AUTOMATIC DETECTION OF PATTERNS IN ROAD NETWORKS – METHODS AND EVALUATION

F. Heinzle *, K.-H. Anders, M. Sester

Institute of Cartography and Geoinformatics, University of Hannover,
Appelstr. 9a, 30167 Hannover, Germany - (frauke.heinzle, karl-heinrich.anders, monika.sester)@ikg.uni-hannover.de

KEY WORDS: pattern recognition, networks, vector data, geometric information

ABSTRACT:

Spatial data play a major role in many areas. They are an important source of information for a wide range of applications. The quality and quantity of data have increased in the last decades, resulting in a huge amount of data in the field of geoinformation. As a result of the multitude of information, various techniques for the automatic interpretation of large quantities of data have been investigated and implemented. Well known spatial data mining techniques, like spatial association rules, are primarily seeking for dependencies between attributes and spatial relations. In our opinion few investigations have been done for the detection of geometrical structures in vector data.
In this paper we will concentrate on the topological and geometrical interpretation of line or route networks. Especially we will focus on the investigation of road networks, but applications and differences to railway and river networks will also be described. Such kind of vector data contains a great potential of knowledge, which is not given by means of explicitly stored geometric elements and their predicates but rather encoded as implicit knowledge in terms of topological connections, relations between elements, and typical structures or configurations of single geometric features. Typical road databases do not explicitly store such implicit information on patterns. Thus, a considerable need for interpretation of this data exists.
First of all three basic patterns are introduced, namely grids, stars and ring roads. We will describe and analyse these patterns in detail and give algorithms to detect them automatically. Therefore we use an extended stroke approach, single source shortest path algorithm, theory of affine invariant moments and we perform a classification of a road network in urban and rural areas. We have developed an interactive control tool to realise these algorithms for pattern recognition and to display the results. Secondly, we will give an evaluation of these three patterns based on a comparison with a user test, in which test persons had to manually detect the described patterns in different test data sets and the test persons had to evaluate the patterns which were automatically detected with our approach.
We will conclude the paper with an outlook to possible further tasks and applications of these patterns. The detected road network structures and their typical characteristics provide an opportunity for knowledge enhancement in domains like map generalisation, automatic map generation, visualisation of data, automatic annotation of databases, as well as information retrieval from the internet.

1. INTRODUCTION

1.1 Motivation

Urban development as well as the construction of transportation routes follow usually structural conditions. Economical rules, security issues or the best possible utilisation of terrain are some basic principles for the emergence of typical structures in road networks. The knowledge of such patterns provides information about the design, history and development of an urban area as well as the functional organisation of a town. This knowledge can be found implicit in road network data and is our motivation for the investigation of patterns in urban road networks.

Urban geography is a domain which studies the inner structure of cities, especially complex patterns of distribution, movement, flows and interaction within cities. It provides various categorisations of types of cities on the basis of their historic development as well as on the basis of their today's appearance, functional and social organisation. Usually roads play an important role in those categorisations, because they constitute the linkage system between different regions and land uses.

Marshall (2005) investigates the road network and its structure of cities and districts of towns. The study is based on graph theory and uses quantitative characteristics like continuity, connectivity and depth of single routes in a road network graph. He also introduces a taxonomy of main general road patterns and distinguishes the following types - “linear”, “treelike”, “radial”, “cellular” or “hybrid” (Figure 1).

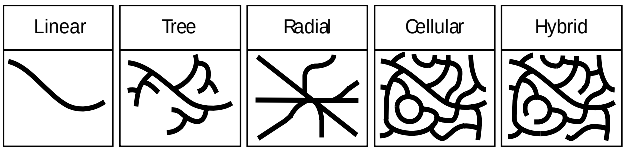

Figure 1. Taxonomy of patterns (Marshall, 2005)

On the basis of this taxonomy we will introduce some basic patterns to be detected in urban road networks. The attention is particularly put on four important patterns:

1. strokes – as a linear configuration
2. grids – as a cellular configuration
3. stars – as a radial configuration
4. ring roads – as a cellular configuration, observed over a wider area.

