Classifying urban structures for mapping purposes using discriminant analysis

Stefan Steiniger

Department of Geography, University of Zurich, CH-8057 Zurich, Switzerland
Tel. +41-44-6355252 Fax +41-44-6356848
sstein@geo.unizh.ch

KEYWORDS: urban area classification, discriminant analysis, map generalisation, visualisation

1. Introduction

Topographic maps at medium scales (1:25’000 - 1:100’000) and derived maps for urban planning often emphasize urban structures. The visualisation of urban area structures, such as inner city districts or industrial sites, should support on one hand map reading and on the other hand initial decision making processes in planning. For instance, the German topographic regional map (Topographische Gebietskarte, 1:100’000) visualises different settlement densities by colour tints. Our aim is to identify different built-up area types using pattern recognition techniques, based solely on building geometry.
With respect to scale, different kinds of urban structures have been investigated, ranging from the classification of different cities (Frenkel, 2004) and identification of settlement types such as towns, villages or hamlets (Boffet, 2001) to the recognition of building alignments and building clusters (Regnauld, 1996; Anders, 2003).

2. Objective and Research Questions

Similarly to Barr *et al.* (2004) we intend to classify urban land-use, starting off from topographic map data. The aim of Barr *et al.* was to show that a mapping between form and function exists. We want to extend their objective aiming to detect specific built-up area types. Based on the analysis of several topographic maps (Swiss, French, German) and with respect to their usefulness for further GIS analysis purposes we decided to specify five types of built-up area: industrial and commercial areas, inner city, urban areas (dense buildings), suburban areas (loosely spaced buildings) and rural area (individual buildings).
Barr *et al.* (2004) described the built-up areas by the morphologic properties area and compactness as well as proximity relations, realized by a Gabriel Graph. We aim to extend their set of morphologic properties, using vector instead of raster based measures and further establish proximity relations by buffering operations instead of using a graph structure. Since the morphologic measures and the buffering measures will result in geometric attribute values (i.e. properties) for every building we can apply pattern classification approaches in feature space. Buildings with similar properties will be close to each other or even form groups in feature space.
Several research questions emerge based on our objective to classify buildings into five structure types with respect to the condition of using classification approaches. The main questions are:

- Which variables and measures describe urban morphology sufficient for our purposes?
- Does a (2D or 3D) representation of the variables exist to support visual analysis of urban structures, e.g. to evaluate the separability of built-up area types in feature space?
- Which classification approach can be applied (discriminant analysis or cluster analysis)?
- Which classification algorithms show good performance?
- Are built-up area patterns from different regions or countries similar to such a degree that we need only one initial definition of prototype buildings or urban structure, respectively, for every structure type? Or do we need sample data for every region?

The next section will address the first three questions, and describe the basic approach to classify the building dataset.

3. Defining Morphologic Measures and Classifying Buildings

3.1 Characterizing Urban Structures

Our assumption is that the measures and classification can be based on morphologic information, rules of perception and subsequently Gestalt theory. In particular, Wertheimer (1923) developed a list of *laws of organisation in perceptual form*. He identified approximately nine 'laws' describing perceptual conditions which allow a human to perceive groups rather than individual objects. Two of them can obviously be applied to define urban area types. The 'law of proximity' describes that object distances among individuals of a group will be smaller than to objects not part of that group. Hence, this law proposes density or distance measures to identify the built-up area structures. We developed three different density indices based on buffer operations for the classification. The second law identifies similarity among the group individuals. Similarity can be defined with respect to four visual variables: colour (object category), size, shape and orientation. The variables colour or category, respectively, and orientation are not very useful in our case (e.g. all industrial buildings will not be aligned to north). But we can use measures of size and shape to extend the set of density measures. Therefore we defined six additional measures: building area, number of corner points, building elongation, the Shape Index from Schumm (McEachren, 1985), building wall squareness, and number of courtyards.

3.2 Searching a 3-D and 2-D Representation for Initial Analysis

After defining the measures two problems appear. First, we have no information if the selected measures are sufficient to separate the five types. Second, we do not know whether a clustering approach or a discriminant analysis approach is best applied. The clustering approach groups objects together which are close to each other in feature space. The discriminant approach divides the feature space into regions corresponding to representative samples of the built-up area types.

