Automatically Assessing the Geometric and Structural Quality of Building Ground Plans

Lemonia Ragia

Institute for Photogrammetry
University of Bonn, Nussallee 15
D-53115 Bonn, Germany
Phone: +49 (228) 732968
FAX: +49 (228) 732712
email: nitsa@ipb.uni-bonn.de

Wolfgang Förstner

Institute for Photogrammetry,
University of Bonn, Nussallee 15
D-53115 Bonn, Germany
Phone: +49 (228) 732714
FAX: +49 (228) 732712
email: wf@ipb.uni-bonn.de

Abstract

The paper develops an approach for assessing the quality of ground plans of buildings. Quality is measured not only by geometrical but also by *structural* differences between an acquired data set and a reference data set. New hybrid techniques for automatically determining quality measures are developed, and shown to be applicable to real data. The uncertainty of the given data is taken into account. Automating quality assessment increases efficiency in checking data, allowing complete checks instead of sampling, moreover it makes quality checks objective. The developped techniques are applicable to sets of 2D regions of any type and internal structure. We also demonstrate the necessity to use the quality of the quality parameters when checking the fullfillment of quality specifications.

Résumé

Le papier développe une approche pour léstimation de la qualité des plans se bâtiments. La qualité est mesurée pur la difference, par comparaison géométrique et structurelle entre un enseble de données extraites et des données de référence. De nouvelles techniques hybrides de mesures automatiques de qualité sont proposées et appliquées à des cas réels. L'incertitude des données mises en jeu est prise en compte. Cette évaluation automatique de la qualité accroit l'efficacité du contrôle puisquélle permet un contrôle exhaustif et non sur un échantillon. Le contrôle est ainsi beaucoup plus objectif. Les techniques développées sont applicables à des ensembles de régions 2D quelque soit leur type et leur structure interne. Nous démontrons également la nécessité de prendre en compte la qualité des paramètres d' évaluation utilisés losquón vérifie que les spécifications de qualité sont remplies.

Keywords: structural accuracy, geometric accuracy, sets of 2D regions, quantitative quality measures, cartographic data acquisition, quality of building extraction.

Mots Clefs: précision structurelle, précision géométrique, ensemble de régions 2D, mesures de qualité quantitatives, acquisition de données cartographiques, qualité des extractions de bâtiments

1. Introduction

1.1 Motivation

The increasing use of geoinformation systems (GIS) requires well controlled data acquisition schemes. Assessing the quality of spatial data is therefore an important issue. Assessment schemes, however, especially when they aim to be generic, are difficult to establish. One reason is, that quality is task dependent which contradicts multiple purpose use intended for most GIS-data. The other reason is the complexity of spatial data which requires a broad range of quality measures to cover all quality aspects and take into consideration all types of acquisition errors. As an example, take the section of an aerial image in Fig . 1 and the two ground plans acquired independently with two different methods in Fig. 2. Observe missing buildings in both data sets, the different neighborhood relations and the differences in geometry. Other differences could show in the aggregation of building parts leading to larger building blocks. Without knowing the specifications it is not clear which data set is better or good enough. If the specification requires a planar accuracy of 1 m, the acquisition of buildings with more than 100 m^2 and

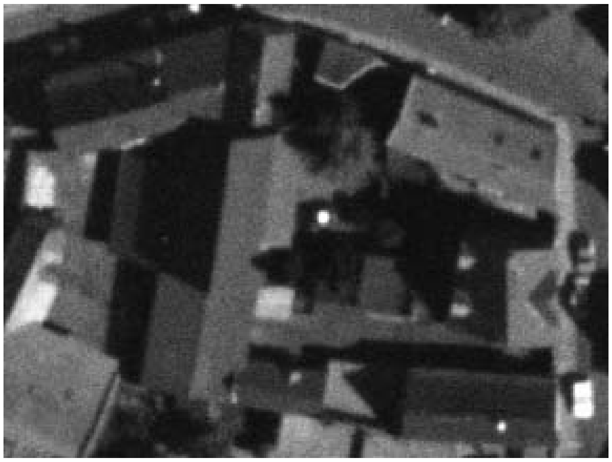

Figure 1: *shows a section of an aerial image with some buildings, ©DeTeMobil GmbH, Bonn, 1998*

separation of buildings in case of height differences larger than 3 m, the geometric differences appear acceptable. However, some small buildings appear to be superfluous and, without explicit reference to the 3D-structure, some building parts might have been better fused.

This paper develops automatic methods for comparing two sets of regions, especially polygons, and the evaluation of their difference. We do not want to refer to a specific application. We want to develop quality characteristics, which might then be used for specification, control or quality check. Automatic quality checks enable complete checking of data and give more objective results.

In spite of polygons being a subset of 2D-data, it already appears to be a complicated task. The envisaged techniques, however, may be used in a much broader context, or even be extended to evaluating 3D structures.

1.2 Previous Work

Quality control and especially the assessment of geometric 2D-data is quite a new research topic. *Geometric precision*, being a subset of quality characteristics, however, is an old issue in geodesy and surveying. Not surprisingly, most articles apply concepts from uncertainty representation to GIS-data ([Kraus and Haussteiner 1993; Bill 1996; Caspary and Scheuring 1992]). We may represent geometric data in vector or raster format, also leading to different representations of uncertainty in 2D.

