Map Generalization of Road Networks

Dr. ing. Jan Terje Bjørke
Norwegian Defence Research Establishment
and
Department of Mathematical Sciences and Technology
The Agricultural University of Norway
FFI, P.O. Box 115
NO-3191 Horten
NORWAY
fax: +47 33 04 78 34, phone: +47 33 03 38 37

email: jtb@ffi.no

ABSTRACT

The present paper introduces a new algorithm for the elimination of arcs in road maps. The algorithm is based on information theory. A case study illustrates how the algorithm works. The perceptual properties of the map reader and the resolution of the display unit are brought into the algorithm by a similarity function.

1 INTRODUCTION

Road networks have the following properties which should be considered in map generalization: (1) topological (connectivity, nodes and arcs), (2) metrical (arc shape and length), and (3) hierarchical properties (main roads and small roads). Generalization of road networks can be based on different principles like: (1) shortest path algorithm (2) minimum spanning tree algorithm (Mackaness and Beard 1993), (3) reduction of the number of vertices by amalgamating a selection of existing vertices (Mackaness and Mackechnie 1999), or (4) elimination of arcs (Bjørke 2003). In general, cartographic generalization is a complex process which is composed of several sub processes like: elimination of map objects, reduction of the amount of detail, enhancement of the appearance of map objects, amalgamation, collapse, exaggeration, typification and displacement (see for example McMaster and Shea 1992, or Jones 1997). The sequence and the combination of the different generalization operators are not obvious, and it represents a challenging task in map generalization. Nickerson (1988) demonstrates how the operators can be combined, and he proposes the three stages in the generalization of linear features: (1) feature elimination, (2) feature simplification, and (3) conflict detection and resolution.

Jiang and Claramunt (2004) present a topology oriented elimination algorithm, but they do not consider the visual separation of the roads. Bjørke (2003) describes a method for the thinning of arcs where the visual distance between them are measured. The method applies information theory. Isaksen (2004) has implemented the algorithm and tested it on road networks in Norway. The present paper summarizes the results of Bjørke (2003) and Isaksen (2004). The generalization algorithm considered is based on two principles: (1) eliminate the conflicting arcs of the network, (2) introduce constraints which maintain the topology and the hierarchical ordering of the different arcs of the network. The simplification of the shape of the arcs or displacement of the arcs are therefore not considered here. The forthcoming procedure can be summarized as: (1) compute an entropy based conflict index for each arc in the network, and (2) eliminate the most conflicting arc. The process terminates when the useful information of the map gets its maximum value (i.e., when the channel capacity is reached). Before an arc is to be eliminated,

Bjørke, J.T. (2005) Map Generalization of Road Networks. In *Visualisation and the Common Operational Picture* (pp. 17-1 –17-8). Meeting Proceedings RTO-MP-IST-043, Paper 17. Neuilly-sur-Seine, France: RTO. Available from: http://www.rto.nato.int/abstracts.asp.

its contribution to the connectivity of the network is evaluated. Very often road maps are used for shortest path travel. Therefore, arcs of high information value or high priority can be associated weights so that they survive in the elimination process.

2 BACKGROUND

The present paper contributes to the elaboration of the cartographic relevance of information theory, and it will be shown how the theory successfully can be applied in the generalization of road maps. Bjørke (1996) presents a framework for the application of information theory in cartography and gives a few examples illustrating the principles of the theory. Up to now information theory is not much used in map generalization, but there are some attempts to define generalization algorithms based on information theory (see for example Bjørke 1997, Bjørke and Myklebust 2001 or Huang 2002).

The difference between the entropy and the equivocation of an information source is termed useful information and is defined as (Shannon and Weaver 1949):

$$R = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X), \tag{1}$$

where the useful information R can be computed on the sender side or on the receiver side. Bjørke (1996) shows how the transition probabilities required for the computation of the map equivocation can be derived from a similarity function. The similarity between two objects x and y is a real number in the interval [0,1] and measures the visual separation between the map objects. It is defined so that the similarity is 1 if the map symbols can't be visually separated at all, and it is 0 when the symbols are clearly separable.

