The Cartographic Journal *Vol. 39 No. 2 pp. 153–166 December 2002*

REFEREED PAPER

Topographic Map Generalization: Association of Road Elimination with Thematic Attributes

Zhilin Li and Yuen Hang Choi

Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong.
Email: Lszlli@polyu.edu.hk

It has been widely recognized that, in order to automate the map generalization process, cartographic knowledge needs to be formalized. One method of knowledge acquisition is to analyse existing maps. Previous studies have concentrated on describing the phenomena on the surface. A typical example is the percentage change of the number of symbols or the percentage of open space on smaller scale maps. This study aims to go one step further, i.e. to analyse the association such changes with thematic attributes. In this study, road features on topographic maps of Hong Kong from 1:1000 to 1:200 000 are studied; six types of thematic attributes, i.e. 'type', 'length', 'width', 'number of lanes', 'number of traffic ways' and 'connectivity', are considered; and two statistic parameters, Lambda and Somers' Delta, are employed. It has been discovered that the dependence of road elimination on these attributes occurs in the following order: type (0.73), length (0.53), number of lanes (0.48), number of traffic directions (0.41), width (0.36) and connectivity (0.19). If these numbers are normalized into percentages, then these values become: type (27%), length (20%), number of lanes (18%), number of traffic directions (15%), width (13%) and connectivity (7%). Such results could be used to formulate an overall weight to determine whether a particular road should be deleted, merged or combined in the generalization process, in order to retain a certain proportion of roads at a smaller scale.

INTRODUCTION

Maps may need to be updated in a 2–3 year cycle due to rapid changes in the environment. The traditional practice is to manually update maps at all scales, which is inefficient and expensive. A better solution, in this digital era, is to frequently update only the largest-scale maps, and then to derive maps at smaller scales by an automated generalization process when needed.

In this area, the acquisition of cartographic knowledge is a major research theme (Weibel, 1995). It has been recognized that knowledge of generalization could be acquired (a) from cartographic experts through interviews, (b) from existing maps through analysis, or (c) via map specifications. A lot of work on this topic has been undertaken (e.g. Buttenfield and McMaster, 1991). However, it has been generally agreed that the knowledge in the cartographic expert's mind is not explicit enough to be extracted, and the map specification normally tells us 'what not to do' instead of 'what to do'. Even where there are specifications telling us 'what to do', they do not specify 'how to do it' in detail.

Efforts have therefore been made to find cartographic knowledge from existing maps. As topographic maps are the most standardized, they have often been selected for study. Töpfer and Pillewizer (1966), through analysis of topographic maps, have formulated the so-called 'Principle of Selection' or 'Radical Laws', expressing the relationship between map scale and the number of features represented on maps. Their formulations are based on the implicit and widely accepted idea that generalization is a process of 'logical diminution'. Yu (1993) has extended this radical law by connecting it to the concept of 'fractal dimension'. A more comprehensive study of knowledge acquisition from existing maps was carried out by Müller (1990). He undertook a detailed analysis of the effect of generalization on the number of features represented on German topographic maps at scales from 1:1000 to 1:500 000. He analysed not only the percentage number of features represented on these maps, but also the ratio between the spaces occupied by different classes of features. His results were represented in two different ways, i.e. a logic representation scheme and charting the relationships between cartographic objects and generalization operations into a relational table. Leiter and Buttenfield (1995) studied the topographic maps of Austria. They analysed the selection factors (percentage of features selected at a smaller scale) for various features from 1:50 000 to 1:200 000 and 1:500 000 scales.

However, in these studies, only the phenomena are described. A typical example of their discoveries is something like this: 'If the scale changes from 1:50 000 to

1:200 000, then the number of individual settlement objects is reduced to 16% of the total at the larger scale.' But one still doesn't know which settlement to eliminate and which to combine. In order to answer these two questions, it seems that thematic (attribute) information could be a key. Indeed, the authors believe that rules based on thematic information should be formulated to control geometric operations, and topological information should be used to validate geometric operations and control the quality of the generalization process. For the reasons given above, the authors attempt to connect the changes of spatial representations at different scales to a set of thematic attributes. However, the discussions will be confined to road features.

RESEARCHING MAP GENERALIZATION

In this study, the authors follow Müller's (1990) advice, i.e. to 'isolate the problems into modules, one for name generalization, one for building generalization, one for road network generalization, independently from one another but in many possible different contexts. Only after, would their interaction be considered'. As a result, only the transformation of the spatial representation of road networks will be considered.

As the topographic map has a long tradition and is regulated by more standards than any other map type (Müller, 1990), the series of topographic maps produced by a local mapping agency, i.e. the Survey and Mapping Office (SMO) of the Lands Department in the Hong Kong Government, is selected for study. The Hong Kong topographic map series covers six scales: 1:1000, 1:5000, 1:20 000, 1:50 000, 1:100 000 and 1:200 000. As it is not meaningful to study all maps covering the whole territory, a portion of Hong Kong Island is selected for this study, because it best represents the highly developed urban areas as well as hilly areas with few houses. The 1:200 000 scale map is shown in Figure 1.

The thematic attributes of road to be considered are 'type', 'length', 'width', 'number of lanes', 'direction of traffic' and 'connectivity'.

The methodology of study consists of three stages:

1. To observe and trace the changes in the representations of road features on different scale maps;
2. to quantify the changes (transformations) for statistical testing using the 'Feature Vanishing Level' (FVL); and
3. to carry out analysis of the dependency of FVL on these thematic attributes.

Statistical parameters will be used to represent the likelihood of a feature with certain attributes being eliminated or preserved. Two statistical parameters, i.e. Lambda (λ) and Somers' Delta (Δ) are used. More detailed discussion of these parameters follows later, in the analysis section.

OBSERVATION OF THE TRANSFORMATIONS OF SPATIAL REPRESENTATION

For typical topographic maps at scales of 1:1000 and 1:5000, many map features are shown in their true dimensions. However, when the scale is further reduced, the resulting map features would be too densely packed and too small to be deciphered. One of the solutions enabling clear presentation of the map features is to symbolize them. At 1:20 000, roads are classified in three types, 'expressway', 'main road' and 'secondary road'. Secondary roads are further divided into five sub-classes, double width road (A), single width road (B), unsurfaced road (C), non-motorable road (D) and lane (E). At increasingly smaller scales, a more general classification is formed. For example, expressways and main roads are grouped together and represented by the same symbol; double width secondary roads and single width secondary roads are reclassified as secondary roads, and unsurfaced secondary roads are separated to form a new sub-class. However, this sub-class is eliminated at 1:100 000 and 1:200 000, and the 'lanes' and 'non-motorable' roads are eliminated at 1:50 000. The transformation process is illustrated in Figure 2. The change in the number of mapped road segments during the scale reduction is shown in Table 1.

However, things are not as simple as indicated in the observations described above. For example, only 25.4%

Figure 1. Topographic map of the Hong Kong Island at 1:200 000 scale (from SMO, HK)

