Housing stock surveys in GIS systems using pattern recognition

S. C. Lannon, D. K. Alexander, P. J. Jones

Welsh School of Architecture
Cardiff University
Bute Building
Cardiff, CF10 3NB
Telephone: +44 2920874000
Fax: +44 2920874623
http://www.cardiff.ac.uk/archi/

1. Introduction

The Energy and Environment Prediction (EEP) model (Jones, Williams and Lannon, 2000) has proven valuable to its users. The EEP model is a computer based modelling framework that quantifies energy use and associated emissions for cities to help plan to reduce carbon dioxide and other emissions. This model has shown that it can produce considerable financial, energy, and Carbon savings through the targeting of energy efficiency measures appropriate to certain house types. It also allows the user, e.g. a local authority, to target those areas with the greatest need.

Regional energy models, such as EEP, require an accurate description of the built characteristic of an area (e.g. the profile of housing type, age, etc.). The survey methods developed in the original EEP project to provide this information required a laborious and expensive 'walk by' survey. When this method was applied to the county of Neath Port Talbot in the United Kingdom, where a total of 55,000 dwellings were surveyed (nearly 100% of the population), it required an investment of 18 man-months. Although the survey provided a level of information that was otherwise unobtainable, an investment in manpower and time on this scale has proved to be a barrier to the further uptake of the model.

In order to allow greater access to such modelling methods as EEP, there is a need to explore and develop more efficient methods for acquiring survey data of building stock. Should these become available, feedback from users indicates that EEP type systems would find wide-scale application in local and regional government.

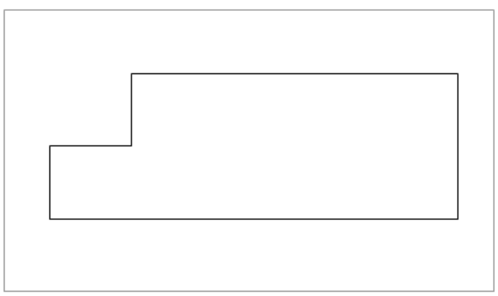
As a first step in this exploration a new method, applying a simple pattern recognition algorithm to the analysis of digital maps, has been developed and tested against the known region surveyed for the EEP programme.

2. Methodology

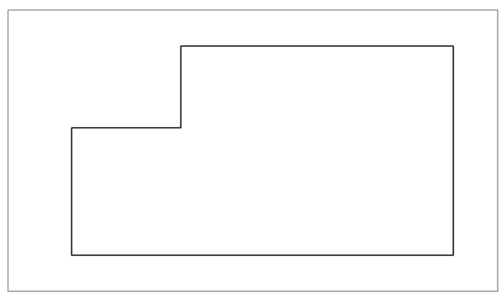
At a larger urban level the use of pattern recognition has been established; Barr and Barnsley (2004), for instance, used Ordnance Survey (OS) maps to infer urban land use and successfully identified areas with similar built ages by considering street layout patterns. However in order to achieve the level of detail necessary for a model such as EEP, data at the level of an individual dwelling is required, and this aspect is unproven.

In regional energy applications it has been found that it is necessary only to identify a building to an era of construction as this will define basic characteristics such as wall construction type and so define heat loss factors required for energy use calculations. As a result the actual date of construction of each building is not needed; rather the dates can be clustered into the eras: pre-1919; 1919-1945; 1945-1965; 1965-1980; post-1980.

An initial subjective comparison of map data and survey data acquired for the Neath Port Talbot region lead to the hypotheses that a) housing of a common era will exhibit a common form, or rather a range of plan forms, and that b) these "standard" forms will alter with the era of construction. A survey based on an analysis of building plan forms should therefore be able to distinguish approximate building age. For example, figures 1a and 1b show typical plan forms of two ages of housing, extracted from the region studied. Although the two eras show a similar plan form, the pre-1919 houses appear to exhibit a consistently higher aspect ratio and so should be robustly identifiable. Other similar distinctions were seen to exist between other housing eras as well.



a) Pre 1919 housing



b) 1919 to 1945 housing

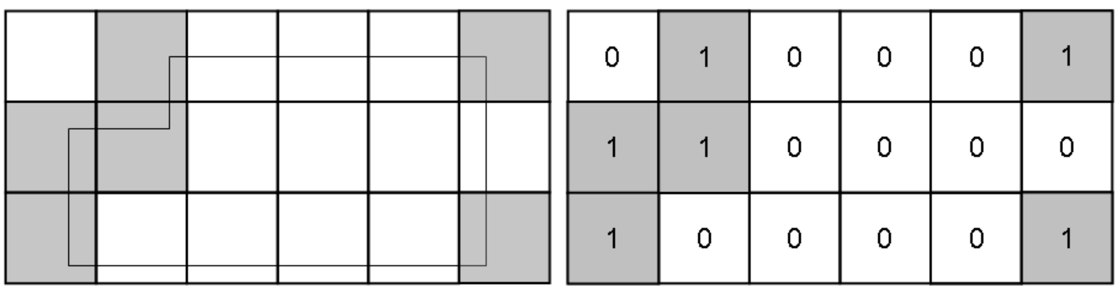
Figure 1. Typical dwelling outline patterns.

