A QUALITY MODEL FOR SPATIAL OBJECTS

Lemonia Ragia
Institute of Photogrammetry, University of Bonn
Nussallee 15, D-53115 Germany
Phone: +49 (228) 732902, FAX: +49 (228) 732712, e-mail: nitsa@ipb.uni-bonn.de

Working Group IC WG IV/III.1

KEY WORDS: Quality control, uncertainty, buildings, topology, analysis

ABSTRACT

This paper presents a concept for analysing the quality of $2\frac{1}{2}$-D spatial objects. The developed method is based on the evaluation of specific quality parameters. These parameters are determined by a topological and geometrical analysis. The quality parameters are classified into three categories: green=accepted, yellow=uncertain, red=rejected, depending on the specifications. We give confidence regions for all quality parameters, especially for completeness, false alarm rate and detection rate. The feasibility of the method is shown by using real examples taking into account the technical specifications.

1 INTRODUCTION

Spatial data are in increasing demand for many applications. Spatial data play an important role in city planning, in mobile phone networks etc. The quality and reliability of the acquired data is essential for any further processing or use.

Information on the quality of data is a major concern for both developers and users of GIS (Chrisman, 1994). The quality of spatial information is multidimensional and complex (Beard et al., 1991). The usefulness of the quality measures depends on the application. It is not always clear to decide on how many quality parameters can be introduced to describe the quality of data. The number of quality parameters can be very large because quality varies spatially and temporally. We want to develop a concept that is flexible enough for a large number of applications and is appropriate for controlling different types of data.

Defining the quality measures is already a very actual topic in the standardisation process. The CEN Meta data Standards of the Commite Europeen de la Normalisation deals with the quality description. The model of ISO is the most complete concept (CEN, 1994). It involves the following aspects: *lineage*, *positional accuracy*, *thematic accuracy*, *temporal accuracy*, *consistency* and *completeness* (Guptill and Morrison, 1995).

Building information, which is important for many applications, needs to be evaluated with respect to given specifications. (Laing and Ruff, 1998). As an example, Fig. 1 shows a section of an aerial image. In Fig. 2(a) and (b) footprints of the buildings are shown, which result from two different data acquisition schemes. There are differences between them e. g. missing parts of buildings or differences in the neighborhood relations. It is not clear which class of quality parameters could cover such types of differences. We therefore need to describe such types of differences and develop adequate quality measures.

Figure 1: *A section of an aerial image with some buildings, ©DeTeMobil GmbH, Bonn, 1999*

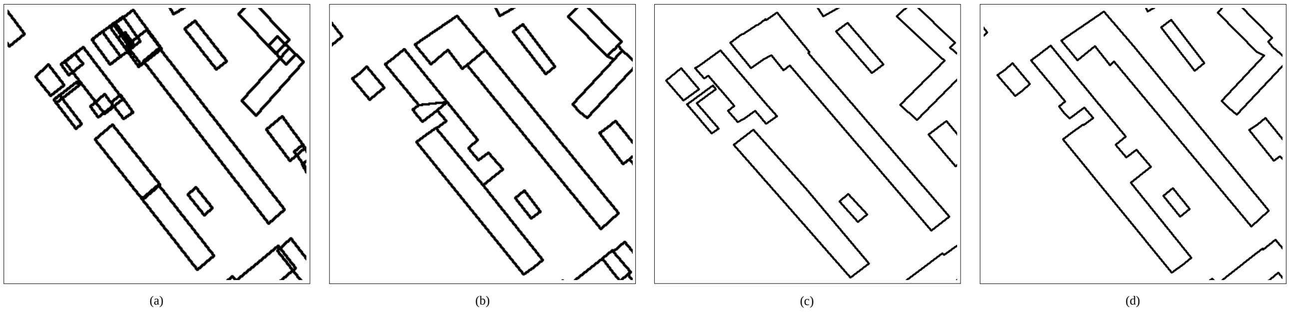

Figure 2: *The footprints of the data sets of the buildings of the image Fig. 1. The figures (a) and (b) show the buildings with their parts and the figures (c) and (d) show the outlines without the partitioning. Observe the difference in the partitioning.*

2 QUALITY MODEL

2.1 The task

We identify and detect the differences and the degree of similarity of two different descriptions of the same spatial area. We assume two independent data sets of spatial objects. In order to reach quality assessment of the two data sets it is assumed one to be the reference to the other. The requirements of the specifications for which the descriptions are going to be used are taken into account. The two descriptions may be extracted from any source but they have to refer to the same coordinate system. The data sets are controlled initially for their homogeneity and consistency, which is not our concern.

