GENERALIZATION BASED ON LEAST SQUARES ADJUSTMENT

Monika SESTER
Institute for Photogrammetry (ifp), Stuttgart University, Germany
Geschwister-Scholl-Strasse 24d, D-70174 Stuttgart
Monika.Sester@ifp.uni-stuttgart.de

KEY WORDS: Generalization, Data Aggregation, Multi-Scale Database, GIS - Spatial Databases.

ABSTRACT

The paper presents solutions for generalization problems using least squares adjustment theory. This concept allows for the introduction of several observations in terms of constraints and for a holistic solution of all these – possibly contrary and competing – constraints. Two examples are used to demonstrate the validity of this approach: the simplification of building ground plans and the displacement of arbitrary cartographic objects. Each approach is verified with several examples; furthermore, the integration of these different approaches is presented, in terms of the fusion of cadastral and topographic data.

1 INTRODUCTION AND OVERVIEW

Objects of our environment exhibit multiscale properties. Depending on the application, different features of an object are important. This concerns both thematic features and geometric features. In order to exploit this property, multi-scale representations of objects have to be available, as well as efficient analysis techniques, which are able to use this structure. The generation of multi-scale representation is a generalization problem. Generalization is needed in order to represent relevant information on an appropriate level of detail. As only a restricted amount of data can be represented on a certain level of detail, different pieces of information have to 'fight' for their representation on a specific aggregation level. This implies that generalization, especially displacement, is an optimization problem, where different goals have to be satisfied simultaneously. Thus, generalization procedures are needed both for the generalization of GIS databases (model generalization) and for cartographic generalization (visualization).

Least Squares Adjustment theory (LSA) is a well known general framework to determine unknown parameters based on given observations. This optimization technique is well founded in mathematics, operations research, and in geodesy. This general concept allows for the integration of different constraints in order to solve an overall, complex problem. This paper proposes to use adjustment theory for generalization. One problem is the set-up of the constraints for the various generalization tasks. The generalization of building ground plans is formulated in terms of a model-based approach, the problem being the determination of the model. In this case it is derived by the application of some rules. The second generalization operation treated with LSA is displacement: different objects have to be displayed on a map – for reasons of legibility certain constraints have to be satisfied, e.g. minimal object sizes and minimal object distances have to be enforced. LSA offers a straightforward framework to introduce different kinds of these constraints. In one step, all these constraints are solved simultaneously, resulting in one optimized solution with the feature that all residuals are distributed evenly among all the observations. Besides this result, quality parameters indicate how well the initial constraints have been satisfied.

The paper is organized as follows: after a review of related work, the simplification of building ground plans using a model-based approach is presented, together with some examples showing the possibilities and the deficiencies. Then the approach for displacement based on least squares adjustment is shown, giving both the theoretical background, and explanatory examples. The integration of the two approaches is demonstrated with the example of the fusion of cadastral information with topographic information. Finally, a summary concludes the paper.

2 RELATED WORK

The paper focuses on the automation of the generation of scale dependent representations. There is a huge number of research efforts in this domain. On the one hand there are approaches dealing with cartographic generalization, on the other hand there are methods for model generalization or database abstraction. In this context, different schemes have been proposed: pragmatic bottom-up approaches use specific algorithms for different generalization tasks. Then there are more holistic ideas, using top-down strategies, to strive at an integrated solution of this complex task. The latter is currently in the focus of most research issues [Mackaness, Weibel & Buttenfield 1997], [Lamy, Ruas, Demazeau, Jackson, Mackaness & Weibel 1999]. These issues have to deal with the integration of different methods for the solution of the compound task.

Simplification aims at a reduction in the number of points, the object is composed of, with the restriction that the characteristic shape of the object is preserved. For arbitrary shapes, algorithms respecting the curvature of the object are typically used. The most prominent algorithm is developed by Douglas & Peucker [1973], taking the maximal distance of points from a hypothetical generalized line into account. Such algorithms cannot simply be applied for special shapes like rectangular structures, that are the main characteristics of man-made objects like buildings. Here, additional constrains like parallelity and rectangularity have to be taken into account. For the building simplification task, e.g. Staufenbiel [1973] developed an algorithm by deriving a set of rules of what to do with a building facade which is too small to be represented. He came up with a detailed processing scheme which later was implemented in the CHANGE software [Powitz 1992]. Similar concepts are currently being implemented in GIS products, like ESRI [Lee 1999]. The algorithm developed at within the AGENT project [Lamy et al. 1999] aims at a concurrent treatment of several competing aims: squaring, rectification, and enlargement of minimal narrow objects inside the building. The approach presented here is similar to a method that has been used for the reconstruction of parametric buildings in aerial images [Sester & Förstner 1989].

