AUTOMATIC QUALITY ASSESSMENT OF GIS ROAD DATA USING AERIAL IMAGERY - A COMPARISON BETWEEN BAYESIAN AND EVIDENTIAL REASONING

M. Gerke

Institute of Photogrammetry and GeoInformation, University of Hannover
Nienburger Str. 1, D-30167 Hannover, Germany - gerke@ipi.uni-hannover.de

KEY WORDS: Database, GIS, Imagery, Networks, Parameters, Quality, Reliability

ABSTRACT

This paper describes the framework for automatic quality assessment of existing geo-spatial data. The necessary reference information is derived from up-to-date digital aerial images via automatic object extraction. The focus is on roads, as these are amongst the most frequently changing objects in the landscape. In contrast to existing approaches for quality control of road data, a common and consistent modeling and processing of the road data to be assessed and the road objects extracted from the images is carried out. A geometric-topologic relationship model for the roads and their surroundings is set up. The surrounding context objects (for example rows of trees, or rows of buildings) support the quality assessment of road vector data as they may explain gaps in the extracted road network. Algorithms are defined for the evaluation of existing relations between extracted objects and the database road objects and thus quality measures are yielded. Mostly, more than one extracted object gives evidence regarding one database object. Therefore, the gained quality measures have to be combined in order to reach an overall quality value for the respective object. In the present work two approaches are used for this reasoning and are compared: a probabilistic one and an approach based on the Dempster-Shafer-Theory of Evidence. Results carried out on real and simulated data show that the overall approach is both reliable and efficient. Both models for the reasoning have major differences, however, differences between the results from both approaches only show up in some cases.

1 INTRODUCTION

This paper describes the framework for automatic quality assessment of existing geo-spatial data. Quality comprises completeness, positional accuracy, attribute correctness and temporal correctness for each object (Zhang and Goodchild, 2002). By means of quality assessment the database objects are compared to the reference: the positional accuracy and the attribute correctness can be checked using the extracted objects. The completeness and temporal aspect is only partly considered, as only commission errors are identified. During a following update process, new or modified road objects not included in the database are extracted. By this means also completeness and temporal correctness are fully considered. In the present paper only the quality assessment is addressed.

The background of this work is given by a project carried out in conjunction with the German Federal Agency for Cartography and Geodesy (BKG). Here a semi-automatic quality control system of the official German spatial reference data is developed. Further information on this project can be found in (Busch et al., 2004).
The necessary reference information is derived from up-to-date digital aerial images via automatic image analysis. The focus is on roads as these are amongst the most frequently changing objects in the landscape. In contrast to existing approaches for quality control of road data, a common and consistent modeling and processing of the road data to be assessed and the road objects extracted from the images is carried out. A geometric-topologic relationship model for the roads and their surrounding context objects is defined. If for instance aerial images are captured in summer, trees along roads hamper the road extraction as the road surface is not directly visible. The extraction and explicit incorporation of those context objects in the assessment of a given road database gives stronger support for or against its correctness.
In this paper the uncertainties inherent in existing geo-spatial data and extracted objects are modeled. The sources of uncertainty are investigated and a statistical model for the given task is defined. The adequate consideration of the statistical properties of the extracted objects is of vital importance, because only by this means it is possible to judge the amount of evidence the extracted objects can give regarding the quality of existing vector data. The focus is on the question of how to combine the evidences given by extracted objects in order to derive a quality measure for a given GIS road object. Two approaches are introduced: a traditional probabilistic one and an approach based on the Dempster-Shafer Theory of Evidence.
Road and context object extraction from imagery goes beyond the scope of this paper; rather the modeling and statistical reasoning for an automatic quality control of given road vector data using such extracted objects is the topic of the present paper.

2 RELATED WORK

With the advent of highly detailed and accurate vector data sets (not being generalized), an automated quality control applying a direct comparison between given objects and extracted objects has become possible. Such vector data sets are for example available in Germany (ATKIS DLMBasis), in France (BDTopo, in the near future to be replaced by RGE) and in Great Britain (OS Mastermap).

In (de Gunst, 1996), a very detailed road model is formulated (focusing on highways). Input data from a road database is used to define the search space for the road extraction. After a detection of road markings, these are grouped into carriageways afterwards. This approach relies on relatively precise data as the only inconsistencies being handled are changes in the road properties: additional carriageways (or a different width from the one registered in the database) and new exits are detected.
A verification and update of the ATKIS DLMBasis is described in (Plietker, 1997). Lines are extracted in the imagery and grouped afterwards. If these lines correspond well to the given vector data in direction and distance, the vector data is assumed to be correct. If this attempt does not give enough evidence, the next step

consists of a verification of the assumed road object by analyzing the region around the line (homogeneity). After this object-based verification, the given network topology is exploited: if a rejected object is connected to two accepted objects, a new connection hypothesis is created considering a possible change of attributes or position. Results from this approach are not presented.
In the German WiPKA-QS[1] project, road objects from the DLM-Basis are verified. The verification system is restricted to open landscape areas (Gerke et al., 2004). Knowledge from the database is used mainly in two ways: firstly, the landscape objects contained in the database are used to define global context regions (open landscape, forest, built-up). Secondly, road objects define the region of interest for the road extraction (considering the nominal positional accuracy of +/- 3m) and support the grouping of extracted lines as they are also used as seed vectors. The developed 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. Results show that the approach works well in open landscape areas if the impact from disturbing context objects is limited.
In (Goeman et al., 2005), a given road network is assessed using image statistics. By means of a buffer overlay algorithm the degree of correspondence between lines extracted in imagery and given road vectors is achieved. The quality of road extraction is estimated using image information; a geometric road model is not used.