Our quality model consists of parameters specifying the following quality aspects: *topology, geometry* and *success*.

2.2 Topology

The topology refers to the *structural differences* of two corresponding spatial objects. These are the differences in partitioning of the buildings, although their outlines are equivalent, and in the topological relations between parts of buildings and building Fig 2.

For deriving parameters characterizing the topological differences we determine the spatial relations of all given objects.

We assume the spatial objects to be represented by the *region adjacency graphs* (RAG). The classification of the topological relations is based on the model of Egenofer (Egenhofer and Herring, 1991). The classification of the topological analysis takes the method of observation and the type of context into account by treating the boundaries as uncertain, as explained below. Due to the consistency checks, only the topological relation *touch* and *overlap* and its alternatives *strong overlap* and *weak overlap* (Winter, 1996) occur. Both RAGs, one for the reference data one for the test data, represent its set of spatial objects together with its spatial relations. Nodes and edges of the RAG have attributes, e. g the number of the holes, the number of the footprint points, or the type of the topological relation (Ragia and Förstner, 1999).

The *region correspondence graph (RCG)* is a bipartite graph, containing all correspondences between regions of the two different data sets. The topological relation can be: *equal, strong overlap, covers, covered by, contains, contained by*. Attributes of the connected components of the *region correspondence graphs* can be used for identifying the interior structure of the sets of regions. This allows to check isomorphic sets of regions or cases where regions are merged or split with respect to the reference data set. Vector and raster format is used leading to a hybrid analysis technique.

As an example in Figure 3 we can see the overlap of two sets of buildings. The identification numbers $\{$1,2,3,4$\}$ characterize the buildings of the first data set and the code $\{$a,...,f$\}$ characterizes the buildings of the second data set. The overlap refers only to the outlines of the buildings, i. e. overlapping buildings are already merged. On the right of the Figure 3 we can see the *RCG* with the building structure represented by the RCG. The nodes represent the building and the edges the correspondences between buildings of the two data sets.

The *degree of partitioning* and the *degree of merging* characterize the topology of the data sets, which can be derived from the RCG similar to the analysis of the transition table used by (Fuchs et al., 1994) for evaluating the result of feature extraction. We may distinguish: the *degree of partitioning* p_j if an object O_j is partitioned into p_j primitives and the *degree of merging* m_i if m_i given objects are merged into one.

