Updating Geodata by Automatic Object Extraction - Modelling Uncertainties

BERND – M. STRAUB[1], MARKUS GERKE[1] & CHRISTIAN WIEDEMANN[2]

In this paper we describe a method, which allows to model the positional and abstraction uncertainty of object's borders stochastically and leads to a probability-based decision of topological relations between area objects. The method was applied to evaluate the results of automatic extracted objects. By means of examples we demonstrate that the method is useful for update and quality description of geodata.

1 Introduction

The integration of automatic object extraction from imagery and Geo-Informationsystems for the update of geodata is an important task in the domain of photogrammetry. The most obvious reasons for this are (1) in many countries large geodata bases were built up in the last decade which have to be kept up to date and (2) the progress in automatic object extraction will lead to practical application.

The update of geodata requires the comparison of objects in order to be able to detect significant changes between an up-to-date and a stored data set. In the case of a manual update of the data, the operator decides if stored object has changed. The automation of the update process requires a method which is able to make a similar decision, which means that the extracted objects have to be compared with the stored objects. The method should be able to take into account, that the objects in the GIS database are influenced by the uncertainty of the measurement process and by the abstraction process. One can say, that the objects have some kind of a typical precision of their geometric representation. For example, the borderline definition of a forest is relatively weak, and therefore it is not reasonable to define a forest border with sub-centimetre accuracy in a GIS database. An accuracy of the forest borderline of a few decimetres seems to be more realistic. But, the accuracy of the same object border in another data set can have a few meter accuracy, for example a classification result from a Thematic Mapper image with 30 m ground sampling distance [GSD]. And even if both objects are correct and valid descriptions of the real world, it is not clear how to compare them.

The knowledge of topological relation is a basis for comparing two datasets. But, if the geometric representation of objects is uncertain, the topological relation between these objects are uncertain, too. Under some pre-conditions, WINTER (1996) prooved, that it is

[1] Bernd-M. Straub/Markus Gerke, Institute for Photogrammetry and Engineering Surveys, University of Hanover, Nienburger Strasse 1, 30167 Hanover, Germany, {straub, gerke}@ipi.uni-hannover.de

[2] Christian Wiedemann, Chair for Photogrammetry and Remote Sensing, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany, wied@photo.verm.tu-muenchen.de

possible to differentiate topological relations based only on the extreme values of the distance function between two objects in question, even if their boundaries are uncertain. We propose to refine the computation of the distance function and their extreme values, in order to overcome some practical problems.

In this paper we give a short description of the applied approach for the comparison between two objects. Afterwards we apply the approach for the comparison of forest and settlement areas.

2 Method for the Comparison of two Objects

2.1 Topological Relations Between Compact Regions

In this section we shortly describe the method that we have applied to assess the topological relations between polygon objects in different data sets. The approach was developed at the University of Bonn. We give a short overview about the concept, proofs and a detailed description are given in (WINTER, 1996; WINTER, 2000). At first we introduce the possible topological relations between two objects, see Fig. 1, these relations are binary. The *conceptual neighbourhood graph* (CNG) (EGENHOFER & FRANZOSA, 1991), gives an overview about the neighboured relations, see Fig. 2. Concerning the edges of the CNG one can see that EQUAL can change to COVERS, but not to TOUCH without passing the OVERLAP relation.

The relations WEAKOVERLAP and STRONGOVERLAP are equivalent from the topological point of view, but with the help of an overlap factor (OF) the conceptual neighbourhood graph can be subdivided into two relation clusters C^1 and C^2. The OF is defined as the ratio between the intersection area of the two objects and the area of the smaller one of the objects, OF is exactly 1 if A and B are EQUAL, and 0 if they are DISJOINT, values ≤ 0.5 lead to the cluster C^1, values > 0.5 lead to C^2 (WINTER, 1996). In the following we focus only on the relation cluster C^2. Consider the possible relations in Fig. 1 and Fig. 2, beginning from EQUAL we will observe the transitions of the topological relations. If the diameter of the object B becomes smaller, all possible distances d_i between the borders of A and B lie inside of A, which means that A contains B – denoted as CONTAINS(A, B) in the following. If one then moves B until the border of B touches the border of A exactly one distance becomes zero and the topological relation changes to COVERS(A, B). Finally, we move B a little bit more and obtain distances lying in A and in B respectively, i.e. the relation changes to STRONGOVERLAP. One can show that the decision between the relations can be made by means of the minimum and maximum distances between the areas A and B. The sign of the distances d is defined such, that d is negative if $d \subset A$, and positive otherwise.