3 APPROACH

The literature review reveals that an approach fulfilling a substantial quality control of road vector data does not exist. De Gunst (1996) shows results using simulated vector instead of information from an existing database, Plietker (1997) does not consider imprecise road vectors at all, and Gerke et al. (2004) use the vectors from the database as seeds without considering their accuracy. Goeman et al. (2005) use a buffer approach for the assessment, which is generally not able to reflect the quality very well; shape and position can not be assessed separately. Moreover, context objects which could explain gaps in the extracted road network are considered only marginally in all recent works for quality control of road vector data.
The approach introduced here has two major characteristics which address the deficiencies of the existing works: a) a sufficiently detailed modeling of roads, context objects and the relations between them, and b) an integrated statistical modeling and reasoning.

The modeling of the relations and the concentration on statistical models and reasoning is of elementary relevance for the assessment of the quality of road vector data, because one does not actually have a real *reference* for this task. The only references one can use are automatically extracted objects. Therefore, an approach which is able to statistically evaluate relations between given road vector data and extracted objects and finally compare them to a given model seems to be a means to overcome the obviously unsolvable problem. The approach presented here follows the maximum likelihood/maximum support principle: if there is more evidence for the conformance of the observations (i.e. extracted objects) and given vector data regarding the model than against it, the respective database object is assumed to be correct (accepted), otherwise it is assumed to be incorrect (rejected).

Objects from a highly detailed vector database like the ATKIS DLMBasis are not generalized and thus must maintain the same relations as in reality. The use of those relations between roads and context objects is a reasonable means for the explanation of gaps in road extraction. Therefore, a geometric-topologic relationship model for the roads and their surroundings is defined (refer to section 4.1).
An object model for the geometry and the uncertainty inherent in extracted objects will be given in section 4.2. This modeling is a necessary requirement for the stochastic determination of whether an extracted object and a given GIS road object correlate with the given relationship model. Algorithms for the calculation of these measures are defined in section 5. As mostly more than one extracted object gives evidence regarding one database object, an approach to reasoning must be able to collect and balance evidences given by the named algorithms and finally infer the quality of the given road vector data (section 6). A focus of this paper is on the question of whether a probabilistic (Bayesian) approach serves better for this reasoning than an approach based on the Dempster-Shafer Theory of Evidence.
Data from the German ATKIS DLMBasis is used as an example. The transfer of the methods to similar data sets is possible without any problems. The function of a road object within the whole road network will not be considered and exploited here. The network properties can be incorporated using an approach as introduced in (Gerke et al., 2004). Questions concerning coordinate system transformations or unknown scale or orientations between given objects are not treated. Therefore, all objects need to be given in the same coordinate system.

4 MODEL

4.1 Object Classes and Relations Between Objects

In the relationship model the geometric and topologic relations between an ATKIS road object, the local context objects and the extracted road objects are given (ref. to Fig. 1). The relationship model distinguishes between objects to be assessed (*ATKIS Carriageway Object*), objects which directly give evidence (*Extracted Road Object*) and context objects (*Local Context Object*). The model is independent from the global context, i.e. the appearance of objects in different environments. Therefore, global context knowledge must be considered by the respective object extraction algorithm.
In contrast to many other definitions geometry does not comprise the position of an object, but the shape and orientation. The relative position of objects is modeled by the topologic relation.
The geometric relations *same shape* and *same orientation* express the fact that the course of an ATKIS road object and the respective extracted object needs to be similar (shape) and that both objects must have the same orientation.
The topologic relation is important for this work due to the fact that for example rows of trees (the stems of the trees) must be located outside the carriageway given in ATKIS whereas an extracted road (the surface of the road) must be contained in the ATKIS carriageway. The topologic relations considered are *disjoint* and *contains*. The latter one is defined relative to the ATKIS object. Besides this qualitative topologic relation one may define side conditions. For *disjoint* it is often desirable to give a minimum and a maximum distance (d_min, d_max). For example a row of trees must have a minimum distance to the carriageway (due to security reasons) and it is also expected that trees having a distance to the carriageway larger than a certain value are not suitable to explain gaps in the road extraction, i.e. they do not cover the road in aerial imagery. For *contains* additionally an *identical width* of objects may be required. The relations between the *ATKIS Carriageway Object* and the *Local Context Object* and the respective values given in the depicted relationship model are

[1] Wissensbasierter Photogrammetrisch- Kartographischer Arbeitsplatz zur Qualitätssicherung (Knowledge-based Photogrammetric-Cartographic Workspace), cf. (Busch et al., 2004)

defined by experience and common knowledge. An alternative way to find the measures would be to incorporate official specifications, for instance from road construction.

