SEMI-AUTOMATIC VERIFICATION OF GEODATA FOR QUALITY MANAGEMENT AND UPDATING OF GIS

P. Helmholz [a, *], C. Becker [b], C. Heipke [a], S. Müller [b], J. Ostermann [b], M. Pahl [b], M. Ziems [a]

[a] IPI – Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover,
Nienburger Str. 1, 30167 Hanover, Germany – (heipke, helmholz, ziems)@ipi.uni-hannover.de
[b] TNT – Institute of Information Processing , Leibniz Universität Hannover,
Appelstr. 9A, 30167 Hanover, Germany – (becker, mueller, pahl)@tnt.uni-hannover.de

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

ABSTRACT:

The usefulness and acceptance of geoinformation systems mainly depend on the quality of the underlying geodata. The aim of this paper is to introduce a system for the verification of geodata. The verification of geodata can be seen as part of quality management, and furthermore, of the update of geodata systems (GIS). The presented system compares GIS data with reference data derived from, e.g. high resolution remote sensing imagery using image analysis operators. After the detection of evidence an object will be accepted or rejected by the system. It works semi-automatic, the final decision is made by a human operator. The verification system is presented in detail, and examples are given for various methods for feature extraction. In addition, we illustrate the potential of the system, but we show the limits of this system, too.

1. INTRODUCTION

An immense amount of decisions in private and public life relies on geospatial information. The automatic management of spatial data is conducted in geoinformation systems which present mostly a special part of the real world.

The basic methodology to present the real world in a geoinformation system (GIS) is to define objects using a data model (e.g. a feature type catalogue) which defines objects to be selected, and their properties, structure and rules. In DIN EN ISO 8402: 1995-08 quality is defined as "Totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs". Hence, first the data model must present the real world with sufficient details and without any contradictions (quality of the model), and second, the data must conform to their specification (quality of the data). This paper will focus on the quality of data, especially the verification as a part of quality management and basis for updates.

In Joos (2000) quality measures are introduced based on four criteria – completeness, correctness, accuracy and consistency. Besides these quality criteria, quality measures must be defined, too. A data error can belong to more than one criterion. Furthermore, the criteria are valid for objects (e.g. road) and their attributes (e.g. road type and width). GIS data must comply with all criteria in order to be considered as correct.

The consistency can be checked automatically, i.e. to check if the data comply with the rules of the data model. In this paper we will focus on the criteria completeness, correctness and accuracy. If disjunct area objects and a full area covering GIS are available, the completeness will be checked implicitly by checking correctness and accuracy.

Reference data such as airborne or satellite images are used here to check these criteria. The availability of high resolution optical satellite imagery appears to be interesting for geospatial database applications, namely for the capture and maintenance of geodata. Among others Büyüksalih and Jacobsen (2005) show that the geometry of IKONOS and Quickbird imagery is accurate enough for topographic mapping.

In this paper, we present a system which has the aim to verify GIS data semi-automatically using IKONOS imagery. The aforementioned criteria correctness and accuracy can be checked in this way. After a general introduction to the system in section 2, we describe the automated procedures method. In the last section we give some conclusions.

2. THE SYSTEM

The system was developed in the project WiPKA- QS[1] which was initiated by the German Federal Agency for Cartography and Geodesy (BKG) together with the Institute of Photogrammetry und GeoInformation (IPI) and the Institute of Information Processing (TNT), both at the Leibniz Universität Hannover. The first version of WiPKA-QS was installed at BKG in 2003 (Busch et al., 2004). Since 2003 the system has been permanently enhanced. In this section we first give some background information of ATKIS[2]. Afterwards, we describe the sources and the workflow of WiPKA-QS. Finally, the knowledge-based image interpretation system GeoAIDA which is part of WiPKA-QS is introduced in detail.

2.1 ATKIS

WiPKA-QS was initiated for the automated verification of the German topographic reference dataset ATKIS. ATKIS is a

[1] Wissensbasierter Photogrammetrisch-Kartographischer Arbeitsplatz - Qualitätssicherung (Knowledge-based photogrammetric-cartographic workstation - quality management)

[2] Amtlich Topographisch-Kartographisches Informationssystem (Authoritative Topographic Cartographic Information System)

trademark of the Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany (AdV). The geometry accuracy is 3m.