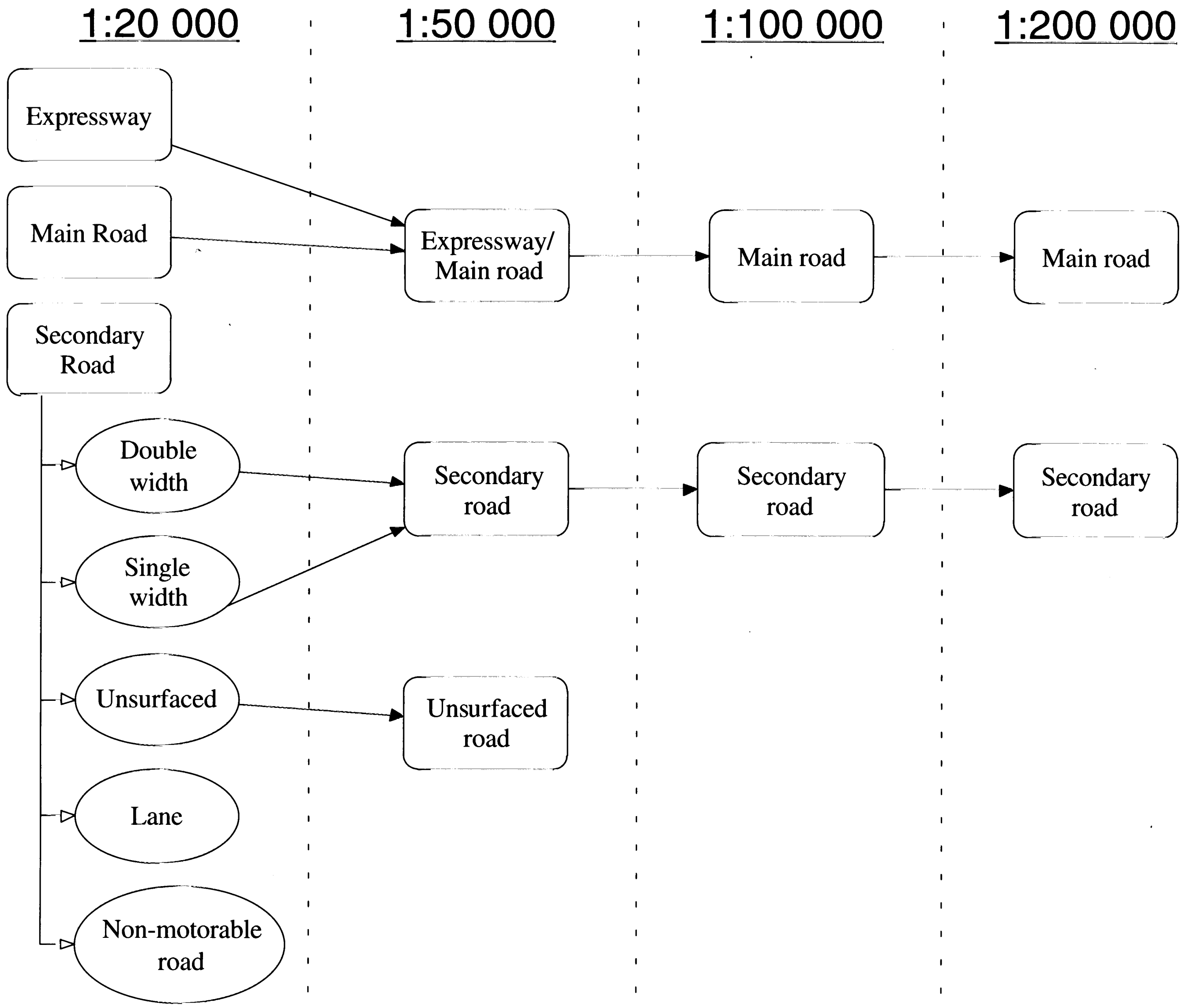

Figure 2. Transformations of road features from 1:20 000 to 1:200 000

of 'secondary roads' are preserved from 1:100 000 to 1:200 000. An important question arises: 'What kind of secondary roads should belong to the 25.4%?' This is certainly not an easy question. The answer could be, 'those secondary road with connections in four or more directions are more likely to be preserved than those with two connections'. The question then becomes something like 'how likely is it that the secondary roads with four or more connecting directions will be preserved?' There are more similar questions. The next two sections will try to answer such questions.

Before discussing the quantification of changes of road representation at different scales, a discussion of the classification systems is required. There are two classification systems used by two separate government departments, i.e. the Surveying and Mapping Office (SMO) of the Lands Department, and the Transport Department. The classification of roads by the Transport Department is

Table 1. Change in number of road segments of different classes from 1:1000 to 1:200 000

	1:1000	1:5000	1:20 000	1:50 000	1:100 000	1:200 000
Main Roads	–	–	79 (100%)	74 (93.7%)	74 (93.7%)	64 (81.0%)
Secondary Roads	–	–	195 (100%)	108 (55.4%)	56 (28.7%)	7 (3.6%)
Total	280	280	274 (100%)	182 (65.0%)	130 (46.4%)	71 (25.4%)

based on function, and roads are classified into two network types, i.e. Major Road Network and Minor Road Network. The Major Road Network includes Expressways (EX), Urban Trunk Roads (UT), Rural Trunk Roads (RT), Primary Distributors (PD), District Distributors (DD), Rural Roads A (RD), and some Local Distributors (LD). The remaining Local Distributors form the Minor Road Network. The classification is based mainly on their designed functions, which are declared by their names and supplemented by the road construction standard and traffic management. The differences in classification between the SMO and the Transport Department can be found by simply matching these two classification schemes. A comparison of the classification of roads between the Lands Department and the Transport Department reveals:

- All the Expressways are classified as 'Expressways'.
- All the Urban Trunk Roads are classified as 'Main Roads'.
- Nearly all Primary Distributor roads (96.6%) are classified as 'Main Roads'. Only one Primary Distributor road is classified as a 'Secondary (type) A' roads.
- About three-quarters of District Distributor roads (73.9%) are classified as 'Main Roads' and the remaining quarter of District Distributor roads (26.1%) are classified as 'Secondary (type) A' roads.
- A few Local Distributor roads (5.1%) are also classified as 'Main Roads'. Most Local Distributor roads are classified as 'Secondary (type) A' roads (84.8%). Some of them are classified as 'Secondary (type) B' roads (7.3%) and the remaining few are classified as 'Lanes' (2.8%).
- Nearly all the Lanes (LE) (84.6%) are classified as 'Lanes', with just a few (15.4%) being classified as 'Secondary (type) A' roads.

However, there is no information about unsurfaced and non-motorable roads in the Transport Department classification. So these two classes cannot be compared, but can only be clarified by their definitions from the Lands Department.

The two systems, considered overall, are reasonably consistent. In this study, the Transportation Department's system is used, because this classification is mainly based on the designed functions of roads, and is thus more related to their thematic attributes.

QUANTIFICATION OF TRANSFORMATIONS OF ROAD REPRESENTATIONS

A total of 280 (at 1:20 000 or larger scale) road segments on topographic maps of Hong Kong Island were studied. In this section, the changes in the representation of road features at different scales are quantified.

Road type serves as a major index of the importance of different roads. At each scale, numbers of road segments grouped by road types were counted. The results are summarized in Table 2, where the numbers in percentages are shown in brackets. A graphic representation of the results in Table 2 is shown in Figure 3. It is clear that all UT (Urban Track Roads) and EX (Expressways) roads are retained, while the number of LD (Local distributors) and DD (District distributors) types decreases more or less linearly with a decrease in scale from 1:20 000. All roads of LE (Lanes) type disappear from 1:50 000.

By observation, it can be seen that roads with different lengths may be erased or retained during scale reduction. There is no threshold to determine the elimination of road segments at each scale. To obtain quantitative observations, the total number of road segments on the maps at each scale was counted. At each scale, the road segments are divided into ten groups, so that the first group occupies the first ten percentages, the second group occupies the first twenty percentages, and so on. The mean values of road length for each group are then computed (Table 3).

Table 2. Number (%) of road segments at different scales in different types

	1:1000	1:5000	1:20 000	1:50 000	1:100 000	1:200 000
Expressways (EX)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)
Urban Track Roads (UT)	7 (100%)	7 (100%)	7 (100%)	7 (100%)	7 (100%)	7 (100%)
Primary distributors (PD)	29 (100%)	29 (100%)	29 (100%)	29 (100%)	28 (96.6%)	27 (93.1%)
District distributors (DD)	46 (100%)	46 (100%)	46 (100%)	42 (91.3%)	37 (80.4%)	27 (58.7%)
Local distributors (LD)	180 (100%)	180 (100%)	178 (98.9%)	103 (57.2%)	52 (28.9%)	9 (5.0%)
Lanes (LE)	17 (100%)	17 (100%)	13 (76.5%)	0 (0%)	0 (0%)	0 (0%)

Table 3. Mean lengths (at different portions) of all of roads at different scales

	1:1000	1:5000	1:20 000	1:50 000	1:100 000	1:200 000
100%	3032.4	3032.4	3082.6	3615.9	4224.4	5134.3
90%	1271.6	1271.6	1329.4	1684.5	2109.2	2489.9
80%	665.8	665.8	690.2	1145.8	1540.5	1918.9
70%	435.1	435.1	457.2	734.7	1089.5	1510.6
60%	267.4	267.4	278.3	538.1	769.6	1174.1
50%	192.2	192.2	201.7	392.4	583.2	857.3
40%	130.9	130.9	134.8	261.7	453.2	603.7
30%	101.9	101.9	106.9	187.8	296.8	447.6
20%	66.9	66.9	70.4	124.6	170.5	205.1
10%	44.1	44.1	47.6	66.3	109.5	114.3