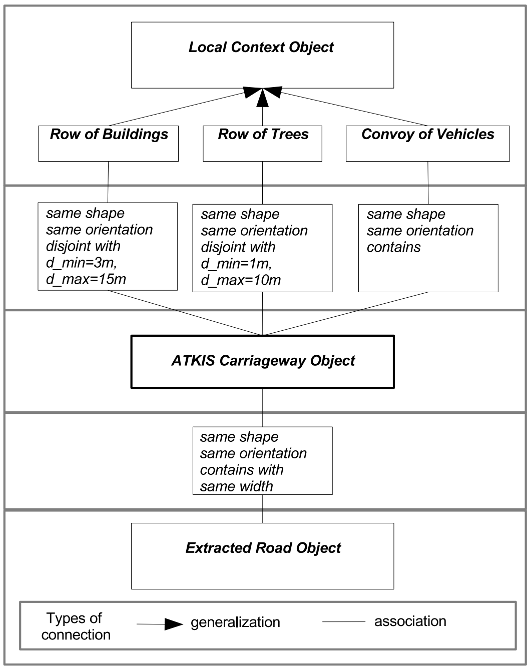


Figure 1: Relationship Model

4.2 Objects: Geometry and Uncertainty

Geometry Only elongated extracted objects are considered in this approach. This enables a direct comparison of the shapes (section 5.1) between those objects and GIS road objects. An extension to arbitrarily shaped extracted objects is possible, but then the definition of geometric relations is ambiguous. Two components are of importance for the assessment: the borders of the objects are used to evaluate topologic relations and the middle-axis is used to compare shapes and orientations. Therefore the geometric model should allow a conversion between these two representations.
In the object model four borderlines $R_{L1,2}, R_{Q1,2}$ and a middle-axis M are defined, refer to Fig. 2. The borders $R_{Q1,2}$ are optional. The derivation of M from given $R_{L1,2}$ is explicit:
Any Point $P_i \in M$ is center of a circle with radius $r = \frac{1}{2}B_i$ which touches R_{L1} and R_{L2}. As a separation between the borders is explicitly given, there will be no junctions inside the middle-axis, compared to the skeleton of a region.

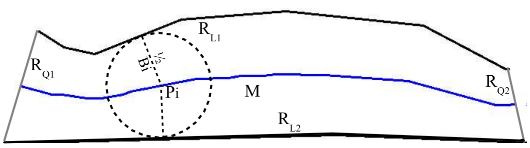


Figure 2: Object Model - Geometry

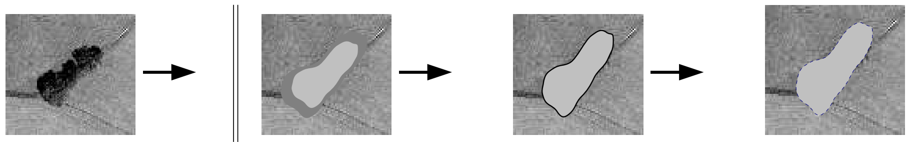
Roads from ATKIS DLMBasis or from many road extraction algorithms are modeled as middle-axes, including a constant width B as attribute. In this case the borders $R_{L1,2}$ are derived by means of moving M in a perpendicular direction by $\frac{1}{2}B$. The borders $R_{Q1,2}$ are the connection between the respective starting/end points of the $R_{L1,2}$. It is reasonable to define the objects in discrete space (i.e. in the raster domain) as they are captured from digital imagery. If required, a conversion into Euclidean space is possible.

Uncertainty In this paragraph the aspects to the uncertainty of extracted objects are identified and a possible means to model them statistically is shown. Though the extraction of objects is not treated in this paper, it makes sense to have a look at possible impacts on the object's uncertainty. The interesting issue here is that not all aspects found in the following have a direct impact to the assessment of both kinds of relations. If an object is shifted by an unknown value (see *virtual object* below), this shift has no impact on the geometric relation (shape and orientation).
According to Glemser (2001), the capture of an object can be classified in three single steps: modeling, abstraction and measurement. Here, the given definitions are slightly modified for 2D image analysis and object extraction (given oriented imagery).
In the given context, the model for the capture comprises the mathematical process of mapping objects from image space to object space, i.e. transformation. For this approach it is presumed that the respective parameters are known accurately enough, but 3D-objects such as houses or trees are also incorporated in the assessment of road objects. These objects have to be given in their projection in 2D (footprints). A resulting offset of the position in 2D has to be considered in the statistical modeling, because some extraction algorithms do extract 3D-objects from imagery where the height of objects above the terrain was not considered during orthogonal projection (leading to an offset in x-y-plane).
The intermediate result after the mapping is called virtual object as up to this step no concrete object definition was applied. This happens in the abstraction step, where an operator or an algorithm has to decide which parts of the mapped virtual object belong to a certain object class. This abstraction can be understood as a generalization, and thus a notion of uncertainty is introduced. The measurements taking place on these abstracted objects propagate this uncertainty. Table 1 summarizes these definitions for areal objects and gives an idea how to model the impacts statistically, i.e. with which type of density function. It is important to note that the assumed distributions are an ideal but also reasonable view of the world. (For instance, the assumption that the final measurement can be modeled by a normal distribution).
The statistical parameters for the objects have to be given by the respective extraction operator (depending on algorithm and input data). Often these values can just be estimated. As the algorithms for the assessment of the relations (next section) need both middle-axis (orientation/shape) and border (topology) the transfer of the statistical measures from the borders to the middle-axis need to be done applying variance propagation. In the case that the objects are not given explicitly as areas, but as middle-axis plus width (for example from road extraction), the statistical measures for the borders also need to be estimated using variance propagation.

