GRAPH BASED APPROACHES FOR RECOGNITION OF PATTERNS AND IMPLICIT INFORMATION IN ROAD NETWORKS

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

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

ABSTRACT

The paper will introduce into the subject of recognition of typical patterns in road networks. We will first describe the design and lay-out of roads. Applications to detect patterns and to use them for finding more knowledge in vector data are shown. We will familiarise the reader with different patterns in road networks and approaches for the automatic detection of those patterns in vector data. The objective of the present work consists in the automatic localisation of a city centre, which can be seen as one possibility to derive new knowledge by analysing existing information. We will show, that the combination and interaction of "simple" patterns – which will be described in the following sections – can lead to higher-ranking patterns or explicit new information, which is hidden in the data.

1. INTRODUCTION

Spatial data sets store an extensive potential of knowledge: on the one hand explicit collected geometrical features and associated non-spatial attributes. On the other hand we can find much more knowledge, which is implicit given in terms of topological information, typical structures between features and relations of their attributes. The localisation of the centre of a city can exemplify typical hidden information. Normally such specifications will be not explicitly stored in data bases.

The objective of the present work consists in the automatic discrimination of an urban place and its surrounding region as well as the localisation of a city centre on the basis of road network, which is given in a GIS. Therefore we evaluate the existing vector data by means of different types of graphs. Starting from the base graph representing the whole road network, we derive so called strokes and analyse their properties in the graph. Further typical patterns in the network will be recognised, like loops, splitting of lanes, motorway interchanges, parallel streets and the like. In the next following step more significant structures will be found, like grid shaped road networks and starlike or ring-shaped configurations. For this purpose we analyse topological properties of the graphs. We present different criteria to determine the region of the city centre as well as first approaches like cluster algorithms to detect differences in the density of the road network and the use of minimum circles around intersection points of stars.

The extracted information regarding the structure of an urban area, its road networks and their typical characteristics provides on the one hand a basis to localise the city centre automatically. On the other hand this gain in knowledge can be used for map generalisation and other visualisation purposes. With the knowledge of patterns in road networks those structures can be considered and preserved during the process of generalisation. For example a starlike centre pointed network could be symbolised as a star. Furthermore the paper can be seen as a contribution to the general field of data mining, because the knowledge recovered in the described context can be used to enhance data bases. The derived attributes, which describe the topological aspects of the data can be evaluated to find subsequent spatial rules. Additionally the derived characteristic properties can act as a fundament to receive higher concepts of knowledge with the help of data mining techniques.

This paper is composed of the following sections. After this introduction we will give a short overview of typical implicit information in data sets. We will identify some possibilities to augment data bases with new knowledge by using spatial rules. In section 3 we investigate typical patterns in road networks, namely node types, strokes, grids, stars and rings. We describe the properties of every individual pattern as well as our implemented algorithms to detect them and give some examples for clarification. These patterns are used in section 4 to visualise the idea of combining that inherent information for the detection of a city centre. In section 5 a small summary and outlook is given and the list of references in section 6 completes this paper.

XXII International Cartographic Conference (ICC2005)

2. IMPLICIT INFORMATION IN SPATIAL DATA SETS

Spatial data sets can be of manifold kind. The two main representatives are images and data bases with vector data. The domain of interpreting digital images is an extensive field of research. This long-existing investigation of basic techniques for image interpretation has produced many standard applications to recognise structures in pictures (e.g. Haralick and Shapiro, 1992). In contrast, the usefulness and necessity of interpreting vector data is still undervalued. In fact it is obviously not necessary to extract topographic features in vector data, because they are already stored in the data bases as geometrical elements. However vector data are often pure "Spaghetti" data in GIS, which means there are points, lines and polygons defined by coordinates but without topological or structural information. Furthermore higher aggregations of topographic features might be only implicitly present in data, e.g. city centre. The matter of our work is the investigation of such kind of vector data in terms of finding implicit structures.

