

# Integration of heterogeneous geospatial data in a federated database

Matthias Butenuth <sup>a,\*</sup>, Guido v. Gosseln <sup>b</sup>, Michael Tiedge <sup>c</sup>,  
Christian Heipke <sup>a</sup>, Udo Lipeck <sup>c</sup>, Monika Sester <sup>b</sup>

<sup>a</sup> *Institut für Photogrammetrie und GeoInformation, Leibniz Universität Hannover, Germany*

<sup>b</sup> *Institut für Kartographie und Geoinformatik, Leibniz Universität Hannover, Germany*

<sup>c</sup> *Institut für Praktische Informatik, Leibniz Universität Hannover, Germany*

Received 28 June 2006; received in revised form 9 March 2007; accepted 9 April 2007

Available online 1 June 2007

## Abstract

The integration of heterogeneous geospatial data offers possibilities to manually and automatically derive new information, which are not available when using only a single data source. Furthermore, it allows for a consistent representation and the propagation of updates from one data set to the other. However, different acquisition methods, data schemata and updating cycles of the content can lead to discrepancies in geometric and thematic accuracy and correctness which hamper the combined integration. To overcome these difficulties, appropriate methods for the integration and harmonization of data from different sources and of different types are needed. In this paper we describe two generic cases including novel integration algorithms, namely the integration of two heterogeneous vector data sets, and the integration of raster and vector data. Both algorithms are linked to a federated database which allows for automatic object matching and for managing *n:m* relationships. We describe and illustrate our work using vector data from topography and the geosciences, as well as multi-spectral imagery.

© 2007 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

*Keywords:* Integration; Spatial infrastructures; Federated databases; Vector data; Raster data; Matching; Image analysis

## 1. Introduction

### 1.1. Data integration—benefits, requirements and applications

With the advance of more and more sensors and automatic data acquisition tools, the number of available digital data sets is ever increasing. This is especially true for geospatial data, which are acquired by different organizations, e.g. administrations like national mapping and environmental agencies, but also private

companies, e.g. in car navigation. In addition to this diversity of data providers, there is a diversity of data models, data acquisitions schemes, as well as spatial data types: vector data are typically structured in some sense and attributed, whereas images are only composed of grey values and have to be interpreted in order to extract explicit geospatial information.

The diversity of available digital data sources bears the chance of integration in order to exploit the relative benefits of each. These benefits are:

- *Integrated analysis with prior information:* data of one data set can be inspected using also data from the other set.

\* Corresponding author. Tel.: +49 511 7624922; fax: +49 511 7622483.  
E-mail address: [butenuth@jpi.uni-hannover.de](mailto:butenuth@jpi.uni-hannover.de) (M. Butenuth).

- *Reference to common geometry*: linking data sets based on common objects allows to geometrically reference a non-referenced data set.
- *Mutual corrections and refinements of geometry*: if the accuracies of the geometric elements of both data sets are known, they can be taken into consideration to generate a new, enhanced geometry.
- *Mutual enrichment with semantic and geometric properties*: attributes and properties of one data set can be transferred to the other.

Obviously, true integration is much more than just overlaying data in a geographic information system (GIS), as it must make the relations between the individual objects in the different data sets explicit. Technically, it also means more than information fusion, if the original data sets should still be available to be used in their own right. This is a common requirement today, as different agencies are interested in maintaining control over the data they are responsible for and they have the knowledge to maintain the data properly.

The requirements for the realization of a true integration of spatial data sets are as follows:

- Corresponding objects of the different data sets have to be identified, so they can be connected by explicit links.
- The links have to be set up by automatic matching algorithms which take semantic and geometric similarities into account.
- It should be possible to match and link arbitrary data types, for instance:
  - Vector with vector data,
  - Raster with vector data,
  - Raster with raster data,
  - Two-dimensional and height data.
- Different thematic and geometric granularities have to be considered, as well as different accuracies of the data sets.
- The matching and linking task is to be carried out in a federated database environment.

Beside the general benefits of data integration, there are a lot of practical applications of integration. One is the verification and update of data sets: in order to check the currency and correctness of a data set, a second or third data set can be used to check the information. The task can be extended to provide also update capability: whenever current information is available in one data set, it can be employed to update the other sets, based on the known relations between the data sets and the known link structure between corresponding objects. Further-

more, integration can be used to provide prior information for a dedicated analysis using one data set: for example if a road network has to be updated using aerial imagery, information from existing road data can be used to partition the space and identify potential areas for new roads. Similarly, land use classifications from imagery can be used as input for a more detailed inspection of dedicated areas of interest. Also, data from one information source (e.g. cadastre) can be used to enrich data from another one (e.g. topography). In this way, topographic road data can be enriched with address information, which is used as an indirect georeference in many other databases.

The remainder of the paper is structured as follows: After briefly describing the application scenario of our work and the data sets employed in our study, the next section gives an overview of the state of the art in data integration. Afterwards, the developed architecture for database supported integration is described. We then present our methods for vector/vector and raster/vector data integration. Subsequently, results demonstrating the potential of the proposed solution are presented, and finally conclusions are drawn and further work is discussed.

### 1.2. Application scenario

Our research aims at developing concepts for the integration of geoscientific data sets. We have developed different ideas and implemented them in a prototype consisting of three modules: a vector/vector integration module, a raster/vector integration module, and a database module as a base for our work.

Describing and representing objects of the same physical reality but with different thematic focus will naturally lead to different data sets. Due to the same spatial context, similar objects usually exist in the various data sets. This is true for the data sets under investigation in our case, namely topographic data and geological data. Although geology and topography have not too much in common, they still share some basic objects, e.g. water bodies. When overlaying the two data sets, these common objects should obviously coincide geometrically. However, this is not exactly true. The reasons are manifold: one is, that different data acquisition methods may have been used, the other – more important – one is that one of the data sets may be out of date and may need to be updated. In the course of the project, we developed a method to first bring semantically equivalent data into correspondence using matching techniques. In order to decide, if a match is correct, the geometric deviations between the objects are

evaluated. The applied geometric alignment is able to compensate small geometric discrepancies. If these discrepancies become too large a change in the data themselves is assumed and no geometric adaptation is conducted.

The second module deals with the integration of images and vector data for the purpose of detecting field boundaries. One application area are the geosciences, for example the derivation of potential wind erosion risk fields, which can be generated with field boundaries and additional input information about the prevailing wind direction, wind shelters and soil parameters. Another area is the agricultural sector, where information about field geometry is important for tasks such as precision farming or the monitoring and control of subsidies. This integration task is an example for the enrichment of the vector data by integration: On the one hand, there is imagery, that – implicitly – contains indications for field boundaries; on the other hand topographic data sets are available that contain objects, which stand in certain relations to field boundaries, e.g. borders of vegetation classes, roads, railways and rivers all represent candidates for field boundaries. The integration task is to adequately model these relationships and integrate them into an image interpretation strategy.

Interestingly, in our modules we used methods for geometric alignment—in the one case the alignment to an image, in the other the alignment to another vector data set. The underlying concepts and methods, however, are similar. This is one reason why we have implemented both modules in the framework of a federated database. This database is firstly used to store all the data. Secondly, it provides mechanisms for data pre-processing, like topology building, that are optimized on the database. Thirdly, it models, represents and maintains the data structure into which many-to-many relationships between corresponding objects are embedded and provides interfaces to access the individual data sets and the linked partners. Finally, it controls the application processes in order to keep track of the housekeeping data.

### 1.3. Used data sets

The vector data sets used in this project are the German topographic data set (ATKIS DLMBasis), the geological map (GK) and the soil science map (BK), all at a scale of 1:25000. The used raster data are high resolution multi-spectral (RGB+IR) pan-sharpened satellite images (IKONOS). We use a ground resolution of 2 m, because in our experiments this resolution has given better results than the original 1 m resolution.

Simple superimposition of different data sets already reveals some differences. For ATKIS the topography is the main thematic focus, for the geoscientific maps it is geology and soil science. Geological and soil science maps have been produced by superimposing the result of geological research (in many cases punctual bore hole measurements) with topographic maps available at the time. Based on the punctual information area objects have been derived using interpolation methods and geoscientific models. On the one hand, both geoscientific data sets are thus georeferenced to the topography, on the other hand, they also contain common objects (e.g. water bodies). While the geological content of these data sets remains valid for decades, the topographic information changes more quickly. Nevertheless, in the geoscientific maps topographic updates are usually not integrated unless new geological information has to be inserted. In addition, the employed data models are different: Geological and soil science maps are single-layered data sets, they consist only of polygons with attribute tables for the representation of thematic and topographic content. ATKIS has a multi-layered data structure with point, line and polygon objects and corresponding attribute tables. As a result, the content of the different vector data sets may differ considerably, and a simple overlay is not adequate for data integration.

In raster data the information is implicitly represented using a regular matrix of grey values rather than explicitly in vector geometry. The task is then to transform the raster into a vector representation rather than matching, prior to setting up the links for data integration. This transformation can be facilitated by using existing vector data, e.g. as a focus of attention. In our work we extract field boundaries from high resolution multi-spectral satellite imagery with the help of ATKIS data as prior information.

## 2. State of the art of geospatial data integration

The integration of vector data sets presented in this paper is based on the idea of comparing two data sets and finding corresponding objects. In the most general case, this task can be tackled using relational matching, see e.g. [Walter and Fritsch \(1999\)](#). [Beeri et al. \(2005\)](#) show on databases how corresponding objects can be found when several point data sets have to be integrated. When data sets stem from data of different spatial resolution, also generalization effects have to be taken into account when selecting possible matching candidates ([Mustière, 2006](#)). Due to the complexity of the integration problem it is very difficult to solve this task

in a closed system. Therefore the development of a strategy based on different modules, which can be adjusted individually according to the requirements was proposed (Yuan and Tao, 1999). The usage of a component-based processing chain allows for the selection of different algorithms for different subtasks to build up the most suitable workflow for a particular problem. In order to geometrically align data sets of different origin, map conflation (Saalfeld, 1988) or rubber sheeting (Doytsher, 2000) is being applied. The integration of cadastral data with simultaneous consideration of geometric constraints has been presented by Hettwer and Benning (2000). The integration of street data sets has been shown (Volz, 2006). While all of these approaches yield successful results in their respective domains, to our knowledge none of them has dealt with geoscientific data.