Components of ATKIS are object based digital landscape models (DLM) encompassing several resolutions. Important DLMs are the DLMBasis with information content comparable to a topographic map of scale 1:25.000, and DLMBund with information content similar to a map of scale 1:50.000. The DLMBasis dataset is free from any kind of cartographic generalisation. In comparison to the DLMBasis in DLMBund some objects are deleted (e.g. small dirt roads), some area based objects are changed into point objects (e.g. power plants), some objects are transformed into other objects (e.g. ditch to watercourse), some objects which are smaller than a threshold are joined with adjacent objects, and some object borders are smoothed. Furthermore, the DLMBund is extended by additional environmental information.

The geodata for ATKIS DLMBasis is collected by every federal state in Germany. The traditional update cycle is 5 years. However, an update of objects with high relevance is to be completed in 3 (e.g. roads), 6 (e.g. airports) or 12 (e.g. wind power stations) months. The BKG merged the geodata of all federal states with the goal of producing a homogeneous dataset for the whole of Germany. At the Geodata Centre of BKG (GDC), the geodata is checked with respect to logical consistency. Furthermore, the BKG is interested in an efficient and independent verification of the data with respect to completeness, correctness and accuracy. This verification is realized in combination with an automated indication of changes in the landscape compared to current ATKIS data. For this reason the BKG has initiated the interdisciplinary project WiPKA-QS.

2.2 Sources in WiPKA-QS

In WiPKA we verify GIS data automatically comparing them with the real world in terms of remote sensing images. To increase the effectiveness airborne and high resolution satellite images are in use. Currently, we mainly employ pan-sharpened IKONOS consisting of orthorectified images with a red, blue, green and infrared channel with a resolution of 1m. Attempts showed that lower resolved images (e.g. 5m) are much less for the verification of ATKIS.

2.3 Workflow of WiPKA-QS

The WiPKA-QS system consists of two components – an interactive GIS component and an automated knowledge-based image analysis component. Furthermore, the interactive GIS component is divided into a pre-processing before and a post-processing step after the image analysis component.

First of all, in the pre-processing step, the necessary sources for the verification systems are defined, which are the ATKIS dataset, the IKONOS images, and in addition a semantic net as knowledge-base for image analysis.

In the automated knowledge-based image analysis component, the verification system compares ATKIS objects of interest with the image data based on the semantic net to collect evidence for the acceptance or rejection of these objects. ATKIS objects of interest are objects which cover large areas (settlement, industrial area, cropland, grassland and forest) or objects where many changes arise (like roads). The knowledge-based image interpretation system GeoAIDA (see section 3) (Bückner et al., 2002) and various methods for feature extraction (see section 3) are the core of the automated procedures.

Currently, the verification of GIS is still far away from being carried out completely automatically. Therefore, the final decision about the rejection of objects is made by a human operator - an interactive post-processing step is necessary. The results of the automatic procedures are passed to the human operator in the form of a traffic light diagnostics. Rejected objects (red) are visually for further editing, whereas it is not necessary for the human operator to take a look at accepted (green) objects.

The workflow of WiPKA-QS is sketched in Figure 1.

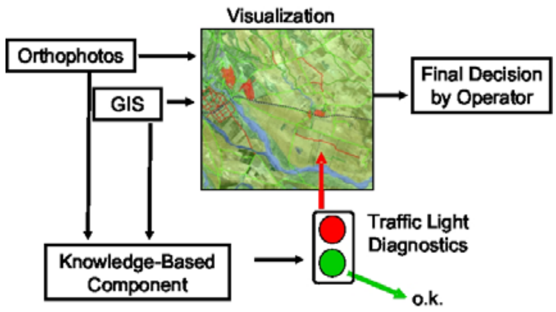

Figure 1: Workflow WiPKA-QS

Further details of WiPKA-QS are available in (Busch et al, 2004).

2.4 GeoAIDA

The GeoAIDA system takes control of the application flow for the automated interpretation of remote sensing images which results in a tree structured symbolic description of the respective scene. Each object detected in the scene corresponds to a node of the tree. The position of an object is represented by a unique identification value in a set of label images. Based on the scene tree and the label images, maps with different level of detail can easily be obtained. An appropriate map is then compared to GIS data to find candidates for objects to be rejected.

The automated recognition of an object of a particular class requires prior knowledge about all classes of objects that are expected to be present or likely to occur in the scene. This knowledge is represented by an object class hierarchy which allows two operators to be assigned to each object class: (1) region of interest operators, i.e. image operators that determine one or more search areas for the respective object class in remote sensing data, and, (2) operators that are applied for evaluation and grouping of child classes.