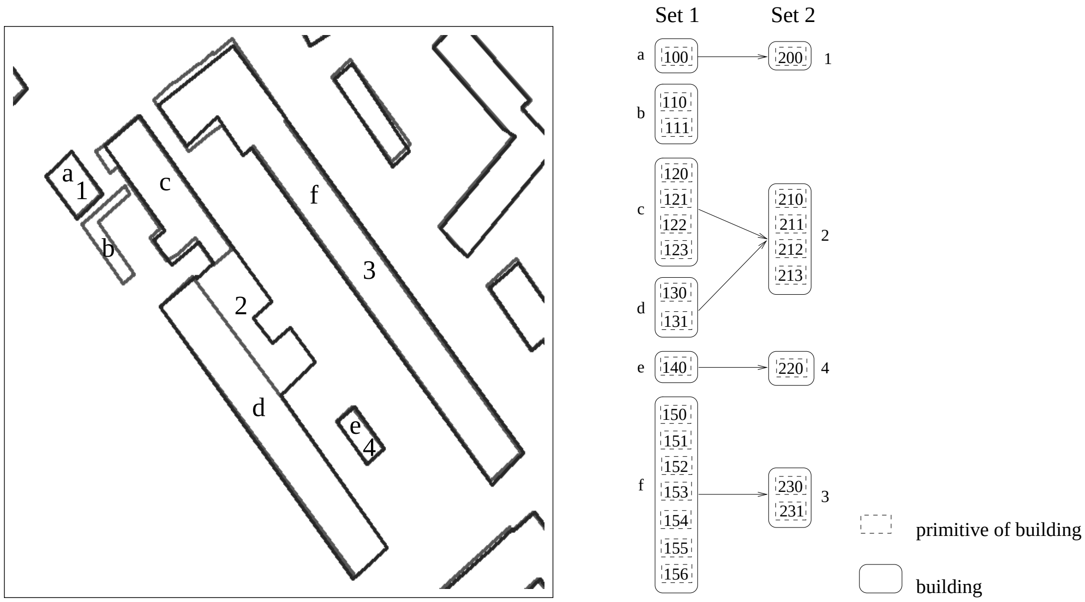

Figure 3: *The overlap of two sets of ground plans of the building structures shown in Fig. 1 is on the left and the corresponding* RCG *is on the right*

2.3 Geometry

2.3.1 Position The geometrical analysis of the footprints of the objects is based on the distance function which characterizes the difference area, i. e. the sliver polygon of two corresponding sets of regions, or formally the symmetric different area of the objects A and B $A \ominus B = A \setminus B \cup B \setminus A$.

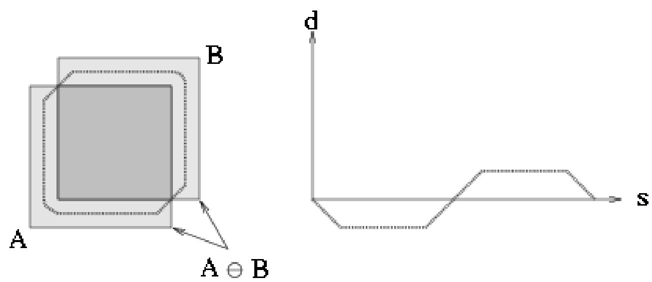

Figure 4: *The zone skeleton of an artificial example and the corresponding distance function*

The distance function $d(s)$ (Winter, 1996) is derived from the so called zone skeleton, including all centers of circles which touch the boundaries of regions of different data sets. The distance function is the radius of those circles as a function of the arc lengths of a zone skeleton Fig. 4 (Ragia and Winter, 1998).

The qualitative classification of the geometrical differences uses the distance function and depends on the users specification namely the degree of required generalization. Two thresholds a and b with $a < b$ are used in order to define three parts: the red part, the yellow and the green one (cf. Fig. 5).

If the skeleton curve lies between $[-a, a]$ then the two objects are classified as equal, the result of the evaluation is *'green'*. If the skeleton curve is between $[-b, b]$ and outside of $[-a, a]$ then the two objects have some differences and they could be checked further, the result of the evaluation is *'yellow'*. If the skeleton curve is outside of [-b, b] then the objects have gross differences, the result of the evaluation is *'red'*.

In the general case the zone skeleton may consist of *more than one closed line*, leading to a *list of distance functions*. Moreover, the zone skeleton and its distance function can also be used to *identify* $n : m$*-relations* between corresponding objects independent of their partitioning and missing parts.

Fig. 6 shows a real example taken from the left part of the section of the aerial image in the Figure 3 with the corresponding distance function in Fig. 7.