To solve these problems a transformation from 9-D feature space into a 3-D and 2-D artificial feature space has been developed with the condition of a low loss of information (see Figure 1). Therefore the transformation parameters are obtained from a Principal Components Analysis (9-D to 3-D transformation) and a plane projection (3-D to 2-D) of an initial sample set of buildings.

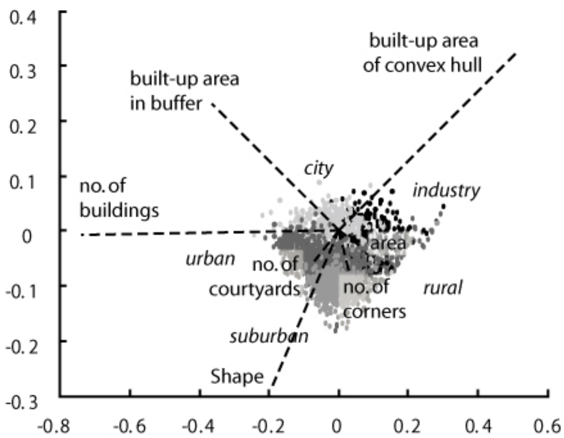

Figure 1. The sample data transformed from 9-D into an artificial 2-D feature space. The lines indicate the direction in which changes of measure values act (e.g. large and small values). The line lengths indicate the weighting of the variables. Two measures – elongation and squareness – are not used for the transformation.

3.2 Classification by Discriminant Analysis Techniques

After application of the transformation to an initial set of sample buildings from 9-D to 3-D, then 2-D the visualisation revealed that buildings of one built-up area type are loosely distributed in the normalized feature space, not forming a cluster (see Figures 1 and 3). Hence, the conclusion is that discriminant analysis techniques have to be applied to detect the built-up area type of a building.

To classify the buildings by defining the regions in feature space for every type we basically have two options. Either we define the type regions in 2-D manually (e.g. by simply drawing an enclosing polygon) or we use machine learning algorithms to detect the borders between the types. We have

Data set	*Cohen's Kappa*	*Total Accuracy* (%)	*Possible improvement by median filter* (%)
Zürich (25k)	0.65	75	+10
Southampton (2.5-10k)	0.52	67	+3-7
Southampton (generalised, 25k)	0.35	57	---

Table 1. Classification results for buffer size of 50m.

realized four different classification techniques: a Batch Perceptron algorithm (see Duda *et al.*, 2000), a Minimum Squared Error algorithm based on Pseudo Inverse, AdaBoost with decision stumps (Schapire, 1999) and a Support Vector Machine (SVM) called SVM^{light} (Joachims, 1999), which offers different approaches by different kernels. The first and second algorithm yield linear class borders in feature space, the latter two can calculate non-linear decision borders.

When classifying the buildings it is very likely that objects close to each other will be of the same built-up area type, that is, spatial auto-correlation appears. Since the classification algorithms can not deal with this additional information we have implemented a spatial median filter, based on buffer operations, to enhance the classification results.

4. Experiment and Results

A number of experiments have been performed using building data from the Swiss Canton of Zurich and OS MasterMap^{TM} data from British Ordnance Survey (OS). The Swiss data are vector data digitised from the topographic map of scale 1:25'000. We defined training and validation areas for the classification algorithms in the areas of Zurich and Southampton. Classification results, evaluated using the total accuracy and Cohen's Kappa index, are given in Table 1 and Figure 2. The dashed circles in Figure 2 mark areas with false classified buildings. From the right image can be inferred that identification of commercial and industrial buildings seems to be more complicate.

In an additional experiment we generalised a set of the Southampton data for the target scale of 1:25'000 - by building aggregation, simplification, elimination and enlargement - to evaluate if decision borders trained on Zurich buildings could be applied. The worse classification results indicate that the building structures for the scale 1:25'000 of Southampton and Zurich are significantly different.