The analysis strategy follows the hierarchical structure of the knowledge representation. In a first step, a scene presented to the system is subdivided according to the search areas obtained from the region of interest operators of all classes in the object class hierarchy. Subsequently, for each class, hypothetical instances of the respective class are generated for each search area obtained by its region of interest operator. This procedure is repeated recursively from the top to the base of the hierarchy, yielding a hypothetical hierarchy of object instances in the scene. In a second step, starting from the base of this hierarchy, all hypothetical instances are evaluated and subsequently grouped into new instances of the parent class applying the evaluation and grouping operator of the respective parent class. Again, this is repeated recursively until the top of the hierarchy is reached. The hierarchy created during the latter step finally represents the model knowledge applied to the remote sensing data of the scene.

Further details of the system are available in (Bückner et al, 2002).

3. AUTOMATED VERIFICATION PROCEDURES

The core of every automated procedure are image analysis algorithms. In this section we discuss the verification of roads, built-up area objects, and furthermore of grassland/cropland objects.

3.1 Road Verification

The road verification module is designed to check the existence and positional accuracy of roads from a given database. In order to solve this task a region of interest is defined for each road object, depending on its geometric description from the database. More precisely, a buffer around the vector representing the road axis is defined, and the buffer width complies with the overall requested geometric accuracy of the GIS and the road width attribute in the database. If the width 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 an automatic control of the parameters considering the scene specific radiometric properties of the roads and 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 in open landscape as linear objects in aerial or satellite panchromatic imagery with a resolution of about 2 m.

The multispectral information is currently used to generate NDVI-channel as well as intensity-channel images. Therefore, the line extractor is applied twice: Firstly, to extract bright and dark lines in the intensity channel and secondly to extract dark lines in the NDVI-channel. In a subsequent step all the line extraction results are fused for further assessment.

A geometric-topologic relationship model for the roads and local context objects is defined for the assessment. Considered local context objects are rows of trees, which occur frequently in German rural areas. They may explain gaps in road extraction. The automatic rows of trees extraction is based on supervised multi-scale, multi-spectral segmentation of (Gimel'farb, 1996). For the assessment, existing relations between road objects from the database and the extracted objects are compared to modelled relations. Thereby, all extracted objects provide some evidence. If the majority of the total evidence argues for the database object and if a certain amount of this database object is covered by extracted objects, the database object is assumed to be correct, i.e. it is accepted, otherwise it is rejected. For further information concerning the modelled quality assessment refer to (Gerke, 2006).

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. A detailed explanation is given in (Gerke et al., 2004). Results achieved with the Road Verification module are shown in (Gerke, 2006).

3.2 Verification of Built-up Area

Components for the verification of built-up area objects are the textural analysis and the automatic building detection component. If the textural analysis algorithm detects a settlement area, and if in addition the automatic building detection algorithm extracts buildings, the object will be verified as built-up area object.

3.2.1 Textural Analysis Algorithm: The textural analysis uses a segmentation algorithm described in (Gimel'farb, 1996). The algorithm was extended to use a multiresolution technique to segment the image. The classification algorithm has to learn the properties of the classes with manually created training regions for the classes.

The learning steps are:

- Learning of texture with the training areas in four subsampling resolution levels resulting in four parameter files.
- Segmentation of the input image in all resolution levels based on the parameter files
- Evaluation of the segmentation for each class in all resolutions.
- Calculation of an evaluation matrix.

As a result of the learning process four parameter files and an evaluation matrix are derived. The segmentation is done by a top-down operator that begins with the lowest resolution and processes the higher resolutions level by level. It uses the parameters derived from the training areas. The steps of the top-down texture operator are:

- Segmentation of the input image in all resolution levels using the parameter files.
- Calculation of a resulting segmentation using the segmentations in the different resolution levels and the evaluation matrix.

The learning step determines the resolution level on which a class gains significant signatures. From the evaluation matrix we derive in which resolution level a texture can be differentiated. The resolution with the best separation characteristic may differ from one class to another; the classification of inhabited areas is, for example, significantly better in the lower resolutions and therefore preferably used.

The learning step is a crucial part for the effectiveness and correctness of the derived results. This step is preferably done by a human operator, who manually defines training areas for the desired classes. The automatic generation of training areas by the use of GIS data is possible. The training areas for the desired classes can be taken from the regions of a GIS and be used to train the classifier. This has to be done for a few areas, whereas the resulting classification definitions can be used for similar images, e.g. the complete set of images of a flight. 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 it is with manually defined areas.