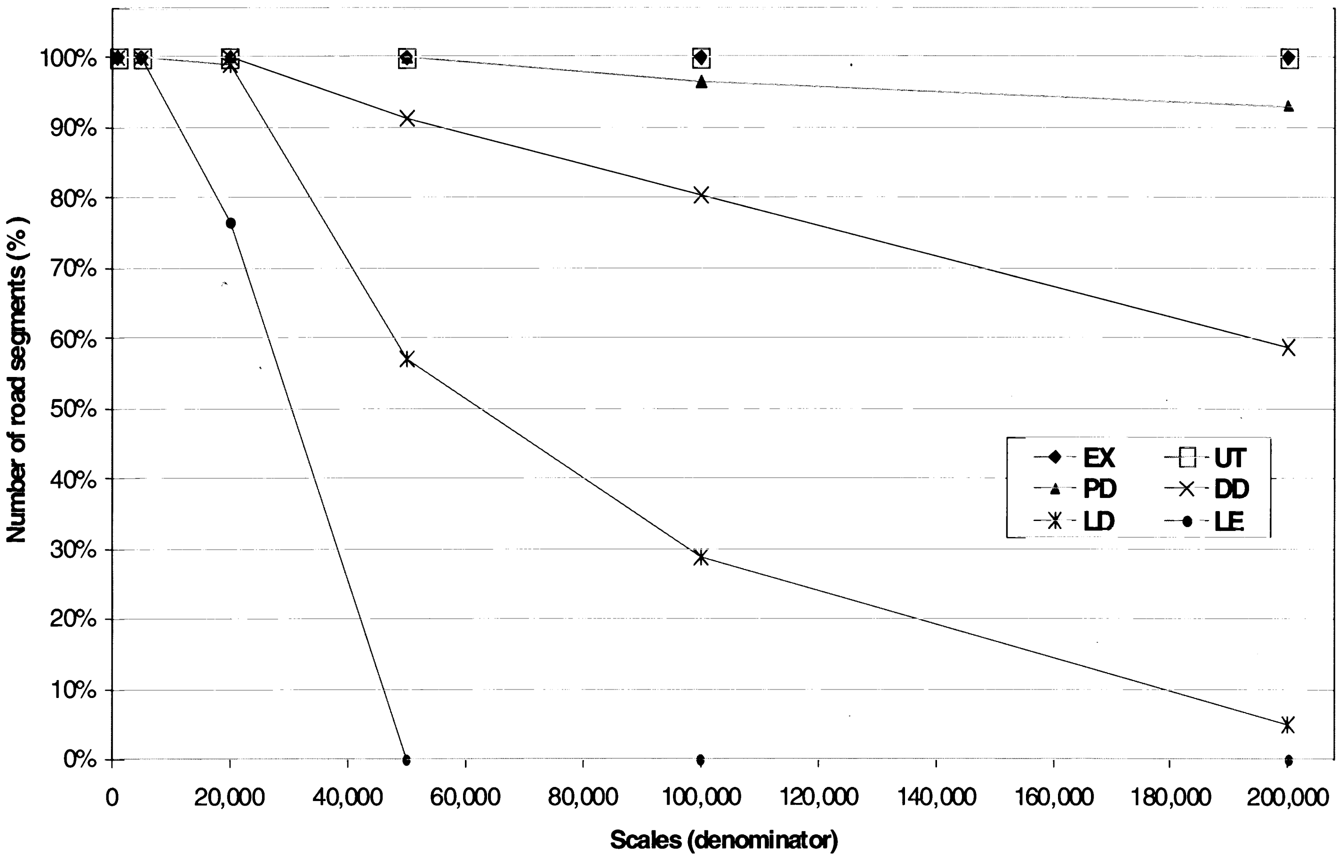

Figure 3. Retention of road segments (in percentage) for each road type at different scales. (EX=Expressways; UT=Urban Track Roads; PD=Primary distributors; DD=District distributors; LD=Local distributors; and LE=Lanes)

The results of this categorization are shown in Figure 4. It is clear that the reduction of short road segments is dramatic with a decrease in scale.

In the case of road width, the situation is similar. Therefore, the methodology used for the study of road width is the same as that used for the study of road length. The means of road widths with different percentages are shown in Table 4 and Figure 5. It appears that the mean widths of the first 30% of road segments represented on maps at the scale 1:20 000, 1:50 000 and 1:100 000 are nearly the same. The occupying rate of wider roads increases dramatically when the scale is 1:50 000 or smaller.

In the study of the association of road representation with number of lanes, non-motorable road segments were excluded. There are 20 such road segments in total. The number of lanes for the remaining 260 roads ranges from 1–11. Table 5 records the total number of road segments shown at each scale for different numbers of lanes. The percentage value is given in brackets. It can be found that roads with fewer lanes are more likely to be eliminated. Elimination takes place mostly when the number of lanes is one or two. No roads with six or more lanes are eliminated. Figure 6 is a graphic representation of the data in Table 5. However, the data for roads with six or more lanes are not plotted because they are all 100% retained.

A road is usually not designed for only one traffic direction. Most roads are designed to be bi-directional (i.e. the traffic moves in both directions). However, some roads are non-motorable, and hence the number of traffic directions is declared as ‘0’. Sometimes, a road is wide

Table 4. Mean widths (at different portions) of all roads at different scales

	1:1000	1:5000	1:20 000	1:50 000	1:100 000	1:200 000
100%	26.4	26.4	26.8	30.7	33.6	38.3
90%	15.5	15.5	15.7	18.2	20.9	25.5
80%	11.8	11.8	12.1	14.3	15.8	20.6
70%	9.9	9.9	10.0	12.0	12.9	16.2
60%	8.2	8.2	8.3	10.1	10.6	13.5
50%	7.4	7.4	7.4	8.8	9.3	12.0
40%	6.7	6.7	6.8	7.6	7.7	10.4
30%	6.1	6.1	6.2	7.0	7.1	9.5
20%	4.8	4.8	5.0	5.4	5.0	7.7
10%	3.3	3.3	3.4	3.1	3.0	5.2

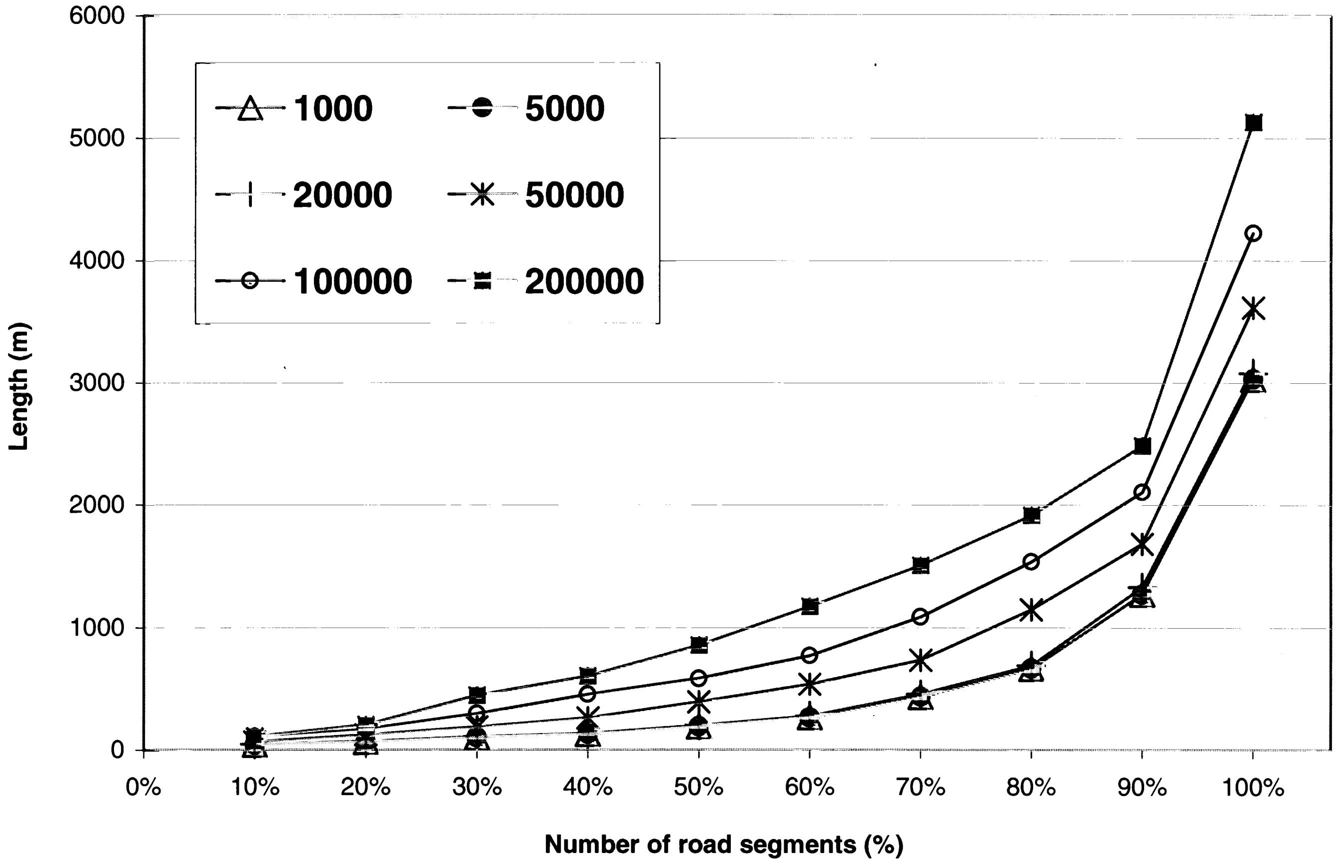

Figure 4. Mean road lengths at different portions of total number of roads for different scales