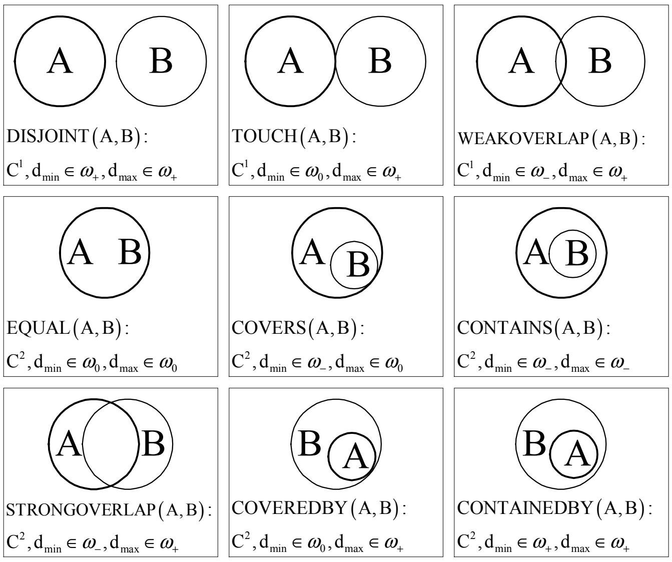


Fig. 1: Possible topological relations between areas

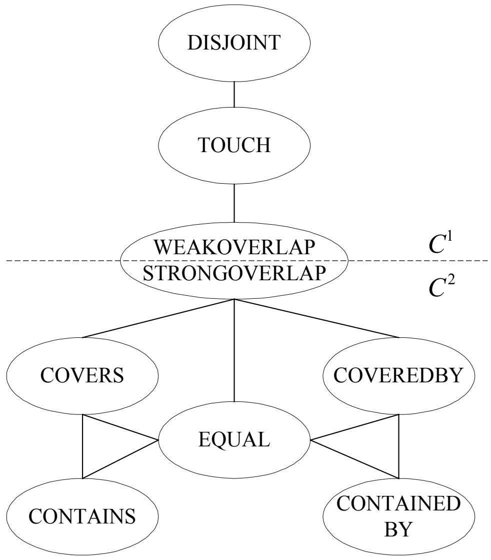


Fig. 2: Conceptual neighbourhood graph

Until now, we are able to deduce from the minimum and maximum distances (d_{min}, d_{max}) the topological relation between A and B. But these distance values are uncertain, they are influenced by the abstraction during the object extraction process and the imprecision of the measurement. As described above we want to introduce these aspects, whereby the abstraction leads to fuzziness and the measurement to imprecision. At first one has to define a range which is interpretable as IS_ZERO. Winter motivates the IS_ZERO concept with an example from a cadastral application: “Consider, e.g. a cadastral dataset with two points nearer than, let us say, 5 cm. Then one has strong support that both entries refer to the same point in real world” (WINTER, 2000, p.14). An interval for IS_ZERO can be determined by asking experts. For our problem we are at this point able to introduce the a priori knowledge about the resolution of the objects’ borders. The imprecision of measurement can be estimated from experience to σ, see also section 3. With these parameters we define a distribution function $D_0 = f([b,c],\sigma)$ for IS_ZERO. This distribution function D_0, is given by a convolution of a constant distribution (equipartition) in the interval [b, c] with a Gaussian (μ,σ) -the middle distribution in Fig. 3. For values outside of the IS_ZERO interval two other closed intervals have to be defined, which are enclosed by the largest possible values, keeping in mind that the objects are of a finite size.