The application of Equation 1 to the generalization of road maps requires that we define the elements to be used in the entropy computation. In our case a fruitful approach is to spread points regularly along the arcs of the network. The points considered will be termed information points. The distance between the information points should be small, but since the computational cost of the forthcoming algorithm depends on the amount of points, their number must be limited.

The useful information of a map is an index which isolated does not bring much value to map design, but the search for its maximum value opens an exciting view. The strategy we will follow eliminates the arc which has the highest equivocation. The elimination goes on until R gets its maximum value.

3 THE NEW ROAD GENERALIZATION ALGORITHM

Based on the information points and the similarity function the entropy and the equivocation of the road map can be computed. The algorithm makes the simplification that the roads are represented as linear features, i.e., the width of the roads is not considered. The local equivocation for an information point x can be computed from

$$H(Y \mid x) = -\sum_{y \in Y} p(y \mid x) \log_2 p(y \mid x), \tag{2}$$

where the conditional probabilities are derived from a similarity function (see Bjørke 1996). In the forthcoming experiment I will apply a linear similarity function μ defined as

$$\begin{aligned} &\mu(y,x) = (T\text{-}s)/T \quad \text{when} \quad s \le T \\ &\qquad\qquad\qquad \text{else} \\ &\mu(y,x) = 0, \end{aligned} \tag{3}$$

where $\mu(y,x)$ is the similarity between the two information points y and x, s their distance and T is the parameter which defines the separation of the points, i.e., T is selected according to the map scale and the perceptual properties of the map user. The relation between similarity and conditional probability is defined by the normalization

$$p(y \mid x) = \frac{\mu(y,x)}{\sum_{y \in Y} \mu(y,x)} . \tag{4}$$

The computational cost of Equation 2 depends on the search area, i.e., how many neighbouring information points are visited. As seen from Equation 3 there is no need to visit points outside a certain range from x, therefore the algorithm will benefit from the utilization of a spatial data structure (e.g. a hash function or a quad tree structure). Note that Equation 4 also computes the similarity between an information point and all the information points. Therefore, we also get the probability that an information point is interpreted as itself. From Equations 2 and 4 we can compute the equivocation for each of the arcs of the network, i.e., we compute the mean value of $H(Y \mid x)$ for all the information points which belong to a certain arc.

Since we in Equation 2 compute the equivocation on the receiver side, we also have to define the useful information on the map reader side. The computation of the entropy at the map reader side, i.e., $H(Y)$, requires higher computational cost than the computation of the map entropy, i.e., $H(X)$. However, in our generalization procedure we can make the approximation $H(Y) \approx H(X) = \log_2 N$, where N is the total number of information points of the road network. Since the important aspect of our application of information theory is the definition of a stop criterion for the arc elimination, and since we can control the degree of generalization by the selection of the value of T, the approximation considered is not crucial. Based on a similar argument we also introduce an approximation of $H(Y \mid X)$ in Equation 1. From Equation 2 we compute an approximation of the equivocation of an arc a as

$$H(Y \mid a) \approx -\frac{1}{N_a} \sum_{x \in a} \sum_{y \in Y} p(y \mid x) \log_2 p(y \mid x) , \tag{5}$$

where N_a is the number of information points in a. In a similar manner the mean value of the equivocation of all the arcs is computed from

$$H(Y \mid X) \approx \sum_{a \in A} \frac{N_a}{N} H(Y \mid a) , \tag{6}$$

where A is the set of arcs in the network. Since we compute the probability of an arc as $\frac{N_a}{N}$, long arcs are associated higher weights than shorter arcs. In that way the spatial extent of the arcs is considered.

