AUTOMATED VERIFICATION OF A TOPOGRAPHIC REFERENCE DATASET: SYSTEM DESIGN AND PRACTICAL RESULTS

A. Busch a, *, M. Gerke b, D. Grünreich c, Ch. Heipke b, C.-E. Liedtke c, S. Müller c

a BKG – Bundesamt für Kartographie und Geodäsie, Richard-Strauss-Allee 11, 60598 Frankfurt, Germany – (andreas.busch, dietmar.gruenerich)@bkg.bund.de
b IPI – Institut für Photogrammetrie und GeoInformation, Universität Hannover, Nienburger Str. 1, 30167 Hannover, Germany – (gerke, heipke)@ipi.uni-hannover.de
c TNT – Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Appelstr. 9A, 30167 Hannover, Germany – (liedtke, mueller)@tnt.uni-hannover.de

Intercommission Working Group II/IV

KEY WORDS: Automation, Change Detection, GIS, Quality, Updating

ABSTRACT:

Quality of an official topographic reference dataset is important since it is the groundwork for many applications. Verification is classified as being part of quality management in this paper. Its main topic is the automated verification of a topographic reference dataset by means of orthoimages. The objects from the dataset are compared to an up-to-date orthoimage in order to obtain information on their quality. The main objects of interest are roads and built-up areas. As it is assumed that most of the objects in the database are correct, the strategy is to use the ATKIS objects stored in the DLMBasis as a starting point. Automatic image operators being able to detect the objects of interest as much prior knowledge as possible from the dataset, such as contextual and geometrical information. By this means inconsistencies between the ATKIS objects and the image features can be detected. Positional accuracy of roads is checked as well. To organise the verification of the data independently from its capture, a semi-automatic working environment has been installed. In the subsequent interactive step the human operator just has to focus on the objects not being found in the automatic run. In this paper we will introduce the whole system, we then focus on the automatic components and their integration in a semi-automatic workflow. Therefore, we can give a detailed report on the system in its daily application as well as an evaluation of the results.

1. BACKGROUND

Verification as referred to within this paper is part of the concept of quality management for topographic reference datasets as realized at BKG (German Federal Agency for Cartography and Geodesy). Therefore, we first present the background of quality and quality management with respect to geoinformation.

1.1 Quality of Geoinformation

Every application based on spatially referenced data or geodata requires a certain knowledge about their quality or at least an idea of the consequence of possible errors and the risk associated with these errors. The level of knowledge about quality needed by the user differs widely depending on applications and specifications given. Many consumers will rate the quality of spatially referenced data by its fitness for use and the effort needed for handling the datasets. In contrast, companies creating new products by processing geodata or by linking their data to a referenced dataset are interested in knowing whether the reference data fulfill the warranted characteristics.

Data quality should be described by a certain set of measures, which express comprehensive and useful criteria. These should enable the user to compare the quality of different data sets. Therefore, quality measures are part of standards or specifications from e.g. ISO, CEN, or the OpenGIS Consortium.

We do not want to go into detail about these specifications but start with a subdivision of quality measures into two categories that are important for practical applications owing to the arguments given below:

- logical consistency, i.e. consistency with respect to the data model,
- consistency as regards content, i.e. consistency of data and reality within the scope of the model.

In this paper we refer to the first category as logical consistency since it is characterized by the fact that it can be checked without any comparison of the data to the real world. We can perform a complete check of this category using solely the data set without additional information. Only routines and functionality within the database or the GIS are needed. Once implemented, the inspection of logical consistency is performed automatically. Format specifications, topological constraints, uniqueness of identifiers, and domains of attribute values count among the criteria for logical consistency. For the second category a comparison of data and reality is required. Basically the comparison can be performed by means of current sensor data or field work. A complete comparison of data and reality requires a lot of effort and cost, but in return it furnishes all the update information for the data. Consistency as regards content is the focal point of this paper.