5 ASSESSMENT OF GIS OBJECTS

Any extracted object E_i gives evidence regarding the question of whether the modeled relations between this object and the given ATKIS road object are maintained. The methods introduced in this section are formulated to achieve those evidence-measures per object. The combination of all evidence given regarding one AKTIS road object is the topic of the next section.
Four interesting measures are identified for the assessment:
$\mathbf{p_{g_i}}$: Probability that the required *geometric* relations – shape and orientation – are kept, given the object specific quality measures.

$\mathbf{p_{t_i}}$: Probability that the required *topologic* relation is maintained,



	Modeling (virtual object)	Abstraction (discrete object)	Measurement (captured object)
In 2D image analysis (given oriented imagery)	Mapping from image space to object space	Definition of points (pixels) to be captured (e.g. classification/segmentation)	Capturing of points/pixels
Impact on uncertainty	Possible unknown offset of the whole segment in case of 3D objects	Fuzzy definition of the object's border	Uncertainty in measurement
Type of possible density function	Uniform distribution, parameter (radius) depends on possible object height and image size	Mix of normal and uniform distribution	Normal distribution
Influences ... relation	topologic	topologic and geometric	topologic and geometric

Table 1: From image to object - possible influences on uncertainty for 2D image analysis

given the object specific quality measures.

$\mathbf{qcov_i}$: Length of the projection of E_i onto the ATKIS road object related to the overall length of the ATKIS road object (cf. Fig. 3). This factor is important in order to limit the impact of E_i for the quality assessment. Imagine an extracted road which covers 80% of the ATKIS road and the geometric and topologic relations fit well to the model. This object should have more impact on the quality assessment than for example a row of trees which just covers about 10% and perhaps indicates less quality.

$\mathbf{pcon_i}$: Confidence measure: many object extraction algorithms apply an internal evaluation of the results. This measure should be used for the assessment.

All measures are defined in $[0, 1]$. The object E_i is only considered for the assessment of a given ATKIS road object if the value of p_{t_i} is larger than zero. By this means, the value of p_{t_i} also serves as assignment criterion.

5.1 Assessment of Geometric Relations

The probability p_g that the geometric relations between a given ATKIS road object and any given extracted object correspond to the model is composed of the two components $p_{g-shape}$ and $p_{g-orientation}$. Finally, if both geometric properties are required, p_g is the product of those two probabilities (independence is assumed). For the assessment of geometric relations the middle-axis M of the objects is considered. Its positional accuracy is given as σ_M, the standard deviation of a normal distribution, see section 4.2.

Shape D_i describes the distance between a point on the extracted object E and its nadir point on the given GIS road object A, cf. Fig 3. The distribution of the D_i about its mean value $\bar{D}$ is a measure for the local similarity of both shapes, called s_D. A theoretical value for this variance can be calculated when identical shapes – considering σ_E and σ_A –are assumed, leading to σ_D. Hence, using s_D and σ_D, the probability $p_{g-shape}$ indicating whether the measured variance of distances is identical to the theoretical one can be calculated.
In Fig. 3 the coverage $qcov_i$, defined above, can also be identified: It is the relation between the length of A' and the length of A. A' is the projection from E onto A.

Orientation In order to be able to calculate piecewise orientations the objects need to be represented by a line-string. This conversion is done by a quantization, where the sample distance

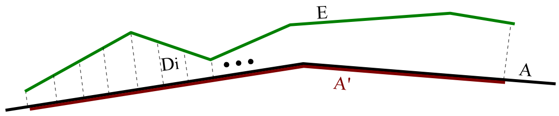


Figure 3: Assessment of shape

depends on local variances as well as on the statistical measures of the middle-axes.
Based on the line-string representation, an overall orientation of the given ATKIS road object and the extracted object is calculated, including a certainty measure (standard deviation). The difference of both orientations needs to be zero or maximum tolerance value, if given. The value of $p_{g-orientation}$ depends on whether the measured orientation difference is larger than this required difference (i.e. it is the outcome of a significance test).