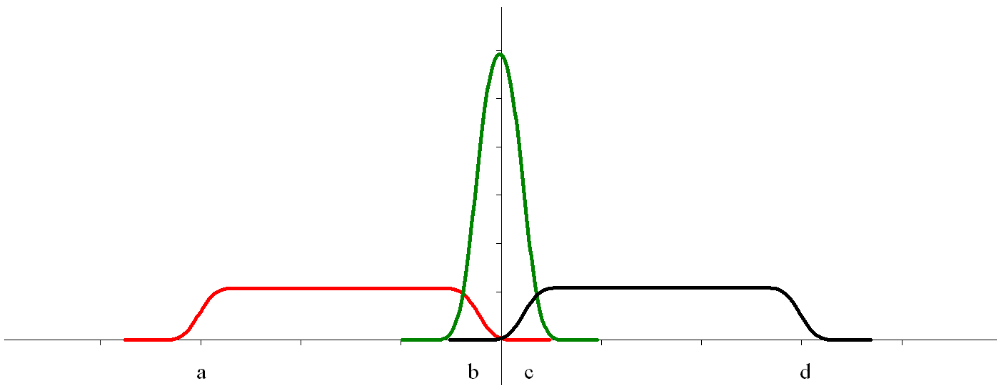


Fig. 3: Example for probability distributions for the possible classes of distances between two objects

D_0 represents a class ω_0 from $\Omega := \{\omega_-, \omega_0, \omega_+\}$. In a similar way we develop the distributions $D_- = f([a,b],\sigma)$ and $D_+ = f([c,d],\sigma)$, representing the probability distributions for the two other classes of distances, ω_- and ω_+ (Fig. 3).

We are interested in the topological relations between the objects A and B, which can be determined by the probabilities of d_{min} and d_{max} being members of one of the three classes from Ω. These probabilities $P_\omega(d)$ can be calculated with Bayes Theorem (Equ. 1) if $P_d(\omega)$ and $P(\omega)$ are known.

$$P_\omega(d) = \frac{P_d(\omega)P(\omega)}{\sum_{\forall\omega} P_d(\omega)P(\omega)} \tag{1}$$

In Equ. 1 $P_d(\omega)$ is given through the distributions D_-, D_0, D_+, and $P(\omega)$ through the intervals of the equipartitions. With Equ. 1 the probabilities $P_\omega(d_{min})$ and $P_\omega(d_{max})$ can be calculated for d_{min} and d_{max}. Under the assumption that d_{min} and d_{max} are stochastically independent the searched probabilities for the topological relations can thus be computed.

2.2 Distance Function

The necessary input values, d_{min} and d_{max}, needed for the estimation of the topological relations, can be found in the distance function from two objects A and B, shown in Fig. 4. With the zones $P=A^C \cap B^C$ and $Q=A \cap B$, where the superscript *C* describes the outer area of the object, we define an uncertain zone $O = \mathbb{R}^2 \setminus P, Q$ representing the area between the boundaries of A and B, see Fig. 5. The needed values d_{min} and d_{max} can be calculated with a morphological distance transformation of O, as done by WINTER (1996). The sign of the distances d_i is negative if $O \subset A$ otherwise positive, see the distance function in Fig. 6.