* Corresponding author.

1.2 Aims

The overall aim of the work is the extraction of implicit information in vector data. Therefore a process chain was designed, which mainly consists of five steps:

1. pre-processing of the data set (extraction of urban areas in a data set)
2. structuring of the data (graph structure)
3. detection of patterns in the road network
4. verification of the detected structures
5. transfer of the new knowledge to other applications

The main focus of the research is laid on the third point – the recognition of patterns in road networks. Existing descriptions of such patterns were used for the development of algorithms, which automatically detect typical structures in road networks. The primary concern is to develop implementations free of human interactions and human controlling of the processes respectively because we aim at automatically recognising implicit information. These automatic tools can then be used for machine-aided enrichment of data sets, accumulation of knowledge in ontologies and GIS and further applications.

1.3 Related Work

The work comprises investigations in topics like urban geography and town planning (Scargill,1979; Short, 1984; Whitehand, 1985), cognition science and Gestalt Theory (Wertheimer, 1925), pattern recognition in images, pattern recognition in large data bases (data mining), graph theory (Diestel, 2005), and graphics recognition (GREC, 1999).

Further approaches to recognise patterns in urban networks or to analyse road networks are described in Jiang & Claramunt (2004), Mesev (2005), Christophe & Ruas (2002), Zhang (2004), Thomson & Richardson (1999), and Li, Claramunt, Ray & Lin (2006).

2. PRE-PROCESSING OF A DATA SET

Our approach focuses on the automatic recognition of the patterns in urban areas. Spatial data sets often contain the whole road network of an administrative region or a broader area, e.g. topographic data sets of a country. Most of the time there are many urban areas inside such regions, which have to be separated to analyse them individually. For this reason a pre-processing of the data set has to be done before investigating patterns in the road network of single towns.

The separation algorithm is based on the assumption, that settlement areas show a greater number of traffic routes per unit of area than rural regions. A higher density of roads implies a smaller size of the areas enclosed by roads. Therefore the algorithm looks into significant differences of the size of such road polygons.

To avoid misinterpretations along highways, which are usually connecting several cities and are composed of small sized polygons, a smoothing of adjoining polygons is done by calculating the average area of all adjacent polygons. By doing this different urban areas can be separated even they are connected by a belt of small sized polygons representing a highway. The results of the separation process are clipping windows containing individual settlement areas (Figure 2).

In a last step we introduce a statistical value to guarantee a centred position of the town within the clipping window. Therefore we use the Tukey depth (Tukey, 1975) which is a measure of the depth of data and the centrality of objects respectively. The depth of every centroid representing a road polygon is determined and analysed regarding the deepest value. Subsequently the centre of the clipping window will be translated to the location of the deepest centroid.

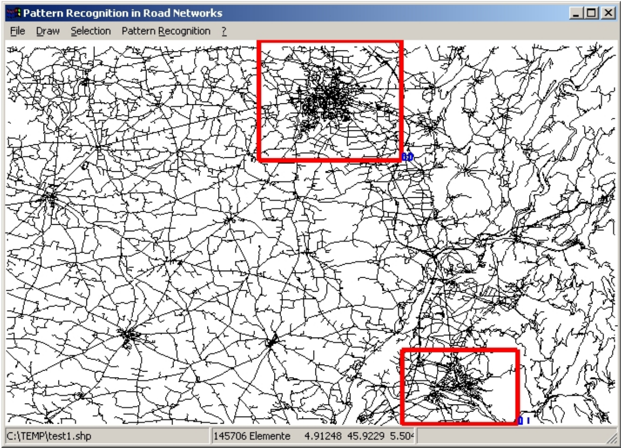

Figure 2. Separation of individual urban areas in a data set.