enough to serve three directions of traffic (e.g. an additional separated lane with a physical barrier leading to a flyover), so the number of directions is declared as '3'. The categorization of road segments by number of traffic ways is presented in Table 6. A graphic view is given in Figure 7. It is clear that almost all non-motorable roads disappear at 1:50 000. Road segments containing three directions of traffic are kept at all scales. In sum, road

Table 5. Number (%) of road segments with different lanes, at different scales

No. of Lanes	Number of road segments appears at different scales					
	1:1000	1:5000	1:20 000	1:50 000	1:100 000	1:200 000
1	86 (100%)	86 (100%)	85 (98.8%)	28 (32.6%)	19 (22.1%)	5 (5.8%)
2	106 (100%)	106 (100%)	105 (99.1%)	88 (83.0%)	58 (54.7%)	30 (28.3%)
3	23 (100%)	23 (100%)	23 (100%)	21 (91.3%)	14 (60.9%)	9 (39.1%)
4	25 (100%)	25 (100%)	25 (100%)	24 (96.0%)	15 (60.0%)	8 (32.0%)
5	8 (100%)	8 (100%)	8 (100%)	8 (100%)	7 (87.5%)	7 (87.5%)
6	7 (100%)	7 (100%)	7 (100%)	7 (100%)	7 (100%)	7 (100%)
7	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)
8	2 (100%)	2 (100%)	2 (100%)	2 (100%)	2 (100%)	2 (100%)
9	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)
10	0	0	0	0	0	0
11	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)

Table 6. Number (%) of road segments for traffic ways from 0 to 3 at different scales

No. of traffic ways	1:1000	1:5000	1:20 000	1:50 000	1:100 000	1:200 000
0	20 (100.0%)	20 (100.0%)	16 (80.0%)	1 (5.0%)	0	0
1	111 (100.0%)	111 (100.0%)	111 (100.0%)	62 (55.9%)	35 (31.5%)	21 (18.9%)
2	147 (100.0%)	147 (100.0%)	145 (98.6%)	117 (79.6%)	88 (59.9%)	48 (32.7%)
3	2 (100.0%)	2 (100.0%)	2 (100.0%)	2 (100.0%)	2 (100.0%)	2 (100.0%)

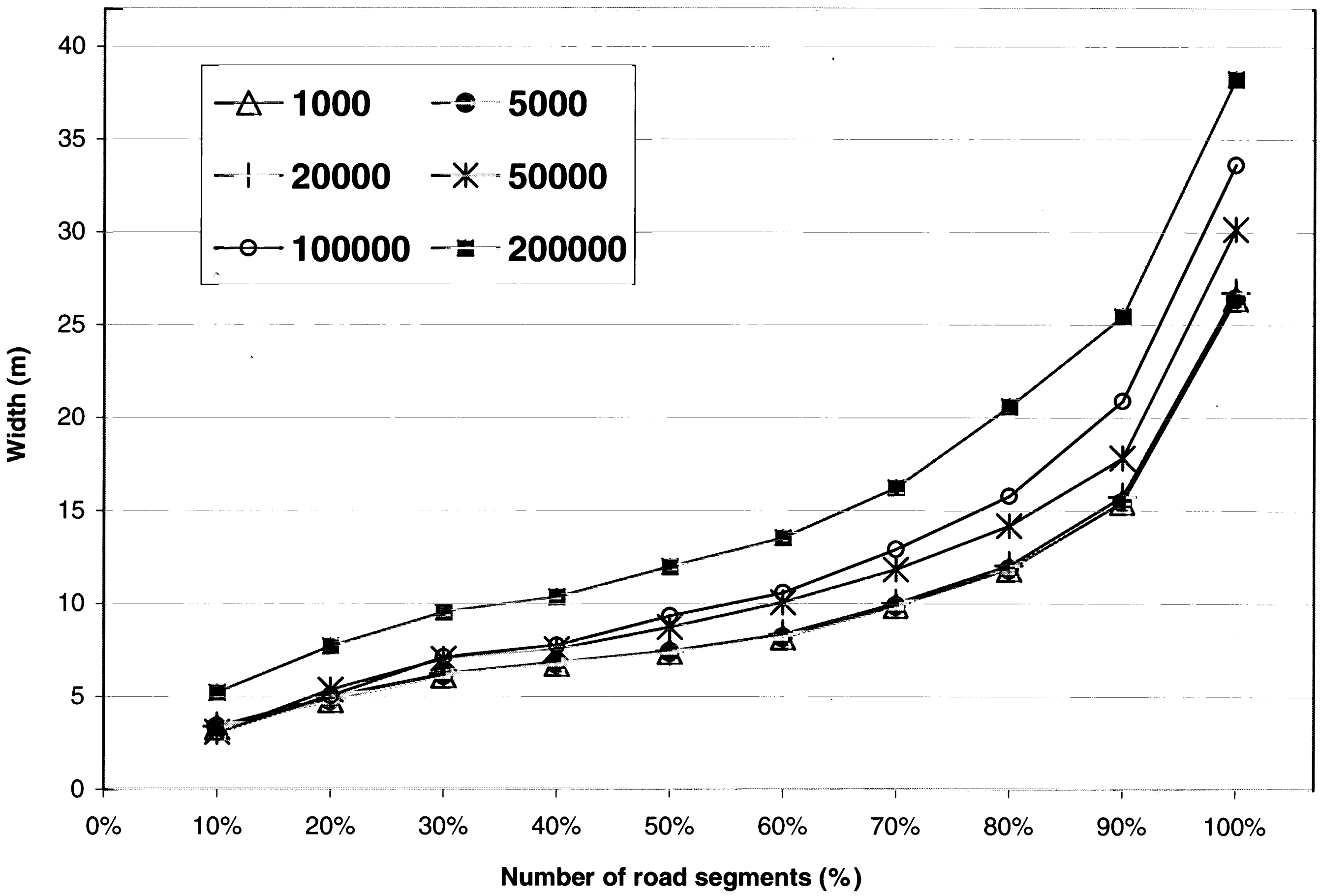

Figure 5. The relation between mean road widths and total number of roads for different scales

segments with more directions are more likely to be shown on maps.

Connectivity is an important concept in network analysis. If a node is connected by two arcs, it is considered as having an order of 2. Similarly, if a node is connected by three arcs, it has an order of 3, and so forth. As a road segment consists of a node at each end, there is a need for a pair of two numbers to identify its order, e.g. '0,0'. The values of these pairs range from '0,0' to '5,5' in this study. The results of the road segments categorized by this system are listed in Table 7a.

The data in Table 7a can be reclassified into five categories, '0,0', 'X,0', '1,1', 'X,1' and 'X,X'. '0,0' means that the road segment is either a floating segment (without connection to other segments) or a self-contained loop. 'X,0' signifies a dangling segment with one end unconnected and the other end connected to any number of road segments, except number '0'. '1,1' means that it is a bridge between two segments; 'X,1' also means a bridge but it is used to connect a segment and several other segments at the two ends; the 'X' in this case can be any number except 0 and 1. Finally, 'X,X' means that the two ends can be connected to any number of road segments, but the number should be equal to or greater than 2. After the data are reclassified, a new table (Table 7b) containing the above five categories is formed, and the relationship between connectivity and the number of roads retained at different scales is illustrated in Figure 8. It is clear that roads with '0,0' connectivity are all eliminated from 1:50 000.