Data registration is also a common goal when integrating raster and vector data (Sester et al., 1998; Dowman, 1998). Since image data contain geospatial information only implicitly, the first step is to make this information explicit, i.e. to extract the objects of interest from the images. For reasons of speed and cost, this step should be automated as much as possible using image analysis methods. For this task, in many cases existing vector data are used as prior knowledge to support object extraction. An integrated modelling of the objects of interest and the surrounding scene exploiting context relations between different objects leads to an overall and holistic description (Baltasvias, 2004). In the past, many approaches have dealt with the automatic extraction and interpretation of man-made objects (e.g.

Mayer, 1999). Similarly, the extraction of trees has been investigated, see Brandtberg and Walter (1998) for an example suitable for forestry. Zhang et al. (2006) have presented an approach for the extraction of wind erosion obstacles in the open landscape. In contrary, the extraction of field boundaries has not received much attention so far: a first approach to update and refine topologically correct field boundaries by fusing raster images and vector map data was presented in Löcherbach (1994). The author focuses on the reconstruction of the geometry and features of land use units, however, the acquisition of new field boundaries is not discussed. In Torre and Radeva (2000) a so-called region competition approach is described, which extracts field boundaries from aerial images using a combination of region growing techniques and snakes. To initialize the process, seed regions have to be defined manually, which is a time and cost-intensive procedure.

As mentioned before, the different data sets usually reside in heterogeneous databases, preventing their integrated use. In order to overcome this limitation, multi-database architectures were discussed for loose coupling, followed by so called federated databases to support closer coupling (Conrad, 1997). Federated databases allow for integrating heterogeneous databases via a global schema, and they provide a unified database interface for global applications. Local applications remain unchanged, as they still access the databases via the local schemata. For database schema integration a broad spectrum of methods has been investigated (Batini et al., 1986), nevertheless identifying objects is typically restricted to one-to-one-relationships. In the context of

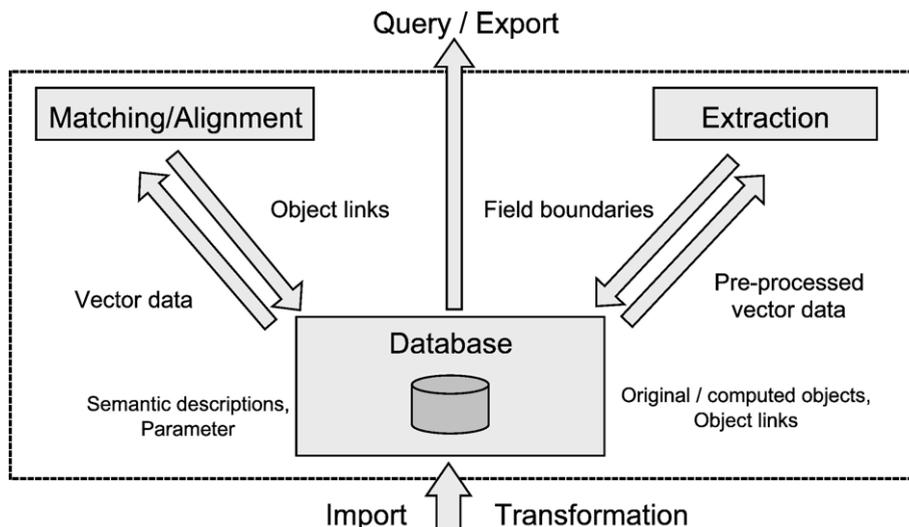


Fig. 1. Architecture overview.

geospatial integration more sophisticated methods are needed to incorporate complex correspondences between objects (many-to-many-relationships), which have not been considered in federated databases up to now. Whereas there are a lot of overview articles of spatial databases (e.g. Rigeaux et al., 2002), federated spatial databases – with the exception of Devogele et al. (1998) and Laurini (1998) – are hardly investigated.

### 3. Integration architecture

In this section, the developed architecture and modelling concepts of the database integration are explained; they provide an organizational framework for the approaches of geospatial data integration given in Section 4. The database architecture is designed to pre-process inputs and to store and export results of the vector/vector and the raster/vector integration steps. Fig. 1 gives a simplified overview of the integration architecture with respect to the interaction between the database and the object matching and extraction processes.

The underlying database architecture is chosen according to the paradigm of federated databases (Conrad, 1997), as it gives a close coupling and at the same time keeps the databases autonomous. Fig. 2 shows this kind of architecture.

In this way, the matching and extraction processes are given an integrated view to the different databases via a global database schema (global applications). Nevertheless, the access to the local individual databases is still possible. For our purpose the known

architecture of a federated database is extended to handle geospatial data and the linkage between multiple data sets. In order to select certain objects satisfying given semantic criteria, we have defined mappings to harmonize the object classes of the different data sets on the schema level. For the object level, we have developed an advanced matching process for identifying objects and a data structure for maintaining many-to-many links between corresponding objects.

The federation service requires an “integration database” (IDB, cf. Section 3.3) on its own to maintain imports and descriptions of the involved data sets (component databases, CDB), and to maintain integrated data sets according to the advance of the integration process: first, it incorporates qualified links between individual objects as the result of the matching process, later further results such as geometrically adjusted and newly extracted objects are added, finally the federated database contains semantically enriched data (i.e. semantic selections).

#### 3.1. Schema adaptation

To make the structurally different data sets accessible to the federation service a generic but flexible export schema was designed based on experience with geospatial data sets containing topographic objects with respect to object-relational databases (Kleiner et al., 2000; Mantel and Lipeck, 2004). The schema contains all objects, object classes, attribute types and attribute values, each of them in one entity type (or table in the relational DBMS). Fig. 3 shows the schema for

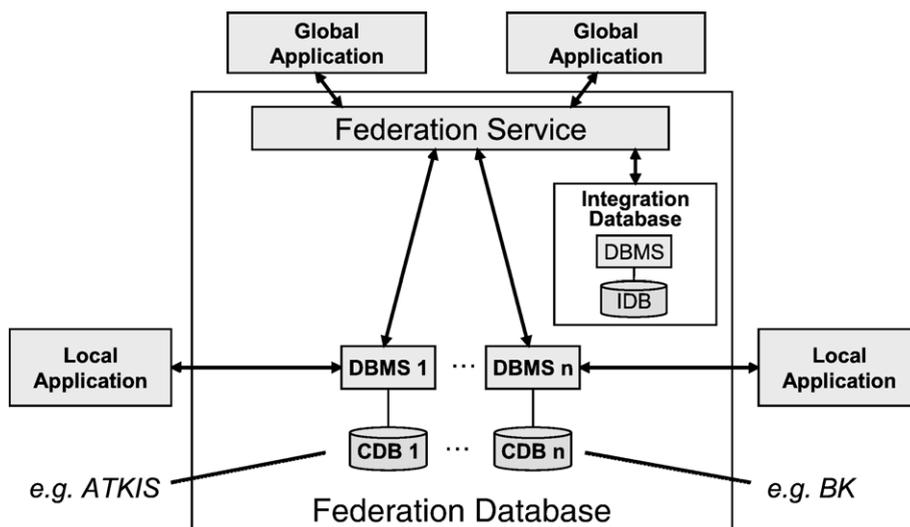


Fig. 2. Architecture of a federated database.

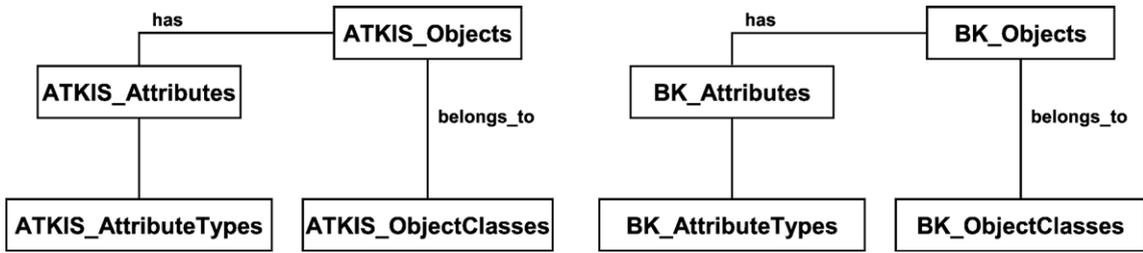


Fig. 3. Export schema for ATKIS data and soil science map (BK).

ATKIS data and the corresponding export view of the soil science map (BK). The other geoscientific data have isomorphic export views, i.e. they have application specific attribute types and object classes according to their own representation model.

A geospatial object of entity type ATKIS\_Objects, e.g. a road, has several entries of type ATKIS\_Attributes, namely (attribute, value)—pairs like e.g. (width, 10 m). The corresponding type of the attributes or the classification of the geospatial objects can be found in the collections ATKIS\_AttributeTypes and ATKIS\_ObjectClasses.

### 3.2. Object linking and semantic selection

Given the structural adaptation of the different data sets, the federated database can be enabled to incorporate correspondences through so called links. In the context of federated databases the process of identifying objects is usually restricted to one-to-one correspondences often based on simple attribute matching (like, e.g. the ISBN of books). Identifying geospatial objects, however, does not only involve simple one-to-one-relationships, as real-world objects can be represented differently in different maps. Fig. 4 shows an instance of three and two objects, respectively, e.g. a section of a water body segmented in two different ways. To represent these many-to-many correspondences, the

database stores attributed one-to-one links between aggregated objects (denoted by dashed lines in Fig. 4).