3. DETECTION OF TYPICAL PATTERNS IN ROAD NETWORKS

In Figure 1 we introduced some typical patterns in road networks. These examples will to be described and the algorithms for their detection will to be illustrated in this section in detail. Geometrical and topological information of a road network (the location and arrangement of junctions and connecting roads) are used as the basic data of our investigation. All the developed algorithms were implemented in C++.

We established a graph based approach to search for patterns in road networks. A hierarchical structure of different graphs to reproduce different levels of detail of the network was implemented. The basic graph contains all nodes and lines of the network, with nodes representing the line intersections and edges corresponding to the lines themselves.

3.1 Strokes

The first pattern to mention here is the structure of strokes, namely long lines in the graph that proceed relatively straight with a continuous flow. The principle of the strokes (Thomson and Richardson, 1999) is based on the assumption that traffic routes are built as curvature-poor as possible. Figure 3 shows a stroke in a data set. Strokes are basic patterns that can also be used for the search of other patterns.

The algorithm for tracing strokes in the graph is well-known in principle. At each node the successor edge is selected, which shows the most continuous direction, whereby a minimum of smoothness must be kept. We extended this approach with the consideration of three important details to make strokes more adaptive to special appearances of roads in networks (Heinzle et al., 2005).

1. splitting of roads in a dual carriageway (Figure 4)
2. roundabouts (Figure 5)
3. curving trend of strokes (Figure 6)

Figure 3. The red line presents a main stroke in a data set.

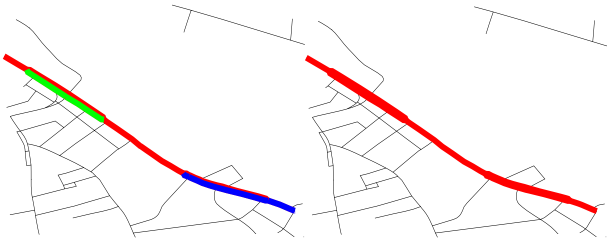

Figure 4. Dual carriageways split the stroke (left). The parts are combined to one stroke (right).

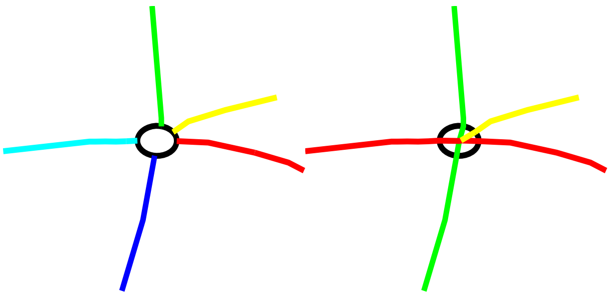

Figure 5. Roundabouts split the strokes (left). The parts are combined in the centre of the roundabout (right).

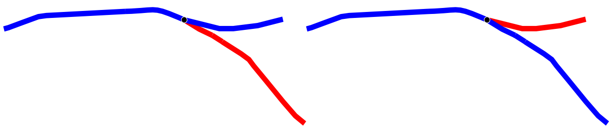

Figure 6. The most continuous direction is the blue line (left). Consideration of the overall trend of the stroke results in another stroke, which is blue coloured as well (right).

3.2 Grids

Another important part of road networks are grids. A grid is characterised by a set of mostly parallel lines, which are crossed by a second set of parallel lines. The algorithm for their detection is based on the arrangement of the centroids of the grid polygons. After translating the centroid in the direction and length of a bordering edge, it is assumed, that the new calculated centroid has to be in the vicinity of the centroid of the neighbouring grid polygon. Figure 7 shows the approach to detect grids and Figure 8 exemplifies some results in a data set of Glasgow.