DEPENDENCY OF 'FEATURE VANISHING LEVEL' TO THEMATIC ATTRIBUTES

The study quantified the strength and nature of the relationships between road elimination levels (REL) and six road attributes: type, length, width, number of lanes, number of directions of traffic and connectivity.

Due to the different natures of the attributes, the data observed belong to two different scales of measurement, i.e. nominal and ratio. Two types of association (dependence) measures were applied, respectively. For the nominal data, the method used was Goodman–Kruskal's Lambda (λ). For the ratio data, Somers' Delta (Δ) is used.

Goodman-Kruskal's Lambda statistic (see Norusis, 1988) is a (proportional) reduction in error (PRE) measure, which looks at how much improvement is the prediction of the value of a dependent variable when the value of an independent variable is known. The formula is expressed as follows:

$$\lambda = \frac{\text{Misclassified in situation 1} - \text{Misclassified in situation 2}}{\text{Misclassified in situation 1}} \quad (1)$$

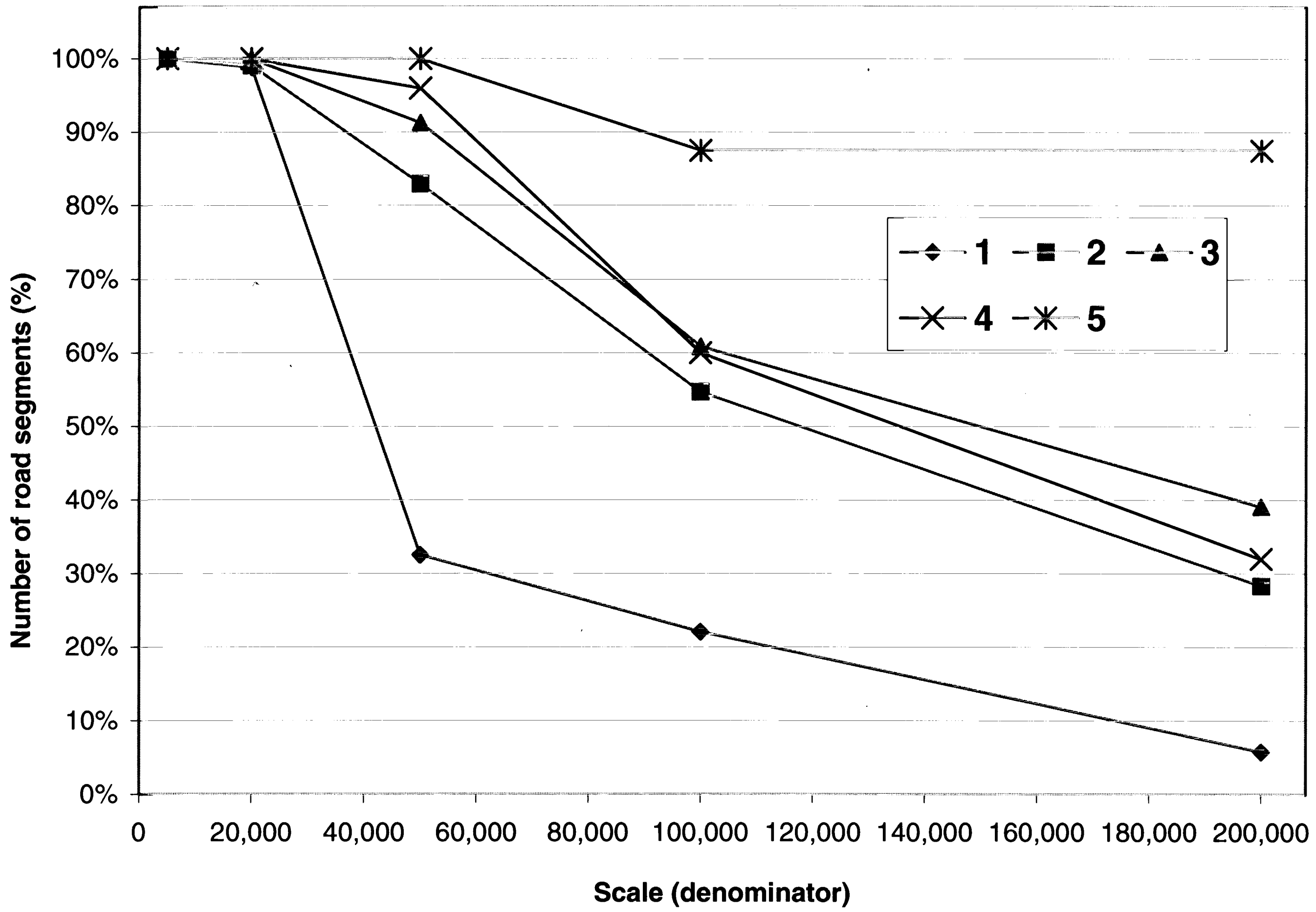

Figure 6. Retention of road segments (in terms of percentage) at different scales for lanes from 1 to 5

In this formula, 'misclassified in situation 1' and 'misclassified in situation 2' can be explained by the example data given in Table 8. This example is used to predict whether one's life is seen as dull, routine or exciting by using information about the highest academic qualification a person has earned. In this example, a total of 5.1% people found life 'dull', 44.2% 'routine' and 50.8% 'exciting'. If one predicts 'exciting' (which is of highest frequency) for everyone, the

Table 7a. Number of road segments appeared in different scales for different orders

Orders in segment	1:1000	1:5000	1:20 000	1:50 000	1:100 000	1:200 000
0,0	3	3	3	0	0	0
0,1	12	11	11	10	6	1
0,2	49	47	47	21	11	4
0,3	3	3	3	2	1	0
0,4	0	–	–	–	–	–
0,5	2	2	2	2	2	1
1,1	23	23	23	21	21	18
1,2	52	52	52	43	33	23
1,3	6	6	6	6	5	3
1,4	4	4	4	4	3	3
1,5	1	1	1	1	1	0
2,2	97	95	95	48	27	12
2,3	19	19	19	15	9	1
2,4	2	2	2	2	2	2
2,5	2	2	2	2	2	2
3,3	3	3	3	2	0	0
3,4	1	1	1	1	1	1
3,5	0	–	–	–	–	–
4,4	0	–	–	–	–	–
4,5	1	1	1	1	1	1
5,5	0	–	–	–	–	–

Table 7b. Number of road segments in different scales categorized by their orders (after reclassified)

	1:1000	1:5000	1:20 000	1:50 000	1:100 000	1:200 000
0,0	3 (100.0%)	3 (100.0%)	3 (100.0%)	0	0	0
X,0	66 (100.0%)	66 (100.0%)	63 (95.5%)	35 (53.0%)	20 (30.3%)	6 (9.1%)
1,1	23 (100.0%)	23 (100.0%)	23 (100.0%)	21 (91.3%)	21 (91.3%)	18 (78.3%)
X,1	63 (100.0%)	63 (100.0%)	63 (100.0%)	54 (85.7%)	42 (66.7%)	29 (46.0%)
X,X	125 (100.0%)	125 (100.0%)	123 (98.4%)	71 (56.8%)	42 (33.6%)	18 (14.4%)

Table 8. Highest degree (academic qualification) received and perception of life

			LIFE			Total
			Dull	Routine	Exciting	
DEGREE (qualification)	High School	Count	35	251	231	517
		Row %	6.8%	48.5%	44.7%	100%
	Bachelor	Count	2	58	97	157
		Row %	1.3%	36.9%	61.8%	100%
	Graduate	Count	1	21	51	73
		Row %	1.4%	28.8%	69.9%	100%
Total		Count	38	330	379	747
		Row %	5.1%	44.2%	50.8%	100%

error for this prediction (i.e. 'misclassified in situation 1') is the total number of misclassified, that is, 368 (i.e. 38+330). However, if one's highest qualification is known, the error in prediction (i.e. 'misclassified in situation 2') could be reduced. One could simply predict 'routine' (which is of highest frequency) for high school graduate and 'exciting' (which is of highest frequency) for both bachelor's and higher degree holders, then the total errors would be