The uncertainty of *point* data can easily be described by the covariance matrix of their coordinates, already proposed by Baarda in the 50's ([Baarda *et al.* 1956]) later leading to his well known concept of criterion matrices, also usable as substitute matrices, which allow a quite compact representation of the uncertainty of point fields with only a few parameters. The uncertainty of *straight lines*, already mentioned in ([Wolf 1968]), later is used also to describe the uncertainty of polygons ([Kraus and Haussteiner 1993]), though not linking the uncertainty of points and line segments into a common representation. The uncertainty of *arbitrary curves* is more involving and in general requires concepts from stochastic processes, which is quite involving ([Kiiveri 1997]). The uncertainty of the geometry of compound object has not been addressed up to now.

Recently the uncertainty of regional data has been addressed from the point of classification uncertainty in remote sensing ([Molenaar and Cheng 1998; Fritsch *et al.* 1998]), representing regions in raster format. The distinction into points, straight lines and curves is not necessary anymore. Regions of any structure, e. g. with holes, can be handled. However, only summarizing characteristics of the uncertainty of regions and boundaries can be handled. Also, parts of regions cannot easily be addressed.

Topological characteristics of regions have been investigated quite early
([Egenhofer and Sharma 1993]). Crisp relations between regions can be derived in a simple manner using set theoretic concepts. The early concepts have been extended to not simply connected regions, to lines and points ([Egenhofer and Herring 1991]). Only recently these concepts have been extended to allow uncertainty ([Winter 1994]), and only restricted to simply connected regions.

In a previous paper we have presented a concept for evaluating sets of regions of arbitrary structure ([Ragia and Winter 1998]). However, only the quality characteristics of the geometry have been described in detail. Here we present the concept more formally and give technical details for the automatic determination of the quality characteristics.

1.3 Outline

We first develop a scheme for describing the quality of planar spatial objects, especially of sets of non overlapping regions, usually given as polygons. It is independent of the representation of the spatial structures. We propose measures for the quality of sets of polygons, especially of building ground plans. Here the special representation, raster and vector, of the underlying data and their use are taken into account leading to hybrid analysis techniques. An empirical study, based on real data of a classical and a semiautomatic procedure for building extraction, demonstrates the usefulness of the proposed measures.

2. Quality of Planar Spatial Objects

2.1 General Structure

2.1.1 The Task

We assume two sets $R^j = r_i^j, j = 1, 2, i = 1, ..., I_j$ of regions r_i^j are given. Each region is described geometrically. They are assume to be not necessarily simply connected. The regions within one set may be related by some typed neighborhood relation. E. g. regions a and b may be disjoint, may touch or may overlap.

The geometry of the regions of two sets are assumed to refer to the same coordinate system. Thus correspondence between the regions in one set and the regions in the other set can be established by comparing their geometric descriptions, without taking any coordinate transformation into account.

The task is to qualitatively and quantitatively measure the structural and geometrical differences between the two sets, and to develop tools for testing

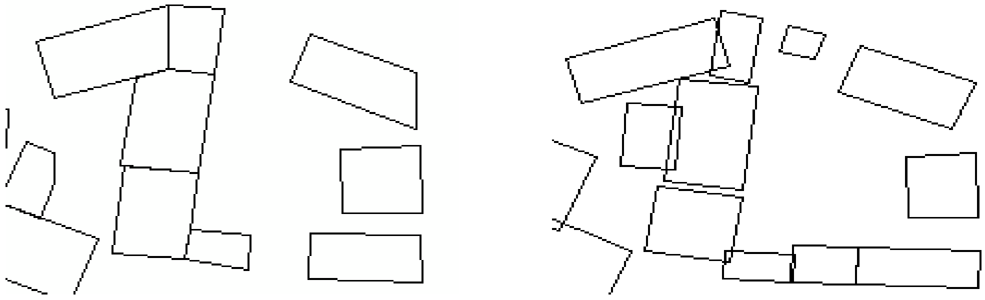

Figure 2: *shows two sets of ground plans of the building structures shown in Fig . 1. The left one is acquired using an analytical plotter, the right one with a semiautomatic digital system. The result of the semiautomatic system (right) is the union of the acquired individual ground plans.*

the equivalence of the two sets or, for detecting and identifying the differences between the two sets.

Finally quality means the difference between two data sets, if the one is the reference for the other.

2.1.2 Classes of Differences of Sets of Regions

We may distinguish several classes of differences of two sets of regions:

1. *geometric* differences.
   In spite that the assumption of the geometric description of the two sets refers to the same coordinate system we need to distinguish between
   (a) differences of *form* and
   (b) differences of *location*
   This enables us to identify e. g. systematic errors between the two sets or irregular generalization errors.
2. *structural* differences. Here we need to distinguish $m:n$ relations.
   (a) differences in *partitioning* while showing equivalent boundaries and
   (b) differences in *existence*, namely missing and spurious regions.

Obviously, these differences may occur simultaneously, making a complete taxonomy of differences impossible. Therefore it appears to be adequate to restrict to the most common cases which involve only a few regions at a time.