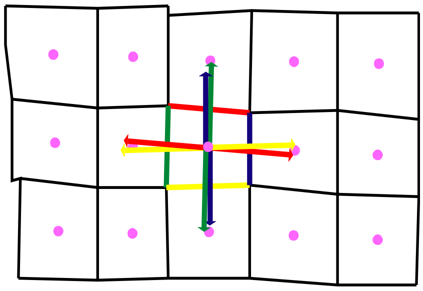

Figure 7. The approach to detect grid structures is based on translating the centroids of grid polygons.

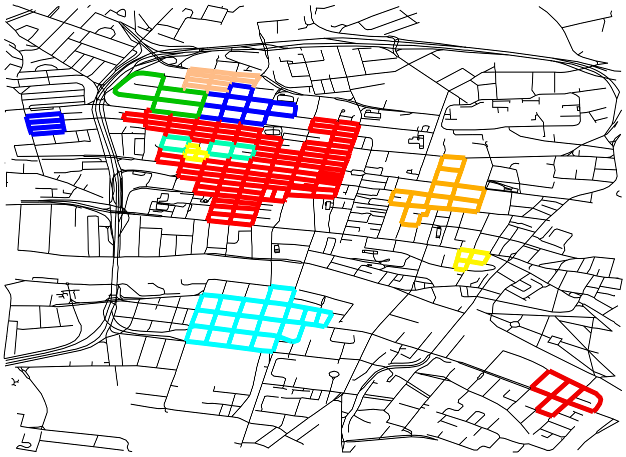

Figure 8. Some detected grids in the city of Glasgow (colours randomly chosen).

3.3 Stars

Starlike configurations are a common structure in road networks and are typical for concentrated urban areas where the infrastructure is dense. It is characterized by a fuzzy centre from which rays radiate. The rays can significantly differ from the ideal radial course. Moreover the equal distribution around the centre may not exist and the rays do not have to have the same length. The approach to detect star patterns is based on the Dijkstra algorithm (Dijkstra, 1959) with the assumption that all rays of the star radiates relatively straight-lined.

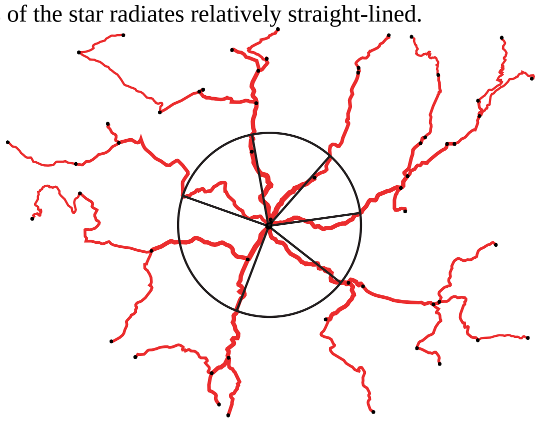

Figure 9. The approach to detect star structures is based on the length of single-source shortest paths compared with radial connections.

We take each node in the graph as potential centre point and the intersection points of the graph with a circle around the centre point. Comparing the length of a single-source shortest path from the centre point to the intersection points with the radius of the circle we can identify possible rays. These ray candidates are verified and the number of rays around the star centre as well as their distribution are the decisive criterion for the classification as star structure. Figure 9 shows the idea of recognising stars and Figure 10 illustrates a detected star in a data set.

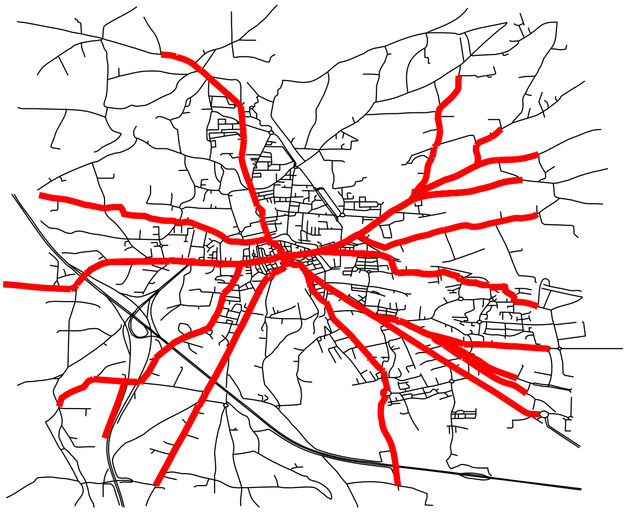

Figure 10. A detected star pattern in a data set of Berlin.