Different strategies to discover implicit information have been developed in dependence of the kind of knowledge to extract. There is the extensive field of data mining, which provides new technologies for knowledge discovery in large spatial data bases. Implicit information is extracted by using spatial rules. The aim is to find regularities in the data by applying those rules (Koperski and Han, 1995, Lu, Han and Ooi, 1993). Another range of applications is the search for patterns in vector data. Especially the field of automated map generalisation is dealing with the problem of pattern recognition and structure modelling (Mackaness and Edwards, 2002, Jiang and Claramunt, 2004, Zhang, 2004).

One aim of our work is the automatic detection of city centres in vector data. Therefore we intend to investigate road networks. A first idea is to find several segmentations or parts, which have characteristic structures. We expect, that the combination of such detected structures – which we will describe in the following – will provide useful evidence for the location of a city centre.

3. PATTERNS IN ROAD NETWORKS

A graph based approach is used to search for patterns in road networks. We introduce a hierarchical structure of different graphs to reproduce different levels of detail of the network. The basic graph contains all nodes and lines of the network, with nodes representing the line intersections and edges corresponding to the lines themselves. The investigation of the topological structure of the graphs leads to the recognition of patterns, which we describe in detail in the following sections. The first four described patterns, namely the investigation of different node types, strokes, grids and star structures are already implemented and the recognition methods are described in detail. Still we are investigating the algorithms to detect the last described pattern – ring-shaped structures. We will present some ideas, but can not come up with good results because of the then explained problems.

3.1. Appearance of intersection points

The design of crossroads can be differentiated according to their node degree and the arrangement of the intersecting lines (Sester, 1995). The investigation of repetitive elements results in some first patterns. I.e. the recurrent occurrence of so called TEE nodes along one street forms a 'comb' structure (Figure 1 – the TEE nodes are red coloured). The arrangement of a series of opposite TEE nodes results in a 'ladder' structure (Figure 2).



Another interesting node type of degree three are the so called FRK nodes, which look like a fork. Crossroads are designed in such a way to enable continuous direction changes of the traffic routes. Such smooth passages often can be found in the vicinity of motorway interchanges or access roads of important roads. The splitting of traffic routes into separate lanes is another typical appearance of FRK nodes.



3.2. Strokes

The principle of the strokes (Thomson and Richardson 1999) is based on the assumption that traffic routes are built as curvature-poor as possible. This principle becomes more obvious the higher level a road has. Main artery roads will have a soft course. One will find here sudden curvature changes only rarely.

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. However the splitting of a route into individual lanes as mentioned in the previous subsection proved problems in real vector data. Permitting only one follow-up edge leads to a fragmentation of the strokes (Figure 5). A deterministic linkage of the individual edges could not be guaranteed, since the allocation would depend on the sequence of edge processing, i.e. from which direction the branch will be approached. For this reason the pairs of branches (as shown in Figure 4) were investigated separately. The composition of the split strokes at the branch points creates one stroke, which reproduces the main line of the route (Figure 6).





3.3. Grids

The grid structure is a typical pattern in road networks. Artificially built settlements based on development schemes frequently show straight-line roads, which enclose smaller internal building areas. A grid is characterised by a passel of mostly parallel lines, which are crossed by a second set of parallel lines. Frequently the sets of parallels intersect themselves approximately right-angled, this is however no compelling characteristic of a grid. Further characteristics of grid structures are rectangular or parallelogram-similar shape of the grid surfaces, straightness of the parallel line set, the convexity of the surfaces, a relatively constant side length relationship as well as a similar area of the rectangles.

In practice the grid structures show rarely such "optimal" or ideal properties. Nevertheless the above mentioned characteristics can be recovered in the data with appropriate deviations. According to the characteristics pointed out above different criteria were considered in order to detect grid structures in vector data automatically. We select those points as starting points of the investigation, at which four edges intersect - so called CRS nodes. Subsequently we investigate the histogram of the edge directions (Otsu 1979, Liao et al. 2001) and examine the grid polygons regarding its similarity to the neighbour polygons. Finally the surface size of the grid polygons are compared, as well as the convexity of the surfaces is examined.