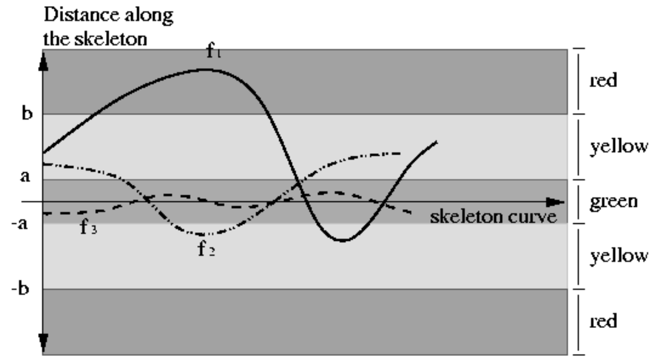

Figure 5: *Three artificial skeleton curves with their classification*

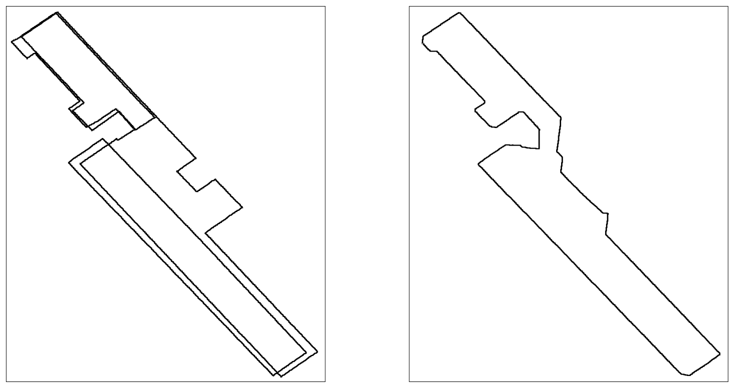

Figure 6: *On the left there is the overlap of an aggregated building and on the right the corresponding zone skeleton*

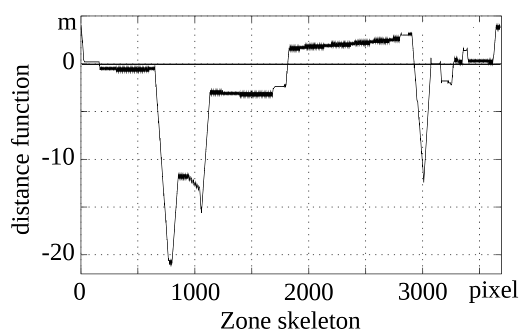

Figure 7: *The distance function of the buildings in the Fig 6*

This is a case of 'yellow' where a further analysis may be needed. We can then with the help of the distance function Fig. 7 characterize the differences between two buildings e. g. there is a missing part, a shift, a magnification etc.

2.3.2 Height The geometrical analysis of the heights is done in a more simple way. Here we obtain parameters characterizing the accuracy of the height but also possible systematic errors, which could be introduced e. g. by errors in the orientation data.

Prerequisites are two data sets with heights as attributes ($2\frac{1}{2}$-D). The points of the objects are classified into three categories: roof-top points, in gutter footprint points and footprints points.

The planar correspondence yielding the RCG is also used to define corresponding heights. The height differences can be evaluated for all corresponding buildings and their parts. Tab. 1 shows the used values for making statements about the quality. Again, two thresholds for the classification of the results are used in order to define three evaluation classes, red, yellow and green. The thresholds depend on the specification.

2.4 Success

We define three global quality parameters depending on the RCG, specifically on the acquired objects and the objects having a correspondence in the other data set.

	systematic errors (median)	statistical errors (robust deviation)
low value	no error	no error
high value	wrong orientation	inaccurate

Table 1: Criteria for statistical values

The reference data set is $\mathcal{O}_1 = \{O_{i1}\}, i = 1, ..., I_1$, I_1 being the number of all objects in the reference data set 1. The data set to be evaluated is $\mathcal{O}_2 = \{O_{i2}\}$ containing I_2 objects.

The set of all *disjoint* objects, which have no corresponding object in the other data set is $\mathcal{D}_j = \{D_{kj}\}$ with $k = 1, ..., K_j$, K_j being the number of disjoint objects in data set j, cf, Fig. 8. Thus we formally have:

$$\mathcal{D}_{j'} = \mathcal{O}_{j'} \setminus \mathcal{O}_{j''} \quad j', j'' \in (1, 2)$$