3.4 Rings

In our applications we search for common ring roads which typically surround a city centre. A ring should approximate a circle even though the appearance of rings in road networks adapt to natural or human made restrictions. Therefore they are usually deformed and stretched. We use two theories in our approach for recognising ring roads – a statistic measure for the depth of data called Tukey depth (Tukey, 1975) and a measure for the evaluation of the shape of a polygon by using affine invariant geometric moments (Voss and Süße, 1995). The algorithm is described in detail in Heinzle et al. (2006). Figure 11 shows an example of a detected ring road.

Figure 11. A detected ring road in a data set of Apt/France.

3.5 Implementation of a wxWidgets-tool for pattern recognition in road networks

The described algorithms and applications for pattern recognition in road networks were integrated in a wxWidget-tool. Satisfying all the necessary steps to detect the presented structures in a vector data set, the tool provides utilities to display the data, the results and some additional information too. Figure 12 shows a screenshot of the tool.

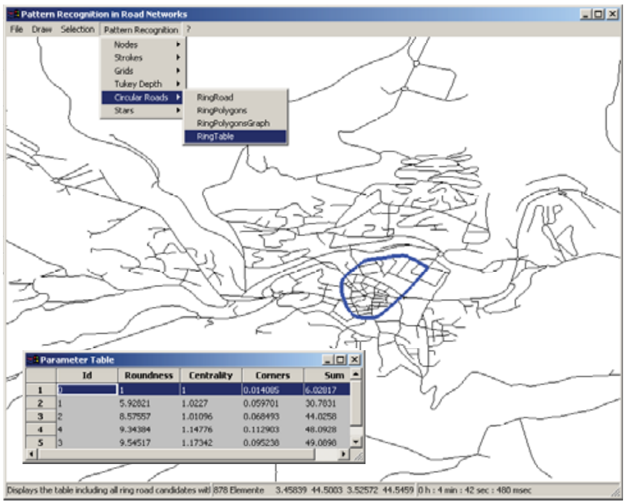

Figure 12. The wxWidget-tool to proceed pattern recognition techniques and to display the results.

With the help of this tool we applied our developed algorithms to a range of data sets. We could analyse the time complexity of the methods and use the results for evaluation purposes as well.

4. EVALUATION

An evaluation of the algorithms and the results have been made in two respects. First of all we asked test persons to identify the described patterns and draw them in real data sets. We used therefore different kinds of data varying from small towns to metropolitan areas of large cities, from smaller sections of the inner city to extended regions covering more than the town itself. Only the mere geometry was used as input information without any attributes, special characteristics like the name or status of a road or explanations of the road categories.

The test persons were given no indications about the type and number of possible patterns in a scene. They had to decide themselves, if there are e.g. grid structures or a ring road and should mark the corresponding road elements in the data set. The evaluation has shown various results. One conclusion was that even humans do not qualify the same structure in the same way. The patterns were defined differently regarding their appearance and their function respectively.

However it was evident, that the probands agree the more on their drawn structures the more the characteristics of the structures are consistent with the postulations we made for the recognition process, e.g. a circle-shaped ring, a star with straight proceeding rays or an absolute uniform rectangular grid. For this reason the test confirmed our approaches for the detection of the patterns. On the other hand it became apparent, that most of the humans tend to generalise the elements or parts of the patterns. For example grid structures were qualified even though single roads, which did not belong to the grid scheme, were existing inside the grid polygons. Another example is the determination of star structures without localising the centre of the stars. The test persons often have drawn only the dominant roads radiating from a core zone of a town but refused to define the explicit routes of the rays inside the city centre. We still have to investigate those "instinctive" generalisation processes to adapt them to our approaches in future.

