Classifying Buildings Automatically: A Methodology

Masroor Hussain; Chris Davies; Robert Barr
School of Environment and Development
University of Manchester, UK

Introduction

In this paper, a pattern recognition based methodology is proposed to automatically classify the buildings by using their morphological properties (i.e. area, perimeter, orientation) and topological spatial relationships (i.e. Adjacency, Neighbourhood). Recognition, analysis and classification of urban structure is an important issue in urban land use modelling (Batty 1992). In order to understand overall urban morphology and land use classification, importance of calculating the structural and behavioural properties of built up land parcels is well studied (Batty 1992; Batty and Longley 1994). Most of the pattern recognition methodologies adopted and applied in the past, to ascertain urban land use and its structural identification, are based on remotely sensed data and driven by image processing techniques (Barnsley, Moller-Jensen et al. 2001). Researchers have applied both: (a) visual interpretation and (b) automated methodologies to extract the built up features and also to infer their uses/functions (Bauer and Steinnocher 2001). One of the fundamental weaknesses of such approaches is that it is difficult to extract land parcel functional information solely from the identification of land surface characteristics and it requires ancillary data to improve the classification (Barnsley and Barr 1997; Barnsley, Moller-Jensen et al. 2001; Bauer and Steinnocher 2001). In the UK there has been a substantial research towards creation of national level land use database at Government level. For example; national land use classification (NLUC) 1975, The land use change statistics (LUCS) early 1980, national land use stock survey during early 1980, and national land use database (NLUD) (Harrison 2000; Tompkinson, Morton et al. 2004). The studies to assess the feasibility of a National Land Use Stock Survey (e.g. Roger Tym and Partners, 1985 and Dun and Harrison, 1992) commissioned by former Office of Deputy Prime Minister (ODPM) currently Department of Communities and Local Government (DCLG) concluded that land use should be calculated and maintained using large scale digital mapping (Harrison 2002; Wyatt 2002). The results of the research work done based on these recommendations have figured out that there is a need for methods to automatically classify built up land parcels (Harrison 2002; Tompkinson, Morton et al. 2004; Wyatt 2004 ; ODPM 2006.). Therefore this paper presents a methodology that uses pattern recognition techniques for automatically classification of buildings.

Problem Definition

The pattern recognition techniques used for identifying the built up land parcels, computing their characteristics and inferring urban land use, utilises human intuition of map or image interpretation (Barnsley, Moller-Jensen et al. 2001; Bauer and Steinnocher 2001). Given a map or satellite image, humans intuitively study the spatial context of topological features (e.g. adjacency, distance, orientation) within overall environment along with the morphological properties (e.g. size, perimeter, shape) and infer the function of features (Barnsley, Moller-Jensen et al. 2001; Tompkinson, Morton et al. 2004). The obvious disadvantage or weakness of this approach lies in the fact that the task becomes manually intensive when applied to anything larger than the local scale. If intuition of this nature could be modelled along the lines of how maps or satellite images are interpreted, it would become possible to classify buildings automatically (Barnsley, Moller-Jensen et al. 2001). The objective of this part of research is to automate the classification of buildings using

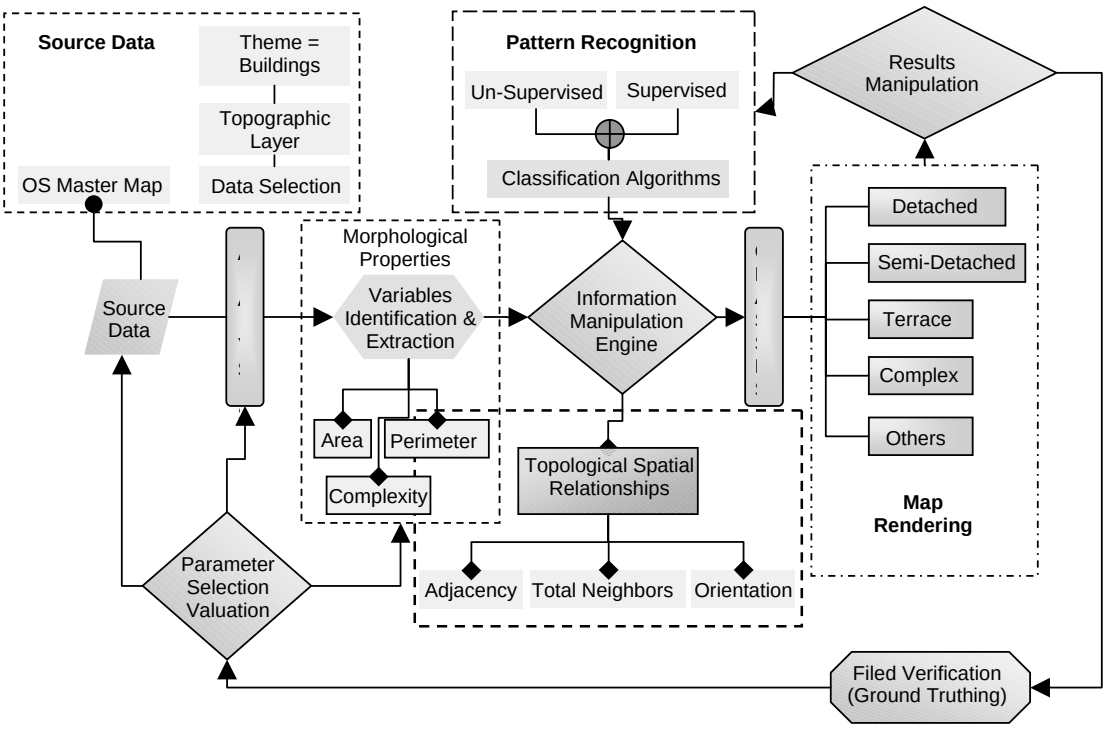
morphological properties and topological spatial relationships. This will then provide evidence for the examination of the emergent properties of building complexes and surrounding land.