5.2 Assessment of Topologic Relations

For the examination of the topologic relations the approach presented in (Winter, 1998) is applied. In that work the topologic relations between imprecise and uncertain regions are assessed, considering any density function for the respective object's borders. Winter shows that all eight topologic relations two objects may undergo can be derived from the minimum and maximum distance between so-called certain zones of both objects. All relations modeled above can be assessed by this approach. Three distance classes (minus, zero, plus) are defined and based on the given density functions for the object's borders the probability for the class membership of the minimum and maximum distances are derived. All topologic relations can be mapped to a concrete class membership of both distances. Its probability can be calculated using the derived class membership probabilities. Hence, the probability p_t that a given pair of objects maintains the modeled topologic relation can be achieved by this means.
The value of p_t is also influenced by the width of the two objects in the case that the side condition *identical width* is given for the relation *contains*. The difference of widths must be zero, but the certainty of the widths measure must also be considered. The probability that this difference is zero is derived, and finally leads to a new value for p_t.

6 COMBINATION OF EVIDENCE

Any extracted object E_i which is assigned to a GIS road object A ($p_{t_i} > 0$) allows a conclusion $\xi_i = 1$ which states that E_i and

A maintain the modeled relations. The probability of whether ξ_i = is true (p_i^+) or false (p_i^-) depends on the collected measures. The most important criterion to describe the similarity of two objects is p_{g_i}. The other quality values serve as weighting factors: $\alpha_i = p_{t_i} \cdot qcov_i \cdot pcon_i$.
This leads to: $p_i^+ = p_{g_i} \cdot \alpha_i$ and $p_i^- = (1 - p_{g_i}) \cdot \alpha_i$
Two hypotheses are defined:
$\mathbf{H}^+$: the GIS road object is correct given the observed data, i.e. the extracted objects and the GIS road object maintain the modeled relations
$\mathbf{H}^-$: the GIS road object is not correct given the observed data, i.e. the extracted objects and the GIS road object do not maintain the modeled relations

An approach combining all conclusions $\xi_1 \ldots \xi_n$ related to a GIS road object A must consider the specific probabilities and finally infer the quality of A, permitting an overall assessment conclusion, i.e. approve H^+ or H^-.

6.1 Probabilistic Approach

The combination of the given ξ_i can be done using an Bayesian approach, although some questions remain (see remarks below). Here a discrete problem is given, as the set of possible values (unknowns) is fixed: $\Theta = \{\theta_1 = H^+, \theta_2 = H^-\}$ and ξ has the constant value 1. The à-priori-distributions $\pi(\theta_j)$ can also be given as probabilities: $\pi(\theta_j) = \pi_i$ for $j = 1, 2$.
The conditional probabilities for the correctness of the statement $\xi_i = 1$ are given by p_i^+ and p_i^-:

$$\begin{aligned} p(\xi_i = 1 | \theta_1 = H^+) &= p_{g_i} \cdot \alpha_i = p_i^+ \\ p(\xi_i = 1 | \theta_2 = H^-) &= (1 - p_{g_i}) \cdot \alpha_i = p_i^- \end{aligned}$$

The ξ_i are assumed to be independent, therefore the combined probabilities for the correctness of θ_1 and θ_2 are:

$$w(\xi_1, \ldots, \xi_n | \theta_j) = \prod_{i=1}^{n} p(\xi_i | \theta_j)$$

Finally, the à-posteriori-probability for the unknowns θ_1 and θ_2 is:

$$\pi(\theta_i | \xi_1 \ldots \xi_n) = \frac{w(\xi_1, ..., \xi_n | \theta_i)\pi_i}{w(\xi_1, ..., \xi_n | \theta_1)\pi_1 + w(\xi_1, ..., \xi_n | \theta_2)\pi_2}$$

Whether a given GIS road object is accepted depends on the fulfillment of $\pi(\theta_1 | \xi_1 \ldots \xi_n) > \pi(\theta_2 | \xi_1 \ldots \xi_n)$ and the attainment of a given minimum total coverage percentage.

Remarks Two issues remain unanswered so far if the Bayesian approach is applied: The choice of the à-priori-probabilities and the consideration of ignorance. Both issues are closely related. The given pieces of evidence p_i^+ and p_i^- must be allocated completely to the possible hypotheses in this Bayesian framework, although in most cases one extracted object cannot describe the quality of the whole given ATKIS road object as it seldom covers the whole object. Thus à-priori probabilities are introduced in order to give an idea about the GIS road quality. This is an interesting issue here: the given object should be assessed objectively and the final result should be independent from assumptions concerning the quality. Regarding the choice of à-priori probabilities, Jeffreys (1961) states (citing from (Kass and Wasserman, 1996)):

> ...if there is no reason to believe one hypothesis rather than another, the probabilities are equal ...if we do not take the prior probabilities equal we are expressing confidence in one rather than another before the data are available ...and this must be done only from definite reason.