A well defined system is spanned by four quality measures, namely completeness, correctness, consistency, and accuracy, which are conceptually independent, i.e. orthogonal (Joos,
In this sense consistency is part of logical consistency as defined above. Some aspects of completeness cover logical consistency, too, e.g. attribute completeness which concerns the question whether all required attributes are stored together with an object. Most aspects of completeness, correctness, and accuracy must be checked in comparison to reality. It has to be verified, whether e.g. all objects are registered in the data set and whether their attributes are set correctly. Accuracy concerns positional accuracy and temporal accuracy, i.e. currency.

1.2 Quality Management at BKG

A major task of BKG consists in providing the geodata of the Authoritative Topographic-Cartographic Information System ATKIS on the territory of the Federal Republic of Germany. ATKIS® is a trademark of the Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany (AdV). The most important components of ATKIS are object-based digital landscape models (DLM) encompassing several resolutions and digital topographic maps (DTK). The ATKIS DLM Basis, i.e. the ATKIS data offering highest resolution, is produced by the 16 surveying authorities of the federal states of Germany and is delivered to the BKG. Here, at the Geodata Centre (GDC) of BKG, the ATKIS DLM Basis is checked with respect to logical consistency, joined to one homogeneous data set for the territory of the Federal Republic of Germany, and stored in a database. Since these data are delivered to customers on the one hand and are used to derive data of smaller scales within BKG on the other hand, a system for quality control of the ATKIS data is essential. For testing logical consistency of the data sets an exhaustive check on the full coverage of the data is performed at the GDC. This check includes establishing logical and geometrical consistency of the different data sets from the surveying authorities at their borders. Thus, the check of logical consistency is done in an operational way within the daily production process. Errors detected during the quality control are reported to the respective federal state. Since the federal states are producers of the data of the ATKIS DLM Basis, they are responsible for the appropriate amendment of the data.

Since nearly all of the changes in the real world are man-made, information concerning the changes is available very early, usually already during the phase of planning. Therefore, the surveying authorities of Germany are forcing topographical information management to gather information about changes and make it available in time for the update of geoinformation.

In addition to this well established, broad inspection of logical consistency of the ATKIS dataset, BKG pushes the increase of quality control with respect to reality. Therefore, BKG has initiated a common project with the University of Hannover to develop a system for automated quality control comparing orthoimages and the ATKIS DLM Basis (Busch and Willrich, 2002). Since in practice the comparison to the real world still is far away from being fully automatic, it is implemented as an interactive procedure based on ArcGIS on the one hand and on automated image analysis methods on the other hand. The knowledge-based image interpretation system GeoAIDA (Bückner et al. 2002) and various methods for feature extraction form the core of the automated procedures. These programs run separately in batch mode furnishing results that are imported to the interactive working environment. While the final decision about errors is reserved to a human operator, the strategy is to detect as many coincidences of ATKIS objects and objects detected in the orthoimages as possible. By filtering these correct situations the human operator can concentrate on the objects where the automated procedure failed. The comparison utilises orthoimages of recent date which are an up-to-date reference of reality and can be used to assess completeness, correctness, positional and temporal accuracy. Our main interest concerns objects where most changes arise and that are important, namely the road network and built-up areas. Other objects and their attributes can be verified, too, if they are visible within the images.

The paper focuses on a method for quality control of the ATKIS DLM Basis that automatically compares the data with reality by means of images from an independent source. Thus, we look at quality control as an independent procedure to rate the quality of geodata by sample and to detect deficiencies within the chain of production. Depending on the type of image, e.g. airborne or satellite imagery, different features and attributes can be verified. Orthoimages of recent date are an up-to-date reference of reality and can be used to assess temporal accuracy, too. Orthoimages are a very suitable source to determine the positional accuracy of features and geometric attributes like the width of a road. Nevertheless, there exist features and attributes that are not detectable in the image data. In these cases quality control has to be based on the topographic information management.