Based on the previous discussion, we arrive at the following stop criterion:

$$\boxed{\text{stop the elimination when } \max(\log_2 N - H(Y \mid X) \text{ is reached,}} \tag{7}$$

where $H(Y \mid X)$ is computed from Equations 5 and 6. The search for the maximum value in Equation 7 requires in principle $O(n!)$ computations (where n is the number of arcs in the initial network). It is possible to limit the search space so that the computing time is reduced to the order $O(n^2)$. The reduction of the computational cost is based on the assumption that what is best locally is also best globally. Therefore, we eliminate the most conflicting arc and stops the elimination when a maximum value of R is reached. Since R can have several local maximum points, the elimination must continue until a global maximum value is reached.

The way we have formulated the entropy computation does not consider the topological properties of the road network. Since connectivity is important in road networks, additional constraints must be added to Equation 1. The additional constrains can be of several types. They can consider the hierarchy of the roads and their connectivity. The hierarchy can be introduced by some weighting functions so that high priority roads tolerate higher equivocation than the minor roads. The topological properties can be of the kinds don't eliminate the candidate arc if this elimination: (1) increases the number of end nodes or (2) brakes the network into disconnected components. The first condition will be termed strong and the second weak topology condition. If a candidate arc cannot be eliminated due to the topological condition, the next arc is evaluated. If all the arcs are evaluated and none of them can be eliminated, the process terminates. Therefore, the algorithm has the two stop criteria:

stop the elimination when the criterion in Equation (7) is met
or
stop the elimination if the topological conditions prevent any arc to being eliminated. (8)

When applying Equation 3, all neighbouring points are evaluated. Therefore, also the "wanted" equivocation is computed. The previous statement sounds remarkable, but it has a simple explanation. Assume a straight line and that we regard the line as composed of a set of dots. When the dots get closer and closer to each other, we cannot separate them and the line looks smooth, i.e., we can talk about wanted equivocation. This kind of equivocation is utilized by raster printers and raster screens. In order to remove the wanted equivocation from the stop criterion in Equation 7, the algorithm filters out the information points which contribute to the kind of equivocation considered.

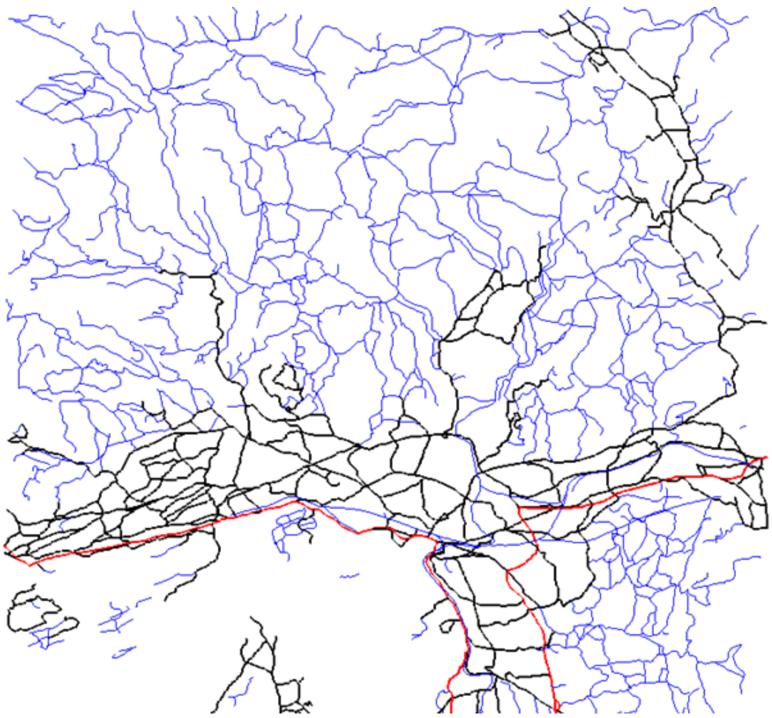

Figure 1: Original Map. Main roads in red, secondary roads in black and other roads in blue.