In order to provide the applications with a model independent and uniform method to access objects with respect to thematic attributes, a mechanism for the semantic description of geospatial objects was developed. It characterizes comparable object sets for the matching process and simplifies object selection for the extraction process. To fulfil these requirements the architecture of federated databases had to be expanded to unify the handling of semantic descriptions. Fig. 5 shows two simplified semantic selections of topographic objects, namely of open landscape and a partitioning network.

Semantic object selection is defined in three stages, see Fig. 5. (1) Coarse semantic classification is achieved through the references to object classes given by the export views. (2) A more precise characterization is provided through the specification of object attributes (fine object classes), i.e. the coarse selection via object classes is restricted by attribute conditions. For instance, road objects in ATKIS appear as both one-dimensional and two-dimensional objects due to modelling rules. In order to build a partitioning network only the one-dimensional road objects are needed. (3) Finally, fine object classes are merged to class sets, which provide semantic selection for the global applications, independent of the semantic

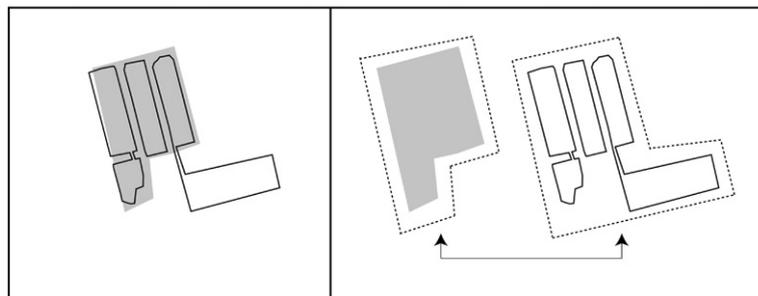


Fig. 4. Realization of a many-to-many-relationship ATKIS (solid line), GK (filled); left: representation in GIS application; right: link between object aggregations.

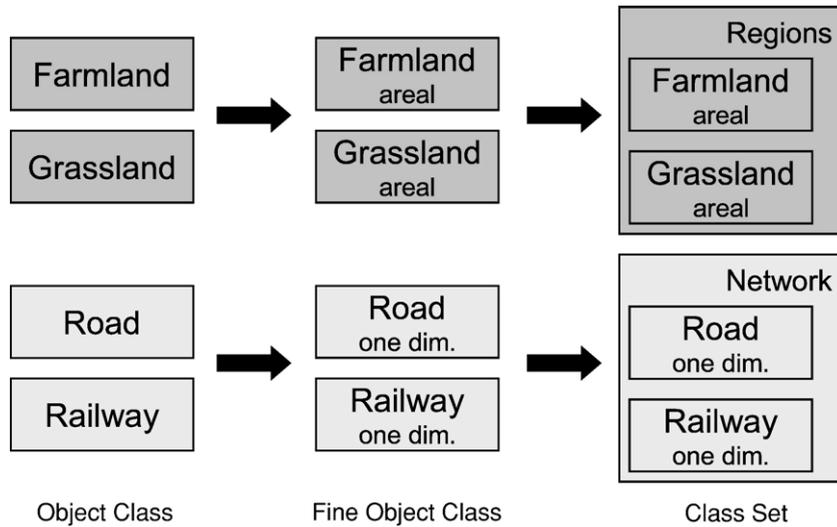


Fig. 5. Semantic selections for regions and networks.

specifications of the original data set. Class harmonization based on class defining attributes is achieved by connecting conforming semantic selections (e.g. regions and network of different data set models).

It is necessary to provide this manually defined semantic description for any representation model only once via an interactive graphical user interface, independent of the number of instances (component databases) of this particular model. The matching and extraction processes only reference these semantic

selections and the federation service provides the requested objects automatically by internally generating the necessary SQL queries.

### 3.3. Integrated schema

Fig. 6 summarizes the schema architecture of the integration database. The component databases are the transformed geospatial data sets according to the export views described above. The term “Objects” stands

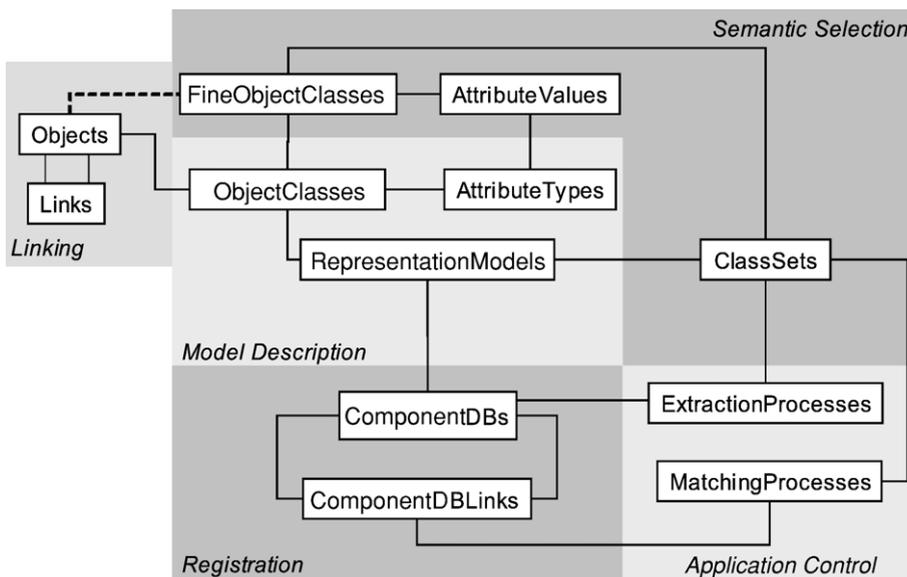


Fig. 6. Overview of the integrated schema.

for all objects of the integration database, including adjusted and extracted geometric objects. The different parts of Fig. 6 show that the federation service supports the following tasks:

1. The *model description* provides a characterization of object classes and attribute types of a certain data set representation model, e.g. topographic, geological and soil science map (cf. Section 3.1.).
2. The *registration* is responsible for registering the component databases according to a certain representation model.
3. The *semantic selection* is supported as described in the previous section.
4. The *application control* stores metadata used for extraction and matching processes, in particular about the utilized semantic selections, links between the involved component databases, advance states, and parameters (e.g. tolerance values).
5. The *linking* of objects from different data sets (i.e. many-to-many-relationships) realizes the integration on data level as a result of the matching process (cf. Section 3.2).

#### 4. Methods of data integration

In this section the developed approaches to the vector/vector and raster/vector integration are described. First, the integration of heterogeneous vector data sets is presented using the matching of the soil science and the geological map as an example. Afterwards, the integration of raster and vector data is highlighted by extracting field boundaries from imagery using vector data from GIS as prior knowledge.

##### 4.1. Vector data integration

The goal of this approach is to match and align corresponding geometries by either determining an intermediate one, or by taking one as a “master” geometry, e.g. due to its higher accuracy or more recent acquisition. If the geometric discrepancies between the data sets are too large, they cannot be compensated by the developed mechanisms. This is an indication that major object changes have occurred and an update of one of the corresponding objects is needed. In all other cases, after matching corresponding individual objects (and also the areas between them, including objects without a partner in the other data set), have to be aligned geometrically using a rubber sheeting algorithm. Thus the approach consists of the three steps: (1) matching, (2) geometric alignment, (3) rubber sheeting.

##### 4.1.1. Creation of links between corresponding objects

In the beginning of the integration process the semantic content of all data sets has to be compared and corresponding object classes in both data sets are identified and selected (in our case water bodies). An area-based matching process is used for the creation of links between object candidates. This matching strategy first tests whether objects from different data sets cover the same area using intersection. In case they overlap, they are stored as potential matching candidates. By using unambiguous object identifiers combinations of up to  $n:m$  can be created. In order to limit the number of correspondences only similar objects are accepted as matching pairs. Similarity is determined using the set operation of symmetric difference (i.e. the union of both geometries minus their intersection), as well as the respective azimuth histograms of the two polygons. These combinations are stored as links in the federated database. Problems during matching occur, if the data sets stem from different aggregations levels. For example, a group of water objects, like a group of ponds, can either be represented as a group of individual objects, or it can be aggregated to fewer or a single generalized object (see Fig. 4). Furthermore, objects can be present in one data set, while they are missing in the other one. All these considerations lead to the following relation cardinalities: 1:0, 1:1, 1: $n$ , and  $n:m$ . After the corresponding relations have been identified, so called *relation sets* are generated through aggregation, which can be handled as 1:1 relations (v. Gösseln and Sester, 2004). Although a relation set is handled as a 1:1 relation in terms of data management, the geometric alignment will take not only a geometric envelope, but also each single element into account.

##### 4.1.2. Geometric alignment of corresponding objects

As the objects in the above mentioned *selection groups* from all data sets represent the same real world objects, they resemble each other in shape and position. Nevertheless, their actual boundaries may differ. Before deciding whether these discrepancies are indications for object changes or are only results of the different data acquisition methods, a geometric alignment has to be carried out, leading to a better geometric correspondence of the objects. For these alignment processes measures and thresholds are required which allow the reduction of discrepancies due to map creation. Revealing discrepancies that are due to object changes happening between the different times of data acquisition is still possible. Two different methods have been implemented, the well known ICP (iterative closed point) algorithm, and a new approach called DIA (dual

interval alignment), both are described in the following section.

**4.1.2.1. Iterative closest point (ICP) algorithm.** The ICP algorithm (Besl and McKay, 1992) has been implemented here using a 4-parameter Helmert-transformation with position, scale and orientation as unknowns. As a result the ICP delivers the best fit between the objects using the given transformation. Evaluation of the transformation parameters allows for classifying and characterizing the quality of the matching: if the registration of the data sets is good, the scale parameter should be close to 1 and rotation and translation should be close to 0. A scale factor different from 1 can indicate differences between two objects that are not based on map creation, but on a change on the real world object, that occurred between the different times of data acquisition (v. Gösseln and Sester, 2004). The result of this transformation is stored as a set of displacement vectors in each point of the object geometry, which it is required in a subsequent step, in which the neighbourhood of the transformed objects is aligned (see Section 4.1.3). ICP is used for a coarse adaptation of the two geometries using a rigid transformation. In order to completely align them, the subsequent DIA-algorithm has been developed, which allows for a distance-dependent adaptation.

