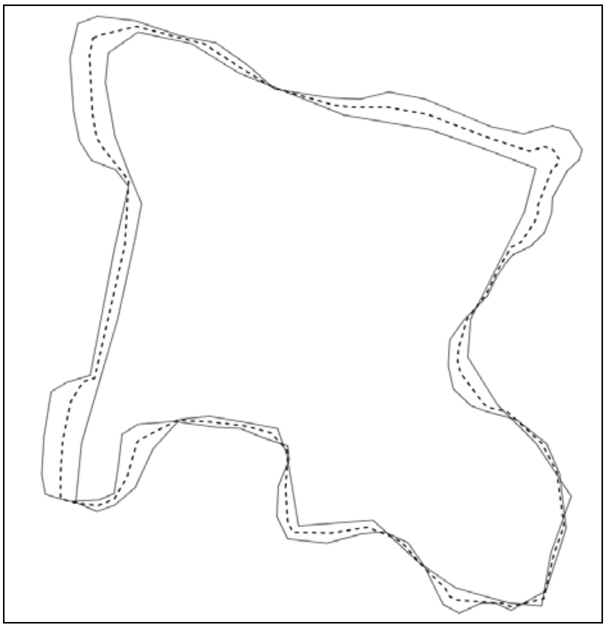

Fig. 6: Derivation of an intermediate geometry by introduction of a weight

In Fig. 5a the resulting relations from the strategy used so far are shown: based on the distance criterion, for every vertex a corresponding target can be revealed, but by assigning more source points to a single target some of the reference points are not assigned a partner, leading to the possibility of an invalid resulting geometry (Fig. 5b). Thus, the selection of the partner vertices has to be refined. One possibility is to include additional criteria like the similarity of object angles. Another is to select only the shortest feasible relation between two points, and remove all other relations (Fig. 5c).

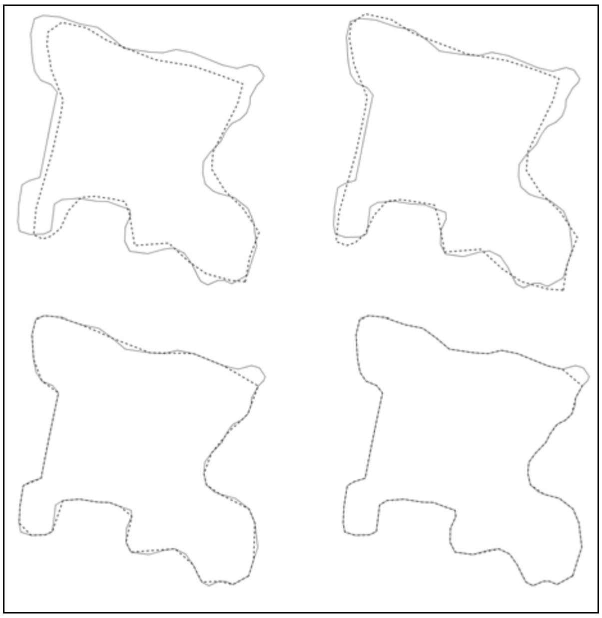

Fig. 7. Comparison of all four stages, from left to right, (first row) Original, ICP, (second row) DIA, DIA 2 with detection of corresponding vertices.

Then, using a direct/indirect strategy for every remaining vertex, the shortest target on the corresponding object will be calculated, and a new vertex will be added on both objects. In Fig. 5d the new point relations can be seen. Assuming a weight had been introduced to derive an averaged geometry (Fig. 6), the new points will be placed using the resulting relations from 5c and 5d. Using this approach, a unique correspondence between all vertices of both data sets to each other is created. and by evaluating the weight information the length of the vector will be calculated.

The visualization of the different stages of the alignment approach are shown in Fig. 7; they show the possibilities of the different strategies. ICP gives a global adaptation of the objects. DIA leads to a local adaptation within the given range thresholds (which, in this case excludes the upper right part of the object). Due to the different spacing of the vertices in the two data sets, however, there are some discrepancies in the boundary. DIA 2, using the adaptive spacing, leads to the desired result of a full correspondence of the two objects. Although the results are already very successful, there are still some problems which have to be solved in the future. If the level of detail in the object geometries differs to much, then false vertex relations can occur which result in an even worse similarity than the unaligned objects (Fig. 8) or even in invalid topologies.