In a second test we evaluated the patterns found automatically with the described algorithms. Therefore the test persons had to assign marks from 1 to 5 to qualify the detected patterns. In summary it can be ascertained that the automatically detected structures received fair to very well grades. It averages 1.9 for the grid, 1.5 for the star and 1.7 for the ring structures.

5. USE OF THE DETECTED PATTERNS IN OTHER APPLICATIONS

The described patterns can be used to obtain additional information about the formation and origination of a town. The presented typical structures provide indications for chronological, spatial and cultural components of the historical background of a city. Especially the combination of the road patterns has to be analysed and interpreted to get implicit information, which can be used in many applications.

5.1 Implicit information and enrichment of GIS

The basis of our research is the assumption, that we are able to infer new implicit information from the road network and its inherent patterns. The exploration of issues like

- classification of more or less important cities
- sphere of influence of cities
- detection of the centre of a city
- determination of tourist areas and attractive destinations
- possibilities of suburban or industrial settlement, urban development, quality of housing
- age determination of a town
- calculations of land values

provides the opportunity of detecting such implicit information. Certainly many of the examples may need more data than the road network and some basic patterns to result in new knowledge. However the proposed way should be a first step to approach the objective. In this regard data mining techniques can also play a major role.

The exemplified knowledge can be discovered in vector data, but it is usually not explicitly stored in geographical information systems. An automated enrichment of GIS with new geometric and thematic information leads to a significant increase of its usability. In recent years the research was especially focused on the semantic enrichment of data bases and the development of intelligent search methods. The presented results of pattern recognition in road networks can contribute to this issue. Examples are the attributive annotation with the importance of a street on the basis of a detected pattern, calculations of the traffic volume or the accessibility of the city centre. Other possibilities are the geographical coding of the city centre in a data base or the automatic enrichment with navigation data.

5.2 Enrichment of ontologies

Ontologies act as representation of knowledge on the basis of a specified concept. The configuration of such concepts and the enrichment of the ontology with entities is often done manually. Semantic interpretation of data sets can help to automate this process. Our research is focused on the learning of spatial concepts by using the patterns detected in the data sets. Indeed the computer does not know the meaning of words like “highway”, “roundabout” or ring road”. However after extracting those structures it is learnt that e.g. a highway is a main connecting route with few junctions and often to find in the vicinity of big cities. By the help of such additional information ontologies can be enriched automatically.

5.3 Generalisation

The object of a generalisation process consists in an aggregation, a simplification or an abstraction of individual data but preserving the dominant structures and the ensemble of the data set. The generalisation of road networks can be supported by the use of the described patterns. In this context we introduce two concepts of graph types distinguishing the distribution of node degrees in the network (Heinzle and Anders, 2007). It is about “random graphs” (Erdös and Rényi, 1960) and “scale-free graphs” (Bollobás and Riordan, 2004). Figure 13 shows typical examples of networks representing a random graph and a scale-free graph respectively.

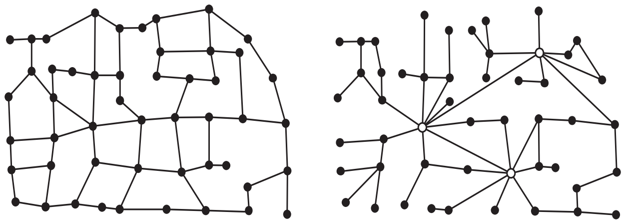

Figure 13. A network representing a random graph (left) and a scale-free graph (right) respectively.