2.2 Structural Differences

2.2.1 The Region Adjacency Graph

The structure of one set R^j of regions may be described by the region adjacency graph (RAG) $G_A = G(R^j, A^j, \rho^j, \alpha^j)$, where the edges A^j represent neighborhood relations between the regions within on set (cf. fig. 3). Both, the regions as well as the relations are attributed by ρ^j and α^j resp. This allows to characterize complex regions, consisting of several (atomic) regions by their neighborhood relations. Complex regions thus lead to connected components of the RAG. This characterization appears useful, as one then might identify missing links or changes in partitioning, which would be difficult to do in case only the geometry is given. Up to now we assume binary relations to be sufficient. On the other hand, the relations may be uncertain due to the uncertainty of the underlying geometric description of the regions, e. g. allowing to have multiple weighted attributes, such as e. g. (touch, 0.3), (overlap, 0.7), indicating the two regions likely overlap, but may also touch. The selection of the neighborhood types and possibly their uncertainty is task dependent. Due to the uncertainty of the original data, connected components in the RAG not necessarily correspond to complex regions.

2.2.2 The Region Correspondence Graph

The region correspondence graph (RCG) $G_C = G(R, C, \rho, \gamma)$ is similar to the RAG. It refers to all regions $R = R^1 \cup R^2$. It also contains typed and possibly uncertain neighborhood relations C^1, C^2, but now between regions of R^1 and R^2. The most important relation here is equal, but also contains, contained by, covers, covered by and strong overlap, as neighboring relations are relevant here. Thus the RCG is bipartite, as it only contains edges between the two disjoint sets R^1 and R^2 of regions.

In the ideal case of no differences each region in R^1 corresponds to exactly one region in R^2. Thus the connected components of the RCG then consist of exactly two regions. Otherwise the connected components of the RCG are the maximal subgroups of regions in R^1 and R^2 which need to be addressed and analyzed with respect to differences in the corresponding subgraphs of the two RAGs (cf. fig. 4). In case the number of regions in these connected components and the number of different neighborhood

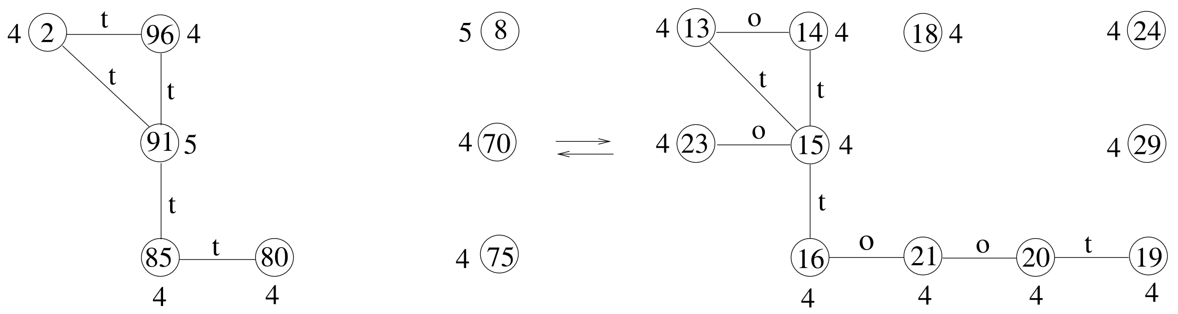

Figure 3: *shows the two RAGs of the building structure of Fig. 2. The relations are o for* overlap, *and t for* touch.

types is small, one could explicitly classify these differences and use them for evaluation.

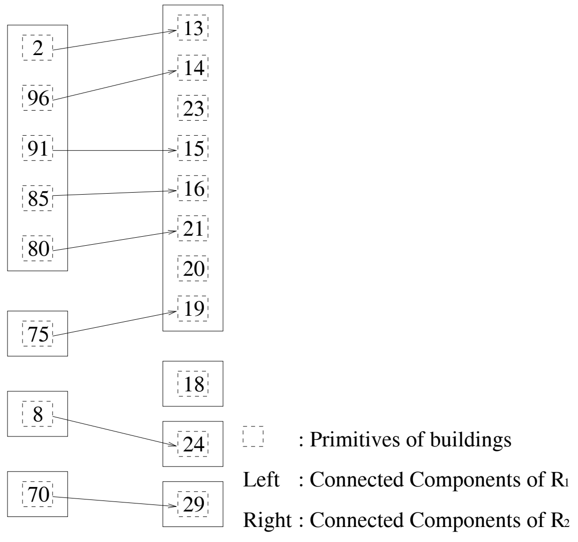

Figure 4: *shows the RCG of the building structures of Fig. 2. Observe the clear grouping onto connected components, from which existence of regions and partitioning of regions may be analyzed.*

2.3 Geometric Differences

Geometric differences may refer to the area covered by the regions of the boundary of the regions.

2.3.1 Differences in Form

In case no systematic errors occur and the geometric descriptions refer to the same coordinate system only form differences occur. In general, they may be of arbitrary nature. However, depending on the application, form differences refer to typical errors in the data acquisition.

The form difference may relate to the area covered by the regions, then the symmetric difference $a \ominus b = (a-b) \cup (b-a)$ of the regions a and b may be analyzed, e. g. by determining their number, their thickness, or their form, especially in case of severe differences. No explicit correspondence between boundary points needs to be established.