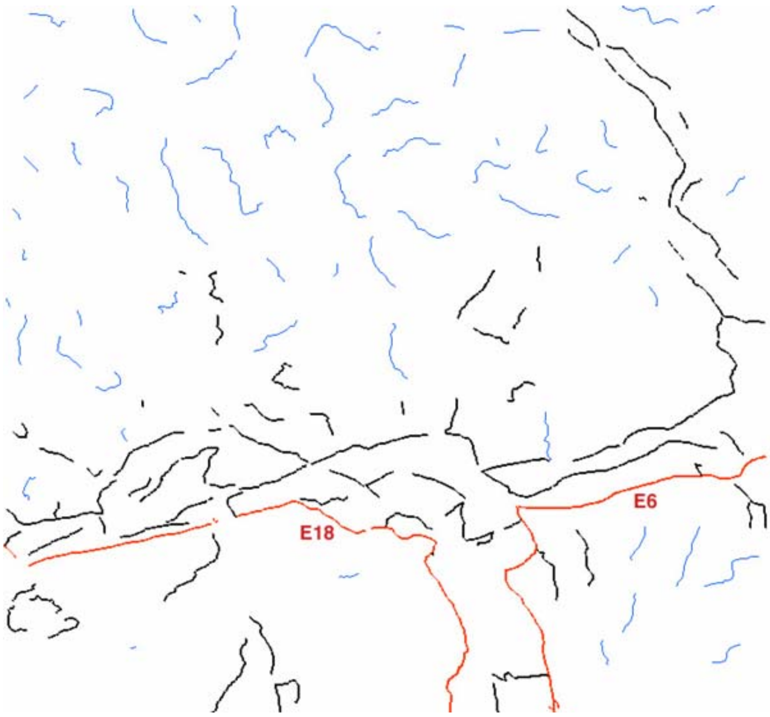

Figure 2: Generalized Map, T=35. The hierarchy of the roads are considered, but no topology constraints are introduced. 83% of the roads are eliminated. Main roads in red, secondary roads in black and other roads in blue.

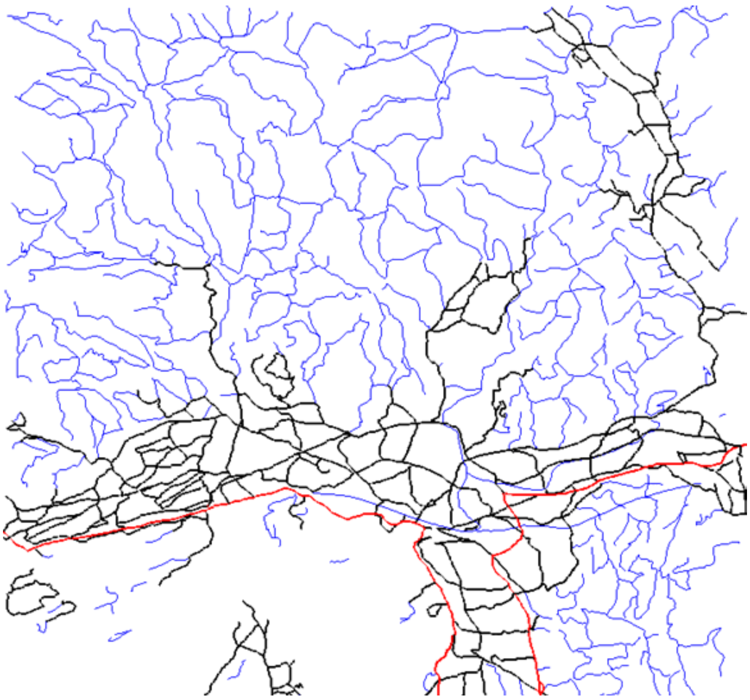

Figure 3: Generalized Map, T=15. The hierarchy and the topology of the roads are considered. 24% of the roads are eliminated. Main roads in red, secondary roads in black and other roads in blue.