A scale-free graph is characterised by some “hubs” with a high node degree (white points in Figure 13). They are very important for the connectivity of the graph. However road networks show usually the distribution of a random graph with a node degree of 3 or 4 for the majority of nodes. For generalisation purposes it would be very useful to transfer the road network in a scale-free graph with some outstanding hubs. Our idea is to identify such hubs by using the pattern recognition process. A striking example is the detection of a star pattern and the localisation of the star centre in one hub.

5.4 Typification

Typification is a generalisation operation where a large number of similar objects is substituted by a few objects while maintaining their original spatial arrangement. A typical example is the generalisation of linear arranged houses along the roads. With the use of detected grid patterns we try to bring the structural knowledge to the typification process. For this purpose we utilise the centroids of the house ground plans and build a connected network by determining the relative neighbourhood graph (Toussaint, 1980). On the basis of this network we can detect grid structures characterising the alignment of houses. Anders (2006) describes an approach to typify those automatically detected grid structures by adjustment and reduction. Figure 14 shows this approach.

6. CONCLUSION

We have shown approaches to detect characteristic patterns in road networks fully automatically, especially strokes, grids, stars and ring roads. The evaluation of the results is promising. We also have shown various application areas of the automatically extracted patterns. Our future research focus will be put on two topics. On the one hand we would like to improve the existing approaches and we would like to concentrate on other patterns too. The search for more detailed structures can give also more implicit information about a city and its development. On the other hand we would like to focus on further exploitation of the patterns to get new implicit information, to learn new concepts and to use them in new applications as well.

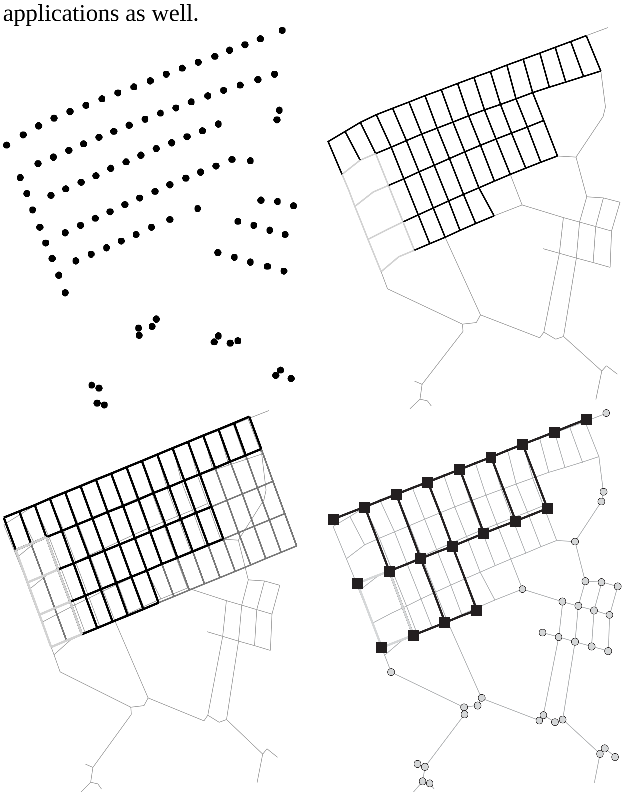

Figure 14. Typification approach: centroids of houses (above left), detected grid in a RNG (above right), grid adjustment (down left) and grid reduction (down right).

REFERENCES

Anders, K.-H., 2006. Grid Typification. In: *A. Riedl, W. Kainz, G.A. Elmes, (Eds.): 'Progress in Spatial Data Handling'*, 12th International Symposium on Spatial Data Handling (SDH), 10-14 July, Vienna, Austria, Springer-Verlag, Berlin Heidelberg New York, pp. 633 - 642.

Bollobás, B., Riordan, O., 2004. The Diameter of a Scale-Free Random Graph. *Combinatorica*, 24(1), pp. 5-34.