3.2.2 Automatic Building Detection Algorithm: The approach is divided into a low-level and high-level image processing step. The low-level step includes image segmentation and post-processing: first, the input image is transformed to HSI and the intensity channel is taken as input for a region growing segmentation. The necessary seed points are set flexibly in a fixed raster under consideration of the red channel to prevent setting seed points in shade areas. The segmentation result is post-processed to compensate effects like holes in the regions and to merge roof regions which are split into several parts. The regions are taken as building hypotheses in the following step.

The high-level step includes feature extraction and classification. First implausible hypothesis are rejected by the region area and colour. Afterwards features are calculated for each hypothesis like:

- geometric features
  - object size: area, circumference
  - object form: roundness, compactness, lengthness, angles, etc.
- photometric features:
  - most frequent and mean hue
  - mean NDVI
- structural features
  - shadow, neighbourhoods

Furthermore, the main axes of the hypothesis are calculated. They define a hexagon describing the region’s contour.

The classification works as follows: First all building hypothesis get an evaluation value of 1. For each feature an expected value range is defined for valid building hypotheses.

All features are considered sequentially and hypotheses with feature values outside the value range are multiplied with a weight less than 1. Hypotheses without neighbours get a reduction of 0.1 at the end. The final decision, if a building hypothesis is taken as a correct building is done by a threshold decision.

The building detection algorithm was created for airborne imagery. We currently enhance the algorithm for use with IKONOS imagery.

The evaluation of the algorithm is published in (Busch et al., 2004). Further details of the algorithm are available in (Müller and Zaum, 2005).

3.3 Verification of Cropland and Grassland

The goal of this step is the differentiation of cropland and grassland. A main differentiation between grassland and cropland is the exploitation of structures caused by the cultivation, which is conducted more frequently in crop fields, compared to grassland. The agricultural machines normally cause parallel straight lines which are observable in the image.

For ATKIS objects, one issue is important: Inside one ATKIS object the existence of more than one land cover class is tolerated if a size threshold is not exceeded. Furthermore, several objects of the same land cover type are permitted. For example, in an ATKIS object “cropland” the existence of a small area of grassland is allowed and it is possible that several crop fields with different cultivation directions are present. Therefore, the object must be segmented into radiometrically homogeneous regions before further processing.

Our approach for the detection of parallel straight lines is divided into three steps: we detect edges which then are transformed into Hough space, and finally the orientation is estimated.

The edge image (image space) is transformed to a proper accumulation space (Hough space). The line parameters in image space are the angle between the normal vector of the line and the x-axis ( ), and the distance of the line from the origin (d). These parameters define the Hough space. Thus, parallel lines are mapped into points vertically above each other, assuming the -parameter is mapped to the horizontal axis in Hough space. By extracting these points of interest in Hough image we focus on salient lines in image space.

In the next step, a histogram of the extracted points along the –axis in Hough space is derived. As a final step we fit a Gaussian to the histogram and investigate the resulting standard derivation . For cropland must lie below a pre-defined threshold t, whereas for grassland is assumed to le larger than t. An example of a cropland object is given in Figure 2.

The whole strategy of this approach fails if

- line structures caused by cultivation are not observable (e.g. maize close to harvest, untilled crop fields)
- lines in crop fields are not straight respectively parallel to each other (e.g. on hillsides),
- grassland possesses parallel lines

Regarding the first point, we have to resort to other radiometric features. The last two aforementioned cases are not very common in Germany. However, the influence of these problems will be investigated. More details are available in Busch et al., 2006.

4. CONCLUSIONS

We presented a semi-automatic verification system which was initiated to verify the German topographic reference dataset ATKIS compared with the real world in terms of remote sensing images. This system is in practical use at the German Federal Agency for Cartography and Geodesy. Besides ATKIS further GIS dataset can also be verified. In practical applications at BKG we achieved a speed up of GIS data verification of a factor of three.

First of all, we gave an introduction to the quality management. In the second section, we described the geodata set ATKIS, the knowledge-based image interpretation system GeoAIDA, and furthermore the sources and the workflow of the verification system WiPKA-QS. WiPKA-QS in a nutshell: the system