**4.1.2.2. Dual interval alignment (DIA).** In order to compensate for local discrepancies in the object boundaries an approach, called Dual interval alignment (DIA) has been developed, aligning the geometry of matched features by calculating the translation of single vertices. The basic idea is to create a unique one-to-one-relation between the vertices in both objects. Based on these links, the adaptation can be calculated. Assuming that two objects have been selected as corresponding, and one is labelled as A and the second one is labelled as B. For every point  $P_A$  in object A a suitable neighbour  $P_B$  in the corresponding object B is determined using different similarity criteria. These criteria are the Euclidean distance  $d$  and the difference  $\gamma$  of the angles  $\alpha$  and  $\beta$  shown in Fig. 7. For  $d$  the ideal value is 0, which means that  $P_A$  and  $P_B$  are identical. A worse case is a value of  $d$  that is equal or larger than the Hausdorff distance between A and B. The ideal value for  $\gamma$  is also 0 and the worst case is a difference close to  $2\pi$ . These criteria are used to qualify each point-to-point relation. For every point in object A a best corresponding point in B is estimated. In the case of 1:n relations, only the one with the highest similarity value is kept, the other relations are deleted.

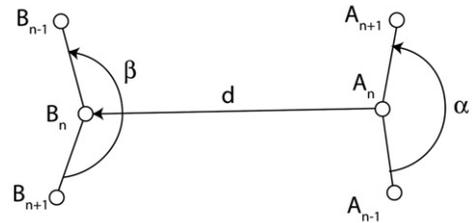


Fig. 7. Derivation of geometric criteria between corresponding points.

This step is performed from object A to object B and vice versa. Subsequently, for every point which has no partner point assigned to it, a suitable target in the corresponding object is interpolated and a point-to-point relation is established. This requires topologic changes in the data set, but it allows for an alignment without gaps or sliver-polygons.

The next step is the alignment of object A as such to the corresponding object B. DIA contains different alignment methods. The first one follows the idea of aligning object A to object B, the latter being considered as reference object by moving every point in A according to the point-to-point relation to its corresponding point in B.

While this is an appropriate solution for aligning one object to a reference set, another idea seemed to be very promising, namely the detection of changes between data sets by evaluating the distance between corresponding points. First, a maximum allowable threshold  $t1$  is manually defined according to map scale and thematic requirements. Points with larger distances are not aligned. By setting this threshold to a distance larger than the Hausdorff distance the complete alignment of one object to a reference object is feasible. However, using a only a single threshold, the alignment process is a step function, which may yield visually unacceptable results. Thus, a second threshold  $t2 > t1$  is introduced, in order to allow for a transition range where between full adaptation and no adaptation. The application of  $t1$  and  $t2$  is shown in Eq. (1), where  $A_{old}$  are the original coordinates of the point in A and  $A_{new}$  describes its aligned position, and  $f(d)$  is a user selected function yielding a value  $< 1$ .

$$(x,y)_{A_{new}} = \begin{cases} (x,y)_{A_{old}} + (\pm \Delta_{AB}(x,y)) \cdot p & ,0 < d < t1 \\ (x,y)_{A_{old}} + (\pm \Delta_{AB}(x,y)) \cdot f(d) \cdot p & ,t1 < d < t2 \\ (x,y)_{A_{old}} & ,d > t2 \end{cases}$$

$$\Delta_{AB}(x,y) = (x,y)_B - (x,y)_{A_{old}} \quad (1)$$

Using this approach the distance range can be separated in three areas. Points with a distance  $d$

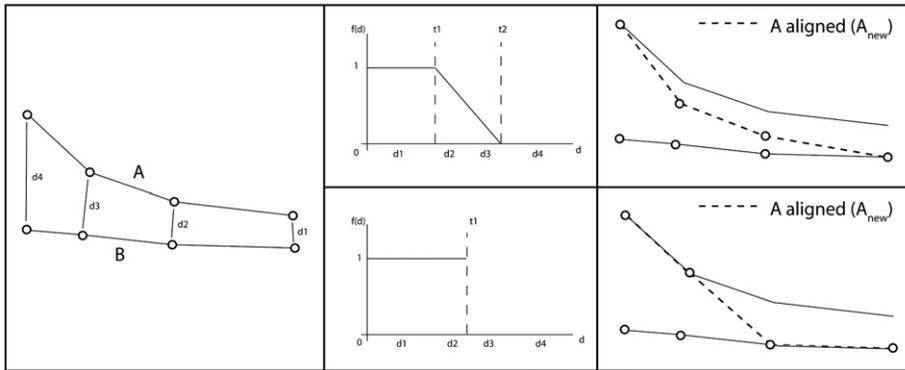


Fig. 8. Results of DIA—left: original situation; centre: two threshold functions  $f(d)$ ; right: result with two thresholds (upper) and with one threshold (lower).

between 0 and  $t_1$  are fully aligned, for  $d$  between  $t_1$  and  $t_2$  the alignment is adjusted using a function ( $f(d)$ ). Points with a distance larger  $t_2$  are not aligned. The effect of the application of one and two thresholds and the function  $f(d)$  can be seen in Fig. 8.

The DIA-algorithm not only allows the adaptation of one data set to the second (considered as master data set), but also to generate an intermediate geometry, taking the relative weight of both data sets into account. This leads to the calculation of a weighted geometry. As described before the allocation of the points between object A and B is unambiguous and therefore a weighted geometry can be derived by multiplication of the translation vector with a defined scalar. To derive a weighted geometry between objects A and B, each distance is multiplied with a constant scalar  $p$  between 0 and 1 (see Eq. (1)). For example, if  $p$  is set to 0.5 the result is the mean geometry. The effect of  $p$  can be seen in Fig. 9.

#### 4.1.3. Neighbourhood transformation using rubber-sheeting

The alignment of individual corresponding objects results in gaps, overlaps or inconsistencies, if the remaining objects of the data set are left unaligned. Therefore, the

neighbourhood of the aligned objects must also be transformed appropriately. We apply a simple rubber sheeting algorithm that uses the individual displacement vectors derived with ICP and DIA: All vectors are used to generate a vector field as a basis of the neighbourhood transformation. Subsequently, the translation of every vertex is derived using a distance weighted interpolation.

#### 4.2. Integration of raster and vector data

The integration of raster and vector data is demonstrated using the extraction of field boundaries from multi-spectral imagery (RGB+IR) exploiting GIS data as prior knowledge. This task is both, scientifically challenging due to the variable nature of the field boundaries and practically relevant due to an increasing need for such data in many applications. First, the integrated modelling and the derived strategy are described, followed by the presentation of fully automatic methods to extract the field boundaries.

##### 4.2.1. Model and strategy

In object extraction from images a semantic model is used to structure the available knowledge in terms of

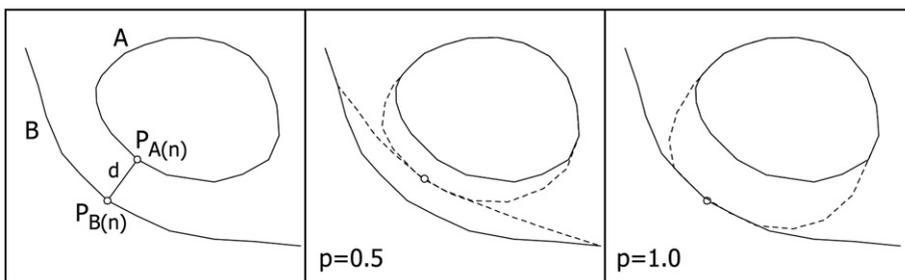


Fig. 9. Example for the influence of weight  $p$ ; left: original situation between objects A and B; centre: alignment between A and B using weight  $p=0.5$ ; right: object A aligned to reference object B. The results are depicted as dashed lines.

objects, their properties – especially visual properties – and relations to each other. The model is differentiated in an object layer, a geometric and material layer, and an image layer (cf. Fig. 10), see also Butenuth and Heipke (2005). It is based on the assumption, that the used images are captured in summer, when the vegetation is in an advanced state of growth. The use of vector data as prior knowledge plays an important role, which is represented in the semantic model with an additional GIS-layer (ATKIS BasisDLM, cf. Section 1.3): Field boundaries are exclusively located in the open landscape, thus, further investigations are focused on this area. As the open landscape is not modelled as a separate object in the ATKIS BasisDLM, all areas are selected, which are not labelled settlement, forest or water body. Additionally, objects of type road, railway, river, tree row and hedge are introduced in the semantic model as field boundaries with a direct relation from the GIS layer to the real world (i.e. a road is a field boundary). This simplification of the boundary extraction problem assumes that the GIS data are correct. Modelling of these geospatial objects in the geometry and material layer together with the image layer is not of interest in our context, because they are already available in the ATKIS BasisDLM and do not have to be extracted from

the imagery; thus, the corresponding parts are represented with dashed lines in Fig. 10. In this way, the model allows for the extraction of additional objects which are not yet included in the GIS database.

In the semantic model the object to be extracted, the *field*, is divided into *field boundary* and *field area* in order to allow for a differentiation in the other layers. The field boundary is a 2D elongated vegetation boundary, which is modelled as a straight line or edge in the image. The field area is a 2D vegetation region, which is a homogeneous region with a high NDVI (Normalized Difference Vegetation Index) value.