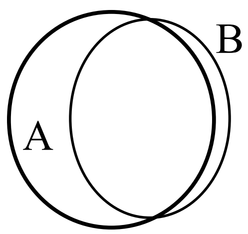


Fig. 4: Objects A and B

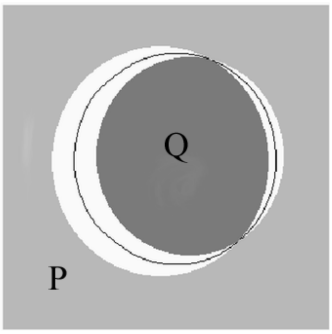


Fig. 5: Uncertain zone with skeleton S

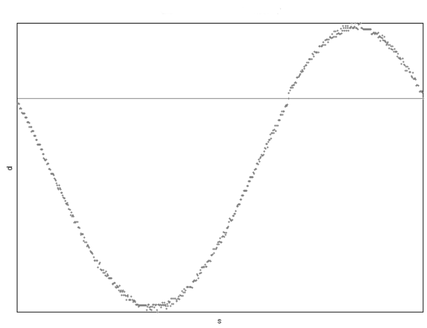


Fig. 6: Distance function along the skeleton S

2.3 Discussion of the Distance Function

In general one has to take into account that the distance function along the skeleton of the uncertain zone O doesn't represent the distance between P and Q in every case. One reason is that irregularities in the borders lead to small branches in the skeleton. Another reason is that the shape of O may lead to wrong distances between P and Q. For example, take the situation in Fig. 7. The value d_s along the skeleton S_2 of O is not a correct estimation of the needed distance. Therefore we use the distance $d'_{(q)}$ along the border of the certain zone Q. This value represents the correct distance which is needed for the topological analysis, see Fig. 8. In general, one has to take into account, that the borders of O and Q are not parallel, an angle dependent correction for $d'_{(q)}$ has to calculated. A detailed description of the calculation of the distance function is given in (GERKE, 2000).

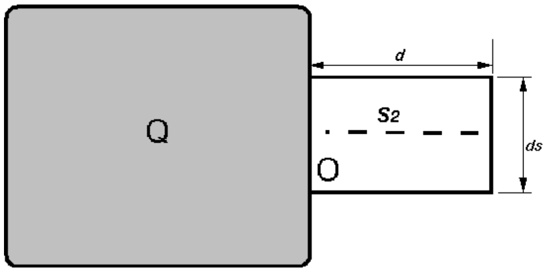


Fig. 7: Skeleton in O

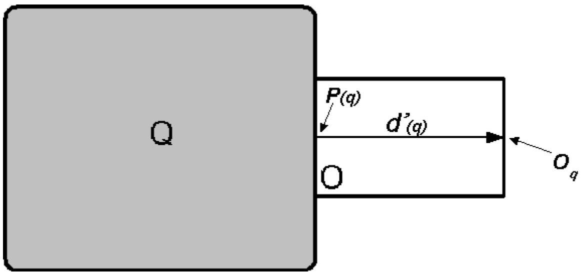


Fig. 8: Distance along the border of Q

In the distance function some irregularities appear as peaks, which can be seen as a kind of gross errors. In order to eliminate these peaks, the distance function is median filtered. Because every distance d_i is assigned to one pixel s_i one can analyse the borderlines of A and B, especially the irregularities.

3 Example

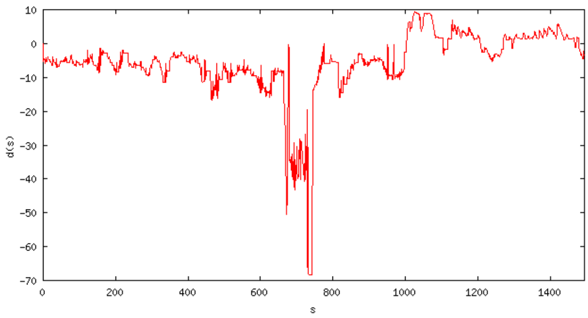
In this section a detailed example for a forest area is given, in order to explain the approach. Furthermore we show numerical results of the evaluation of settlement and forest areas, which were extracted automatically. A description of the used approach for the automatic object extraction is given in (STRAUB ET AL, 2000). The application of the approach related to buildings are given in (RAGIA & WINTER, 1998; RAGIA, 2000).