The algorithm can be explained as follows: The basis is a graph, where nodes are the crossroads of the road network, the edges are the connections between the intersections. All the below mentioned thresholds were determined by investigations of real test data.



ENDFOR

The algorithm has to be able to deal with interruptions of the grid, such as crossing roads, slight changes in the main directions of the parallel line set etc. Figure 6 and figure 7 show detected grids in the city of Zurich and Barcelona.



Another approach is the application of a Hough transform (Hough 1959) to the road segments of the grids. This leads to the detection of the straight lines which show the main directions in the grid. With the example of the town of Manhattan (Figure 9) the dominant straight lines in the grid can be identified (Figure 11). Figure 12 represents the distribution of all line directions in a diagram.



Fig. 9: The structure of Manhattan





Fig. 10: Detected grids in Manhattan (different grids are differently coloured)





The example shows the differences of both approaches. First described algorithm detects all the grids. However small changes of the grid properties lead to the identification of separate grid structures (Figure 10). Whereas the Hough transform approach provides a more generalised result. All elements of the two sets of parallel lines are considered as eligible candidates of one grid. Differences in the distances of the parallel lines does not cause separation.

3.4. Stars

A star-shaped arrangement of roads in a network is typical for concentrated urban areas where the infrastructure is dense. Cities have been and will be planned in a star-shaped structure due to the necessity of fast reachability of the inner areas of a city. The rules of economy and usefulness are forming such kind of structures.

The properties of a star-shaped structure can be described as follows: It exists a centre from which rays radiate. The ideal star pattern has one unique point as centre, the rays have all the same length and are equal distributed around the centre. Stars in real road networks differ from that ideal pattern. The rays do not need to intersect in one single crossing point. The middle point will be rather a fuzzy centre, which means that all crossing points of the rays are inside of a central minimal circle. The diameter of this minimal circle is a measure for the fuzziness of the centre point of the star pattern. In many cases central places, e.g. a market place or a small roundabout, which the rays meet, represent this fuzzy area of a star centre.

Another problem in the real world is that usually the rays of the star pattern are not perfectly linear. Due to the existence of curves, small sharp bends and crossings with other roads the shape of the rays can significantly differ from the ideal radial course. Further the rays may differ continuously in one direction over a larger distance like a spiral. Moreover the equal distribution around the centre does not exist as well as the rays do not have the same length. Some radial roads can end at crossings with circular roads and others can actually reach neighbour cities. A further substantial characteristic of the star structure is the occurrence in any level of detail. That means that we can find stars within a very small range, e.g. a star-shaped pedestrian precinct with the radius of 100m. On the other hand one can also find star patterns, which take the radius of the entire city or even an entire region.

We take any node in the graph as potential centre point. Using the Dijkstra algorithm (Dijkstra 1959) a single-source shortest path is computed from this node to all other nodes. After computing all shortest paths they will be intersected with a circle around the centre point with a certain radius (Figure 13). At all intersections the length is computed along the shortest path from the centre point to the intersection. With the assumption that all rays of the star radiates relatively straight-lined, the computed length should be approximately the same as the search radius. 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.



Fig. 13: Circular search region in the Dijkstra graph with blue coloured radial connections

Fig. 14: Four intersection points in the fuzzy centre of the star (Fig. 13 enlarged)

The definition of the search radius is one of the arising problems in this algorithm. Since there is no a priori knowledge of the size of such a pattern as described above, we must repeat this process for different search radii. In the example above four stars were found. The reason is the existence of four nodes at close quarters in the middle of the star (Figure 14). Each one presents a possible centre point. The centre points belonging together to the same star pattern were detected by a cluster algorithm (Anders 2004). The found star structures of the four points were merged corresponding to the clustered points. The algorithm permits a certain fuzziness in the meeting place of the rays.