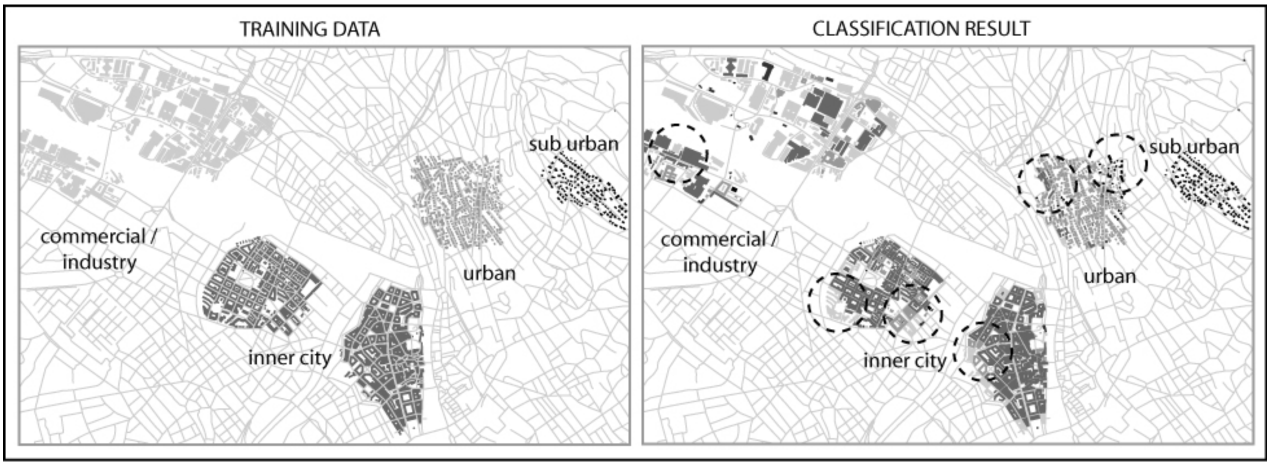

Figure 2. A part of the Zurich training data (left), where every group of buildings represent one urban structure type. Classification results for the same data are shown in the picture on the right side. Correctly classified buildings have the same grey value as in the left picture. Dashed circles mark false classified zones. Data: VECTOR 25, reproduced by permission of swisstopo (BA057008).

5. Discussion, Conclusions and Perspectives

We will now summarise the main findings with respect to the current state of research and address the remaining research questions. The different classification accuracy of about 10 percent between Swiss and British buildings, and even more following the median filter processing, could have several reasons. One may be that the selection of training areas was not optimal. A second, the measures used may have been inappropriate for British urban structures.

The developed transformation of the 9-D feature space into two and three dimensions for visualisation has proven in first tests to be a useful tool for structural analysis of urban morphology. But it is difficult to evaluate the explanatory power for Southampton data since the transformation has been developed based on Zurich samples. Comparing the 2-D visualization of the OS Data and the swisstopo data we can clearly see differences in position of buildings of the same type (Figure 3). This is further supported by the bad classification results of generalized Southampton data (25k) with Zurich type borders (25k). Thus, we infer that the urban morphology of the Southampton and Zurich region is so different that in general every region has to be classified based on its own sample set.

A final performance evaluation of the different classification algorithms is still in progress but current results with respect to the Zurich data indicate that none of the algorithms seems to perform significantly better than the others. For the OS data differences in accuracy are larger whereas a well parameterised Support Vector Machine shows best results. Further research will evaluate the influence of buffer size on structure type separability, correlation among the measures and the influence of map generalisation.

Exploration of detected built-up area types, assigned to every building, is a task of future research. The application areas are manifold. The enriched data could be used in map generation with areal depiction of built-up areas as indicated in the introduction. It could be further used in automated map generalization of topographic maps to support algorithm selection.