There is a long history of automatic methods for displacement: Nickerson [1988] developed an approach for linear objects, which are displaced according to the degree of overlap of their corresponding signatures. The use of so-called displacement mountains is proposed by [Jäger 1990]. The importance of an object is coded in a raster image, with important objects producing high mountains. This mountain allows for the extraction of the degree of displacement and the direction. Mackaness [1994] proposes a radial displacement of objects around a given point. His approach allows for the degradation of the degree of displacement depending on the distance from the given point. Another approach is given by Ruas [1998] aiming at an incremental displacement. Objects producing the greatest conflicts are treated first. As with all local operators, secondary conflicts can be induced by solving one conflict, thus there is a need for iterative operations, and a control strategy to decide, when to stop. Only recently, an emergence of using 'force models' to solve displacement can be observed: Burghardt & Meier [1997] use a snake approach to displace linear elements in maps (e.g. road and rail network). Hojholt [1998] came up with the idea of applying Finite Element Theory as a framework for displacement. Bobrich [1996] also uses a mechanical analogue as a model, namely springs. Ware & Jones [1998] use simulated annealing as an optimization technique. Independently, Harrie [1999], Sarjakoski & Kilpelainen [1999] and Sester [1999] developed ideas to solve displacement with Least Squares Adjustment theory. These approaches aim at a global reduction of all spatial conflicts.

3 SIMPLIFICATION OF BUILDING GROUND PLANS

The main constraints involved in the simplification of building ground plans are preservation of right angles, collinearity and parallelity. These characteristics do not only have to be preserved, they even have to be enhanced or exaggerated, in order to give the observer the impression of seeing a characteristic building. This also concerns typical building parts, that have to be enlarged if they are considered as important. The factor which triggers simplification is the minimal length of a facade which can be represented at a certain scale. All the building sides smaller than this critical length have to be replaced.

The algorithm treats individual polygons one after the other. It works locally, trying to replace polygon sides which are shorter than the critical length. These sides are substituted according to some given rules, depending on the geometry of the adjacent building sides. This results in a simple, but not necessarily correct model of the original situation. Thus, in an adjustment process, the original building shape is adjusted to the building model. The advantage of this procedure is the fact that the rules can be simple and coarse, as they yield only approximate values, which are refined in the adjustment process. Furthermore, also additional parameters can be introduced in the adjustment, like the fact that the building size has to be preserved.

3.1 Algorithm

The first step of the algorithm results in a simplified ground plan, which is used as a model. This model is described in terms of building parameter, that are in turn adjusted to the original ground plan in the subsequent adjustment process. The decision of how to substitute a short facade depends on the geometry of the neighboring sides:

- ▷ Intrusion / extrusion: the angle between preceding and subsequent side is approx. 180 °. The small side is just set back to the level of the main facade.
- ▷ Offset: the angle between preceding and subsequent side is approx. 0 °. The longer one of the adjacent building sides is extended, and the shorter side is dropped.

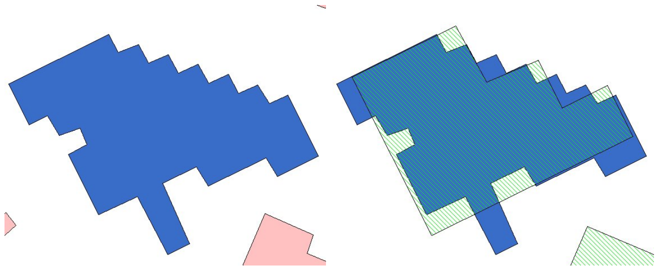

Figure 1: Application of simplification to staircase-building. Situation before (left) and after the simplification (right, shaded area).