1.3 Automated Verification and Update of Spatial Databases by Means of Imagery

Automatic and semi-automatic feature extraction has been a focus of international research in photogrammetry and computer vision for a few decades (e.g. Balsavias et al. 2001, Heipke et al. 2004). As a consequence the results are now starting to enter into the commercial market. Obviously algorithms particularly give good results if applied to well-defined application areas. The reason is that all approaches need additional knowledge in the form of appropriate models, which can more easily be formulated for restricted situations. Since any automatic feature extraction algorithm will show a certain error rate, it has to be integrated to an interactive workflow leaving final decisions to a human operator. For achieving an efficient workflow, the algorithms have to be equipped with an appropriate and reliable self-diagnostics allowing the operator to concentrate on situations where the automatic procedure failed. Walter (2004), for instance, developed a system that supports the operator in quality control of region and line objects in ATKIS by automatically extracting land cover classes from satellite imagery by multi-spectral classification, and comparing them to the corresponding ATKIS objects. He uses prior information derived from the existing ATKIS dataset for defining training sets for a supervised classification. ATKIS objects showing a high probability of differences to the extracted object classes are indicated as presumed changes. They are visualized for supporting the human operator’s final interactive analysis. GIS data in general can provide a valuable source of prior information (e.g. Vosselman 1996) and can be used to stabilize the image interpretation tasks.

The basic concept relies on a knowledge-based system for image interpretation. Knowledge-based systems have proven to be a suitable framework for representing knowledge about objects and exploiting it during the recognition process. Our system models structural dependencies by semantic networks (Bückner et al. 2002). It has been successfully applied to land cover interpretation by means of orthoimages, laser DEMs, and prior information from a GIS.
In our system the comparison of the ATKIS DLM Basis and reality, also called update, comprises two steps, namely verification and the acquisition of change. Verification is characterised by the following features:

- The image analysis process is guided by information from a GIS about the object to be verified, i.e. the algorithm makes use of the information stored in the GIS to detect the image object.
- If there is a certain degree of consistency of image features and information from the GIS, the object is accepted.
- Otherwise the object is labelled as not accepted.

Thus, verification is suited to determine specific quality measures. For updating, in contrast, information about a new object not yet stored in the dataset or information about changes of the old object have to be extracted from the image, too. Hence, a reasonable process chain starts with a verification step. In the event of any inconsistency other feature extraction algorithms can be triggered to derive more detailed information that is of value for the following update process.

2. Ideas Behind the System

2.1 Concept of the Prototype

The concept of the research and development project of BKG and the University of Hannover is characterized by the following main ideas:

- Transfer of knowledge-based image interpretation techniques to an operational solution for practical applications (cf. Sec. 1.3).
- Development of a prototype for comparing the model ATKIS DLM Basis to reality given in form of digital orthoimages.
- Efficient integration of the prototype into an interactive workflow.

The system development is embedded in a broader concept of a knowledge-based workstation, which provides functionality from photogrammetry, GIS, and cartography for the acquisition, and maintenance of geoinformation. A major goal of this concept is to integrate several components performing different tasks within the framework of a knowledge-based system.

Currently, we are automating the interactive quality control step by step using procedures that have been developed in the research and development project with the University of Hannover. The automated procedures consist of automatic steps that are started by an operator and return a result that requires further interaction of the operator.

The fully automatic part attains to solve the bigger part of the quality control unassistedly and to focus the human operator to those objects where the algorithms detect ambiguous situations. Thus, the goal is to reduce the amount of human interaction by automatically completing routine work which is a time consuming part in the quality control process chain. Challenging situations are afterwards analysed and solved by the human operator in a separate step.

The results of the automatic procedure are passed to the human operator in the form of a so-called traffic light diagnostics, i.e. the results are displayed by means of red and green colour. An attribute corresponding to the traffic light diagnostics and indicating the result of the automatic procedure is attached to each inspected object. If the algorithm is able to detect and locate the corresponding image object without observing inconsistencies, the ATKIS object is marked with green colour. Otherwise the object is labelled red, i.e. not accepted, since the algorithm was not able to establish full correspondence. The human operator can access more parameters of the diagnostics whenever necessary.