The general strategy for the extraction of field boundaries is derived from the modelled characteristics of the fields and their surrounding boundaries taking into account the realization of an automatic processing flow. Imagery and GIS data are used as input data to initialize the process. First, the open landscape is derived from the available GIS data. In addition, within the open landscape, so called regions of interest are selected using the roads, railways, rivers, tree rows and hedges as borderlines (cf. Section 4.2.2). The borders of these regions of interest are considered as fixed field boundaries. The following image analysis methods are then focused to field boundaries within these regions of interest.

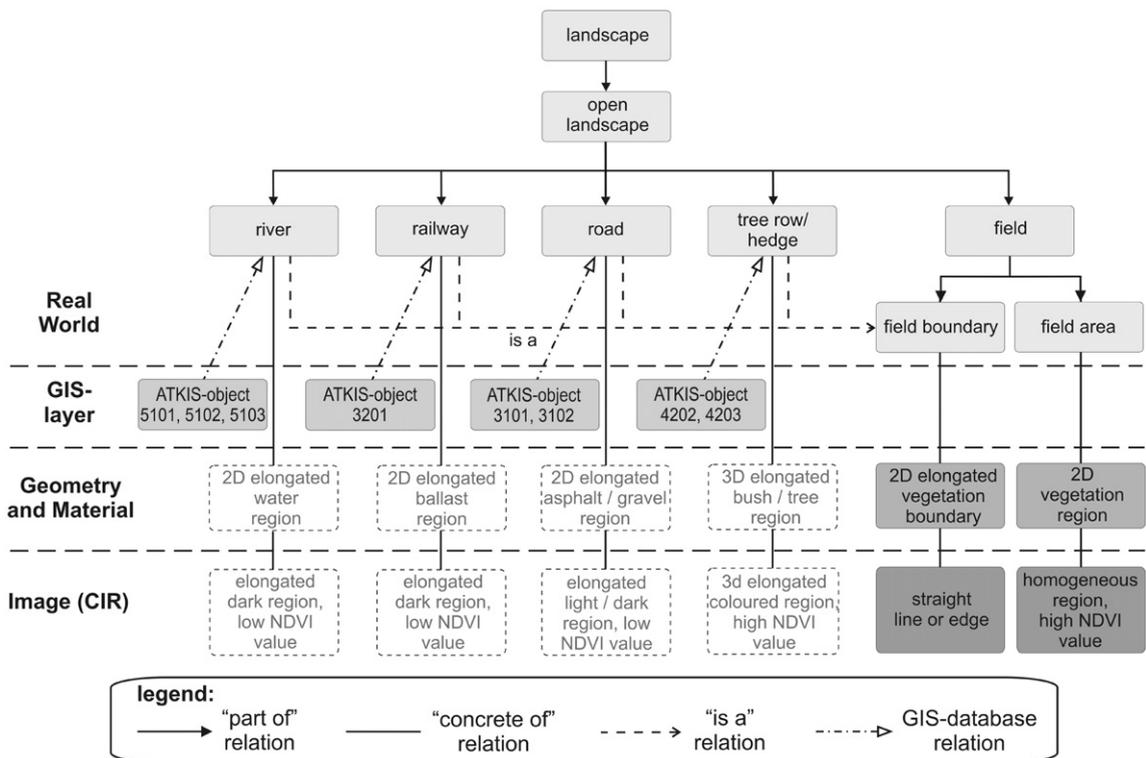


Fig. 10. The semantic model.

The main approach for extracting field boundaries within the regions of interest is divided into two parts: First, a segmentation is carried out in a coarse scale ignoring small disturbing structures and thus exploiting the relative homogeneity of the vegetation within each field. The aim is to obtain a topologically correct result, even if the geometric correctness is not very high. Subsequently, network snakes are used to improve the preliminary results. These two steps are described in detail in Section 4.2.3.

#### 4.2.2. Preparation of GIS data

As described in the previous section the regions of interest are primarily derived from roads, rivers and railways available in the GIS data set, as far as these objects are located in the open landscape. Rows of trees and hedges are also used, taking into account that they often do not extend to the very boundary of a field, and thus form so called undershoots in a topological sense (Ubeda and Egenhofer, 1997). Up to a certain threshold these objects are extended to the next linear object.

These linear objects generate networks and thus partition the open landscape (cf. Fig. 11a). In general, a region of interest contains multiple ATKIS area objects. In order to detect only the borderlines of the regions of interest a topological data model is generated in the database (Egenhofer et al., 1989), which consists of an embedded graph structure containing the open landscape ATKIS objects and the linear objects mentioned above (cf. Fig. 1, pre-processing of vector data). The removal of all edges from the topological data model, which are not borders of the regions of interest, implies merging of certain adjacent ATKIS objects (Fig. 11b).

#### 4.2.3. Extraction of field boundaries

The extraction of field boundaries starts with a segmentation within each region of interest exploiting the modelled characteristics of each field (cf. Section 4.2.1).

A buffer containing the border area of each region is masked out due to disturbing heterogeneities, which are typical for field borders and violate the modelled relative homogeneity within a field. A multi-channel region growing is carried out using all available channels with a resolution of few meters. Neighbouring pixels are aggregated into one and the same field region, if the difference in colour does not exceed a predefined threshold.

If the appearance of neighbouring fields in the image is very similar, field boundaries may be missed. In order to overcome this problem, the standard deviation of the grey values within a small mask is computed with the assumption that high values typically belong to field boundaries. Lines are extracted from the standard deviation image within larger field regions, these lines are then evaluated in terms of length and straightness. Positively evaluated collinear lines are linked and extended to the borders of the regions. Subsequently, they are used to split the initially generated field regions into smaller segments. This segmentation typically leads to a topologically correct but geometrically inaccurate graph representing preliminary field boundaries.

To improve the geometric accuracy of the preliminary field boundaries while maintaining the topology we use so called *network snakes* (Butenuth, 2006). Parametric active contours (snakes) were originally developed by Kass et al. (1988) as an iterative mid level image analysis algorithm, which combines geometric constraints with the extraction of low level features from images. The concept of snakes is shortly summarized here (see Eqs. (2)–(10)) in order to provide a basis for the enhancements concerning network snakes. A traditional snake is a parametric curve (Kass et al., 1988)

$$v(s,t) = (x(s,t),y(s,t)), \quad (2)$$

where  $s$  is the arc length,  $t$  the time or iteration number, and  $x$  and  $y$  are the image coordinates of

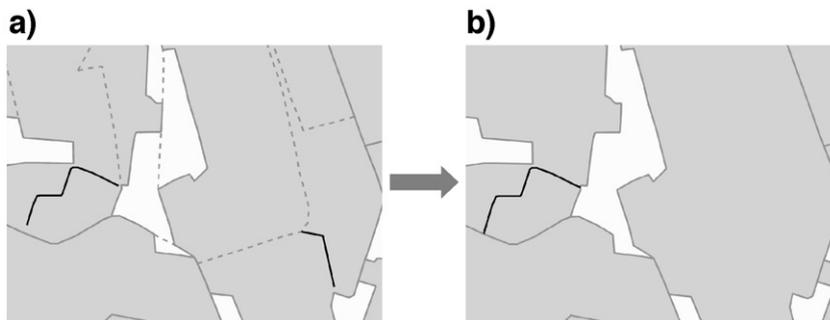


Fig. 11. (a) GIS data: linear objects (solid grey lines), ATKIS borders not belonging to region of interest borders (grey dashed lines), rows of trees (black lines); (b) generated regions of interest.

a closed 2D-curve. The image energy  $E_{\text{img}}$  is defined as

$$E_{\text{img}}(v(s)) = -|\nabla I(v(s))|^2, \quad (3)$$

where  $I$  represents the image grey values,  $|\nabla I(v(s))|$  is the norm or magnitude of the gradient image at the coordinates  $x(s)$  and  $y(s)$ . The internal energy  $E_{\text{int}}$  is defined as

$$E_{\text{int}}(v(s)) = \frac{1}{2} \left( \alpha(s) \cdot |v_s(s)|^2 + \beta(s) \cdot |v_{ss}(s)|^2 \right), \quad (4)$$

where  $v_s$  and  $v_{ss}$  are the first and second derivative of  $v$  with respect to  $s$ . The function  $\alpha(s)$  controls the first-order term of  $E_{\text{int}}$ , the elasticity. When the aim is to minimize  $E_{\text{int}}(v(s))$  and  $v(s)$  is allowed to move, large values of  $\alpha(s)$  let the contour become straight between two points. The function  $\beta(s)$  controls the second-order term of  $E_{\text{int}}$ , the rigidity. Large values of  $\beta(s)$  let the contour become smooth, small values allow the generation of corners.  $\alpha(s)$  and  $\beta(s)$  need to be predefined based on given shape characteristics of the object of interest. The total snake energy  $E_{\text{snake}}^*$ , to be minimized, is defined as

$$\begin{aligned} E_{\text{snake}}^* &= \int_0^1 E_{\text{snake}}(v(s)) ds \\ &= \int_0^1 [E_{\text{img}}(v(s)) + E_{\text{int}}(v(s)) + E_{\text{con}}(v(s))] ds. \end{aligned} \quad (5)$$

The additional external energy  $E_{\text{con}}(v(s))$  is introduced in Kass et al. (1988) as an external constraint, which provides the opportunity for individual forces at particular parts or points of the contour to be introduced. With constant weight parameters  $\alpha(s)=\alpha$  and  $\beta(s)=\beta$  a minimum of  $E_{\text{snake}}^*$  can be derived by solving the Euler equation:

$$\frac{\partial E_{\text{img}}(v(s))}{\partial v(s)} - \alpha v_{ss}(s) + \beta v_{ssss}(s) = 0. \quad (6)$$

The derivatives are approximated with finite differences since they cannot be computed analytically. Converted to vector notation with  $v_i=(x_i, y_i)$  and with  $\partial E_{\text{img}}(v(s))/\partial(s)=f_v(v)$  the Euler equations can be written as

$$\begin{aligned} \alpha_i(v_i - v_{i-1}) - \alpha_{i+1}(v_{i+1} - v_i) \\ + \beta_{i-1}(v_{i-2} - 2v_{i-1} + v_i) - 2\beta_i(v_{i-1} - 2v_i + v_{i+1}) \\ + \beta_{i+1}(v_i - 2v_{i+1} + v_{i+2}) \\ + f_v(v) = 0 \end{aligned} \quad (7)$$

and can be rewritten in matrix form as

$$Av + f_v(v) = 0. \quad (8)$$

$A$  is a pentadiagonal matrix, which depends only on the functions  $\alpha$  and  $\beta$ . Eq. (8) can be solved iteratively by introducing a step size  $\gamma$  multiplied with the negative time derivatives  $\partial v/\partial t$ , discretized as  $(v_t - v_{t-1})$ . It is assumed that  $f_v(v)$  is locally constant from one time step to the next, i.e.  $f_v(v_t)=f_v(v_{t-1})$ . The resulting equation then reads

$$Av_t + f_v(v_{t-1}) = -\gamma(v_t - v_{t-1}). \quad (9)$$

The time derivatives vanish in the equilibrium. Finally, a solution can be derived by matrix inversion:

$$v_t = (A + \gamma I)^{-1}(\gamma v_{t-1} - \kappa f_v(v_{t-1})), \quad (10)$$

where  $I$  is the identity matrix and  $\kappa$  is an additional parameter in order to control the weight between internal and image energy.