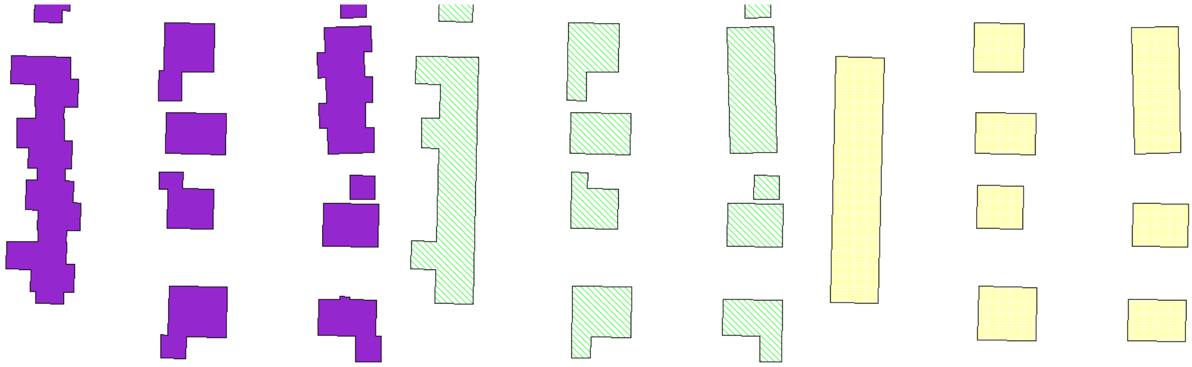

Figure 2: Application of algorithm to some buildings of a small village. Use of small (3m) and larger minimum facade length (7m), respectively.

- ▷ Corner: the angle between preceding and subsequent side is approx. 90°. The adjacent facades are intersected.

These rules are iteratively applied to all the small sides of a building, starting with the shortest ones. This has the positive effect, that also special buildings structures can be generalized nicely, e.g. step-form-buildings (cf. Figure 1).

In some cases, building structures have to be exaggerated. Consider e.g. a building part that is very long, however very narrow. If the length of the narrow side is below the minimal side, this structure is dropped. If, however, the total area of the structure is greater than a threshold it is considered to be an important part of the building which has to be preserved. This is achieved by enlarging the smaller side to the minimal side.

The approximate building model is then transformed into a parametric representation by transforming it into its two main axes. The different width and length parameters can directly be derived: A simple rectangular building thus can be described by two parameters (width and length), whereas an L-shaped building needs four parameters. These form-parameters (plus two additional position parameters) are the unknowns introduced in the least squares adjustment. The observations are the original building edges, that are represented as a functional of the parameters. The stochastic model describes the accuracy of the observations, namely the original building edges. It is modeled in terms of their overlap with the model edges and describes the accuracy of the start and endpoint, as well as the accuracy of the edge as a whole. In this way, a short edge has a lower accuracy as opposed to a longer one, resulting in the fact that the adjusted model line lies closer to the longer edge. In addition to the building edges, also the parameters are introduced as so-called additional parameters; this allows to give them a certain accuracy. This is especially important in the case that a certain building parameter has to be enforced, as it has been exaggerated. In this case, the accuracy of the parameter has to be set very high. In all the other cases, they are low, namely in the range of the minimum facade side.

Figure 2 shows an example of an extract of buildings in a village. Depending on the minimum facade lengths, different generalizations are achieved. The simplification operates on individual buildings. However, in dense city areas, additional constraints appear, as neighboring objects have to be simplified in common. Imagine a ground plan of a city with adjacent objects. Either the objects share a common side (topological adjacency), or they are so close that their distance is no longer noticeable at that scale (i.e. below the critical value specified above). In this case the objects have to be aggregated to form building blocks. After that, the new polygons are simplified. The following example shows an extract of the city of Stuttgart[1]. In order to make a transition from the given cadastral situation to the scale 1:25.000, the minimal facade length is 0.3mm, corresponding to 7.5m. Thus the threshold is set to 7.5m. As there are many adjacent buildings, as well as buildings separated by small gaps, in a first step neighboring buildings are aggregated. This was accomplished

[1] Building ground plans by courtesy of the Stadtmessungsamt of the city of Stuttgart.

by a morphological dilation followed by an erosion, realized in ArcView with a buffer operation. The result of the simplification is given in Figure 3.