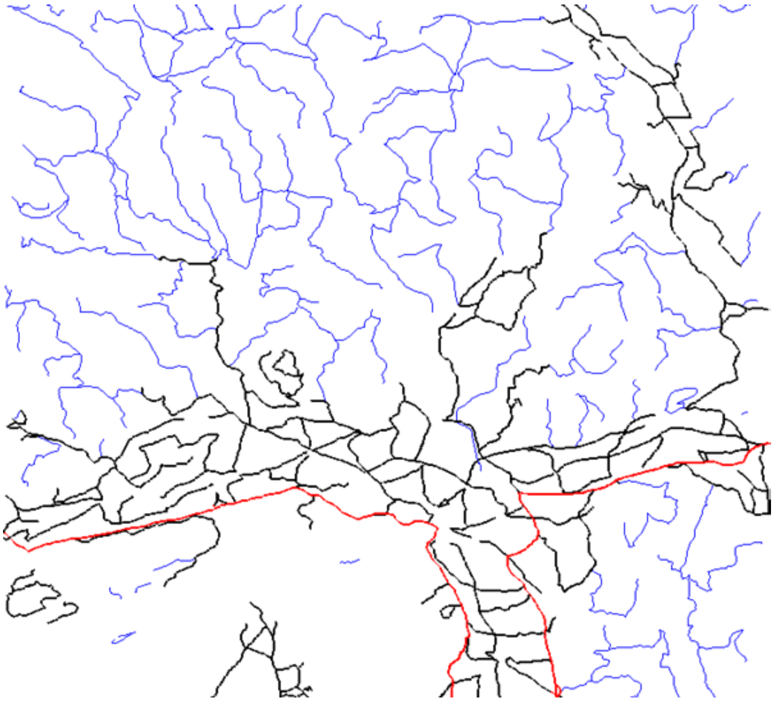

Figure 4: Generalized Map, T=35. The hierarchy and the topology of the roads are considered. 49% of the roads are eliminated. Main roads in red, secondary roads in black and other roads in blue.

4 CASE STUDY ROAD DATA FROM NORWAY

In order to evaluate and demonstrate the cartographic relevance of the proposed algorithm, a case study will be presented. The road network of Oslo is selected. Figures 1 to 4 give some map examples. The original map is shown in Figure 1. The next figure shows a generalized map where no topology constraints are introduced, but the hierarchy of the roads are considered. The map illustrates that if no topology constraints are considered, the generalized road network is broken into unconnected paths. In Figures 3 and 4 topology constraints are introduced. Here we can see that the connectivity of the network is maintained during the generalization process. The two maps illustrate how the threshold value T influences the number of eliminated arcs, i.e., T controls the degree of generalization. The computing time for the generalized maps varies from 2 to 350 seconds dependent on the kind of constraint introduced. In order to keep the computing time at a reasonable level, spatial data structures are applied and for each time the R-value of the map is computed, only the neighbouring arcs of the eliminates arc are recomputed.

5 DISCUSSION

The case study illustrates how the degree of generalization increases as the similarity function catches a greater area, i.e., when we increase the value of T. Therefore, it is clear that the similarity function controls the degree of generalization and its design becomes a key problem in the presented method. When we have knowledge about the perceptual properties of the map reader, the resolution and the size of the display, we assume it is possible to find an appropriate similarity function. The information theoretic approach does not solve the problem of how to design the similarity function. On the other side we can argue that the similarity function brings the perceptual domain into the algorithm and therefore represents a degree of freedom in the map design.

The topological constraints prevent the network from being broken into disconnected peaces. The effect of these conditions becomes more distinct as parameter T increases. In addition to the topological constraints, the hierarchy of the roads is considered. The hierarchy can be based on the type of roads and their relevance. For example, shortest path from one destination to another is often of great interest in mobile map services. Therefore, the roads of the shortest path can be associated high weights.

The algorithm maximizes the useful information in a road map. The described search for the maximum value of R doesn't evaluate all the combinations of the arcs. Since the sequence of the arcs in the elimination may have impact on the final result, the solution is an approximation to the channel capacity of the map. However, we assume that the approximate solution is a good estimate for the solution we are looking for. If all the combinations of the arcs should be evaluated, the computational cost would be of the order $O(n!)$. The strategy we have selected reduces the computational cost to $O(n^2)$, which is manageable.