In *network snakes* a whole graph, whose topology is kept constant, is considered compared to individual polygons only. This extension is useful here, because the segmentation typically yields a topologically correct graph (see above). In network snakes the graph nodes with a degree  $\rho(v) \neq 2$  of the preliminary contour  $v(s)$  are specially considered: nodes with a degree  $\rho(v)=1$  represent end points and nodes with a degree  $\rho(v)>2$  represent the nodal points of the contour. The introduction of the network topology into energy minimization causes a problem when solving Eqs. (7) and (8): the derivatives approximated by finite differences are not defined for nodes with a degree  $\rho(v)=1$  or  $\rho(v)>2$ , because the required neighbouring nodes are either not available ( $\rho(v)=1$ ) or exist multiple times ( $\rho(v)>2$ ). Thus, the shape control cannot be accomplished at these parts of the contour in a traditional way.

Let  $v_a$ ,  $v_b$  and  $v_c$  represent three contours, each ending in a common nodal point  $v_n$  with a degree  $\rho(v)=3$ . Regarding Eq. (7), the first term, weighted by  $\alpha$ , cannot support the control of the internal energy in the vicinity of  $v_n$  during energy minimization when using network snakes: the finite differences of the first term approximating the derivatives are only available for the two nodes  $v_{n-1}$  and  $v_n$ , but not for  $v_{n+1}$ . Thus, no shape control is possible and the first term is not considered. The second term of the internal energy, weighted by  $\beta$ , is rewritten using the available finite differences controlling the curvature of the contour. Consequently, for network snakes the control of the total energy at the common nodal point  $v_n=v_{a_n}=v_{b_n}=v_{c_n}$  is defined as follows:

$$\begin{aligned} \beta(v_{a_n} - v_{a_{n-1}}) - \beta(v_{a_{n-1}} - v_{a_{n-2}}) + f_{v_a}(v_a) = 0 \\ \beta(v_{b_n} - v_{b_{n-1}}) - \beta(v_{b_{n-1}} - v_{b_{n-2}}) + f_{v_b}(v_b) = 0 \\ \beta(v_{c_n} - v_{c_{n-1}}) - \beta(v_{c_{n-1}} - v_{c_{n-2}}) + f_{v_c}(v_c) = 0. \end{aligned} \quad (11)$$

All three contours intersect in the common nodal point without interacting concerning their particular shape. The energy definition of Eq. (11) allows for energy minimization to control the shape of each contour separately while exploiting the network topology. Matrix  $A$  of Eq. (8) is adapted accordingly at the nodal points and their neighbours to fulfil the new definition of the internal energy, i.e. some parts of the banded structure are omitted, and other matrix elements are added to represent the connections between different parts of the contour.

Finally, snakes have the tendency to shorten during energy minimization due to the first term of Eq. (4). A shortening of contours with an open ending can be avoided by chaining the end points to a topologically neighboured object allowing for movement only along the object border.

## 5. Results

In this section some results of test areas in northern Germany are presented to demonstrate the capability of the described method. First, results are shown for the alignment and change detection for updating vector data sets. Subsequently, results of the field boundary extraction are given. It should be noted that both methods are linked via the federated database as described in Section 3.

### 5.1. Results of the vector data integration

Vector data integration is demonstrated using water objects of ATKIS data and of the geological map GK 25. An area of 200 km<sup>2</sup> in Hannover, Germany, including 3.5 km<sup>2</sup> of water area (the Maschsee and its surroundings) has been selected. Data integration is carried out separately for each pair or object clusters. The alignment was designed in the way that ATKIS was chosen to be the master geometry and the GK25 was aligned to it. The results are presented graphically (see Fig. 12) and

numerically (see Table 1). The described rubber sheeting algorithm is not discussed further, because in this paper we focus on the alignment of the individual matched objects, not on the appearance of the whole map after alignment.

The ICP algorithm based on a four-parameter transformation has been found to be very suitable for the coarse alignment of objects. The decision, whether the result is correct is then taken based on an inspection of the transformation parameters. If the scale parameter is close to 1 and the rotation and translation values are small, the objects are assumed to coarsely fit. In these cases the DIA approach is applied to handle the remaining partial discrepancies.

Results of the described alignment strategy are shown in Fig. 12, selected numeric results of the alignment processes are given in Table 1. The symmetric difference is used to evaluate the alignment results. For a successful alignment the symmetric difference should be close to zero. A higher value can be the result of an unsuccessful alignment or of major geometric discrepancies between the two objects. The symmetric differences for the two data sets are given as absolute values before alignment and as percentages of the original situation after applying the ICP and DIA. The values for  $t_1$  and  $t_2$  have been selected through evaluation of the Hausdorff-distances, which have been calculated for every object for which a suitable candidate (1:1) had been found in the matching process.

As it can be seen in Fig. 12, for objects  $a$  and  $c$  the ICP results indicate a good match, and after the alignment of DIA a very suitable result has been achieved. For objects, which are representations of the same real world objects, but are given with different cardinalities (object  $d$ ), a very good alignment can be achieved also, but due to the different components a higher symmetric difference is the result of the alignment process.

The scale from the ICP algorithm for object  $b$  differs compared to the results of the other objects, which can

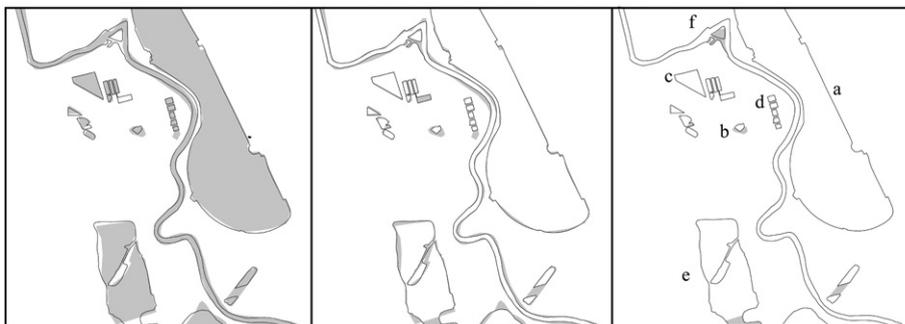


Fig. 12. Left: overlaid ATKIS (dark line) and GK 25 (filled grey) data; centre: symmetric difference (filled grey) between ATKIS and GK 25 without alignment; right: remaining areas (filled grey) of GK 25 after the alignment using ICP and DIA.

Table 1. Detailed results for representative objects (see Fig. 12)

ID (Fig. 12)	Symmetric difference before	ICP			Symmetric difference after ICP	DIA		Symmetric difference after DIA
	ATKIS—GK25 (m <sup>2</sup> )	Distance (m)	Scale	Rotation(rad)	ATKIS—GK 25 (m <sup>2</sup> )	$t_1$ (m)	$t_2$ (m)	ATKIS—GK 25 (m <sup>2</sup> )
<i>a</i>	28,755.5	2.2	1.003	−0.1	86.6%	35.0	80.0	0.2%
<i>b</i>	3646.9	7.9	0.575	0.3	15.4%	DIA not run due to scale in ICP		
<i>c</i>	2123.3	3.1	0.973	0.1	39.2%	35.0	80.0	0.1%
<i>d</i>	5385.2	5.6	0.928	−0.1	80.3%	35.0	80.0	15.5%
<i>e</i>	30,399.5	12.1	1.004	0.1	76.4%	35.0	80.0	22.1%
<i>f</i>	89,665.2	6.4	1.005	0.2	92.0%	35.0	80.0	10.4%

be interpreted either as an indication of a topographic change in the real world or as an error in one of the original data sets. In this case DIA was not executed.

Object *e* has been aligned very well as judged from Fig. 12, but the symmetric difference after the whole process is still rather large, indicating somewhat different resulting geometries. This can be explained by comparing the size of both objects.

The DIA algorithm was very successful in compensating local discrepancies which cannot be corrected using the four-parameter ICP, as it aims at an optimum alignment of the whole object. This is the case for large extended natural objects like rivers (object *f*). Note that in the current implementation of the approach for objects containing holes, only the exterior boundary of object *f* is aligned.

The combined strategy of using ICP for pre-processing and DIA as the second step showed promising results. By evaluation of all objects for which a suitable partner has been found in the corresponding data set the symmetric difference between ATKIS and GK was reduced to 11%. The same strategy was also applied to another data set. Water areas between ATKIS and a GDF (Geographic Data Files, data used in navigation systems) data set have been adapted. For this configuration the symmetric difference (between corresponding objects) has been reduced to 7%.