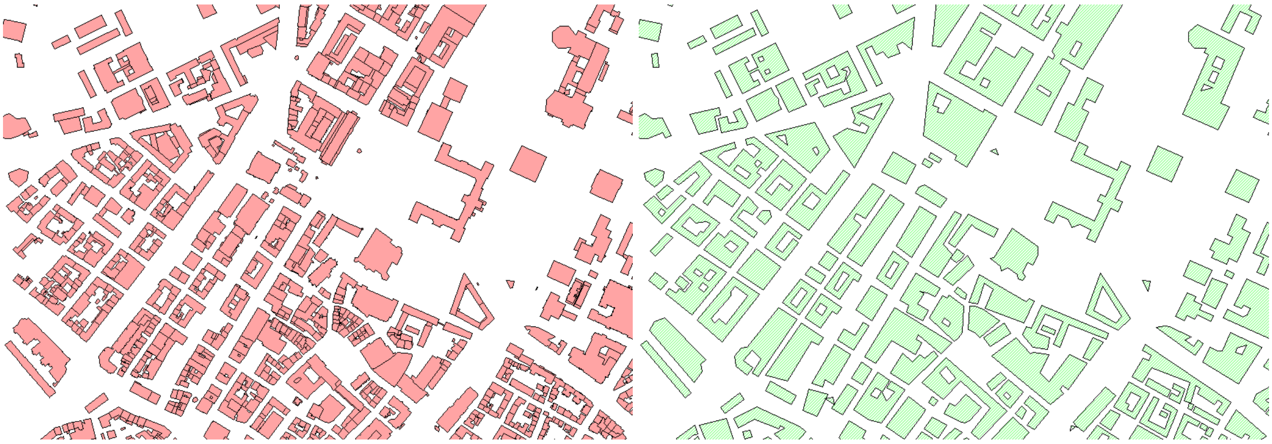

Figure 3: Application of simplification to dense city area. Left: original situation; right: situation after aggregation and simplification.

This approach is quite straightforward and offers appealing results. However, there are also some deficiencies that the current version of the algorithm cannot handle. The courtyards inside a building are treated as separate objects, which could have the effect that they intersect the outer building boundary after the simplification. Furthermore, interior minimal constraints can not be treated. If a building has a narrow interior part (e.g. the bar of a h-shaped building), this part needs to be enlarged in order to be still visible. As the narrow distance is not modeled explicitly in terms of the building sides, it cannot be respected in this current implementation. In order to take it into account, also interior (artificial) sides have to be introduced. One of the advantages of the algorithm is that squaring is implicitly enforced by the transformation of the building into its two main axes. If a building, however, has more than two axes, this results in the fact that the rectangularity cannot be guaranteed any more. On the other hand, such buildings usually have no squared shape.

4 DISPLACEMENT

Displacement is modeled as an optimization procedure, where the position of objects has to be optimized with respect to some given constraints. Thus the unknowns in this process are the coordinates; the constraints are described as functions of the coordinates (functional model) that can be observed with a certain accuracy (stochastic model). Constraints can be two-fold: on the one hand there are the exterior constraints in terms of minimal distances between objects that have to be enforced. On the other hand are the internal constraints of the objects, namely form parameters of the objects. The following set of constraints is introduced in the system:

- ▹ Form parameters: object sides, angles, orientation,
- ▹ distances between objects: minimal distance that has to be enforced and critical distance, that indicates that the objects have to be merged (by setting the distance to zero),
- ▹ additional parameters: the coordinates.

All these observations l are introduced into the conventional least squares adjustment. They form the Jacobean Matrix $\mathbf{A}$. As these functions are not linear, they have to be linearized with respect to given approximate values. Each observation has a corresponding accuracy (or weight), described in the matrix $\mathbf{P}$. These weights describe how well the observation has to be enforced. They can be used to describe different object properties: the objects can be movable or fix, or they can be deformable or stiff. The stiffness is e.g. ensured by assigning a high weight to the internal form parameters of the object (object sides and angles). The introduction of the unknowns as additional parameters allows the assignment of accuracies. This can be used if an object is considered non-movable by assigning high weights to its coordinates.