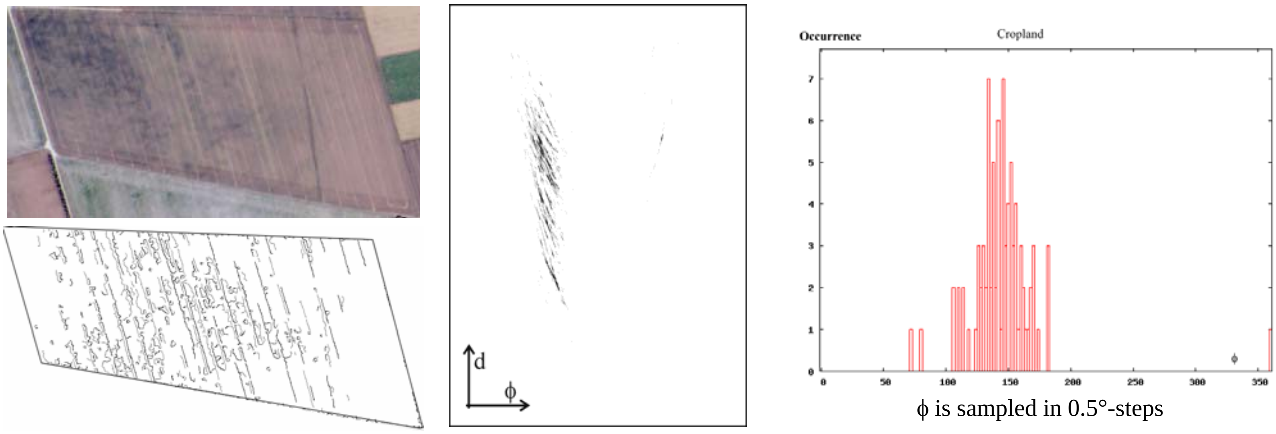

Figure 2: Image (a), Edge Image (b), Hough Space (c) and Histogram (d) of a cropland object

consists of an automated knowledge-based image analysis component and an interactive GIS component. The goal of the image analysis component is to reduce the amount of human interaction which is a time consuming part in the quality control process. Afterwards, challenging situations are analysed and solved by the human operator in a separate step as a part of the interactive GIS component.

Then we focused on the automated procedures method, in detail the road verification, the verification of built-up area objects, and in addition, the verification of cropland and grassland.

We close with a brief outlook to the next steps. The current task in WiPKA-QS is to extend the verification system regarding the discrimination between deciduous and coniferous forests. For the verification of these object classes we use explicit radiometric features as well as structural features as mentioned before by the verification of cropland and grassland.

Furthermore, the possibility of updating is in process. If the system detects a new object of interest, this object will be visualized for the human operator in the interactive post-processing step.

REFERENCES

Bückner, J., Pahl, M., Stahlhut, O., Liedtke, C.-E., 2002: A knowledge-based system for context dependent evaluation of remote sensing data. In: L. J. v. Gool (ed.), *DAGM-Symposium*, Lecture Notes in Computer Science, Vol. 2449, Springer, Zurich, Switzerland, pp. 58–65

Busch, A., Gerke, M., Grünreich, D., Heipke, C., Helmholz, P., Liedtke, C. E.; Müller, S., 2006: Automated Verification of a Topographic Reference Dataset using IKONOS Imagery. In *IntArchPhRS, Band XXXVI/4,* Goa, 2006, S. 134-139

Busch, A., Gerke, M., Grünreich, D., Heipke, C., Liedtke, C.-E., Müller, S., 2004. Automated verification of a topographic Reference dataset: system design and practical results. In: *International Archives of Photogrammetry & Remote Sensing*, Vol.XXXV, Part B2, pp.735–740

Büyüksalih, G. and Jacobsen, K., 2005: Optimized geometric handling of high resolution space images. In *ASPRS annual convention, Baltimore*, 9p (on CD-ROM)

Gerke M., 2006. Automatic quality assessment of road databases using remotely sensed imagery. PhD thesis, Deutsche Geodätische Kommission. Reihe C, Dissertationen, Nr. 599.

Gerke, M., Butenuth, M., Heipke, C., Willrich, F., 2004. Graph supported verification of road databases. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3-4), pp. 152–165.

Gimel'farb, G.L., 1996, Texture Modelling by Multiple Pairwise Pixel Interactions. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, 1110–1114.

Joss, G., 2000: Zur Qualität von objektstrukturierten Geodaten. In *Schriftenreihe Studiengang Geodäsie und Geoinformation, Universität der Bundeswehr München*, Heft 66, Neubiberg, 2000

Müller, S., Zaum, D. W., 2005: Robust building detection in aerial images. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXVI, Part B2/W24, pp. 143-148

Wiedemann, C., 2002. Extraktion von Straßennetzen aus optischen Satellitenbildern. Deutsche Geodätische Kommission. Reihe C, Dissertationen, Heft Nr. 551.

Wiedemann, C., Ebner, H., 2000. Automatic completion and evaluation of road networks. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIII, Part B3/2, Commission III, pp. 979–986.