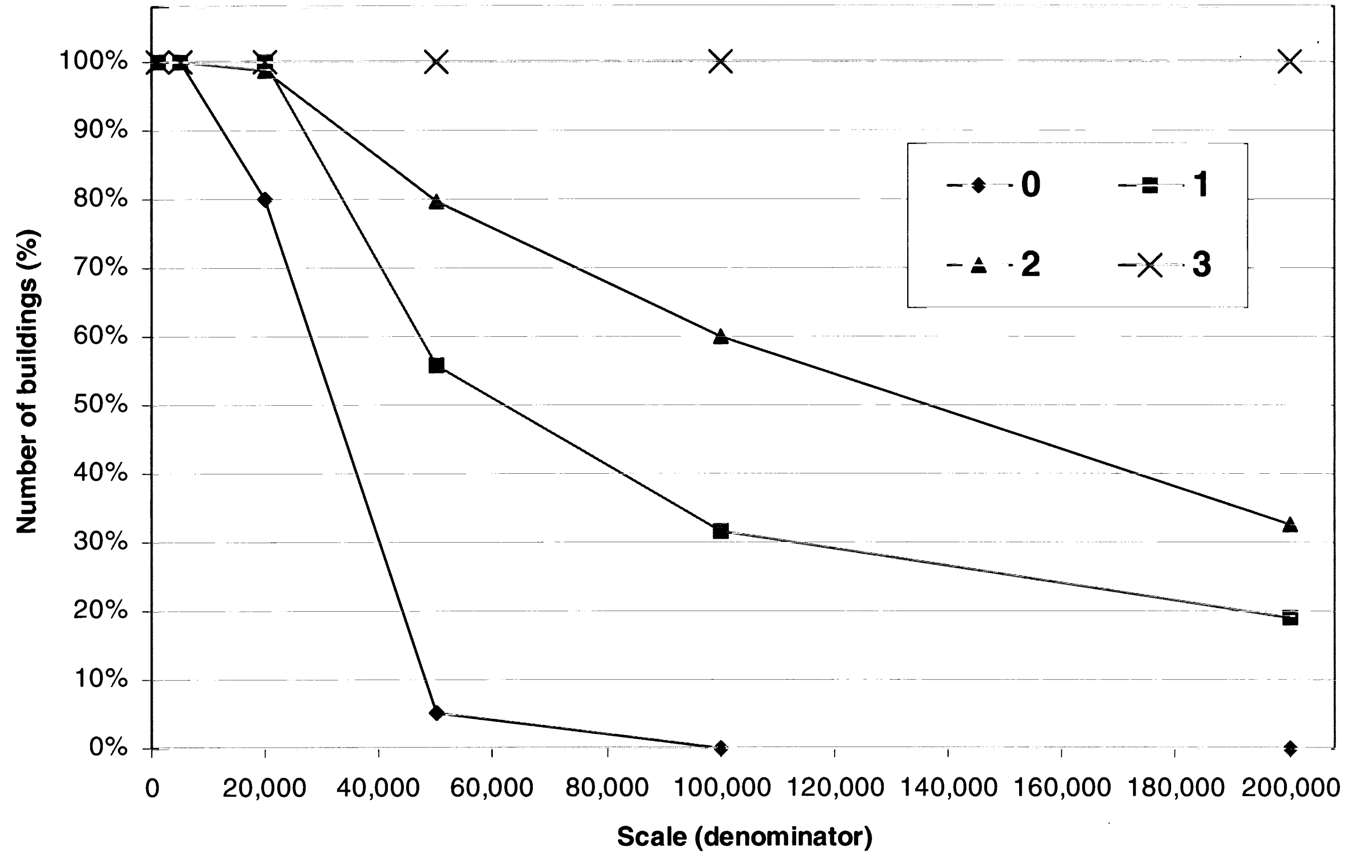

Figure 7. Retention of road segments (in terms of percentage) at different scales for traffic ways 0 to 3

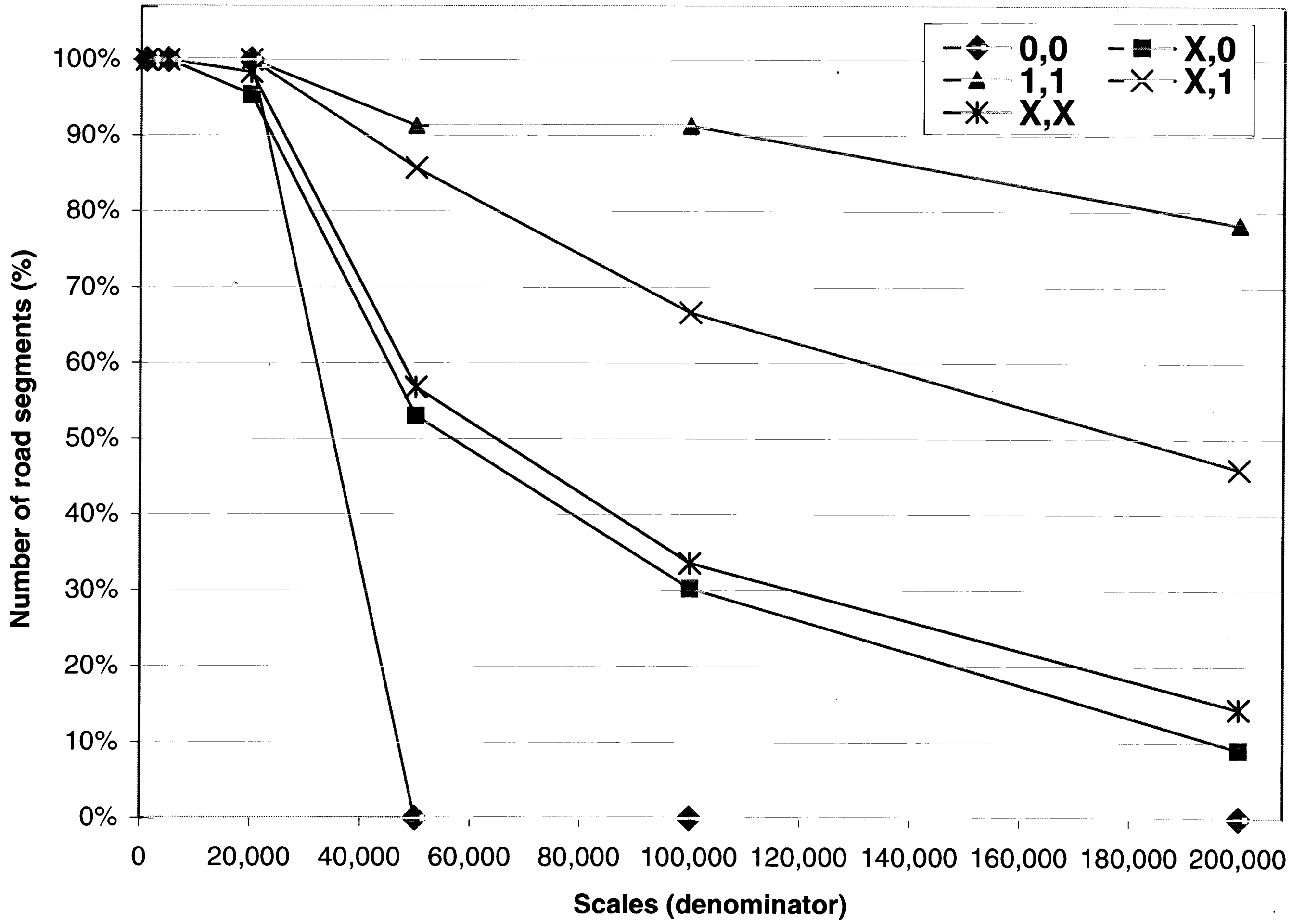

Figure 8. Elimination of road segments at different scales for each connectivity type

266 (i.e. 35+231)+60 (i.e. 2+58) and 22 (i.e. 1+21). Comparing these two errors, the improvement is

$$\lambda = \frac{368 - 348}{368} = \frac{20}{368} = 0.0543 = 5.43\%$$

That is, by knowing the highest degree a person has earned, the error is reduced by 5.4%. Therefore, a person's perception of life is somehow dependent on the highest degree s/he holds.

A value of 0 for Lambda means that the independent variable is of no help in predicting the dependent variable. When two variables are statistically independent, the value of Lambda is 0. However, a Lambda of 0 does not necessarily imply statistical independence.

Somers' Delta (Δ) is a kind of ordinal measure of association that is based on the difference between the number of concordant pairs and the number of discordant pairs. It is defined as follows:

$$\Delta = \frac{\text{Number of concordant pairs} - \text{Number of discordant pairs}}{\text{Sum of all pairs of cases that are not tied on the independent variable}} \quad (2)$$

where concordant pairs, discordant pairs and all other pairs are explained by the example given in Table 9, which contains a cross tabulation of the values of 'life' (the view of life variable) and 'degree' (the education variable) for three cases.