Since the human operator decides on acceptance or rejection in case of the red objects only, the decision of the automatic procedure has to be reliable in particular for objects labelled as green. The different situations that can occur when comparing decisions from a human operator and diagnostics from automatic procedures are classified in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Automatic</th>
<th>Green</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Operator</td>
<td></td>
<td>True Positive</td>
<td>False Negative</td>
</tr>
<tr>
<td>Green</td>
<td>False Positive</td>
<td></td>
<td>True Negative</td>
</tr>
</tbody>
</table>

Table 1. Confusion matrix of decisions: human operator vs. automatic procedure, terminology.

2.2 Components of the System

The system is designed as a knowledge-based workstation, which provides functionality from knowledge-based photogrammetric image analysis and cartography for the production of geoinformation. It consists of three major parts: the GIS component, the knowledge-based component, and the image analysis component (q.v. Figure 1).

![Figure 1. The components of the system for quality control](image)

The GIS component is based on ArcGIS. It acts as an automatic pre-processor of the ATKIS data, as an interface to the database and to the image processing system, as the environment for interactive post-processing of automatically derived results, or generally spoken as the user interface for handling the whole workflow and for visualising the overlay of orthoimages and ATKIS dataset.

The knowledge-based component is designed for making knowledge about topographic objects available, for transferring it to the image analysis component in a suitable way, for handling the results from the image analysis component, for deriving scene descriptions, and for controlling the complete automatic workflow.

The image analysis component comprises the automatic image feature extraction modules and the comparison with the original
ATKIS dataset leading to quality measures. Both tasks are triggered by the ATKIS data being a valuable source of prior information. The image analysis component interacts with the knowledge-based component.

In our implementation the interactive GIS component runs on a different computer than the automatic knowledge-based and image analysis components which are implemented under Linux. The automatic modules may be processed in batch mode, communicating with the GIS component via data files.

3. AUTOMATIC IMAGE ANALYSIS MODULES

The verification of the topographic dataset is done by use of the knowledge-based image interpretation system GeoAIDA (Bückner et al. 2002, Müller et al. 2003) developed at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover. GeoAIDA is based on a semantic network that represents the scene to be analyzed, it was designed for the interpretation of complete scenes, and within this cooperation was modified and expanded for GIS verification purposes.

3.1 Verification of Roads

The first step in road verification consists in defining a region of interest for each road object from the database. More precisely, a buffer around the vector representing the road axis is defined, and the buffer width complies with the geometric accuracy of the road object and the road width attribute in the ATKIS database. If the latter value fails a plausibility test or is not available at all, a predefined value is taken. Subsequently, an appropriate road extraction algorithm to be executed in the image domain of the buffer is selected. The selection includes a control of the parameters considering the knowledge about the given context region. We currently use the road extraction algorithm presented in (Wiedemann and Ebner, 2000; Wiedemann, 2002). This approach models roads as linear objects in aerial or satellite imagery with a resolution of about 1 to 2 m. It should be noted that this algorithm was designed for rural areas. Therefore, the following discussion and the results for road extraction refer to rural areas only.

In the course of road extraction, initially extracted lines (applying an approach given in (Steger, 1998)) are evaluated by fuzzy values according to attributes like length, straightness, constancy in width and constancy in grey values. The evaluation is followed by a fusion of lines originating from different channels. In our case we are using panchromatic imagery, but the line extractor is applied twice: Firstly, using a bright line model (line is brighter than the background) and secondly using a dark line model (line is darker than the background). The last step in road extraction as applied for verification is the grouping of single lines in order to derive a topologically correct and geometrically optimal path between seed points according to some predefined criteria. The decision, if extracted and evaluated lines are grouped into one road object, is taken corresponding to a collinearity criterion (allowing a maximum gap length and a maximum direction difference). All significant and important parameters for road extraction can be set individually. We adapted the described road extraction software to our specific tasks, especially by applying individual parameters for the given context areas and the extraction for each road object separately. A road object from the ATKIS database will be accepted, if the described road extraction in the region of interest was successful and rejected otherwise.