Proposed Solution

In order to automate the classification of buildings, a methodology is proposed in this paper which applies the pattern recognition algorithms on morphological properties and topological spatial relationships for building features to infer their function (useage). The first step requires extracting and computing the morphological properties and topological spatial relationships.

For the purpose of this research, the UK has been chosen as study area. In the UK, all the topographic features both natural and man made (e.g. roads, land parcels, buildings etc) are captured, managed, maintained and updated in digital format by Ordnance Survey (OS). The digital product that provides the spatial and non-spatial information about real world features is called OS MasterMap (Harrison 2002; OS 2006; ODPM 2006.). Other countries use an equivalent for example Digital Cadastral Data Base (DCDB) in Australia. OS MasterMap has been used in this research to extract all the building features. The main advantage of OS MasterMap is that the building features are planimetrically accurate and can be selected as single theme. However one major disadvantage is that the theme based organization in OS MasterMap does not provide any information about the function of buildings but can be used as minimum mapping unit for building sub-classification (ODPM 2005; OS 2006; ODPM 2006.). Once the buildings are extracted from OS MasterMap, next step involves the computing the morphological properties and spatial relationship. Finally applying both supervised and un-supervised pattern recognition algorithms for segregating the buildings into separate classes. The results then rendered onto map and visual inspection performed and model calibrated. The need for calibration will be discussed in coming section. Figure 1 shows the proposed model in block diagram.

Figure 1: Proposed Buildings Classification Model



Implementation

A Geographical Information System (GIS) was used to compute the morphological properties and topological spatial relationships. Figure 2 describe the process involved in building feature extraction from OS MasterMap topographic layer. The extracted layer implicitly contains the some of morphological properties i.e. area and perimeter for each building. However, topological spatial relationships and complexity of the buildings needs to be computed explicitly. For both of the computations, ESRI ArcGIS package was used and some of programs were written in Visual Basic for Application (VBA) using ArcObjects (ESRI). Figure 3 explains the two topological spatial relationships that were computed i.e. "Adjacency" and "Total Neighbours". There is no prior information that which of the pattern recognition technique would be suitable for this application. Therefore, both *supervised* and *un-supervised* classification algorithms were used for segregating the buildings into appropriate groups. *K-Means* and *Self Organizing Maps (SOM)* are two algorithms that were used to cluster the buildings with similar characteristics by feeding in the morphological properties and topological spatial relationship information. Where as discriminant analysis was used for supervised classification. The discriminant analysis approach divided the buildings into classes corresponding to representative sample of the buildings function type.

Adjacency:	This gives the measure about how many buildings are sharing common border or side;
Total Neighbours:	Describe how many buildings are present in local cluster of buildings all sharing at least one side. This variable is helpful in identification of single building and group of buildings;

Figure 2: Extraction of Buildings from Topographic Database (Digital)

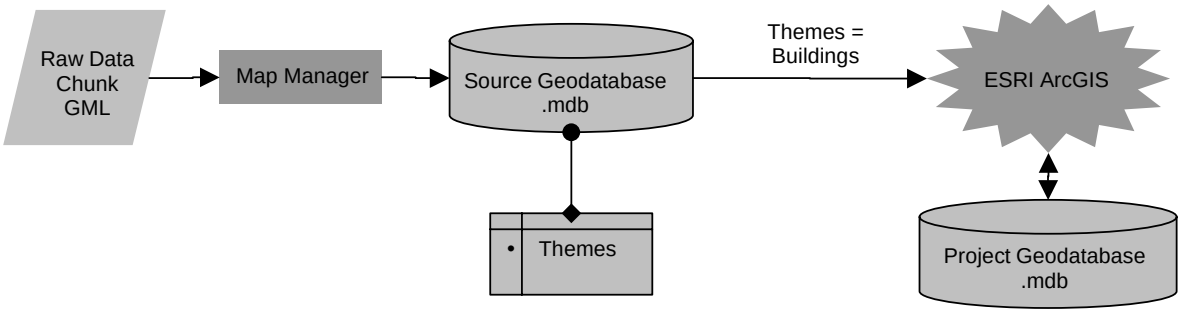
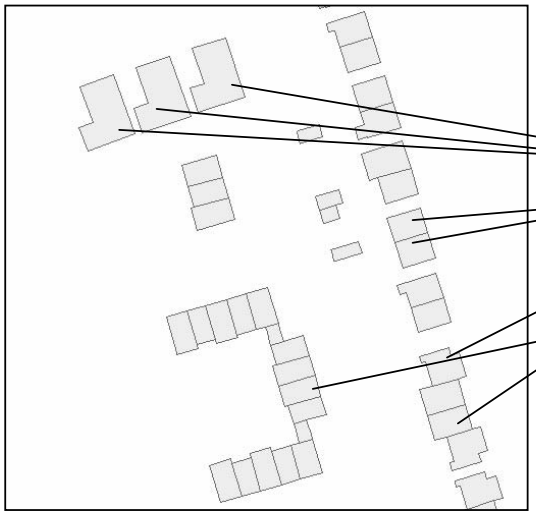