### 5.2. Results of the raster and vector data integration

The described field boundary extraction algorithm has been tested in an area of 25 km<sup>2</sup> in Lower Saxony, Northern Germany using RGB+IR satellite imagery with a ground resolution of approximately 2.0 m. Reference data for the field boundaries were generated manually from the images.

The segmentation result, superimposed on the intensity channel of the image, is shown in Fig. 13. Exploiting the GIS data, settlements, forests and water bodies are masked out (shown as white areas), the

boundaries of the regions of interest representing fixed field boundaries are depicted as thick white lines. The results were compared to manually derived reference data in a 10 pixel wide buffer: correctly extracted field boundaries are depicted as thin white lines in Fig. 13, missing boundaries as thin black and falsely detected boundaries are depicted as thin dashed (black/white) lines. We achieved a completeness of 80% and a correctness of 93%, and geometric accuracy of the segmentation as expressed in the horizontal root mean square value between extracted and reference result amounts to 5.5 m or about 2.8 pixel. The quality of the results is promising, but as expected the geometrical correctness is not very high. Therefore, the second step – the network snakes algorithm – was applied to improve the geometric location of the field boundaries.

For this task a region section was selected (cf. dotted white rectangle in Fig. 13). The initialization of the network snake – equivalent to the result of the segmentation – is shown in Fig. 14a. Note, that the geometrical correctness of the segmentation has been artificially decreased to emphasize the following steps in a better way. The result of the segmentation is used to derive the topology, given in Fig. 14b: The individual snakes forming the network are linked to each other in the nodal point (grey), the end points (black with white hole) are chained to the boundary of the processed region of interest. Since the objects to be extracted are rather straight, parameter  $\beta$  is set to a large value compared to  $\alpha$ . Thus, image noise and small disturbances have relatively small effects. The movement of the snake superimposed to the standard deviation image is shown in Fig. 14c, the final result superimposed on the intensity channel is depicted in Fig. 14d. Within this selected region of interest, the root mean square value decreased from 6.7 m (about 3.3 pixel) after the segmentation step – of course calculated without artificial manipulation – to 3.7 m (about 1.8 pixel) after the optimization step. The example demonstrates that network snakes are a useful tool for improving the

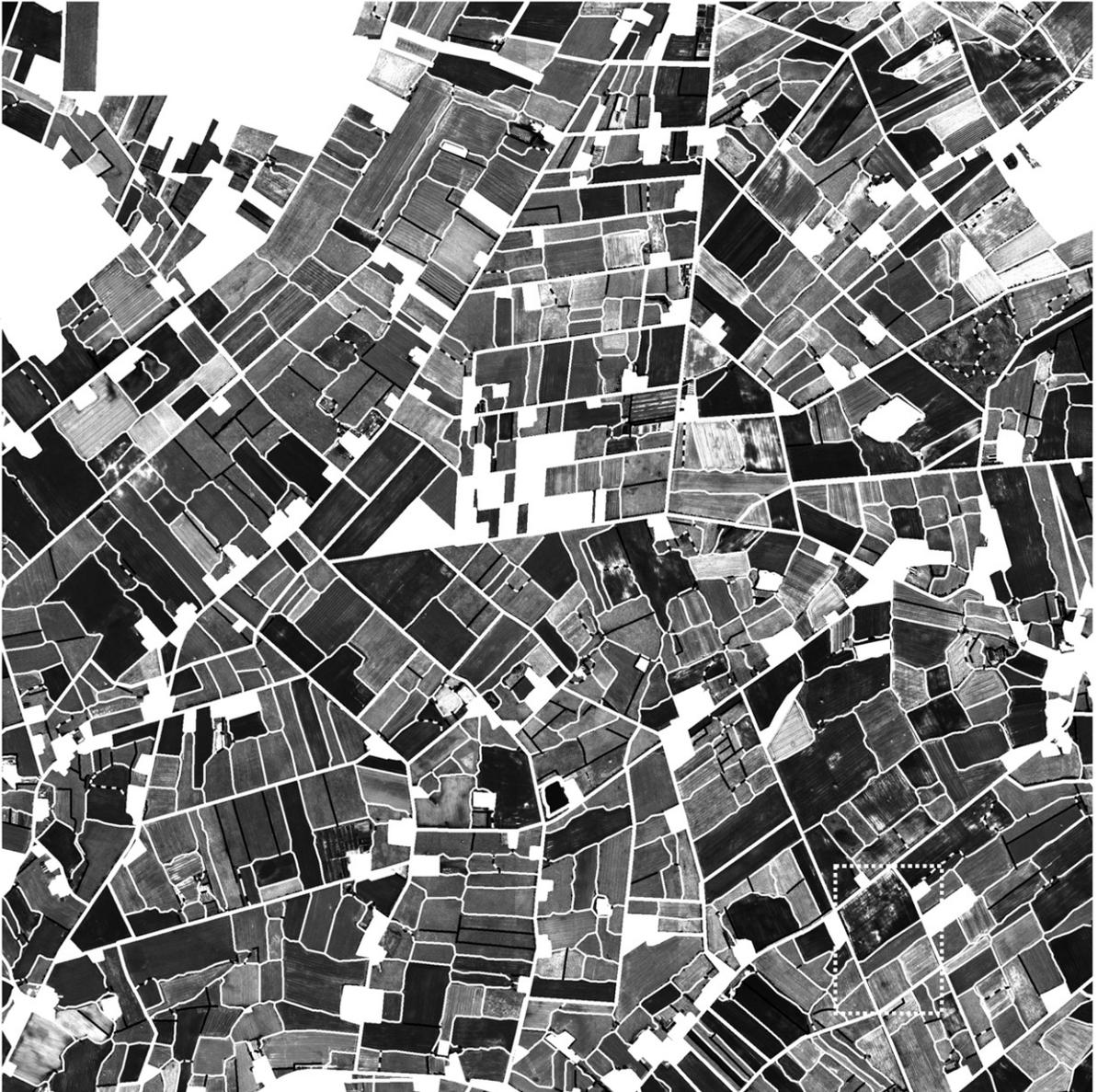


Fig. 13. Result of the segmentation within the regions of interest (thick white lines): correctly extracted field boundaries are depicted as thin white lines, missing boundaries as thin black lines and falsely extracted boundaries as thin dashed lines.

geometrical correctness of topologically correct but geometrically inaccurate results.

## 6. Conclusions

In this paper we have discussed various components which are necessary to perform the integration of various types of geospatial data, and we have described three distinct modules: two data integration approaches, one for matching and aligning vector data sets and a second one for integrating raster and vector data, and –

as a base for any kind of complex data handling – a federated database. We have implemented our ideas in a small prototype integration system comprising all three modules and have presented sample results generated with the prototype. It should be noted that while the data integration approaches could be seen as separate entities, it became very clear during our work that in order to manage and process larger amounts of data in an organized way, the database is indispensable.

As for the individual modules, the developed geospatial federated database provides a uniform access

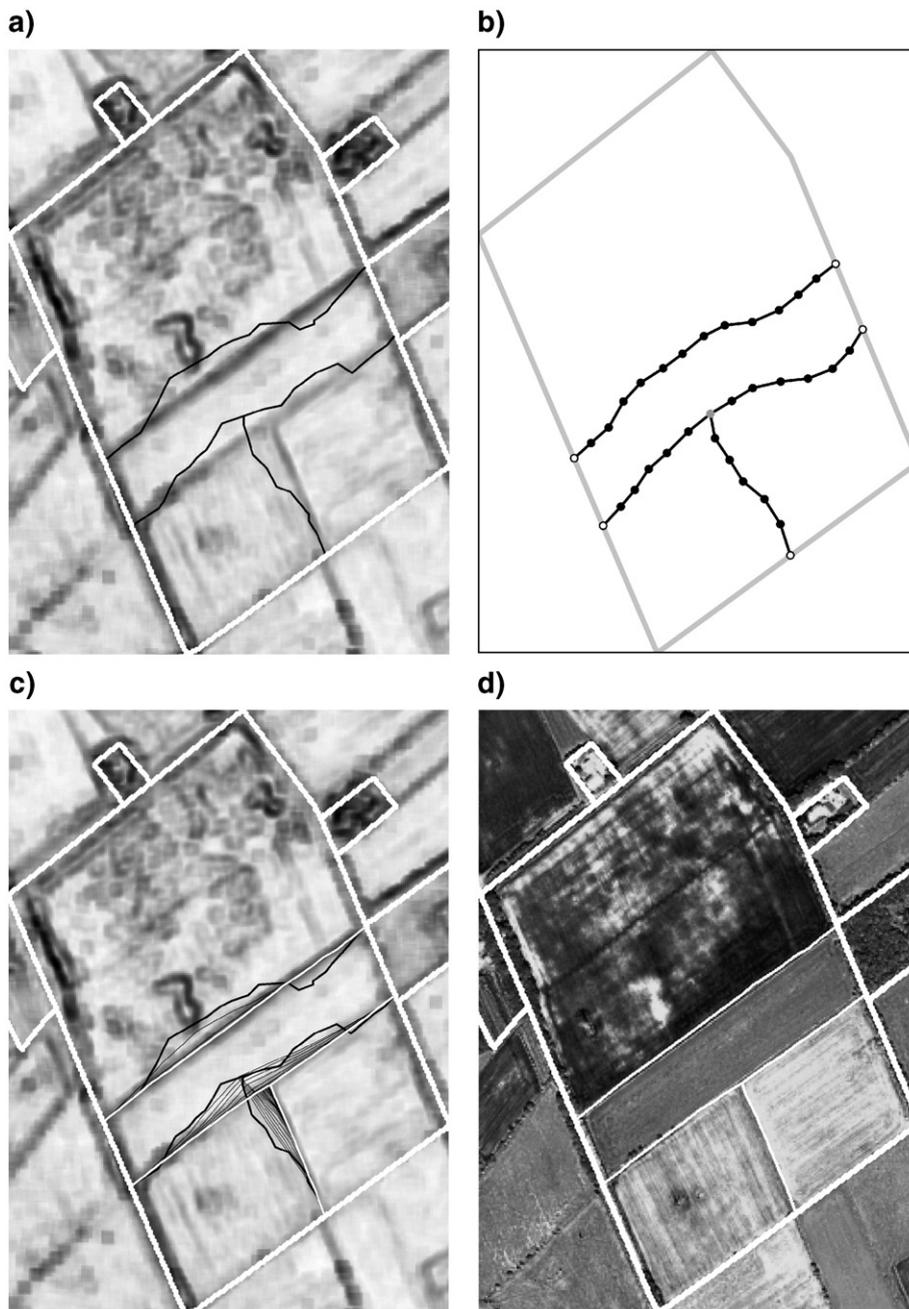


Fig. 14. Result of the use of network snakes: (a) initialization of the network snake (black), (b) topology, (c) initialization (black), movement (thin black) and result of the snake (white), (d) extracted field boundaries superimposed to the intensity channel of the image.