The form difference also may relate to the boundary line, which then requires some correspondence between the boundary lines of the two regions in concern. Based on corresponding points the average or maximum distance between the boundaries may be determined in order to yield some measure of closeness. In case of large form differences it may be that not all boundary points have a corresponding one, then the length of these non matching boundary segments may be used to characterize the differences.

2.3.2 Differences in Location

In case of systematic errors in the data acquisition one has to expect small location differences, usually small translations, sometimes also rotations. They may be determined by minimizing the difference in form, e. g. by minimizing the area of the symmetric difference or the maximum distance between the boundaries.

2.4 Quality Characteristics

Based on the analysis of the RAG, the RCG and the geometric properties of corresponding regions we may derive a set of *quality* characteristics for a *test data set*, which has to be evaluated, with respect to a *reference data set* of regions, e. g.:

1. structural characteristics
    (a) # of missing regions (completeness)
    (b) # of spurious regions
    (c) degree of partitioning
    (d) degree of aggregation

   These characteristics may e. g. be shown in a table indicating how many possibly connected regions in the reference data set are represented by how many possibly connected regions in the test data set ([Fuchs *et al.* 1994]).

A more detailed analysis may take into account partial equivalence of complex regions and only reporting partial deviation, e. g. in case two regions in a complex region with five regions correspond to one region, whereas the other three have equivalent (equal) correspondences.

2. geometric chararacteristics Geometric characteristics are only meaningful for two region which have been identified as corresponding.
    (a) average, maximum distance of the boundaries
    (b) average, maximum difference in orientation, curvature of the boundaries
    (c) # of regions in the symmetric difference of the regions
    (d) size and form of the regions in the symmetric difference
    (e) difference in location, rotation
    (f) difference in geometric shape parameters, e. g. parameters of rectangle in vector representation, size of minimal bounding box, etc.
    (g) # of boundary points, in case of vector format.

The usefulness of these measures highly depends on the application or the specification for data acquisition. Therefore we specialize the discussion and give more detailed information on quality measures for evaluating ground plans of buildings.

3. Measuring the Quality of Building Ground Plans

3.1 Representation of Building Ground Plans

Ground plans of buildings show restrictions, both, topological as well as geometrical. Individual buildings usually have a polygonal ground plan, sides often being orthogonal or parallel. They do not overlap, but often touch. In case data acquisition refers to building parts, however, they may overlap, e. g. when representing a L-shaped building. Depending on the degree of generalization, the number of polygon points may severely differ, without changing the overall shape too much.

We use raster and vector representation. The data usually are given in vector format. In the experiments one of the acquisition techniques actually represents the ground plan of a building as the union of primitive regions: $r = \cup_i r_i$. Some of the analysis appears to be easier in raster format, e.g. comparing the form of two boundaries represented as polygons is much more complicated in vector format than in raster format. We actually use a hybrid raster format, where 2-cells represent areas, 1-cells represent boundaries and 0-cells represent corners. Thus each object is represented by three matrices containing the 0-, the 1- and the 2-cells. This allows easy built up of the union of simple regions, easy topological reasoning as well as determining distances between boundaries independent on the complexity of region.

3.2 The Neighborhood Graphs

The determination of the RAG and the RCG is performed in raster format. Each region, i. e. building primitive, therefore is transformed into the hybrid raster format. For speeding up the computation of the neighborhood relations, the bounding box of each region is used.

The RAG's $G_A^j = G(R^j, A^j, \rho^j, \alpha^j)$ contain for each set of regions R^j all pairs of regions $r^j_{i'}, r^j_{i''}, \in R^j$ which do not show the relation $r^j_{i'}\,\mathsf{disjoint}\,r^j_{i''} = a^j_{i'i''} \in A^j$. This is done using the 9-intersection, containing the intersection of the interior, the boundary and the exterior of both regions.

The regions are attributed, specifying their form and possibly structure. Attributes ρ^j of the regions e. g. are:

- List of boundary points
- area
- # of holes

The relations A^j are also attributed. Attributes α^j e. g. are:

- type of neighborhood relation, i. e. touch, overlap, equal etc.
- degree of overlap $|R^1_{i'} \cap R^2_{i''}| / |R^1_{i'} \cup R^2_{i''}| \in [0,1]$

The RCG $G_C = G(R, C, \rho, \gamma)$ contains for both sets R^1 and R^2 all pairs of regions $a \in R^1$ and $b \in R^2$ which do not show the relations $a\,\mathsf{disjoint}\,b$ or $a\,\mathsf{touch}\,b$, thus which at least overlap to some extent. The attributes of the regions are the same as in the RAG's, whereas the attributes γ_i of the relations $c_i \in C$ describe the type of neighborhood.