In order to determine the neighborhood of the objects, and thus to identify possible conflicts, a triangulation net using a Constrained Delaunay Triangulation of all the object points is established (cf. also the generalization approach of Bundy, Jones & Furse [1995], that is based on the so-called simplicial data structure (SDS)). Thus, all the objects are embedded in a network and distances between them can be computed.

4.1 Functional and stochastic model of the observations

Object sides: The object side s_{12} is derived by the Euclidean distance between the endpoints:

$$s_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The weight of the observation is high, when the object is considered as stiff, otherwise it is low.

Object orientation: The object orientation ϕ_{12} is the direction of one object side s_{12}.

$$\phi_{12} = \arctan\left[\frac{x_2 - x_1}{y_2 - y_1}\right]$$

If the orientation has to be preserved, then this observation gets a high weight.

Object angles: The object angle $angle$ is derived by the difference of adjoining direction angles ϕ:

$$angle_{pn} = \phi_{previous} - \phi_{next}$$

The weight of the observation is high in order to stabilize the object form.

Distances: The distances are derived as heights in the respective triangle, that two different objects share. If the distance is below a critical value $crit\_dist$, then the distance is set to zero. This has the effect of merging objects that are very close to each other instead of shifting them apart. If it is below the minimal distance min_dist then it is set to that distance. These distances have to be enforced, therefore the observations get a high weight. Distances greater than the minimum distance can be preserved, and are allowed to move up to the minimum distance. In the equation, the object edge s_{12} is the link between points 1 and 2, whereas point 0 is the third point in the triangle belonging to the other object.

$$dist = \frac{1}{s_{12}} * [(x_0 - x_1)(y_2 - y_1) - (y_0 - y_1)(x_2 - x_1)]$$

Additional parameters: The coordinates are introduced as additional parameters: if objects are fixed, then these observations get a high weight.

4.2 Least Squares Adjustment

After all the observation equations are set up, the solution of the unknown parameters $\hat{x}$ is gained by solving the following equation:

$$\mathbf{\hat{x}} = (\mathbf{A^T P A})^{-1} \mathbf{A^T P} (\mathbf{l} - \mathbf{f}(\mathbf{x_0}))$$

where $\mathbf{A}$ is the Jacobean Matrix of the derivations of the functions according to the unknowns, $\mathbf{P}$ is the weight matrix, l are the observations and $\mathbf{f}(\mathbf{x_0})$ is the value of the function calculated at the approximate values $\mathbf{x_0}$.

As distance constraints between all the objects are formulated, this ensures that a global solution is found, where a displacement of one object occurs in accordance with all its surrounding objects. The result can be analyzed with respect to different criteria: the change of length of the objects' sides gives an indication for their deformation. The same holds for the change of the internal angles. As a measure for the absolute positional accuracy, the change in the coordinates can be used. In this way, these measures can be used to evaluate the quality of the result and possibly initiate other generalization operations.

4.3 Examples

Figure 4 visualizes the process of displacement for a situation of streets and buildings. The streets are assumed to be fix, whereas the buildings are allowed to move; a minimum distance of 7m between the objects is required. The situation is displaced, based on the neighborhood derivation from a Delaunay-Triangulation (4b)). The result of the displacement is given in Figure 4c). It can be seen, that the objects have been displaced against each other in order to enforce the minimum distances. If there is not enough space to move for the objects, they can be deformed. The colors in Figure 5a) indicate the degree of deformation the buildings have undergone. In the first example the deformations are in the range of 30cm (in object space), which is very good. The situation changes, however, if the constraints are too big: imposing a minimum distance of 15m leads to serious deformations of the buildings between the two streets, it also shifts the streets by some degree – although they were set to be fix. The high degree of deformation, visible in Figure 5b) is a good indicator for a situation, where the constraints are too high and where an interaction is necessary: Either the constraints have to be loosened, or space for the objects to move to has to be provided by eliminating some buildings. Figure 5c) shows the effect of also allowing the streets to move – now the buildings push against the streets, which in turn are shifted. Now the deformations are again in the range of 30 cm. The advantage of this approach is the possibility to derive quality measures that indicate to what degree the constraints could be satisfied.