to the different data sets involved and to the results of the matching and extraction processes. With the matching of geospatial objects a sophisticated process is given, which exceeds simple object identification in former federated databases. As each incorporated geoscientific data set model is semantically described, the application processes (matching and extraction) are

independent of particular models and therefore applicable to different further geoscientific data sets as well. Even though the architecture enhancements like object linking and semantic selection were developed for the matching and extraction processes, they are process independent parts of the federated database. For future work this database serves as a base not only for querying

linked objects but also for update propagation. Future work will also exploit the potential of the already involved topology building on the database: it can help to assure topological consistency during geometric alignment and to support a more powerful matching based on graph structures directly on the database.

The geometric matching and the derivation of object links, together with the ICP and DIA alignment show good results. In the example this strategy was used with one data set as reference which remains unchanged, while a second data set is aligned to the reference, but as shown in Section 4.1 our work can also be adapted to other vector-based conflation tasks requiring an intermediate geometry. Depending on the selected thresholds large discrepancies of the shape boundaries can be considered as outliers and can be treated accordingly in the subsequent analysis step. While matching can be performed automatically, there are still some steps during geometric alignment which require the decision of a human operator (e.g. Fig. 12, object b), but the high degree of automation reduces manual interaction considerably. Future work will concentrate on developing a strategy to also automate these processes, and on the adaptation of a triangulation-based rubber-sheeting strategy which meets the requirements of topographic and geoscientific data sets.

The method for integrating raster and vector data for the extraction of field boundaries from imagery exploiting prior GIS knowledge also shows encouraging results. The exploitation of the prior GIS data provides a useful framework for the segmentation and subsequent optimization of the field boundaries. Concerning the first step of the strategy, the segmentation, we currently experiment with an additional texture channel in region growing to reduce the number of wrong field boundaries. They occur mainly when large heterogeneities within one field exist. The second step, the improvement of the preliminary field boundaries using network snakes comprises a complete shape control of the contours. The exploitation of the topology turned out to be a powerful method to deal with noise and disturbances in the imagery. The extraction of other objects such as road networks and other questions such as the update of GIS data with an already given topology are additional applications for network snakes.

In addition, the semantic model provides the opportunity to introduce further objects, which could be of interest for other or enhanced application scenarios. For example, detailed information about wind erosion obstacles (tree rows, hedges), which are currently not included in the ATKIS BasisDLM, could be extracted from raster data and integrated in the federated database. Together with the geoscientific data and the already

extracted field boundaries the suggested integration system enables a more detailed result concerning the potential wind erosion risk fields.

In summary, besides helping in data organisation, the adopted database centric approach opens up possibilities to combine vector/vector and raster/vector integration in a concise and compatible way. It becomes thus possible, for example, to use the image analysis algorithms with rasterized vector data (e. g. the network snakes approach for vector data alignment) and the matching algorithms in a GIS update procedure with recently acquired imagery, without having to redesign or change the whole system. This clear synergy between the different data types has been a driving force behind our work and will be a focus of future research.

## Acknowledgements

This work is part of the Geotechnologien project funded by the Federal Ministry for Education and Research (BMBF) and the German Council (DFG) under contract no. 30F0374A. The IKONOS images are provided by EC-MARS Programme—distributed by European Space Imaging, EUSI, 2004. The support is gratefully acknowledged.

## References

- Baltsavias, E.P., 2004. Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3–4), 129–151.
- Batini, C., Lenzerini, M., Navathe, S.B., 1986. A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys* 18 (4), 323–364.
- Beeri, C., Doytsher, Y., Kanza, Y., Safra, E., Sagiv, Y., 2005. Finding corresponding objects when integrating several geo-spatial datasets. *Proc. 13th ACM International Symposium on Advances in Geographic Information Systems*, Bremen, Germany, 4–5 November 2005, pp. 87–96.
- Besl, P., McKay, N., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (2), 239–256 (Special issue on interpretation of 3-D scenes - part II).
- Brandtberg, T., Walter, F., 1998. Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple-scale analysis. *Machine Vision and Applications* 11 (2), 64–73.
- Butenuth, M., 2006. Segmentation of imagery using network snakes. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3), 1–6.
- Butenuth, M., Heipke, C., 2005. Network snakes-supported extraction of field boundaries from imagery. In: Kropatsch, Sablatnig, Hanbury (Eds.), *Pattern Recognition. Lecture Notes in Computer Science*, vol. 3663. Springer, pp. 417–424.
- Conrad, S., 1997. *Föderierte Datenbanksysteme*. Springer, Berlin.
- Devogele, T., Parent, C., Spaccapietra, S., 1998. On spatial database integration. *International Journal of Geographical Information Science* 12 (4), 335–352.

- Doytsher, Y., 2000. A rubber sheeting algorithm for non-rectangular maps. *Computer & Geosciences* 26 (9–10), 1001–1010.
- Dowman, I., 1998. Automated procedures for integration of satellite images and map data for change detection: the archangel project. *International Archives of Photogrammetry and Remote Sensing* 32 (Part 4), 162–169.
- Egenhofer, M.J., Frank, A.U., Jackson, J.P., 1989. A topological data model for spatial databases. In: Günther, Buchmann, Wang, Smith (Eds.), *Design and Implementation of Large Spatial Databases*. Lecture Notes in Computer Science, vol. 409. Springer, pp. 271–286.
- Gösseln, G.v., Sester, M., 2004. Integration of geoscientific data sets and the German digital map using a matching approach. *International Archives of Photogrammetry and Remote Sensing* 35 (Part B4), 1249–1254.
- Hettwer, J., Benning, W., 2000. Nachbarschaftstreue Koordinatenberechnung in der Kartenhomogenisierung. *Allgemeine Vermessungs-Nachrichten* 107 (6), 194–197.
- Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: active contour models. *International Journal of Computer Vision* 1 (4), 321–331.
- Kleiner, C., Lipeck, U., Falke, S., 2000. Objekt-Relationale Datenbanken zur Verwaltung von ATKIS-Daten. In: Bill, R., Schmidt, F. (Eds.), *ATKIS-Stand und Fortführung*, Verlag Konrad. Wittwer, Stuttgart, pp. 169–177.
- Laurini, R., 1998. Spatial multi-database topological continuity and indexing: a step towards seamless GIS data interoperability. *International Journal Geographical Information Science* 12 (4), 373–402.
- Löcherbach, T., 1994. Fusion of multi-sensor images and digital map data for the reconstruction and interpretation of agricultural land-use units. *International Archives of Photogrammetry and Remote Sensing* 30 (Part 3/2), 505–511.
- Mantel, D., Lipeck, U.W., 2004. Datenbankgestütztes Matching von Kartenobjekten. Arbeitsgruppe Automation in der Kartographie—Tagung Erfurt 2003. BKG, Frankfurt, pp. 145–153.
- Mayer, H., 1999. Automatic object extraction from aerial imagery—a survey focusing on buildings. *Computer Vision and Image Understanding* 74 (2), 138–149.
- Mustière, S., 2006. Results of experiments on automated matching of networks at different scales. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 2/W40), 92–100.
- Rigeaux, P., Scholl, M., Voisard, A., 2002. *Spatial Databases with Application to GIS*. Morgan Kaufman Publishers.
- Saalfeld, A., 1988. Automated map compilation. *International Journal of Geographical Information Science* 2 (3), 217–228.
- Sester, M., Hild, H., Fritsch, D., 1998. Definition of ground-control features for image registration using GIS-data. *International Archives of Photogrammetry and Remote Sensing* 32 (Part 3), 537–543.
- Torre, M., Radeva, P., 2000. Agricultural field extraction from aerial images using a region competition algorithm. *International Archives of Photogrammetry and Remote Sensing* 33 (Part B2), 889–896.
- Ubeda, T., Egenhofer, M.J., 1997. Topological error correcting in GIS. In: Voisard, Scholl (Ed.), *Advances in Spatial Databases*. . Lecture Notes in Computer Science, vol. 1262. Springer, pp. 283–297.
- Volz, S., 2006. An iterative approach for matching multiple representations of street data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 2/W40), 101–110.
- Walter, V., Fritsch, D., 1999. Matching spatial data sets: a statistical approach. *International Journal of Geographical Information Science* 13 (5), 445–473.
- Yuan, T., Tao, C., 1999. Development of conflation components. In: Li, B., et al. (Ed.), *Geoinformatics and Socioinformatics—The Proceedings of Geoinformatics '99 Conference*, Ann Arbor, USA, 19–21 June 1999, pp. 1–13.
- Zhang, Y., Heipke, C., Butenuth, M., Hu, X., 2006. Automatic extraction of wind erosion obstacles by integration of GIS data, DSM and stereo images. *International Journal of Remote Sensing* 27 (8), 1677–1690.