The presented procedure is embedded in a two-stage graph-based approach, which exploits the connection function of roads and leads to a reduction of false alarms in the verification. In the first phase the road extraction is applied using a strict parameter control, leading to a relatively low degree of false-positive road extraction, but also a high number of roads will be rejected although being correct. For the second phase the latter objects are examined regarding their connection function inside the road network. It is assumed that accepted roads from the first phase are connected via a shortest path in the network. All rejected roads from the first phase fulfilling important network connection tasks are checked again in a second phase, but with a more tolerant parameter control for the road extraction.

For further information concerning the road verification refer to (Gerke et al., 2004). In (Gerke, 2004) a new approach to road data verification is introduced. It incorporates local context objects such as rows of trees to support the assessment of ATKIS road objects and will be integrated into our system in the near future.

3.2 Verification of Built-Up Areas

During the analysis two different complementary approaches are followed. A textural analysis of the scene takes place to decide between the classes agriculture, forest, industrial area and settlement. A structural analysis of the image is carried out, which searches for the most important items of settlement and industrial areas, houses and industry halls.

3.2.1 Textural analysis The textural analysis uses a segmentation algorithm described in (Gimel'farb, 1997), it was extended to multiresolution technique. Here the classes acreage and grassland are combined to one unitary class agriculture, because of similar texture. First, the algorithm has to learn the properties of the classes with classified training regions, the result of the learning process are four parameter files and an evaluation matrix. The learning step determines the resolution level on which a class has significant signatures. From the evaluation matrix we derive in which resolution level a texture can best be differentiated. The classification operator segments the input image level by level with use of the 4 parameter files. The resulting classification is a combination of all resolution levels weighed with the evaluation matrix.

The learning step is a crucial part for the effectiveness and correctness of the derived results. It is not necessary to train the operator for each image, one time learning for a complete set of images of a flight is sufficient. This step is preferably assisted by a human operator, who manually defines and classifies training regions for the desired classes. The borders of the training regions can be taken from a GIS to speed up the learning process. Since the fully automatic derivation of training areas sometimes leads to training areas containing a mixture of classes, the separability of the classes is not as good as with manually defined areas. Another possibility to train the classification operator is to take the borders and classes of a GIS, this is only possible if the sample is large enough. The advantage is a higher level of automation which is an important feature of the system. For the results presented in Section 5 manually defined training areas were used.

3.2.2 Structural analysis The structural analysis is based on finding buildings, which are modelled as complex structures consisting of different parts. An illumination model is assumed,
thus a shadow has to exist next to a building. The azimuth and zenith angles of the sun position are taken as a priori knowledge. First hypotheses for shadows and roofs are generated using two different image segmentation operators. To validate these hypotheses instances of roofs are grouped with one or more instances of shadow. Shadow hypotheses of buildings are derived with a threshold decision in the image. Since shadows are not well visible in a green colour channel, the green colour has been masked during shadow detection. Additional shadows have a limited area, so shadows e.g. near a forest can be excluded. Roof hypotheses are generated in a more complex procedure. Here the so-called colour structure code (Rehrmann and Priese 1998) is used to segment the entire image. Additionally greenish areas are masked out. Only roof hypotheses of a plausible size are selected, additionally the compactness and orthogonality of roof labels are validated. In the last step the grouping of instantiated shadow and roof labels to validated buildings is performed. The neighbouring position of a shadow to a roof has to fulfill the illumination model. The resulting building hypotheses are divided into houses and industry halls using the area of the objects as criterion.