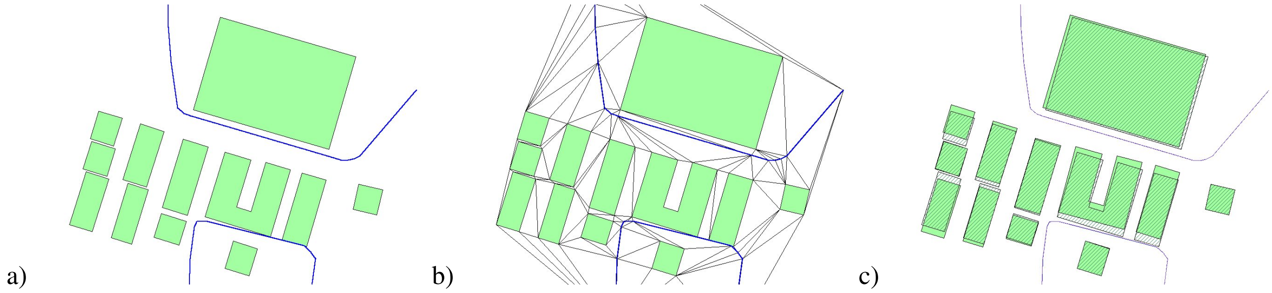

Figure 4: a) Original Situation with buildings and streets, b) Triangulation, c) Displacement with minimum distance of 7m, keeping the streets fixed.

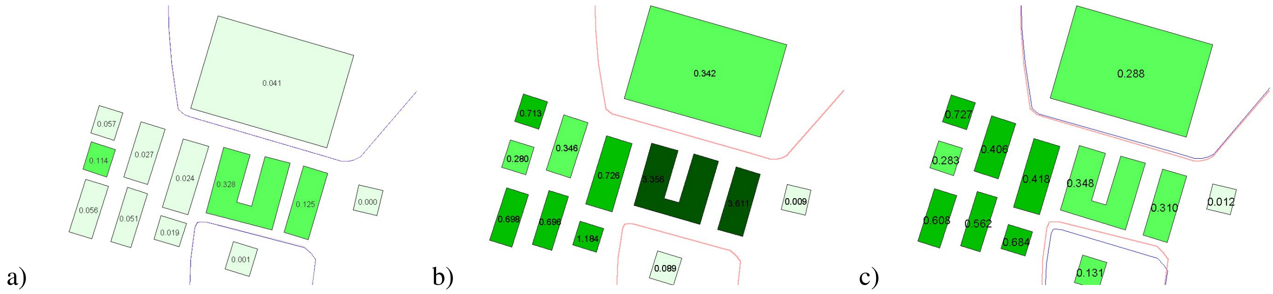

Figure 5: a) Visualization of the deformation of the buildings, b) Displacement with a minimum distance of 15m, keeping streets fixed, c) Displacement with minimum distance of 15m, allowing streets to move.

Figure 6 shows a typical displacement problem: the situation where streets, railways and a river are 'fighting' for space. The result shows that the objects are nicely displaced against each other in the conflict regions.

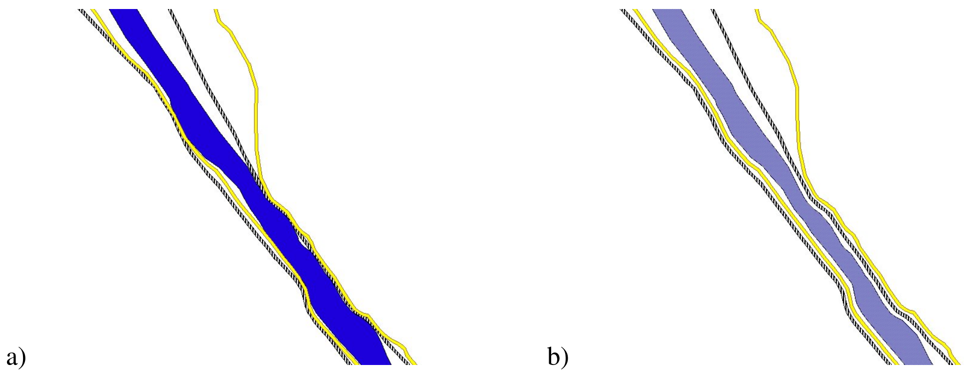

Figure 6: a) Original Situation with the river Rhein, streets and railroads on each side, before displacement, b) Situation after displacement and signaturation.