Fig. 9: Borderline of object A (black line) overlayed to object B (grey area)

Fig. 10: Detected gross errors along the borderline of object B

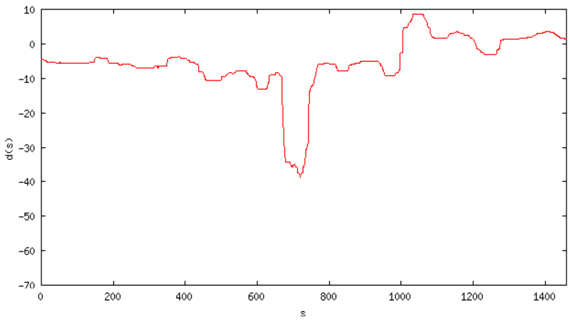
The objects that we want to compare can be characterised as follows. Object A was captured manually from an image with a resolution of 1 m, we assume a standard deviation of 0.5 m for the object's border line. The other object was extracted automatically, the used image had a resolution of 12 m, in this case we assume a standard deviation of 0.5 pix for the imprecision of measurement, which leads to a standard deviation of 6 m for the Gaussian distribution, see Fig. 9.

Assuming, that the automatic image processing steps lead to an uncertainty of two pixels in the borderline of the extracted forest objects the IS_ZERO interval is set to [-12 m, +12 m], which corresponds to two pixels in the used image data. This can be interpreted as the fuzziness of abstraction, introduced in section 2.1. The minimum and maximum possible values for the distributions $D_{-/+}$ are set to -/+150 m, greater than the largest value from the distance function. The distance function printed in Fig. 11 has peaks, which were interpreted as gross errors. The filtered distance function without gross errors is printed in Fig. 12. As the spatial relation between the peaks in the distance function and the position in the image is known, there a local analysis at these image positions should be performed. This analysis can be carried out by a human operator or by a further automatic processing step. In Fig. 10 the gross errors are visualized. The squares at the border of object B show the points which are not taken into account for the topological analysis.



s and d(s) given in pixel, 1 pixel = 3 m

Fig. 11: Distance function for the objects A and B



s and d(s) given in pixel, 1 pixel = 3 m

Fig. 12: Median filtered distance function

Numerical results related with extracted forest and settlement areas are shown in Tab. 1, the forest area, which is printed in Fig. 9, has got the ID 01 in this table. The objects B in Tab. 1 were extracted automatically, and compared with manual captured reference data (A). Both data sets were captured on the base of the same image, but the topological relation between the objects is not EQUAL, as expected. Therefore we have to refine the extraction process by means of an extended model. For example, we have to model the shadows of the trees along the borderline of the forest, which was not done until now.

In the case of settlement areas the IS_ZERO interval is set to [-45 m, 45 m] caused by the weak definition of the borderline. We argue, that even for a human operator it is difficult to define the borderline of a settlement area in an image, the fuzziness begins with the backside of the last building at the border and ends, perhaps, with a fence. In fact, the borderline of a settlement area is not really visible in an image, it is more related to an administrative border. Under this pre-conditions some of the settlement objects fit to the reference, they are assigned as equal.

	ID	d_{min} [m]	class	$P_{dmin}(\omega)$	d_{max} [m]	class	$P_{dmax}(\omega)$	Topological Relation: $P_{top\ Rel}$
Forest	00	-60.0	ω_-	1.00	6.6	ω_0	0.81	COVERS(A, B): 0.81
	01	-120.0	ω_-	1.00	25.1	ω_+	0.99	STR. OVERLAP(A, B): 0.99
	02	-108.0	ω_-	1.00	6.9	ω_0	0.79	COVERS(A, B): 0.79
	03	-22.2	ω_-	0.95	25.8	ω_+	0.99	STR. OVERLAP(A, B): 0.94
	04	-216.0	ω_-	1.00	10.2	ω_0	0.62	COVERS(A, B): 0.62
	05	-15.9	ω_-	0.74	14.1	ω_+	0.63	STR. OVERLAP(A, B): 0.74
Settlement	00	-36.0	ω_0	0.90	23.4	ω_0	1.00	EQUAL(A, B): 0.90
	01	-141.0	ω_-	1.00	63.0	ω_+	1.00	STR. OVERLAP(A, B): 0.82
	02	-36.0	ω_o	0.85	84.0	ω_+	1.00	COVERED BY(A, B): 0.85
	03	-135.9	ω_-	1.00	150.0	ω_+	1.00	STR. OVERLAP(A, B): 1.00
	04	-33.0	ω_0	0.97	36.0	ω_0	0.91	EQUAL(A, B): 0.89
	05	-81.0	ω_-	1.00	102.0	ω_+	1.00	STR. OVERLAP(A, B): 1.00