For the given task this means $\pi_1 = \pi_2$. Thus, these à-priori-values are not considered in the calculation of the à-posteriori-probabilities. However, these non-informative priors can not represent ignorance. An idea of whether this modeling is nevertheless adequate for the given problem, is given with the examples.

6.2 Evidential Approach

The Hint-Theory (H-T) is an approach to the Dempster-Shafer Theory of Evidence; its fundamentals can be found in (Kohlas and Monney, 1995). The measure to what extent a hypothesis is proved by the Hint $\mathcal{H}$ is called support (degree of certitude). The extent to which there is no disagreement to a hypothesis is called plausibility. The interpretations of support and plausibility are very close to Dempster's theory of upper and lower probability. One interesting difference from the Bayesian approach is the possibility of formulating ignorance and therefore a specification of à-priori knowledge is not required.

In H-T a so-called frame of discernment Θ is defined which contains all possible answers to a certain question. A Hint $\mathcal{H}$ is defined as the quadruple $\mathcal{H} = (\Omega, P, \Gamma, \Theta)$; $\Omega = (\omega_1, \ldots, \omega_m)$ represents the set of all possible interpretations of the information contained in the Hint. Each interpretation permits restricting the possible answers to a non-empty subset $\Gamma(\omega_i)$ of Θ. These sets $\Gamma(\omega_i)$ are called focal sets of the Hint. The precision of every interpretation ω_i is represented in its probability $p_i \in P$. The probabilities for the interpretations given by one Hint must sum to 1.

Here Θ contains both hypotheses: $\Theta = \{H^+, H^-\}$. Any given conclusion $\xi_i = 1$ can be interpreted as a Hint $\mathcal{H}_i$, cf. Table 2.

Ω	Γ	P
ω_i^+	$\{H^+\}$	$p_{g_i} \cdot \alpha_i = p_i^+$
ω_i^-	$\{H^-\}$	$(1 - p_{g_i}) \cdot \alpha_i = p_i^-$
ω_i^Θ	Θ	$1 - p(\omega_i^+) - p(\omega_i^-) = 1 - \alpha_i$

Table 2: Hint $\mathcal{H}_i$

The last interpretation (ω_i^Θ) represents the ignorance. By means of applying Dempster's Rule all Hints referring to a GIS road object can be combined into an overall Hint:

$$\mathcal{H}_{c1\ldots n} = (\ldots(\mathcal{H}_{c12} \oplus \mathcal{H}_3) \oplus \mathcal{H}_4) \oplus \mathcal{H}_5 \ldots) \oplus \mathcal{H}_n$$

Finally, the support sp and plausibility pl for both hypotheses can be derived:

$$\begin{aligned} sp(H^+) &= p(\omega_{c1\ldots n}^+), sp(H^-) = p(\omega_{c1\ldots n}^-) \\ pl(H^+) &= 1 - sp(H^-), pl(H^-) = 1 - sp(H^+) \end{aligned}$$

Similar to the probabilistic approach, the final decision of whether a given GIS road object is accepted depends on the fulfillment of the condition that $sp(H^+) > sp(H^-)$ and the attainment of a given minimum total coverage.

7 RESULTS

Two sets of ATKIS road data have been prepared: set A only contains objects with a correct geometry. For set B the correct objects have been rotated in order to obtain incorrect geometries. Each set contains 125 ATKIS road objects. The width of the road objects is given as an attribute in ATKIS. In both sets not all values for the width are correct. It is also a purpose of this test to check if the approach is able to find the incorrect ones.

The extracted road objects are obtained by the approach presented in (Gerke et al., 2004). The examples are restricted to open landscape areas, because the road extraction algorithm is not able to reliably extract roads in built-up areas. The parameters are trimmed for a very strict road extraction, because the influence from artifically inserted road segments (due to automatic gap bridging) should be very low. Those gaps are often caused by vegetation and the intention of the following experiments is to test if explicitly inserted context objects give adequate evidence. The rows of trees representing context objects here are captured manually (line representation). As the positions of the stems are crucial for the topologic relation, the width is set to $1\,m$. In Table 3 the assumed statistical properties of all involved object classes are given, where Δ stands for the radius of uniform distribution and σ for the standard deviation of a normal distribution, given in $[m]$. The values given for the extracted roads are related to the ground sample distance of the used imagery, which is $1\,m$. The reliability of all extracted and captured objects $pcon$ is set to 1. The *width* is also observed by the road extraction algorithm.