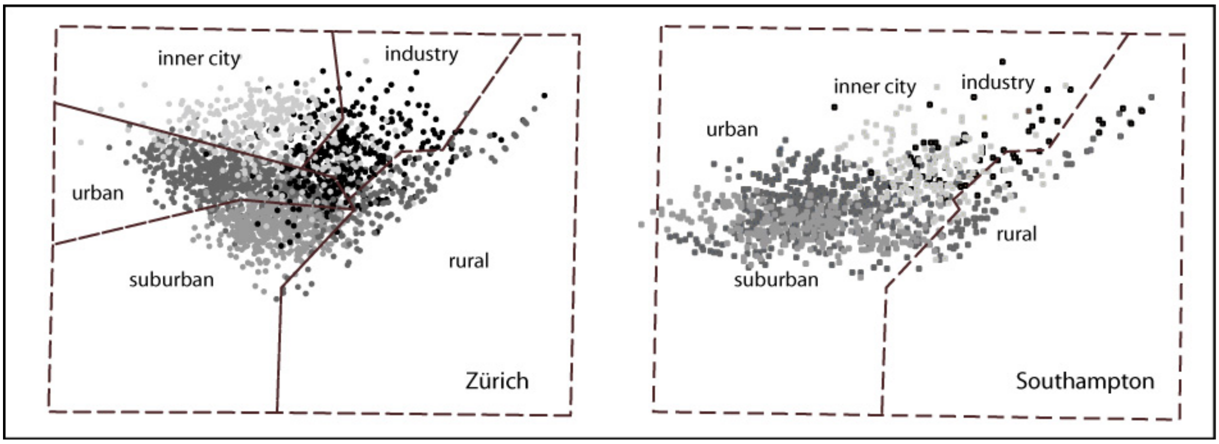

Figure 3. The Zurich (app. 2000, scale 25k) and generalized Southampton buildings (app. 1000, scale 25k) in the artificial 2-D feature space. The sample data of the same structure types have different positions and distributions for both cities. The structure types of Southampton overlap to a greater extent, which may explain the inferior classification results.

6. Acknowledgements

The research reported in this paper was funded by the Swiss NSF through grant no. 20-101798, project DEGEN. I especially like to thank Tillman Lange from the Machine Learning Group at ETH Zurich for providing code of classification algorithms and answering questions on the subject. Further I am grateful to the Canton of Zurich, swisstopo, and the Ordnance Survey for the support with maps and digital data.

References

Anders, K. H. (2003), A Hierarchical Graph-Clustering Approach to find Groups of Objects. *Proceedings 5'th ICA Workshop on Progress in Automated Map Generalization*, Paris (France).

Barr, S. L., Barnsley, M. J. and Steel, A., (2004), On the separability of urban land-use categories in fine spatial scale land-cover data using structural pattern recognition. *Environment and Planning B*, Vol. 31 (3), pp. 397-418.

Boffet, A. (2001). *Methode de creation d'information multi-niveaux : pour la generalisation cartographique de l'urban.* Ph.D. thesis, Université de Marne-la-Vallée.

Duda, R. O., Hart, P. E., and Stork, D. G., (2000), *Pattern Classification*. 2nd edition, John Wiley, New York.

Frenkel, A. (2004), Land-use patterns in the classification of cities: the Israel case. *Environment and Planning B*, Vol. 31 (5), pp. 711-730.

MacEachren, A. M. (1985), Compactness of geographic shape: comparison and evaluation of measures. *Geografiska Annaler* 67, pp. 53-67.

Regnauld, N. (1996). Recognition of Building Cluster for Generalization. In: M. J. Kraak and M. Molenaar (eds.). *Advances in GIS Research II* (Proceedings of 6th SDH'96 Delft), pp. 4B.1-4B.14.

Schapire, R. E. (1999), A brief introduction to boosting. *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*.

Joachims, T. (1999), Making large-Scale SVM Learning Practical. In: B. Schölkopf, C. Burges and A. Smola (eds.). *Advances in Kernel Methods - Support Vector Learning*, MIT-Press.

Wertheimer, M. (1923), Laws of organization in perceptual forms. English translation published in W. Ellis (ed.), 1938. *A source book of Gestalt psychology*. Routledge & Kegan Paul, London, pp. 71-88.

Biography

Stefan Steiniger is a Ph.D. student within the GIS Division, Department of Geography, University of Zurich. He obtained a MSc in Geodesy from the Technical University Dresden. His research focuses on map generalization and more specifically on generalization algorithms, signal processing, process modelling and classification.