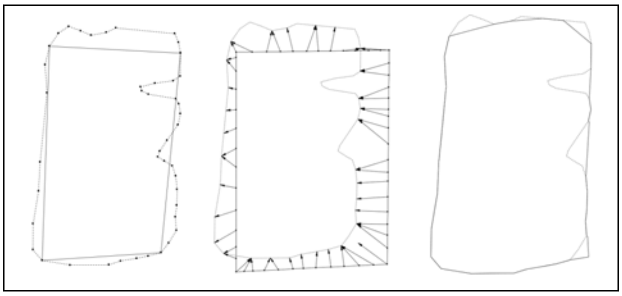

Fig. 8: Poor results alignment results due to insufficient vertex distribution.

6. CONCLUSION

The strategy of using the two alignment procedures yields promising results. As it was stated before the strength of the ICP approach is its fast and very reliable results, with the drawback of not taking local areas into account. DIA at its current stage of development is working vice-versa, it is capable in aligning local differences, but its effectivity fades with the increasing distance of two corresponding objects. Therefore the combination of these strategies is very promising. In a first step the relation set, which consists of a reference and a candidate object is handled with the ICP strategy. The four resulting parameters for the transformation can be used as a first indicator for the plausibility of the alignment process. In the second step the DIA approach in combination with the new preprocessing (vertex addition and segment based simplification) is performed to align the local geometric differences.
The module-based workflow offers the possibility of adopting it to different situations: the integrated weight function allows for the adaptation of one data set to a reference (weight 0 : 1) or can be used to derive a new consolidated data set out of two data sets with various weights assigned (weight a : 1-a).

7. REFERENCES

Besl, P. & McKay, N.,1992. A Method for Registration of 3-D Shapes, Trans. PAMI, Vol. 14(2), pp. 239-256.

Butenuth, M. & Heipke, C. Modelling the integration of hetergeneous vector data and aerial imagery. In: Proceedings of ISPRS Commission IV Joint Workshop, Stuttgart, Germany, September 8-9, 2003.

Cobb, M., Chung, M.J., Miller, V., Foley, H.I., Petry, F.E., and Shaw, K.B., 1998. A Rule-Based Approach for the Conflation of Attributed Vector Data, GeoInformatica, 2(1): p. 7-35, 1998.

Kohonen, T., 1997. Self-Organizing Maps. Springer, 1997.

Kuhn, W., 2003: Semantic Reference Systems, International Journal of Geographical Information Science.

Gösseln, G. v.: A matching approach for the integration, change detection and adaptation of heterogeneous vector data sets, in: Proceedings of 22nd International Cartographic Conference, 9. - 16. July 2005, La Coruña/Spain.

Gösseln, G. v. & Sester, M.: Integration of geoscientific data sets and the german digital map using a matching approach, In: International Archives of Photogrammetry and Remote Sensing, Vol. 35, ISPRS, Istanbul, 2004.

Hettwer, J. & Benning, W. Nachbarschaftstreue Koordinatenberechnung in der Kartenhomogenisierung, Allg. Verm. Nachr. 107:194-197, 2000.

Sester, M., Hild, H. & Fritsch, D., 1998a. Definition of Ground-Control Features for Image Registration using GIS-Data. In: T. Schenk & A. Habib, eds, 'IAPRS', Vol. 32/3, ISPRS Commission III Symposium on Object Recognition and Scene Classification from Multispectral and Multisensor Pixels, Columbus/Ohio, USA, pp. 537-543.

Sester, M., Anders, K-H and V. Walter, 1998b. Linking objects of Different Spatial Data Sets by Integration and Aggregation. GeoInformatica, Volume 2, No 4, pp. 335-358.

Uitermark, H., 2001. Ontology-based geographic data set integration. Doctor thesis, Deventer, Netherlands.

Walter, V. & Fritsch, D,. 1999. Matching Spatial Data sets: a Statistical Approach, International Journal of Geographical Information Science 13(5), 445–473.

Volz, S., 2005: Data-Driven Matching of Geospatial Schemas. In: Cohn and Mark (ed.): Proceedings of the 8th International Conference on Spatial Information Theory: COSIT '05, Ellicottville, New York, September 14-18, 2005.

Walter, V. & Fritsch, D., 1999: 'Matching Spatial Data Sets: a Statistical Approach', International Journal of Geographical Information Science 13(5), 445–473.