3.3 Structural Differences

The structural analysis starts with the analysis of the complete graph $G_A \cup G_C = G(R^1 \cup R^2, A^1 \cup A^2 \cup C)$. Connected components contain pairs (R'^1, R'^2) of subsets of regions in R^1 and R^2 which may be equivalent. The bipartite subgraph $G'_C = G(R'^1 \cup R'^2, C')$ then directly gives the following information:

1. *missing regions* are elements in R'^1 with degree 0.
2. *spurious regions* are elements in R'^2 with degree 0.
3. *split ted regions* are elements r^1 in R'^1 with degree n connected to n regions $r^2_i \in R'^2$ with degree 1 and whose boundaries ∂r^1 and $\bigcup_i (r^2_i)$ are

geometrically equivalent. This is a $1:n$ relation between the regions. The equivalence relation is a weak version of `equal` and is defined below.

4. *merged regions* are elements $r_i^j \in R^1$ with degree 1 connected to the same region r^2 in R^2 and whose boundaries $\partial(\bigcup_i r_i^1)$ and ∂r^2 are equivalent. This is a $n:1$ relation between the regions.
5. all other situations show a $m:n$ relation. This may be analyzed further. E. g. all pairs (i', i'') for which the regions $r_{i'}^1$ and $r_{i''}^2$ are equivalent could be eliminated from the analysis, as they show identical subsets of the subset in concern. This would e. g. allow to identify missing or spurious subparts.

If not m:n relations then the transition table is proposed ([Fuchs *et al.* 1994]). Observe that there are four cases, which are important: (1) Both, the internal structure and the boundary are equivalent, in the case of single regions only the equivalence of the boundary is relevant; (2) the internal structure may be the same, however, the boundaries are not equivalent; (3) In spite of different internal structure the boundaries are equivalent, which contains the above mentioned special cases of splitted and merged regions; (4) both, structure and geometry are different, which may not allow a further specification of the differences.

Also the connectedness of the regions could be compared. This would allow to identify missing or spurious holes. It is not the scope of this paper, to exploit all these situations, but to demonstrate the potential of such a structural analysis.

Observe, that even for a simple region with equivalent boundary the two boundaries need not contain the same number of points, indicating structural differences, caused e. g. by small generalization processes. The geometric equivalence of regions with approximately the same boundary but with different number of boundary points is a typical example, why a raster based analysis at least initally is simpler to realize.

3.4 Geometric Differences

It appears quite complicated to check for the identity or closeness of two polygons or to identify differences between two polygons if they are given in vector format. Therefore the geometric analysis is performed in raster format. This allows easy realization. The operations, however, can also be realized in vector format, which might be preferable for performance reasons.

3.4.1 The Zone Skeleton

The comparison of the boundaries is based on their pointwise correspondence at places where the boundaries are not too different. The basis is the so called *zone skeleton.* It is defined as the skeleton of the symmetric difference $a \ominus b$ excluding those points with shortest distance to the same boundary, a or b. It is attributed by the *distance function* $d(s)$ giving the double distance of each point of the zone skeleton to the boundaries ∂a and ∂b, i. e. the distance function of two parallel boundaries lying 1 m apart, has the value 1 m.

As the zone skeleton is the set of maximal circles touching both boundaries, it at the same time establishes a *partial correspondence* between the boundaries, excluding extreme intrusions or protrusions. This appears very useful in our context, as we have an intuitive marked distance measure, indicating whether one region is locally outside or inside the other.

The last property might be a disadvantage if at least one of the boundaries is very rough as the extrema are not taken onto account. But in our application, this does not occur very likely.

Observe that the zone skeleton need not be simply connected, especially in case the regions have holes.

3.4.2 Analyzing the Zone Skeleton

The distance function can easily be used for checking the geometric equivalence of two regions. Two regions are called *geometrically equivalent* if the distance function $|d(s)| < t$ is smaller than a prespecified threshold t. This criterion is identical to requiring the boundaries to lie in the t-buffers-zone of the other boundary, however, only if complete correspondence between the regions can be established, i. e. if no large intrusions or protrusions occur.

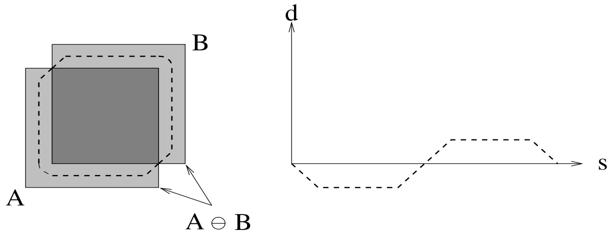

Figure 5: *shows the zone skeleton (left) of two strong overlap areas of a sketch and its distance function (right)*

In order not to depend too heavily on the choice of the threshold t, we use *two thresholds*, a small one t_s and a large one t_l. This leads to three cases, corresponding to a traffic light result:

1. *green*: In case $\max|d(s)| < t_s$ the regions certainly are equivalent.
2. *yellow*: In case $t_s \leq \max|d(s)| < t_l$ the regions may be equivalent, but showing moderate differences.
3. *red*: In case $\max|d(s)| \geq t_l$ the regions certainly are not equivalent, showing large differences.

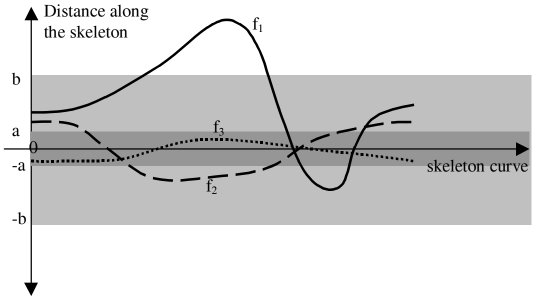

Figure 6: *shows the classification of the distance function*

In the uncertain case *yellow* a detailed analysis is performed with respect to form differences, to missing parts, to single outliers, to a shift, or do generalizations.