Table 9. Values of variable life and degree

	Life	Degree
Case 1	1	2
Case 2	2	3
Case 3	3	2

In the situation of Cases 1 and 2, both values (Life and Degree) at Case 2 are larger than those at Case 1. Such a pair of cases is referred to as being 'Concordant'. In contrast, in the situation of Cases 2 and 3, the value of 'Life' in Case 3 is larger than that in Case 2, but the values for 'Degree' are reversed. Such a pair is called 'Discordant'. If the value for one variable is the same, the pair is referred to as being 'tied'. In summary, there are five possible combinations, i.e. concordant, discordant, tied on the first variable, tied on the second variable and tied on both variables. If most of the pairs are concordant, the association between the two variables is said to be positive. If concordant and discordant pairs are equally likely, it can be concluded that there is no association between these two variables. Thus '+1' indicates a perfect positive relationship, '−1' indicates a perfect negative relationship, and '0' indicates no relationship.

To carry out the statistical analysis of the topographic maps, the observed data are categorized into six different ranked levels, i.e. six feature vanishing levels (FVL). As the features to be studied in this project are road features, a more specific term, i.e. road elimination levels (REL), is used. The categorization is as follows:

REL 1: Features eliminated at scale 1:5000
REL 2: Features eliminated at scale 1:20 000
REL 3: Features eliminated at scale 1:50 000
REL 4: Features eliminated at scale 1:100 000
REL 5: Features eliminated at scale 1:200 000
REL 6: Features retained at all scales

Two statistical parameters are worthy of particular attention. The first one is the value of 'FVL Dependency', indicating the strength of the association measurement. The larger the value, the stronger the association. The second one is the 'Approximation Significance' value, showing the necessity to reject the null hypothesis of the statistical tests, i.e. 'FVL being independent of the tested variable'. The smaller the value, the more strongly one should reject the null hypothesis.

With respect to the association of FVL with road attributes, the relevant data are recorded in Tables 10–15. The Somers' Δ and Lambda tests were carried out, and the results are shown in Table 16. It can be seen that the value of the approximate significance is less than 0.0005, so the null hypothesis, i.e. 'FVL is independent of road attributes', is strongly rejected. The strength of the association is indicated by either Delta or Lambda values. In this test, the results show that all the associations are positive. However, the values range from 0.733 to 0.193, from strong to weak. More precisely, the dependency of road elimination on these attributes is in the following order: type (0.73), length (0.53), number of lanes (0.48), number of traffic directions (0.41), width (0.36) and connectivity (0.19).

The result in graphic form is as shown in Figure 9. However, it should be noted that Δ and λ have different meanings and therefore the values may not be comparable.

POTENTIAL APPLICATIONS IN MAP GENERALIZATION

After the dependency of road elimination on the road attributes is found, the next step should be to formalize rule from these for the implementation of map generalization systems. However, this lies outside of this paper and therefore only a brief discussion on the potential application is given.

One way of implementing such systems is to assign an overall weight to each road segment individually, so that those segments with larger weights can be retained. For

Table 10. Number of road segments at different FVL in different portions of length

Road length (l) in kilometers	Feature Vanishing Levels (FVL)						Total
	1	2	3	4	5	6	
$0<l\leq0.1$	0	5	40	14	2	1	62
$0.1<l\leq0.2$	0	0	36	11	8	10	65
$0.2<l\leq0.3$	0	1	11	12	9	2	35
$0.3<l\leq0.4$	0	0	2	6	3	3	14
$0.4<l\leq0.5$	0	0	3	3	6	3	15
$0.5<l\leq0.6$	0	0	1	3	7	4	15
$0.6<l\leq0.7$	0	0	0	2	3	5	10
$0.7<l\leq0.8$	0	0	0	0	2	0	2
$0.8<l\leq0.9$	0	0	0	0	3	5	8
$0.9<l\leq1.0$	0	0	0	1	1	4	6
$1.0<l\leq1.1$	0	0	0	0	1	0	1
$1.2<l\leq1.3$	0	0	0	0	1	4	5
$1.3<l\leq1.4$	0	0	0	2	0	2	4
$1.4<l\leq1.5$	0	0	0	0	1	2	3
$1.5<l\leq1.6$	0	0	0	0	0	2	2
$1.6<l\leq1.7$	0	0	0	2	2	3	7
$1.7<l\leq1.8$	0	0	0	0	1	1	2
$1.8<l\leq1.9$	0	0	0	0	1	2	3
$1.9<l\leq2.0$	0	0	0	0	1	1	2
$2.2<l\leq2.3$	0	0	0	0	0	3	3
$2.3<l\leq2.4$	0	0	0	0	1	1	2
$2.4<l\leq2.5$	0	0	0	0	0	2	2
$2.6<l\leq2.7$	0	0	0	0	0	3	3
$3.4<l\leq3.5$	0	0	0	0	1	0	1
$3.5<l\leq3.6$	0	0	0	0	0	2	2
$3.6<l\leq3.7$	0	0	0	0	0	1	1
$3.7<l\leq3.8$	0	0	0	0	0	1	1
$4.3<l\leq4.4$	0	0	0	0	0	1	1
$6.9<l\leq7.0$	0	0	0	0	0	2	2
$8.4<l\leq8.5$	0	0	0	0	0	1	1
Total	0	6	93	56	54	71	280

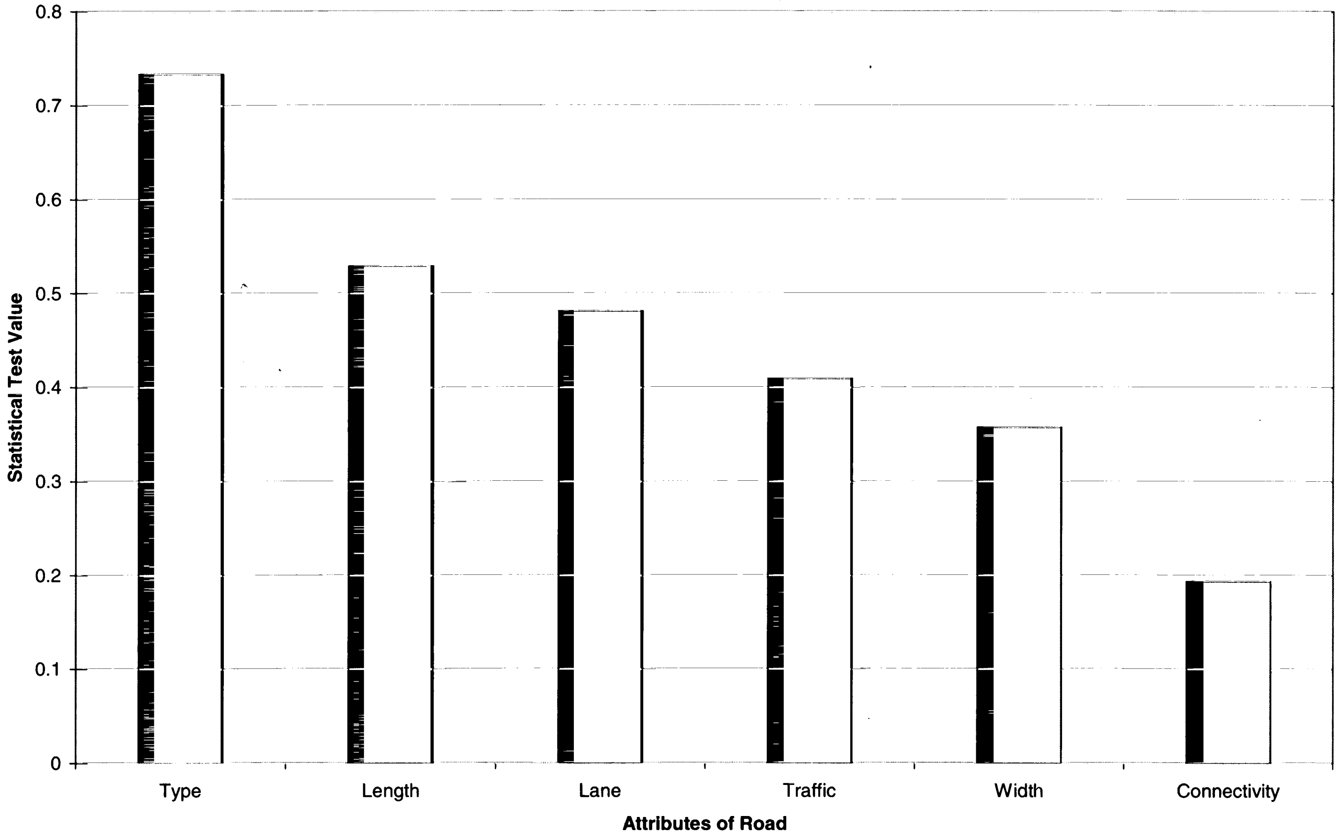

Figure 9. Dependency of the change in road representation on road attributes: Statistical results