The experiments illustrate that there may remain visual conflicts in the map after the thinning of the arcs. Therefore, there may bee a need to supplement the elimination with other generalization algorithms like displacement. In the literature there are proposed solutions to the implementation of displacement algorithms (see for example Ruas, 1998, Højholt 2000, Harrie and Sarjakoski 2002).

6 CONCLUSIONS

The information theoretic approach can be used to reduce the number of arcs in a road map. Since Shannon information theory does not contain a spatial concept, constraints must be introduced in order to take care of the topological properties of the network. The algorithm presented is automated and requires only the parameter of the similarity function to be defined á priori. Since the method is (1) automated, and (2) can be formulated so it is computational effective, it can be applied in mobile map services. Mobile map services often require flexibility in terms of adjustment to the changing context. The presented algorithm offers this kind of flexibility by the introduction of constraints for the priority and the connectivity of the roads. The additional constraints can be utilized in the selection of important information, e.g., the results of graph algorithms like shortest path or travelling salesman analyses can be integrated into the map generalization.

REFERENCES

Bjørke, J.T. 1996. Framework for entropy-based map evaluation. *Cartography and Geographic Information Systems*, 23(2), pp. 78-95.

Bjørke, J.T. 1997. Map generalization: An information theoretic approach to feature elimination. In Lars Ottoson (editor), *Proceedings of ICC97, 18th International Cartographic Conference,* vol. 1, pp. 480-486, Stockholm 23- 27 June 1997, Swedish Cartographic Society.

Bjørke, J.T. and I. Myklebust 2001. Map generalization: Information theoretic approach to feature elimination. In J.T. Bjørke and H. Tveite (editors), *Proceedings of ScanGIS'2001, 8th Scandinavian Research Conference on Geographical Information Science,* pp. 203-211, Ås, Norway, 25-27 June 2001.

Bjørke, J.T. 2003. Generalization of road networks for mobile map services: An information theoretic approach. In P*roceedings International Cartographic Conference (ICA),* Durban, South Africa 2003.

Harrie, L. and T. Sarjakoski 2002. Simultaneous graphic generalization of vector data sets. *Geoinformatica* , 6(3), pp. 233-261.

Højholt, P. 2000. Solving space conflicts in map generalization: Using a finite element method. *Cartography and Geographic Information Science,* 27(1), pp. 65-73.

Isaksen, E. 2004. Generalization of road networks by the application of information theory. *Master thesis* Department of Mathematical Sciences and Technology, The Agricultural University of Norway.

Jiang B. and C. Claramunt 2004. A structural approach to the model generalization of an urban street network. *GeoInformatica,* 8(2), pp. 157-171.

Jones, C.B. 1997. *Geographical information systems and computer cartography*. Essex: Addison Wesley Longman Limited.

Li, Z. and P. Huang 2002. *Quantitative measures for spatial information of maps.* International Journal of Geographical Information Science, 16(7), pp. 699-709.

Mackaness, W.A and G.A. Mackechnie 1999. Automating the detection and simplification of junctions in road networks. *Geoinformatica*, 3(2), pp. 185-200.

Mackaness, W.A and M.K. Beard 1993. Use of graph theory to support map generalization. *Cartography and Geographic Information Systems*, 20(4), pp. 210-211.

McMaster, R.B. and K.S. Shea 1992. *Generalisation in digital cartography.* Washington D.C., 1710 16th Street NW: Association of American Geographers.

Nickerson, B.G. 1988. Automated cartographic generalization for linear features. *Cartographica*, 25(3), pp. 15-66.

Ruas, Anne 1998. A method for building displacement in automated map generalisation. *International Journal of Geographical Information Science,* 20(8), pp. 789-803.

Shannon, C.E. and W. Weaver 1964. *The mathematical theory of communication*. Urbana: The University of Illinois Press.