4. An Empirical Study

4.1 The set up for the empirical study

We applied our approach to real data of building ground plans including two sets of polygonal data of the same test area. One set is taken from an analytical plotter containing non overlapping general polygons; the second one is produced by a semi-automatic system for building extraction which has been developed at the Institute of Photogrammetry, University of Bonn [Lang and Schickler 1993], [Gülch 1997], [Gülch and Müller 1997], [Müller 1997] and contains overlapping rectangles, representing the ground plans as the union of primitives in a CSG manner.

The aerial images have a scale of 1:15000 and were scanned with 7μm pixel size, which corresponds to a ground resolution of 1dm/pixel.

4.2 The results of the test

As two data sets are generated from the same image data and following the same specification – buildings should be larger than 50 m^2 and be with in a tolerance of 1 m – none of them can be taken as reference data set with significantly superior accuracy. The first data set has 43 buildings and the second one 48. From these buildings 40 have been matched. For determining the internal neighborhood relations in the RAG we took the required accuracy of 1 m into account. For determining the correspondences in the RCG we used two thresholds $t_s = 1$ m and $t_l = 3$ m, the lower one again reflecting the specification for the data acquisition.

4.2.1 Characteristics of the data sets

The two data sets have the following characteristics:

- The number of regions, i. e. building primitives in the two data sets are 94 and 126, i. e. 220 in total.
- The number of points per polygon is between 4 and 28 in the first data set and 4 in the second data set, as it only contains rectangles.
- Analyzing the RAG's, in the first data set we only observed the topological relation *touch* (and, of course, disjoint). In the second data set we observed *overlap* and *touch* (Fig. 9).
- The number of the connected components, i. e. complex buildings in the data sets are 48 and 59, making a total of 107 complex buildings acquired.

4.2.2 Differences of the data sets

We first discuss the differences of the two data sets.

- **Structural differences** We observed the following structural differences resulting from the analysis of the RAG's:
  - There are more points per primitive in the first data set. Thus single buildings show a higher degree of generalization in the second data set (Fig. 9), which is to be expected from the type of data acquisition.
  - Overlap between the primitives appears quite often in the second data set (Fig. 9)(Fig. 2), which is intended.
  - There is a higher partitioning level in the second data set, i. e. the number of elements per connected component in R^2 is larger than in R^1.

Figure 7: *shows the original complex building ©DeTeMobil GmbH, Bonn, 1998*

An analysis of the correspondences yielded to the following results:

  - Of all connected 48+59 components in R^1 and R^2 40 components, i. e.

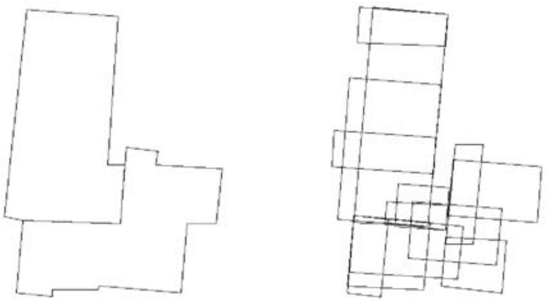

Figure 8: *shows two sets of ground plans of the building structures shown in Fig. 7. The left one is taken from an analytical plotter, the right from a semiautomatic digital system. The result of the semiautomatic system (right) is the union of the acquired individual ground plans.*

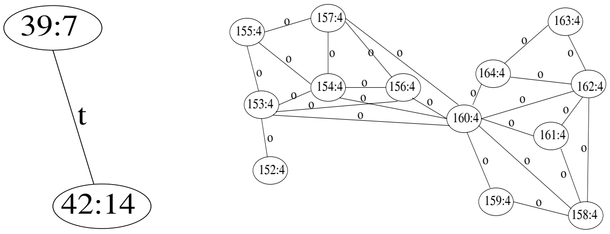

Figure 9: *shows two region adjacency graphs of the ground plans in Fig. 8. The left one corresponds to the from analytical plotter and the right one from a semiautomatic system.*

2×40/107=75 % have been matched. From those 40 components 23 i. e. 57% have the same number of regions, without necessarily having the same form.

  - From the other 17
    - there are 4 i.e 10% have split in the second data set with a partitioning grade of two,
    - there are 2 i.e. 5% have split with a relation 3:4,
    - there are 5 i.e. 12% have split with a relation 2:3,
    - there are 1 i.e. 3% have split with a relation 2:13,
    - there are 3 i.e. 7% have merged with a grade of two,
    - there is 1 i.e 3% has merged with a relation 3:2,
    - there is 1 i.e 3% has merged with a relation 5:4,

- **Geometry** For the geometrical differences we have the following results:
  - *green*: In total 40 components 16 i. e. 40% of the buildings are accepted, as they only show small or no differences (Fig. 10),
  - *yellow*: 15 i. e. 37% buildings have moderate differences (Fig. 11),
  - *red*: The remaining 9 i. e. 23% have big differences (Fig. 12).