Yet another problem in extracting star shaped structures is the deficiency of knowledge about the range of the configurations. The discovery of ring-shaped structures around the star can give an idea about the expansion of the star. The ring-shaped structure is described in section 3.5. However each star is not always surrounded by a ring. So far we assumed a relatively uniform distribution and a similar length of the rays. If this is not the case other measurements for limitation of the ray length are necessary.

Figure 15 shows an example of a city in southern France. The search radii are determined by the average length of all edges in the road network and the minimum, maximum and average lengths of the strokes. A minimum of five rays equally distributed around the centre is the pre-condition to be classified as a star.



Fig. 15: Detected star in a data set of a town in southern France

At the moment we are investigating two kinds of improvement. On the one hand we want to speed up the shortest path calculations. Road networks are mostly sparse networks. Thus we are investigating pruning approaches like described in Wagner and Willhalm (2003). Another possibility to detect star patterns is the use of the found strokes. We will determine the straightness of the strokes. Candidates, which comply with a certain level of straightness are consulted to calculate their intersection points. The size of the minimal enclosing circle of the intersection points could be an indication for the existence of a star centre.

3.5. Ring-shaped structures

Ring-shaped patterns occur in road networks in any places and any styles. There are common ring roads, small circleshaped pedestrian areas, streets along old town walls, roundabouts inside the city as well as in the countryside and big traffic circles of motorways in the vicinity of important towns. The expansion of such ring-shaped structures depends primarily on their functionality and vary from very small entities to larger elements. Often they exist in connection with starlike patterns, but this is not always the case.

A ring should approximate a circle with all the characteristic properties of a circle. In road networks the appearance of ring shaped patterns adapt to natural or human made conditions like town walls, courses of rivers, altitude differences or existing building areas (Figure 16). Therefore ideal circles can be found rarely. Usually they are deformed and stretched. Correspondingly the properties like the centre point, radius or periphery also change. Figure 17 shows the ring shaped structure of a city in southern France.



and interruptions

The typical characteristics of a circle-shaped ring can be described by the compactness and the convexity of the ring structure. The compactness C is a measurement of the jaggedness of the object contour. If the compactness refers to a circle then it is defined as $C = P^2/4\pi A$ where P is the perimeter and A the area of the object. In case of a circle the compactness is optimal and has the value one. C is a measure for the roundness of an object. The convexity of an object is determined by the ratio of the area of the convex hull to the area of the object. In the optimal case of a circle it has the value one. To find ring-shaped contours in a data set it is necessary to detect objects with a compactness and a convexity close to one. Additional constraints could be the size of the objects or the minimal number of the edges belonging to the ring. Trivial rings like small rectangles should be filtered out.

A first idea to detect ring-shaped structures is a Hough transform on the base of circle and ellipse equation (Duda and Hart 1972). First attempts were unsatisfying, because a ring is a higher level pattern which can not be found on small grained individual road sections. We expect better results when working on a simplification of the road network. This, however, remains to be done. Another idea is an exhaustive graph search for all existing simple cycles, which have to be filtered by above mentioned characteristics. A cycle in a graph consists of a sequence of successively incident edges and their end vertices, where the terminating vertices are identical. Because of the exhaustive search this brute force approach is also unpractical for large road networks.

4. DETERMINATION OF A CITY CENTRE

There are various definitions for city centres like historical city centres, economic city centres respectively business and financial centres, centres of the town development, but also politically based or traffic related definitions. A standard (consistent) picture of the "real" centre of a city does probably not exist. However the definitions and the accordingly identified localisation of the centre mostly overlap, since the individual parameters of the definitions are correlated with each other.