	Modeling	Abstraction	Measurement
ATKIS road objects	-	$\Delta = 3$	-
Extracted Roads	-	$\Delta = 0.5$	$\sigma = 0.5$
Rows of Trees	$\Delta = 3$	$\Delta = 2$	$\sigma = 0.3$

Table 3: Statistical properties of objects for experiments

In Table 4 the results for the assessment are shown. The first experiment (upper half) was carried out to analyze if the maximum likelihood, respectively the maximum support rule does reflect the object quality. In practice the decision on whether a GIS road object is accepted or rejected is made based on this rule and on the requirement that a minimum overall coverage has to be reached. Therefore in the second experiment (lower half) this threshold has been set to 90%. The results shown are not only separated by the type of applied reasoning (Bayesian/Evidential) but also by the type of objects giving the evidence: *green* denotes that just the extracted road objects are considered; *yellow* denotes that the context objects (rows of trees) were additionally included in the assessment. Moreover the experiments have been applied twice for each set of ATKIS data: with and without the requirement that the widths of the extracted road object and the given ATKIS road object need to be equal.

The results allow a closer look at some aspects:

Efficiency: When identity of the widths is not required for every object from set A the probability for H^+ is higher than for H^- (first row) for every object. This number decreases when identity of the widths is required (second row). In the simulation of a practical application case where the threshold for the minimum overall coverage is set to 90%, the number of accepted objects decreases to about 65% (green) and to about 70% (yellow). In most cases this can be explained by the road extraction algorithm and the chosen parameters: in order to reduce false positive extractions the contrast between roads and background objects must be relatively high for road detection. The upper image in Fig. 4 shows a typical example: only about 23% of the ATKIS road object are covered by extracted road object, the rows of trees cover about 54% of the AKTIS road object. The probability of hypothesis H^+ and the support for this hypothesis is higher than for H^- if extracted objects are considered and if rows of trees are incorporated. However, this object was rejected as the overall coverage is less than 90%.

Bayesian vs. Evidential Reasoning: In some cases where context objects are involved the Evidential Approach seems to be more stringent, but in fact these are discrepancies due to the different handling of ignorance. An example is given with the lower image of Fig. 4. The three rows of trees cover about 96% of the ATKIS road object. The row of trees in the middle gives evidence against the correctness of the ATKIS object as both shapes differ significantly; p_g is 0.03 (cf. Tab. 5). Moreover, it covers about 51% of the ATKIS object. As p_t is higher than the p_t from the other two objects it is reasonable to reject the ATKIS object. Following the Bayesian approach, the ignorance contained in the observations is allocated uniformly to both Hypothesis, this leads here to a maximum probability for H^+, although the measured evidences do not support this decision. In contrast, in the Evidential approach the ignorance is propagated through the measures (overall ignorance from the rows of trees for the object is $1 - sp(H^+) - sp(H^-) = 1 - 0.06 - 0.11 = 0.83$) and thus the maximum support rule leads to a rejection of the respective ATKIS object.

Consider maximum likelihood/support:

	Bayesian Combination		Evidential Combination	
	green	yellow	green	yellow
SET A: correct ATKIS road objects				
ident. width req.:no	125	125	125	121
ident. width req.:yes	109	109	109	102
SET B: incorrect ATKIS road objects				
ident. width req.:no	14	18	14	15
ident. width req.:yes	12	16	12	13

... and minimum total coverage of 90%:

	Bayesian Combination		Evidential Combination	
	green	yellow	green	yellow
SET A: correct ATKIS road objects				
ident. width req.:no	81	89	81	87
ident. width req.:yes	68	76	68	72
SET B: incorrect ATKIS road objects				
ident. width req.:no	0	3	0	3
ident. width req.:yes	0	3	0	3

Table 4: Assessment results for 125 ATKIS road objects. Upper half: number of objects where H^+ is more likely, lower half: number of objects, where H^+ is more likely and total object's coverage is larger than 90%

Row of trees - No.	$\mathbf{p_{g_i}}$	$\mathbf{p_{t_i}}$	$\mathbf{qcov_i}$
1 (left)	0.93	0.11	0.31
2 (center)	0.03	0.25	0.51
3 (right)	0.55	0.11	0.15

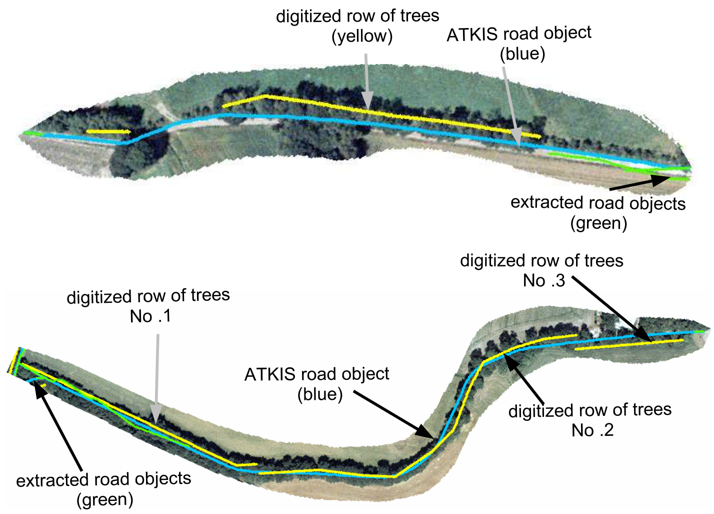
Table 5: Quality measures from rows of trees (cf. bottom object in Fig. 4)

Reliability: Even though the probability for the 'wrong' Hypothesis is higher than for the correct in some cases, all incorrect objects have been rejected in the second experiment (green). Some incorrect objects have been accepted if the rows of trees are incorporated (yellow), because although the ATKIS road dataset has been rotated, some rows of trees fit well enough to some objects.