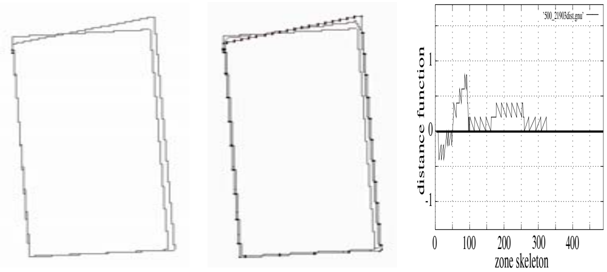

Figure 10: *shows the overlap of two buildings with small or no differences (left) and their overlap with the zone skeleton (middle), and the distance function (right).*

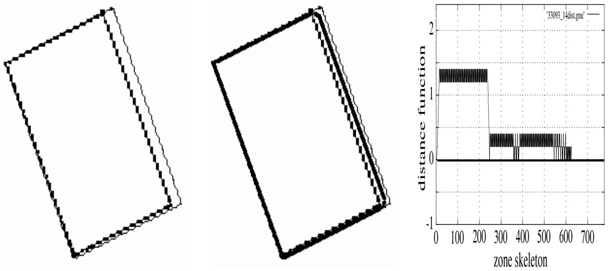

Figure 11: *shows the overlap of two buildings with small or no differences (left) and their overlap with the zone skeleton (middle), and the distance function (right).*

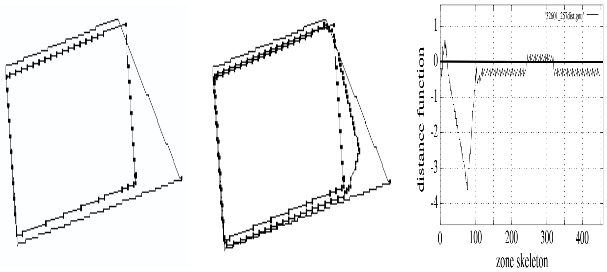

Figure 12: *shows the overlap of two buildings with small or no differences (left) and their overlap with the zone skeleton (middle), and the distance function (right).*

- **Result:** ***yellow***: Further analysis of the buildings with moderate differences shows that:
  - Four (27%) of the buildings with moderate differences show a reduction or a magnification,
  - Four (27%) have a missing part,
  - Three (20%) have more than one missing part,

  - Three (20%) have differences in the measurement of one or two points,
  - One (6%) has a shift.
- **Topology-Geometry** From the 23 subgraphs which have the same partitioning 13 i. e. 56% have no or small geometrical differences (obtained by visual comparison). That means 32% of the whole data set have no geometrical and structural differences.

		Geometry		
		16 green	15 yellow	9 rot
Topology	23 isomorph	13	8	2
	17 not isomorph	3	7	7

Figure 13: *shows the results of the geometrical and structural differences*

5. Quality Result

The results of the analysis can only be evaluated with respect to some specifications. We do not want to speculate about possible specifications, or use an existing one.

In order to give a simple example we assume, only the success rate is to be evaluated.

We distinguish between two cases:

A: Dataset A, is reference
B: Dataset B, is reference

In table 1 the success rate and the 95 % confidence interval are given.

1. *Specifications S1*:
   Let us first assume, the specification is: The success rate must be larger than 90 %.

   Then the results of case B would appear to be acceptable. The success rate, however, actually is an estimate: the true success rate can lie in the 95 % confidence interval [86 %,98 %]. This follows from the Binomial distribution $B(n, p)$ with $n = 43$, being the number of samples being tested, and $p = 40/43$ the success rate and assumes the success rate to be constant for all buildings.

   The achieved result does not really prove the success rate to be above 90 %!

2. *Specifications S2*:
   Therefore we could give more detailed specifications: The 95 % confidence interval for the estimated success rate should lie in the range 90 % to 100 %.

   Obviously these specifications cannot be achieved in both cases. The specifications could be reached for case B with success rate 93 % if the confidence interval would be smaller. The large confidence interval is due to the too low number of used data for the evaluation.

   Assuming, the success rates to be true, we would need at least five times more data to fulfill the specifications, leading to a confidence interval of [90.1 %, 95.8 %]. In this case we could also argue, that we have proved the specifications S1.

   The uncertainty of the achieved success rate is confirmed by the fact, that one additional detected building would lead to an acceptance of the specifications, with success rate 95.3 % and 95 % confidence interval [90.7 %,100 %].

Similar statements could be established for other types of specifications. *The quality of the parameters obviously needs to be used when proving the fulfillment of specifications.*

1	one eye reference	stereo
Buildings	48	43
Buildings	40	40
2	83%	93%
confidence interval 95%	[75%-92%]	[86%-98%]

Table 1: *This table shows the success rate and the confidence interval*

6. Conclusions

The paper developed an approach for assessing the quality of ground plans of buildings. Quality is measured not only by geometrical but also by *structural* differences between an acquired data set and a reference data set.