Table 11. Number of road segments at different FVL in different portions of width

Road width (*w*) in meters	Feature Vanishing Levels (FVL)						Total
	1	2	3	4	5	6	
$0<w\leq 5$	0	5	17	7	16	3	48
$5<w\leq 10$	0	1	64	28	26	20	139
$10<w\leq 15$	0	0	10	13	7	21	51
$15<w\leq 20$	0	0	2	8	5	8	23
$20<w\leq 25$	0	0	0	0	0	8	8
$25<w\leq 30$	0	0	0	0	0	4	4
$30<w\leq 35$	0	0	0	0	0	2	2
$35<w\leq 40$	0	0	0	0	0	3	3
$45<w\leq 50$	0	0	0	0	0	1	1
$50<w\leq 55$	0	0	0	0	0	1	1
Total	0	6	93	56	54	71	280

Table 12. Number of road segments at different FVL for different connectivity of road

Connectivity	Feature Vanishing Levels (FVL)						Total
	1	2	3	4	5	6	
0,0	0	0	2	0	0	0	2
X,0	0	4	28	15	14	6	67
1,1	0	0	2	0	3	18	23
X,1	0	0	9	12	13	29	63
X,X	0	2	52	29	24	18	125
Total	0	6	93	56	54	71	280

Table 13. Number of road segments at different FVL for different types of road

Type	Feature Vanishing Levels.(FVL)						Total
	1	2	3	4	5	6	
Non-transport use	0	0	2	0	0	0	2
Lane	0	4	9	0	0	0	13
Local Distributor	0	2	77	51	41	7	178
District Distributor	0	0	5	4	10	29	48
Primary Distributor	0	0	0	1	3	27	31
Trunk Road	0	0	0	0	0	8	8
Total	0	6	93	56	54	71	280

Table 14. Number of road segments at different FVL for different number of traffic way

Number of Traffic way	Feature Vanishing Levels (FVL)						Total
	1	2	3	4	5	6	
0	0	4	15	1	0	0	20
1	0	0	49	27	14	21	111
2	0	2	29	28	40	48	147
3	0	0	0	0	0	2	2
Total	0	6	93	56	54	71	280

Table 15. Number of road segments at different FVL for different number of lane

Number of lane	Feature Vanishing Levels (FVL)						Total
	1	2	3	4	5	6	
1	0	1	57	9	14	5	86
2	0	1	17	30	28	30	106
3	0	0	2	7	5	9	23
4	0	0	1	9	7	8	25
5	0	0	1	0	0	7	8
6	0	0	0	0	0	7	7
7	0	0	0	0	0	1	1
8	0	0	0	0	0	2	2
9	0	0	0	0	0	1	1
11	0	0	0	0	0	1	1
Total	0	2	78	55	54	71	260

example, it was discovered that only 28.9% of local distributors will be retained from 1:50 000 to 1:100 000. In this case, the top 28.9% in the overall weight list may be retained and the rest eliminated, merged or combined.

Table 16. Dependence of FVL on road attributes by Somers' Δ and Lamda tests

Tests	Value	Approximate Significance
FVL Dependency on road length	$\Delta = 0.529$	0.0005
FVL Dependency on road width	$\Delta = 0.357$	0.0005
FVL Dependency on road type	$\Delta = 0.733$	0.0005
FVL Dependency on number of traffic ways	$\Delta = 0.409$	0.0005
FVL Dependency on number of lanes	$\Delta = 0.481$	0.0005
FVL Dependency on connectivity	$\lambda = 0.193$	0.0005

In the determination of overall weight, one or more attributes may be used. For example, one may wish to make use of only 'type' and 'number of lanes'. Also, there could be a higher correlation between 'number of lanes' and 'width', therefore, only one of them may be used in the computation of weights.

DISCUSSION AND CONCLUSIONS

In the study, topographic maps at six different scales, ranging from 1:1000 to 1:200 000, were selected for the study. Through observations and analyses, some empirical facts were discovered about how roads were represented on maps at different scales. The changes in the representation of road features at different scales can be described as follows:

- In the 1:1000 and 1:5000 maps, roads are represented in their true dimensions by curb lines. The process of selective omission of road features starts at scale 1:50 000. The number of road segments remains unchanged until scale 1:50 000; it has been reduced by 35% at that scale. The number of road features then continues to decrease: about 46% of road features are retained at scale 1:100 000, and only about 25% at scale 1:200 000.
- At 1:20 000, roads are classified and symbolized into five classes: Expressway, Main Road, Secondary Road (which includes five subclasses), Elevated Road and Road Under Construction.
- At 1:50 000, the expressways and main roads are grouped together, the double width and single width secondary roads are classified as secondary roads, the unsurfaced secondary roads are independent and form a new class, and the lanes and non-motorable roads are erased.
- At 1:100 000 and 1:200 000, the unsurfaced secondary roads are erased, and only the main and secondary roads remain.

These are general statements. Their usefulness for rule formalization is limited. To obtain useful information, quantitative analysis is employed. In this study, the findings were quantified into several levels, termed 'Feature Vanishing Levels', or FVL for short. Statistical analysis was conducted on the association of the change of the representation of road features with road attributes. Six attributes were used, i.e. type, length, number of lanes, number of traffic directions, width and connectivity. Two statistical tests were employed, i.e. Somers' Delta and Lambda.

The attribute 'Road Type' is the most important factor for the selection of road features to be retained or erased during scale reduction, whereas 'Road Length', 'Number of Lanes', 'Number of Traffic Ways' and 'Road Width' are similar in significance and, surprisingly, 'Connectivity' is the least important factor. As shown in Table 16, the value of Somers' Delta of 'Road Type' is 0.733, which is outstanding (the greatest) from all the others, while the value of Goodman–Kruskal's Lambda for 'Connectivity' is 0.193, and the values for the others are from 0.3 to 0.5.

If these numbers are normalized into percentages, then these values become: type (27%), length (20%), number of lanes (18%), number of traffic directions (15%), width (13%) and connectivity (7%). Such results could then be used to formulate an overall weight in determining whether a particular road should be deleted, merged or combined, in the generalization process, in order to retain a certain percentage of road at a smaller scale.

It must be noted that this is a first attempt to measure the dependency of changes in the representation of road features on road attributes. This work certainly has limitations. It is recommended that some other attributes be considered and other statistical measures explored. The next step is to make use of these findings for rule formation for the selection/elimination process in map generalization.

ACKNOWLEDGEMENTS

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. PolyU 5073/01E) and the Hong Kong Polytechnic University (350/760). The authors would also like to thank the Lands Department of the Hong Kong SAR Government for providing the digital map data for this study.

REFERENCES

Buttenfield, B. P., and McMaster, R. B. (eds) (1991). **Map Generalization: Making Rules for Knowledge Representation**, Longman Scientific & Technical: Essex, UK.

Leitner, M., and Buttenfield, B. P. (1995). 'Acquisition of Procedural Cartographic Knowledge by Reverse Engineering', **Cartography and Geographic Information Systems**, 22(3), 232–41.

Müller, J. C. (1990). 'Rule based generalisation: potentials and impediments', **Proceedings of the 4th International Symposium on Spatial Data Handling, Zurich**, 1, 317–34.

Norusis, M. J. (1998) **SPSS 8.0 Guide to data analysis**, Prentice-Hall, Inc.

Töpfer, F., and Pillewizer W. (1966). 'The principles of selection', ***The* Cartographic Journal**, 3(1), 10–16.

Weibel, R. (1995). 'Three essential building blocks for automated generalization', in **GIS and Generalization**, 56–69, ed. by Müller, J.-C., Lagrange, J-P., and Weibel, R. Taylor & Francis, London.

Yu, Z. (1993). **The effects of scale change on map structure**, Doctoral Thesis, Department of Geography, Clark University.