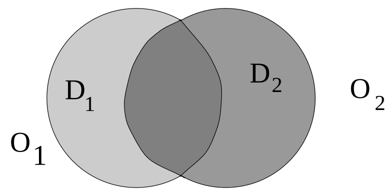

Figure 8: *Two sets $\mathcal{O}_1$ and $\mathcal{O}_2$ of objects and the definition of the sets $\mathcal{D}_1$ and $\mathcal{D}_2$*

2.4.1 Completeness The *completeness* c is defined as the ratio of all missing or spurious objects in relation to all acquired objects:

$$c = \frac{|\mathcal{O}_1 \cap \mathcal{O}_2|}{|\mathcal{O}_1 \cup \mathcal{O}_2|}$$

2.4.2 Detection rate The *detection rate* d has a commonly used definition (McKeown et al., 1997), (Nevatia, 1999) and represents the number of the objects that are detected only in the reference data set in relation to the number of reference objects:

$$d = \frac{|\mathcal{O}_1 \cap \mathcal{O}_2|}{|\mathcal{O}_1|}$$

2.4.3 False alarm rate The *false alarm rate* f (McKeown et al., 1997), (Nevatia, 1999) represents the number of the objects that are not detected in the reference data set in relation to the number of objects to be evaluated:

$$f = \frac{|\mathcal{D}_2|}{|\mathcal{O}_2|}$$

The three quality parameters are mutually dependent as only the ratio of magnitude of the three sets $\mathcal{D}_1$, $\mathcal{O}_1 \cap \mathcal{O}_1$ and $\mathcal{D}_2$ is relevant, as e. g.

$$c = \frac{1}{\frac{1}{d} + \frac{f}{1-f}}$$

3 EMPIRICAL TEST

3.1 Selection of representative test areas

We want to demonstrate the usefulness and the flexibility of this method checking its applicability on real examples. We have at our disposal some data sets from twelve cities. Because of the great number of the buildings per city (e. g. more than 10.000) a selection of representative test areas takes place. The representativity of the chosen test areas is decisive for the evaluation of the given data sets. In order to show and cover many types of these differences we select areas of cities with quite different type of structure.The test areas are chosen from:

- areas with isolated buildings,
- central city areas with simple block structures,

- typical old city areas with complicated building structure,
- mixed building areas,
- complex building areas, e. g. industrial parks.

In this way we capture the different building characteristics dealing with the complexity of the individual buildings and the complexity of the building areas.

3.2 The set up for the empirical study

In each of the twelve test areas we have two data sets: one data set is given by the user of the data and the second one is the reference data set and produced by a semi-automatic system for building extraction (Gülch, 1997), (Gülch and Müller, 1997), (Müller, 1997). The test areas have about 150 buildings.

The tolerances *a* and *b* are to be chosen with regard to the specification. We use a standard deviation in planimetry of 1 *m* for determining the topological relations in the *RAG* and the thresholds *a* =1 m and *b* = 3 m to describe the geometrical differences. For classification of the other quality parameters we have defined thresholds, which are based on the *median* value $\text{med}_i(v_i)$ of all the values v_i pro quality parameter and the *robust standard deviation*, namely the median absolute deviation $\text{med}_i(|v_i - \text{med}_j(v_j)|)$.

3.3 Results of the test

The result of the quality analysis of all cities is shown in the table 2 (Ragia and Laing, 2000). It contains for all quality parameters the number of cities in the classes *green*, *yellow* and *red*.

The user of the data can change the thresholds for the classification of the results in three parts (colors) and then the results can be totally different. The user can use one or more quality parameters for making decisions in different applications. E. g. if only the geometry (position) is taken into account then we have 5 cities of 12 are not accepted or if only the geometry (height) is regarded then we have 8 cities in green part 1 in the yellow and 3 in the red part.