New hybrid techniques for automatically determining quality measures are developed. The basis is the region adjacency graph (RAG) of each data set revealing the internal structure of the buildings. The region correspondence graph (RCG) allows a detailed analysis of the structural differences of corresponding complex buildings. The uncertainty of the given data is taken into account when building up the RAG and the RCG, which allows to include tolerances given in the specifications.

We applied the technique to real data and showed it to be useful for checking the quality of data acquisition of complex buildings.

Automating quality assessment increases efficiency in checking data, allowing complete checks instead of sampling, moreover it makes quality checks objective. The developed techniques are applicable to sets of 2D regions of any type and internal structure.

Our final goal is to automatically produce a quality report which includes all details on the differences between an acquired and a reference data set. In case of multiple acquisition techniques the user may decide on which of them is more useful for his special application.

Acknowledgement

We would like to thank DeTeMobil GmbH, Bonn for giving us permission to use their data. We wish to acknowledge Stephan Winter for the use of his programs and Eberhard Gülch, Thomas Läbe and Hardo Müller for their support in this work.

References

[Baarda *et al.* 1956] BAARDA, W; GROOT DE, D.; HARKINK, F. (1956): *Handleiding voor de technische werkzaamheden van het kadaster (Manual for the Technical Activities of the Cadastral Service in the Netherlands.* Gravenhage, 1956.

[Bill 1996] BILL, RALF (1996): Metainformation und Datenqualität in Geo-Informationssystemen. In: *Workshop Datenqualität und Metainformation in Geo-Informationssystemen*, pages 141–154. Universität Rostock, 1996.

[Caspary and Scheuring 1992] CASPARY, WILHELM; SCHEURING, ROBERT (1992): Error-Bands as Measures of Geometrical Accuracy. In: *EGIS 92*, pages 226–233, Utrecht, 1992.

[Egenhofer and Herring 1991] EGENHOFER, MAX J.; HERRING, JOHN 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, 1991.

[Egenhofer and Sharma 1993] EGENHOFER, MAX J.; SHARMA, JAYANT (1993): Topological Relations Between Regions in R^2 and Z^2. In: ABEL, D.; OOI, B. (Eds.), *Advances in Spatial Databases*, pages 316–336, New York, 1993. 3rd Symposium on Large Spatial Databases SSD '93, Springer LNCS 692.

[Fritsch *et al.* 1998] FRITSCH, DIETER; GLEMSER, MICHAEL; KLEIN, ULRIKE; SESTER, MONIKA; STRUNZ GÜNTER, OBERPFAFFENHOFEN (august 1998): Unsicherheit von Vektor- und Rasterdaten. *GIS*, (4):26–35, august 1998.

[Fuchs *et al.* 1994] FUCHS, C.; LANG, F.; FÖRSTNER, W. (1994): On the Noise and Scale Behaviour of Relational Descriptions. In: EBNER, HEIPKE, EDER (Ed.), *Int. Arch. f. Photogr. and Remote Sensing, Vol. 30, 3/2*, Band XXX, pages 257–267, 1994.

[Gülch and Müller 1997] GÜLCH, E.; 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*, 1997.

[Gülch 1997] GÜLCH, E. (1997): Application of Semi-Automatic Building Acquisition. In: GRÜN, A. (Ed.), *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II).* Birkhäuser, Basel, 1997.

[Kiiveri 1997] KIIVERI, H.T. (1997): Assessing, representing and transmitting positional uncertainty in maps. *International Journal of Geographical Information Science*, 11(1):33–52, 1997.

[Kraus and Haussteiner 1993] KRAUS, KARL; HAUSSTEINER, KARL (1993): Visualisierung der Genauigkeit geometrischer Daten. *GIS*, 6(3):7–12, 1993.

[Lang and Schickler 1993] LANG, FELICITAS; SCHICKLER, WOLFGANG (sep 1993): Semiautomatische 3D-Gebaeudeerfassung aus digitalen Bildern. *ZPF*, 61(5):193–200, sep 1993.

[Molenaar and Cheng 1998] MOLENAAR, M.; CHENG, TAO (1998): Fuzzy spatial objects and their dynamics. In: *Proceedings ISPRS Comm. IV Symposium, Stuttgart, Sept. 7-10*, 1998.

[Müller 1997] MÜLLER, H. (1997): Designing an object-oriented matching tool. In: *3D Reconstruction and Modelling of Topographic Objects*, Band 32 der Reihe International Archives of Photogrammetry and Remote Sensing, pages 120–127. ISPRS Commission III/IV Workshop, Stuttgart, September 17-19., 1997.

[Ragia and Winter 1998] RAGIA, L.; WINTER, S. (1998): Contributions to a quality description of areal objects in spatial data bases. In: *Proceedings ISPRS Comm. IV Symposium, Stuttgart, Sept. 7-10*, 1998.

[Winter 1994] WINTER, STEPHAN (1994): Uncertainty of Topological Relations in GIS. In: EBNER, H.; HEIPKE, C.; EDER, K. (Eds.), *Proc. of ISPRS Comm. III Symposium* Spatial Information from Digital Photogrammetry and Computer Vision, pages 924–930, München, 1994.

[Wolf 1968] WOLF, HELMUT (1968): *Ausgleichungsrechnung nach der Methode der kleinsten Quadrate.* Dümmler, Bonn, 1968.