3.2.3 Evaluation of an ATKIS region The results of the structural and textural approach are combined to verify or falsify the checked ATKIS regions. The two approaches lead to different measures of quality. The texture classification, which is a holistic operator, leads to a pixel-wise assignment of classes. The structural approach identifies complex objects by using a combination of different clues and the structure of objects. The considered ATKIS regions are evaluated with an evaluation catalogue, that was designed by use of the ATKIS catalogue and with use of the experience of human operators. Both the structural and textural conditions have to be passed for a validated ATKIS region. Settlement and industrial areas are verified by means of the found buildings. Vegetation classes are falsified if houses or buildings are found in the region. This decision is based on the definition in the ATKIS objects catalogue. Within the vegetation classes forest can be identified and verified. Agriculture is not supposed to contain forest.

4. COMPLETE WORKFLOW

The main feature of the complete workflow for automated quality control is the interaction of automatic procedures with the interactive steps or the decisions taken by the human operator. Orthoimages and the ATKIS DLMBasis are available for the automatic procedures as well as for the human operator. The results of the automatic procedures are passed to the human operator in the form of a traffic light diagnostics. A speed-up of the automated workflow in comparison to a purely interactive workflow is attained by relying upon the objects accepted by the automatic procedures, i.e. the ATKIS objects highlighted with green colour. The human operator has to concentrate exclusively on his final decision for the rejected ATKIS objects, i.e. the red ones. The consequences of this approach are described in Table 2. The percentage of corresponding acceptance when comparing an ATKIS object to the orthoimage indicates the efficiency of the system. Objects that have been accepted by the automatic procedure, but would have been rejected by the human operator, will result in undetected errors when using the workflow as depicted in Figure 2. Their percentage has to be as small as possible since only very few errors should remain undiscovered. For avoiding false alarms, i.e. false negatives according to Table 1, all ATKIS objects rejected by the automatic procedure have to be checked finally by the human operator.

<table>
<thead>
<tr>
<th>Automatic</th>
<th>Green</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Operator</td>
<td>Efficiency</td>
<td>Interactive Final Check</td>
</tr>
<tr>
<td>Green</td>
<td>Undetected Errors</td>
<td>Interactive Final Check</td>
</tr>
</tbody>
</table>

Table 2. Confusion matrix of decisions: human operator vs. automatic procedure, effects (cf. Table 1).

5. EXAMPLES AND RESULTS

We have compared the diagnostics of the automatic system with the results that have been obtained by a human operator without automatic image interpretation. Figure 3 shows an example of a situation that has been detected by the automatic road verification. The paths in the scene have been acquired improperly. Paths are of secondary interest. Nevertheless they are analysed by means of the automatic procedure demonstrating its ability to detect errors. Table 3 subsumes the results of the road verification.
Table 4. Statistics for the verification of built-up areas based on 14 images each of a size of 2 km × 2 km. The percentage relates to the number of ATKIS objects tested.

<table>
<thead>
<tr>
<th></th>
<th>Automatic</th>
<th>Green</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Operator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green</td>
<td></td>
<td>65.7 %</td>
<td>32.0 %</td>
</tr>
<tr>
<td>Red</td>
<td></td>
<td>0.76 %</td>
<td>1.53 %</td>
</tr>
</tbody>
</table>

Table 3. Statistics for road verification. The results are based on the number of ATKIS road objects within rural areas. The test area comprises 14 images each of a size of 2 km × 2 km.

Figure 4 depicts an example showing a development area that has been detected automatically. The statistics for the verification of built-up areas are revealed by Table 4. It indicates a comparatively high number of false positives (cf. Table 1). Most of these cases originate from single buildings only covering a small part, e.g. less than 10 %, of the parcel. We want to solve this by an adaptation of the procedure to these situations.

Figure 4. Example for the verification of built-up areas. Orthoimage and ATKIS DLMBasis (left), automatic detection of errors (right).

6. CONCLUSIONS AND OUTLOOK

The automated verification method as described in this paper is characterised by an efficient workflow of automatic procedures and a final interactive inspection of situations which turned out to be critical in an automatic self-diagnostics. We expect to achieve further improvement of efficiency and of the number of undetected errors since some problems have been identified and solutions are in process. Additionally we plan to test our method with images from high-resolution satellites.

REFERENCES