The user can combine the quality parameters and can take into account two or more of them. An example is shown in table 3 where a combination of results of two quality parameters are given. E. g. When we regard the geometry (position) and geometry (height) than we have 3 cities in green part, 7 in yellow and 2 in red. This can be further extended by giving different weights.

evaluation	green #cities	yellow #cities	red #cities
degree of partitioning	5	5	2
degree of merging	6	4	2
geometry (position)	3	4	5
geometry (height)	8	1	3
completeness	5	4	3
detection rate	7	3	2
false alarm rate	5	7	0

Table 2: The final estimation of the twelve cities, green: collection of data based on the given criteria is accepted, yellow: making a decision is uncertain, rot: collection of the data is not accepted.

evaluation	green #cities	yellow #cities	red #cities
geometry (position) - geometry (height)	3	7	2
Completeness - geometry (position)	5	2	5

Table 3: The estimation of the twelve cities regarding a combination of two parameters.

3.4 Quality results

The results of this model can only be evaluated taking into account some specifications.

Let us assume we have three specifications:

1. The first one need to have a completeness of more than 80%, a green part of geometrie (position) of more than 30% and a detection rate of more than 70%. The results of the city number 1 appear to be acceptable. According to the Binomial distribution $B(n, p)$ the values lie in the 95% confidence interval [75%-93%], [19%-57%] and [60%-88%].

2. The second specification takes into account only the quality parameter completeness. It has to be more than 82% guaranteed. Then the answer is the city number 3 because the completeness is 92% and the 99% confidence interval [82%-100%].

3. The third takes into account only the red part of the geometrie (position). It must be smaller than 25% guaranteed. Then the answer is the city number 2 because the value and the confidence interval is smaller than 25%.

city	1	γ=95%	γ=99%	2	γ=95%	γ=99%	3	γ=95%	γ=99%
p	11%	[0%-22%]	[0%-26%]	3%	[0%-5%]	[0%-6%]	17%	[1%-32%]	[0%-34%]
m	7%	[0%-16%]	[0%-20%]	0%	[0%]	[0%]	0%	[0%]	[0%]
c	85.5%	[75%-93%]	[73%-95%]	82%	[76%-87%]	[74%-88%]	92%	[84%-99%]	[82%-100%]
d	74%	[60%-88%]	[56%-91%]	69%	[61%-77%]	[58%-80%]	89%	[77%-100%]	[73%-100%]
f	3%	[0%-9%]	[0%-11%]	1%	[0%-3%]	[0%-4%]	4%	[0%-11%]	[0%-14%]
G_{gr}	38.4%	[19%-57%]	[14%-62%]	25.3%	[40%-62%]	[37%-65%]	10%	[0%-21%]	[0%-25%]
G_{yel}	34.6%	[16%-53%]	[11%-59%]	60.3%	[49%-70%]	[37%-65%]	52%	[31%-73%]	[24%-80%]
G_r	27%	[10%-44%]	[5%-49%]	15.2%	[7%-23%]	[5%-25%]	38%	[18%-58%]	[11%-65%]
H	0.67	-	-	0.33	-	-	0.67	-	-

Table 4: The values of the quality parameters with the confidence interval, p:degree of partitioning, m:degree of merging, c:Completeness, d:Detection rate, f:False alarm rate, G_{gr}:Geometry (position) green, G_{yel}:Geometry (position) yellow, G_r:Geometry (position) red,

4 CONCLUDING REMARKS

In this paper we proposed a method for quality control of spatial objects. The concept we proposed was implemented and tested empirically. The user needs only to specify the threshold values. The system processes automatically and results either in the acceptance of the buildings data or it comes back with the areas where there is uncertainty. The problem areas, for instance in buildings with inner courts are further processed manually.

The following extensions and improvements can be considered.

- The processing can be further improved with more topological and geometric analysis. It is especially important in testing to analyse the individual objects in order to generate a concrete description of data errors. To achieve this result we will need an exact description and classification of possible errors.
- There is a need to obtain the sensitivity of the method. For instance, it is interesting to clarify the dependence of the results on the choice of threshold parameters, on the complexity of the buildings, on the choice of the testing data e.t.c
- A self diagnosis of the method needs to be developed. This can be obtained by using boot-strap technics (Cho et al., 1997) which provide parameter free statistical information on the quality of results.

The previous application of the method has shown that it is flexible. The experience of the analysis of twelve testing areas have been useful for improving the method. They also give insight for further improvements. The openess of the quality model allows the integration of further parameters and the flexible specification of boundary conditions. In this way given knowhow in quality evaluation of city plans can be taken into account.