5 INTEGRATED APPLICATION: FUSING CADASTRAL AND TOPOGRAPHIC INFORMATION

The figures visualize the integration of the two modules to solve generalization problems in a larger spatial scene: buildings from cadastral information system are integrated into a topographic data set (scale 1:25.000). In order to guarantee legibility, first the buildings have to be simplified, and then they have to be displaced. Figure 7, left, shows the original situation, where the buildings are too detailed, and some distances between buildings and streets and among some buildings are too small. The result after simplification and displacement is given in Figure 7, right.

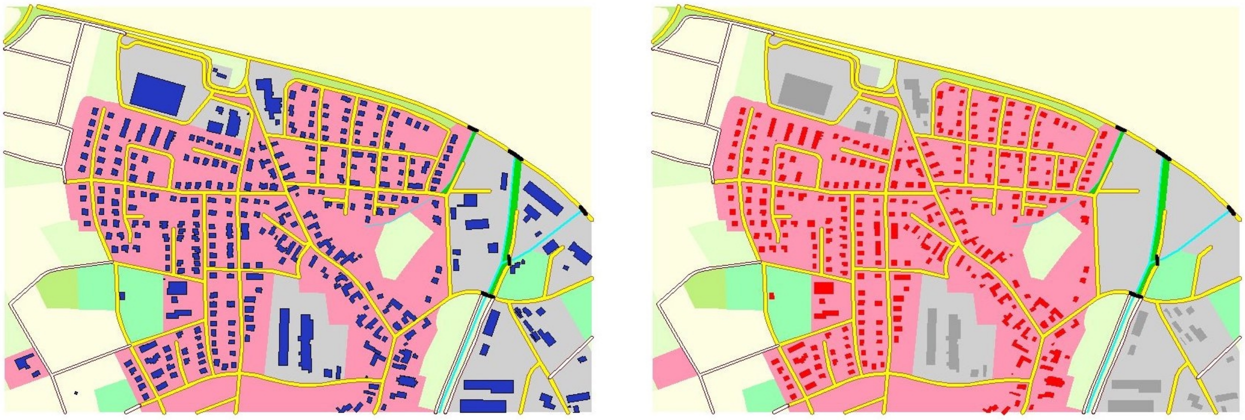

Figure 7: Original and generalized Situation.

In general, the result gives a much clearer and less complex impression. There are however situations, which are not satisfying from a cartographic point of view: a look at the deformation values of the buildings reveals, that in the center of the village a row of buildings that exhibits a typical cluster structure has been displaced and deformed in a way, that the original structure is no longer visible. In this case, there is simply not enough space to place all objects adequately, so another generalization operation has to be carried out before the displacement. A possible solution is a prior reduction, amalgamation or a typification. An approach for this is given in Anders & Sester [2000].

6 CONCLUSION

The approaches presented in this paper rely on a general mathematical framework, namely the optimization technique of least squares adjustment. It is applied to two different problems in generalization, namely simplification and displacement. Although the set-up of the observations differ considerably, the mathematical framework is the same. The advantage of the simplification of ground plans is the possibility to apply relatively simple rules and leave the optimization to the adjustment process. The advantage of the displacement approach is the fact that it establishes a global conflict reduction of the whole situation. As all objects are linked with each other, the effect of displacing a neighboring object is inherently taken care of. The algorithm treats point, linear and areal objects. Furthermore, it can easily be extended to integrate also the generalization operations enlargement and reduction.

Algorithms like the ones presented here have to be part of a larger generalization environment. Generalization as such is a task that is very much context and object dependent, involving elaborated strategies for the individual generalization of objects in accordance with their context. Here global structure recognition algorithms have to govern the subsequent (local) generalization steps. If, however, only geometric conflicts are left to be solved, the displacement based on least squares adjustment offers a very appealing solution, since it is general and global.