Tab. 1: Evaluation results

The topological relations together with the distance functions can be looked upon as some kind of object specific indicators pointing to problems in the extraction process on the one hand, and on the other hand they can be used to specify the quality of the object in question.

4 Summary And Outlook

A method for the classification of topological relations between two objects, proposed by WINTER (1996), was implemented and tested. In order to be able to detect local inconsistencies between the objects, we have modified the calculation of the distance function. This modification allows us to access local inconsistencies between the objects borders. Nevertheless the analysis of the distance function is a critical point in the automation of the evaluation.

Summarising one can say, that the applied method seems to be suited to characterize the performance of a given approach for object extraction. The method can be seen as a step to a theoretical founded evaluation of extraction results, which was often demanded in the past (FÖRSTNER, 1996).

Furthermore the applied method gives us the possibility to find problems in the extracted objects automatically even if they have different geometric accuracy. This should be seen as an important information for the design of further processing steps.

In the future we plan to investigate the force of expression of this approach for a whole GIS data set. Together with quality measures like completeness and correctness (WIEDEMANN ET AL., 1998) we will try to get statements like “Object A and B together are covered by Object C”.

5 Acknowledgement

Parts of this work were developed within the IST Programme CROSSES financed by the European Commission under the project number IST-1999-10510. Thanks to Stefan Winter for his research in uncertain topological relations, which was a base for this paper. Many thanks, also, to Christian Heipke for the long discussions concerning the scope for development.

6 References

EGENHOFER, M. J. & FRANZOSA, R. D. (1991): Point-set topological spatial relations, International Journal of Geographic Information Systems, **5**(2), pp. 161-174.

FÖRSTNER, W. (1996): 10 pros and cons against performance characterization of vision algorithms, Workshop on "Performance Characteristics of Vision Algorithms", Cambridge 1996, 22 pages.

GERKE, M. (2000): Topologische und geometrische Analyse zum Vergleich ungenauer Flächen, Master thesis, Institute for Photogrammetry and Engineering Surveys, University of Hanover. 92 pages, unpublished.

RAGIA, L. & WINTER, S. (1998): Contributions to a quality description of areal objects in spatial data sets, International Archives of Photogrammetry and Remote Sensing, **32/4**, Fritsch, Englich & Sester (eds.), Stuttgart, pp. 479-486.

RAGIA, L. (2000): A Quality model for spatial objects, International Archives of Photogrammetry and Remote Sensing, **33**(B4), Amsterdam, pp. 855-862.

STRAUB, B.-M., WIEDEMANN, C. & HEIPKE, C. (2000): Towards the automatic interpretation of images for GIS update, International Archives of Photogrammetry and Remote Sensing, **33**(B2), Amsterdam.

WIEDEMANN, C., HEIPKE, C., MAYER, H &, JAMET, O. (1998): Empirical Evaluation of Automatically Extracted Road Axes, Bowyer Phillips (eds.), Empirical Evaluation Methods in Computer Vision, Los Alamitos, California, IEEE Computer Society Press, pp. 172-187.

WINTER, S. (1996): Unsichere topologische Beziehungen zwischen ungenauen Flächen, Ph.D. thesis, Landwirtschaftliche Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn, Deutsche Geodätische Kommission DGK-C, **465**, Munich.

WINTER, S. (2000): Uncertain topological relations between imprecise regions, International Journal of Geographical Information Science, **14**/5, pp. 411-430.