Roads affect the development of a city to a great extent. Economical principles and conceptualisation by humans are reasons, that typical patterns emerge in road networks. We can use these patterns to determine a city centre. On the one hand special arrangements point the location of the centre out, like the middle point of a ring or star pattern. On the other hand it is possible to find regularities and to combine different rules or occurrences of special structures. In the following we will describe some possible criteria to define a city centre and to trace the border to the surroundings. The complexity will be clarified by the few criteria listed below, which are only the most prominent features. Furthermore the pattern "city centre" naturally depends in each data set on the selected scale. In small scales the search for the city centre will be limited to the search of a central point. In the case of a more detailed data set the centre will rather be a polygon like a market place or an area within the city wall.

A road network reflects very well the historical growing of a city. At smaller cities or villages one can frequently differentiate between the typical old village forms. Nucleated village, meadow village, ribbon-built village and other types have a very remarkable appearance, e.g. the central arrangement around a market place or an elongated village along a road. Larger cities are often shaped by recent construction activities and older smaller structures are overlaid. Nevertheless one finds also here typical development patterns. Following criteria are first ideas, which could be used for the investigation of city centres:

- 1. In the core of a city the road meshes, i.e. the internal building blocks which are enclosed by roads, often are smaller in their size compared to the ones in the outskirts.
- 2. In many cases there are main traffic routes, which approach radial to a centre or the inner core of the city. This does not have to mean, however, that these rays really meet in the centre, the intersection may be only virtual.
- 3. Frequently ring-shaped roads exist, which enclose more or less the inner city. In addition, these rings arise often together with case 2 but can also occur without the appropriate arterial roads.
- 4. Many times cities are surrounded by a more or less broad belt of suburbs, which can represent subordinated cities themselves.
- 5. In the vicinity of a city the settlement proceeds along radial arterial roads. This fact can be useful to determine the extension of a city, assuming that inside the city the complete region between such radial structures should be a built-up area, outside the city the built-up areas are only along that radial roads.
- 6. The city centre is typically a central point, which means that all other regions of the city have short distances to the centre.

In extension of the above described characteristics a further detail must be considered for the search of a city centre. The arrangement of the city centre and the assumptions and rules exemplarily specified above depend not only on the scale of the regarded data, but on the scale of the city itself. That means the appearance of the pattern is mainly dependent of the size of the city (Heinzle and Sester 2004). The investigations of this paper are focused on larger towns.

The following examples show some results of the city centre investigation. Figure 18 shows a result of inspecting the first criteria, namely the density of building blocks enclosed by roads. The blue coloured points are the centroids of every block. The areas with different density were clustered by an algorithm of Anders (2004). The part of the city with the highest density is located in the centre. Figure 19 shows the most dense area.





Fig. 18: The road network of Aberdeen. The blue coloured points are the centroids of all road meshes.

Fig. 19: The black framed area is the most dense region, where the size of the road meshes is small.

According to the second criteria the minimal circle around the intersection points of a found star pattern (like the four intersection points described in section 3.4. and Figure 14) can approximate the city centre. Figure 20 shows a town in northern Germany with the red coloured core of the city.



Fig. 20: A minimal circle around the star centre indicates a city centre.

5. CONCLUSION

We have shown some approaches to detect basic patterns in road networks. The paper has presented methods to recognise geometrical structures, which provide indications of the city configuration. The results of the recognition process of patterns like strokes, grids and stars are very promising. These components can be used to localise a city centre. However the applications have identified, that more efficient algorithms are needed to process large data sets in acceptable time. Another task is the improvement of the robustness of the algorithms. At the moment the recognition results depend on the right selection of parameters like the search radius of the star pattern. In future the algorithms have to be independent of those input variables. They should be determined automatically. Furthermore the detection of the ring-shaped structure is still a problem to solve.

We have described the idea of using patterns in road networks to determine more complex information like the above described city centre. However such kind of knowledge requires better techniques to combine the information. In addition to rule based systems data mining methods are further investigated to receive implicit information.