Christophe, S., Ruas, A., 2002. Detecting Building Alignments for Generalisation Purposes. In: *Symposium on Geospatial Theory, Processing and Applications*, Ottawa, Canada.

Diestel, R., 2005. *Graph Theory*. Third Edition, Springer-Verlag, Heidelberg.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1, pp. 269-271.

Erdös, P., Rényi, A., 1960. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5, pp. 17–61.

GREC, 1999. Graphics Recognition, Recent Advances, Third International Workshop. ed. Chabra, A.K., Dori, D., Springer, LNCS

Heinzle, F., Sester, M., Anders, K.-H., 2005. Graph-based Approach for Recognition of Patterns and Implicit Information in Road Networks. In: *Proceedings of 22nd International Cartographic Conference*, 9. - 16. July 2005, La Coruña/Spain.

Heinzle, F., Anders, K.-H., Sester, M., 2006. Pattern Recognition in Road Networks on the Example of Circular Road Detection. In: *M. Raubal, H.J. Miller, A.U. Frank, M.F. Goodchild (Eds.): 'Geographic Information Science'*, GIScience 2006, Münster, Germany, LNCS 4197, pp. 253-267.

Heinzle, F., Anders, K.-H., 2007. Characterising Space via Pattern Recognition Techniques: Identifying Patterns in Road Networks. In: *W.A. Mackaness, A. Ruas, T. Sarjakoski,(Eds.): 'Generalisation of Geographic Information: Cartographic Modelling and Applications'*, Elsevier Ltd., pp. 233-253.

Jiang, B., Claramunt, C., 2004. A Structural Approach to the Model Generalisation of an Urban Street Network. *GeoInformatica*, 8(2), pp. 157-171.

Li, X., Claramunt, C., Ray, C., Lin, H., 2006. A Semantic-based Approach to the Representation of Network-Constrained Trajectory Data. In: *A. Riedl, W. Kainz, G.A. Elmes, (Eds.): 'Progress in Spatial Data Handling'*, 12th International Symposium on Spatial Data Handling (SDH), 10-14 July, Vienna, Austria, Springer-Verlag, Berlin Heidelberg New York, pp. 451–464.

Marshall, S., 2005. *Streets & Patterns*. Spon Press, Taylor & Francis Group, New York.

Mesev, V., 2005, Identification and characterisation of urban building patterns using IKONOS imagery and pointbased postal data. *Computers, Environment and Urban Systems*, 29, pp. 541–557.

Scargill, D.I., 1979. *The Form of Cities*. London: Bell and Hyman.

Short, J.R., 1984. *An introduction to urban geography*. London (Routledge & Kegan Paul).

Thomson, R., Richardson, D., 1999. The 'good continuation' principle of perceptual organization applied to the generalization of road networks. In: *Proceedings of the ICA*, Ottawa, Canada, Session 47B.

Toussaint, G., 1980. The relative neighborhood graph of a finite planar set. *Pattern Recognition*, 12, pp. 261–268.

Tukey, J., 1975. Mathematics and the picturing of data. In: *Proceedings of the International Congress of Mathematicians*, Vancouver, Canada, pp. 523–531.

Voss, K., Süße, H., 1995. *Adaptive Modelle und Invarianten für zweidimensionale Bilder*. Verlag Shaker, Aachen.

Whitehand, J.W.R., 1985. Urban geography: the city as a place. *Progress in Human Geogr.*, 9(1), pp. 85-98.

Zhang, Q., 2004. Modeling Structure and Patterns in Road Network Generalization. Technical report. *ICA Workshop on Generalisation and Multiple Representation*, Leicester, UK.

ACKNOWLEDGEMENTS

This work was supported by the EU in the IST-programme Number 2001-35047. We thank the National Mapping Agency of the state Lower Saxony in Germany (LGN) for providing the ATKIS data, the National Mapping Agency of France (IGN) and TeleAtlas Germany for providing several topographic data sets.