Figure 3: Explanation of Topological Spatial Relationship calculated



Adjacency	Neighbours
0	0
1	1
1	4 (>2)
2	4 (>2)
2	17 (>2)

The visual representation "*Adjacency" and "Total Neighbours"*, two topological spatial relationships that are used in proposed classification model.

Conclusion

This paper has presented a methodology for automating the classification of buildings. The proposed solution suggested deploys a supervised and un-supervised classification technique by using morphological properties of each building feature and topological spatial relationships.

The automated classification approach proposed in the model proves to be feasible for deriving the function of building features (Masroor 2007). The essential step in this process is the adjustment of the number of parameters and clusters for un-supervised algorithms (K-Means and SOM) and more accurate prior classification of training dataset for supervised algorithms (Discriminant). By testing different configurations of morphological characteristics and spatial relationship it is possible to populate a rule base for building classification. The same rule based would be applied in next step to automatically segregate all the buildings in appropriate groups. The main types of housing are picked up quite nicely i.e. Terraced, Mid-Terraced, Semi-Detached and Detaches housing. However a building type that caused major problem were garages. It was concluded that it is necessary to identify garages and isolate then before applying algorithms to identify any other building type. *Address Point* is another dataset within suite of OS product and holds key information. Address Point is set of points representing the location where post is delivered. It is assumed that no mail is delivered to a garage. Therefore it is possible to identify and isolate garages by integrating this information with existing one. Initial verification of the results produced by the model is based on the visual interpretation of the MasterMap with an overlay of AddressPoint layer. However, to further test the accuracy of the out puts (classification) of the model, site visits are needed to be arranged to compare the out puts with the reality. Also, further tests are planed to be carried at final stage of research (Masroor 2007) to gage the sensitivity of method to classify similar cluster groups.

It can be seen that automation of buildings classification process can be achieved using a Pattern Recognition Methodology. The methodology has proven to be flexible, adaptable and allows some level of evolution of the approach. However the adoption and evolutionary processes maintain the integrity of the fundamental concept of buildings classification.

References

Barnsley, M. J. and S. L. Barr (1997). "Distinguishing urban land-use categories in fine spatial resolution land-cover data using a graph-based, structural pattern recognition system." Computer, Environment and Urban Systems **21**(3/4): 209-225.

Barnsley, M. J., L. Moller-Jensen, et al. (2001). Inferring Urban Land Use by Spatial and Structural Pattern Recognition. Remote Sensing and Urban Analysis. J.-P. Donnay, M. J. Barnsley and P. A. Longley, Taylor and Francis**:** 115-144.

Batty, M. (1992). "Urban modeling in computer-graphic and geographic information system environments." Environment and Planning B **19**(6): 663-688.

Batty, M. and P. A. Longley (1994). Fractal cities : a geometry of form and function London : Academic Press.

Bauer, T. and K. Steinnocher (2001). "Pre-Parcel Land Use Classification in Urban Areas Applying a Rule-based Technique." GeoBIT/GIS **6**: 24-27.

Harrison, A. (2002). Extending the dimensionality of OS MasterMap™: land use and land cover. the agi conference at GIS.

Harrison, A. R. (2000). The National Land Use Database: developing a framework for spatial referencing and classification of land use features. AGI Conference at GIS, Olympia.

Masroor, H. (2007). A Framework for the Analysis of Urban Change; An Automated Geography Approach. PhD (In Progress), University of Manchester.

ODPM (2005). Generalised Land Use Database Statistics for England. DCLG**:** 11.

ODPM (2006.). National Land Use Database: Land Use and Land Cover Classification. DCLG.

OS (2006). OS MasterMap® user guide: Part 1. Southhampton, Ordnance Survey**:** 206.

Tompkinson, W., D. Morton, et al. (2004). An optimised semi-automated methodology for populating a. national land use dataset. GISRUK. University of East Anglia.

Wyatt, P. (2002). Creation of an Urban Land Database, Education Trust of the Royal Institution of Chartered Surveyors**:** 51.

Wyatt, P. (2004). "Constructing a land-use data set from public domain information in England." Planning Practice and Research **19**(2): 147 - 171.