6. REFERENCES

ANDERS, K.-H., 2004, Parameterfreies hierarchisches Graph-Clustering-Verfahren zur Interpretation raumbezogener Daten. Ph.D. Thesis, Institute for Photogrammetry, University of Stuttgart, http://elib.uni-stuttgart.de/opus/volltext/2004/2024/, (urn:nbn:de:bsz:93-opus-20249).

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

DUDA, R. O., and HART, P. E., 1972, Use of the Hough Transformation to Detect Lines and Curves in Pictures. Comm. ACM, 15, 11-15.

JIANG, B., and CLARAMUNT, C., 2004, A Structural Approach to the Model Generalization of an Urban Street Network. GeoInformatica, 8:2, 157-171.

HARALICK, R.M., and SHAPIRO, L.G., 1992, Computer and Robot Vision. Addison-Wesley, Vol. I & II.

HEINZLE, F., and SESTER, M., 2004, Derivation of Implicit Information from Spatial Data Sets with Data Mining. International Archives of Photogrammetry and Remote Sensing, 35, 335- 340.

HOUGH, P.V.C., 1959, Machine Analysis of Bubble Chamber Pictures, Proceedings of International Conference on High Energy Accelerators and Instrumentation, CERN, 554-556.

KOPERSKI, K., and HAN, J., 1995. Discovery of Spatial Association Rules in Geographic Information Databases, *in:* M. J. Egenhofer & J. R. Herring, eds, 'Advances in Spatial Databases '95', Vol. 951 of *Lecture Notes in Computer Science*, Springer Verlag, Heidelberg, pp. 47–66.

LIAO, P.-S., CHEN, T.-S., and CHUNG, P.-C., 2001, A Fast Algorithm for Multilevel Thresholding. Journal of Information Science and Engineering, 17, 713-727.

LU, W., HAN J., and OOI B. C., 1993. Discovery of General Knowledge in Large Spatial Databases. Proc. Far East Workshop on Geographic Information Systems, Singapore, pp. 275-289.

MACKANESS, W., and EDWARDS, G., 2002, The Importance of Modelling Pattern and Structure in Automated Map Generalisation. Joint Workshop on Multi-Scale Representations of Spatial Data, Ottawa, Canada.

OTSU, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. on Systems, Man, and Cybernetics, 9(1), 62-66.

SESTER, M., 1995, Lernen struktureller Modelle für die Bildanalyse. Ph.D. Thesis, Institute for Photogrammetry, University of Stuttgart, Deutsche Geodätische Kommission, Reihe C, Nr. 441, München.

THOMSON, R., and RICHARDSON, D., 1999, The 'good continuation' principle of perceptual organization applied to the generalization of road networks. Proceedings of the 19th International Cartographic Conference, Ottawa, 1215-1223.

WAGNER, D., and WILLHALM, T., 2003, Geometric Speed-Up Techniques for Finding Shortest Paths in Large Sparse Graphs. Konstanzer Schriften in Mathematik und Informatik, Nr. 183

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

7. ACKNOWLEDGEMENT

This work is 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.

BIOGRAPHY

Frauke Heinzle (Dipl.-Ing.), born in 1971, studied Geodesy at the Technical University of Berlin and obtained the Master's degree (Dipl.-Ing.) in 1997. Her master thesis was about "investigations in relational matching regarding applications of photogrammetric point classification". Between 1997 and 1999 she attended a trainee on public service. Until 2002 she was the head of a department of public land clearance projects. Since 2002 she is a scientific assistant at the Institute of Cartography and Geoinformatics, University of Hannover. She is working in the EC-project SPIRIT and her primary interest lies in spatial data interpretation and data mining.

Dipl.-Ing. Frauke Heinzle Institute of Cartography and Geoinformatics University of Hannover Appelstraße 9a 30167 Hannover

e-mail: frauke.heinzle@ikg.uni-hannover.de phone: +49511/762-5255 page: www.ikg.uni-hannover.de