REFERENCES

Anders, K.-H. & Sester, M. [2000], Parameter-Free Cluster Detection in Spatial Databases and its Application to Typification, *in:* 'IAPRS', Vol. 33, ISPRS, Amsterdam, Holland.

Bobrich, J. [1996], *Ein neuer Ansatz zur kartographischen Verdrängung auf der Grundlage eines mechanischen Federmodells*, Vol. C455, Deutsche Geodätische Kommission, München.

Bundy, G., Jones, C. & Furse, E. [1995], *Holistic generalization of large-scale cartographic data*, Taylor & Francis, pp. 106–119.

Burghardt, D. & Meier, S. [1997], Cartographic Displacement Using the Snakes Concept, *in:* W. Förstner & L. Plümer, eds, 'Smati '97: Semantic Modelling for the Acquisition of Topographic Information from Images and Maps', Birkhäuser, Basel, pp. 114–120.

Douglas, D. & Peucker, T. [1973], 'Algorithms for the reduction of the number of points required to represent a digitized line or its caricature', *The Canadian Cartographer* **10**(2), 112–122.

Harrie, L. E. [1999], 'The constraint method for solving spatial conflicts in cartographic generalization', *Cartographicy and Geographic Information Systems* **26**(1), 55–69.

Hojholt, P. [1998], Solving Local and Global Space Conflicts in Map Generalization Using a Finite Element Method Adapted from Structural Mechanics, *in:* T. Poiker & N. Chrisman, eds, 'Proceedings of the 8th International Symposium on Spatial Data handling', Vancouver, Canada, pp. 679–689.

Jäger, E. [1990], Untersuchungen zur kartographischen Symbolisierung und Verdrängung im Rasterdatenformat, PhD thesis, Fachrichtung Vermessungswesen, Universität Hannover.

Lamy, S., Ruas, A., Demazeau, Y., Jackson, M., Mackaness, W. & Weibel, R. [1999], The Application of Agents in Automated Map Generalization, *in:* 'Proceedings of the 19th International Cartographic Conference of the ICA', Ottawa, Canada.

Lee, D. [1999], New Cartographic Generalization Tools, *in:* 'Proceedings of the 19th International Cartographic Conference of the ICA', Ottawa, Canada.

Mackaness, W. [1994], 'An algorithm for conflict identification and feature displacement in automated map generalization', *Cartography and Geographic Information Science* **21**(4), 219–232.

Mackaness, W., Weibel, R. & Buttenfield, B. [1997], Report of the 1997 ICA Workshop on Map Generalization, 19-21 June, Technical report, http://www.geo.unizh.ch/ICA/, Gävle, Sweden.

Nickerson, B. [1988], 'Automatic cartographic generalization for linear features', *Cartographica* **25**(3), 15–66.

Powitz, B. [1992], Computer-assisted generalization – an important software tool for GIS, *in:* 'International Archives of Photogrammetry and Remote Sensing', Vol. 30/4, ISPRS Commission IV Symposium, Washington, USA, pp. 664–673.

Ruas, A. [1998], 'A method for building displacement in automated map generalization', *International Journal of Geographical Information Science* **12**(8), 789–804.

Sarjakoski, T. & Kilpelainen, T. [1999], Holistic Cartographic Generalization by Least Squares Adjustment for Large Data Sets, *in:* 'Proc. ICA', International Cartographic Association, Ottawa, Canada.

Sester, M. [1999], Generalization based on Least Squares Adjustment, Presentation at the International Cartographic Conference, ICA, Ottawa.

Sester, M. & Förstner, W. [1989], Object Location based on Uncertain Models, *in:* 'Mustererkennung 1989', Vol. 219 of *Informatik Fachberichte*, Springer Verlag, Hamburg, pp. 457–464.

Staufenbiel, W. [1973], Zur Automation der Generalisierung topographischer Karten mit besonderer Berücksichtigung großmaßstäbiger Gebäudedarstellungen, PhD thesis, Fachrichtung Vermessungswesen, Universität Hannover.

Ware, J. & Jones, C. [1998], 'Conflict Reduction in Map Generalization Using Iterative Improvement', *GeoInformatica* **2**(4), 383–407.