Acknowledgments

The author gratefully acknowledges the contributions of Prof. Dr.-Ing. Wolfgang Förstner, Dr.-Ing. Eberhard Gülch, Dipl.-Ing. Ralf Laing, Dipl.-Inform. Thomas Läbe and Dipl.-Phys. Hardo Müller.

REFERENCES

Beard, M. K., Buttenfield, B. P. and Clapham, S. B., 1991. NCGIA Research Initiative 7 — Visualization of Spatial Data Quality. NCGIA Technical Paper.

CEN, 1994. Data description: Quality. Draft of Definitions Document N 15 rev. 4 (august), CEN TC 287 WG 2.

Cho, K., Meer, P., Member, S., IEEE and Cabrera, J., 1997. Performance assessment through bootstrap. In: Proc. Conf. on Transactions on Pattern Analysis and Machine Intelligence, Vol. 19number 11, IEEE Computer Society, IEEE Computer Society Press, pp. 1185–1198.

Chrisman, N. (ed.), 1994. Datenqualität in Geographischen Informationssystemen. Geographisches Institut der Universität Zürich-Irchel, Winterthurerstrasse 190, CH-8057 Zürich.

Egenhofer, M. J. and Herring, J. R., 1991. Categorizing Binary Topological Relationships Between Regions, Lines, and Points in Geographic Databases. Technical report, Department of Surveying Engineering, University of Maine, Orono, ME.

Fuchs, C., Lang, F. and Förstner, W., 1994. On the Noise and Scale Behaviour of Relational Descriptions. In: H. Ebner, C. Heipke and K. Eder (eds), Proc. of ISPRS Comm. III Symposium Spatial Information from Digital Photogrammetry and Computer Vision, SPIE, München, pp. 257–264.

Gülch, E., 1997. Application of semi-automatic building acquisition. In: A. Grün (ed.), Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser, Basel.

Gülch, E. and Müller, H., 1997. Object-oriented software design in semiautomatic building extraction. In: Proceedings Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III, Orlando, Florida, April. SPIE Vol. 3072.

Guptill, S. C. and Morrison, J. L. (eds), 1995. Elements of Spatial Data Quality. Elsevier Science.

Laing, R. and Ruff, B., 1998. 3D-Stadtmodelle für den Mobilfunk. Der Vermessungsingenieur 49(5), pp. 241–248.

McKeown, D., Cochran, S., Gifford, S., Harvey, W., McGlone, J., Polis, M., Ford, S. and Shufelt, J., 1997. Research in the automated analysis of remote sensed imagery. In: 1995:1996, Proceedings DAPRA Image Understanding Workshop, New Orleans, L.A., pp. 779–812.

Müller, H., 1997. Designing an object-oriented matching tool. In: 3D Reconstruction and Modelling of Topographic Objects, International Archives of Photogrammetry and Remote Sensing, Vol. 32, ISPRS Commission III/IV, pp. 120–127.

Nevatia, R., 1999. On evaluation of 3-d geospatial modeling systems. SFPT Societe Francaise de Photogrammetrie et Teledetection (153), pp. 15–21.

Ragia, L. and Förstner, W., 1999. Automatically assessing the geometric and structural quality of building ground plans. SFPT Societe Francaise de Photogrammetrie et Teledetection (153), pp. 22–31.

Ragia, L. and Laing, R., 2000. Qualitäsprüfung von gebäudedaten aus photogrammetrischen auswertungen. PFG, Photogrammetrie Fernerkundung Geoinformation.

Ragia, L. and Winter, S., 1998. Contributions to a quality description of areal objects in spatial data bases. In: Proceedings ISPRS Comm. IV Symposium, GIS-Between Visions and Applications, Stuttgart, Vol. 32, pp. 479–486.

Winter, S., 1996. Unsichere topologische Beziehungen zwischen ungenauen Flächen. PhD thesis, Landwirtschaftliche Fakultät der Universität Bonn, DGK-C 465, München.