8 CONCLUSIONS AND OUTLOOK

This paper describes a framework for the assessment of existing road databases. In contrast to existing works, it includes a detailed statistical and relation modeling of all involved objects. Regarding extracted objects, not only road objects are considered, but also context objects which are able to explain deficient road extraction are incorporated in the quality assessment.

Results show that the goals striven for have been reached: high reliability and efficiency. The differences arising from using a probabilistic or an Evidential approach for the combination of evidences given by extracted objects are explainable by the different kind of handling ignorance contained in the observations.



Measure	Object 1 (upper)	Object 2 (lower)
Extracted roads		
Total coverage	0.23	0.22
$\pi(H^+\|\xi_1 \ldots \xi_n)$ / $\pi(H^-\|\xi_1 \ldots \xi_n)$	≈ 1 / ≈ 0	≈ 1 / ≈ 0
$sp(H^+)$ / $sp(H^-)$	0.029 / 0.0041	0.03 / 0.003
Rows of trees		
Total coverage	0.53	0.96
$\pi(H^+\|\xi_1 \ldots \xi_n)$ / $\pi(H^-\|\xi_1 \ldots \xi_n)$	0.96 / 0.04	0.94 / 0.06
$sp(H^+)$ / $sp(H^-)$	0.034 / 0.024	0.06 / 0.11
Extracted roads and rows of trees		
Total coverage	0.76	0.98
$\pi(H^+\|\xi_1 \ldots \xi_n)$ / $\pi(H^-\|\xi_1 \ldots \xi_n)$	≈ 1 / ≈ 0	≈ 1 / ≈ 0
$sp(H^+)$ / $sp(H^-)$	0.06 / 0.03	0.09 / 0.11

Figure 4: Exemplary results for three objects, probability / support values separated by type of extracted object. Values smaller than $1 \cdot 10^{-8}$ are listed as ≈ 0, values larger than $1 - 1 \cdot 10^{-8}$ are listed as ≈ 1.

In case the observations are quite balanced regarding the question whether the ATKIS road object is correct, this different modeling is the crucial factor; the Evidential approach is more realistic (correct) here.
The examination of context object extraction algorithms is not treated in this paper, but nevertheless a very important issue for the future.
The presented results are not only interesting for road data assessment, also an efficient automatic road data update benefits from this approach: one may assume that new roads are connected to the existing ones. The quality values gained here can be directly incorporated into this process.

ACKNOWLEDGMENT

This work is funded by the German Federal Agency for Cartography and Geodesy (BKG).

REFERENCES

Busch, A., Gerke, M., Grünreich, D., Heipke, C., Liedtke, C.-E. and Müller, S., 2004. Automated Verification of a Topographic Reference Dataset: System Design and Practical Results. In: IAPRS, Vol. 35, pp. 735–740. Part B2.

de Gunst, M. E., 1996. Knowledge-based Interpretation of Aerial Images for Updating of Road Maps. PhD thesis, Netherlands Geodetic Commission Publications on Geodesy, TU Delft. No. 44.

Gerke, M., Butenuth, M., Heipke, C. and Willrich, F., 2004. Graph Supported Verification of Road Databases. ISPRS Journal of Photogrammetry and Remote Sensing 58(3-4), pp. 152–165.

Glemser, M., 2001. Zur Berücksichtigung der geometrischen Objektunsicherheit in der Geoinformatik. PhD thesis, Deutsche Geodätische Kommission. Series C, Vol. 539.

Goeman, W., Martinez-Fonte, L., Bellens, R. and Gautama, S., 2005. Using Image Statistics for Automated Quality Assessment of Urban Geospatial Data. In: IAPRS, Vol. 36. 8/W27.

Jeffreys, H., 1961. Theory of Probability. 3. edn, Oxford University Press.

Kass, R. E. and Wasserman, L., 1996. The selection of prior distributions by formal rules. Journal of the American Statistical Association 91(435), pp. 1343–1370.

Kohlas, J. and Monney, P., 1995. A Mathematical Theory of Hints. An Approach to Dempster-Shafer Theory of Evidence. Lecture Notes in Economics and Mathematical Systems, Vol. 425, Springer-Verlag.

Plietker, B., 1997. Automatisierte Methoden zur ATKIS-Fortführung auf der Basis von digitalen Orthophotos. In: D. Fritsch and D. Hobbie (eds), Photogrammetric Week, Herbert Wichmann Verlag, pp. 135–146.

Winter, S., 1998. Uncertain topological relations between imprecise regions. Technical report, Fachbereich Geoinformation, TU Wien.

Zhang, J. and Goodchild, M., 2002. Uncertainty in Geographical Information. Taylor and Francis.