The analysis of the building plan forms can be undertaken via software analysis of digital map information. Individual building outlines are now available digitally on OS Master Map™ files as polygons. Using GIS methods and the Ordnance Survey's Address Point™ system, each polygon can be associated with an individual address and assessed to provide information on, for instance:

- Geometrical plan shape,
- Number of vertices,
- Aspect ratio,
- Floor area,
- Size of 'garden plot',
- Density of housing.

In order to extract information to assess building age, the polygon information representing building plan form is first normalised for orientation and for longitudinal dimension. Then, allowing for mirroring and distortions, those polygons are analysed, to identify and cluster "similar" shapes.

In order to do this, in the GIS system a simple fixed grid is laid over the normalized building plan outline, with the building centered in the grid. Internal GIS functions are used to locate corners, or "nodes" of the building shape; where a node lies in a grid cell this is marked as occupied otherwise the cell is empty. Figure 2a shows the building outline with an example 6x3 grid overlaid. The grey cells represent the presence of a building node. The occupied and empty cells are converted into a numerical code which identifies the shape of the building outline. Figure 2b show the code formed using 1 for node cells and 0 for empty cells; in this case the building outline translates the code 010001110000100001.



a) building outline with grid b) cell codes

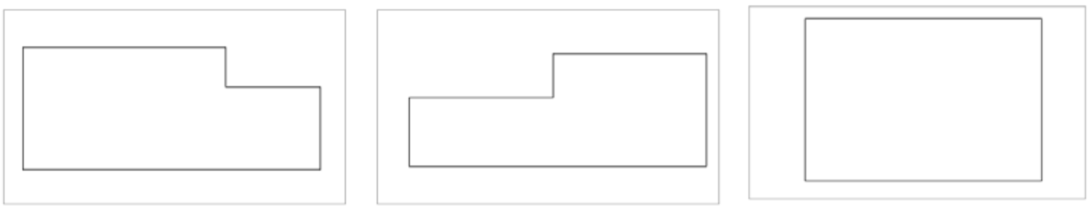
Figure 2. Typical dwelling outline patterns.

This coding method was applied map data for a selection, chosen to contain a wide range of housing age, of over 2,000 domestic buildings from the detailed survey for the Neath Port Talbot region; the actual building characteristics of this sample were therefore known due to the earlier manual survey. The resulting codes that were generated using this method were analyzed to compare the codes to the known building types. The aim of the analysis was to determine level of ability and reliability in the identification of building age, and to determined if further development was worthwhile.

3. Results

The analysis showed that the building "signature", e.g. the numerical code, could indeed identify building era to a certain degree. The effect of grid density was investigated; form signatures generated by 6x3, 8x4, and 8x8 grids were examined. Of these the 8x4 grid produced the best results; success rates dropped considerably for the others.

Using an 8x4 grid produced 491 signatures from the sample. However many of these were produced by "solo" building plan forms; that is, no others in the sample shared that code. Excluding those, it was found that 83 signatures uniquely matched to a building era, while 49 were seen to contain a mix of eras. Thus over 60% of the buildings in the sample could be placed in an era. Sample result plan form "signatures" are shown in figure 3. The best agreement was found for pre-1919 housing; for this era 69% of the stock in the sample was attributed to unique forms. This was considered highly encouraging as this era marks a crucial housing type often characterised by solid walls with a correspondingly poor energy performance; the "hard to heat" home. There were similar clusters found for other ages but with lower success rates; for instance 36% of post-1980 housing and 26% of 1945-1965 housing were correctly identified, however only 12% of 1965-1980 housing were attributable to unique plan forms.



a) b) c)

Figure 3. Typical plan forms generated by the method; a and b uniquely define an era, while c) covers a number of eras.

These success rates, though too low to be useful in practice, are considered encouraging enough to continue to develop the method. In order to become a useful method, the identification success rate must be increased. This initial attempt used only a very simplistic shape matching algorithm. It is considered that an improved success rate could be achieved through the use of more sophisticated matching algorithms; e.g. using methods and algorithms such as those identified by Jain et al. (1988) and Belongie et al. (2002). Inspection of the unidentified and incorrectly identified properties may show common characteristics which the method used was too coarse to detect. In addition, improvements may be made through the incorporation of data from other sources; this may be contextual data (as supplied, for instance, from the map itself, e.g. the neighbourhood road or estate layout or distance from the road centre, or from the analysis of aerial or high resolution satellite images providing information on e.g. roof forms), or from other databases, for instance census data. There may well be an upper limit of success available from this approach; as only ground plan form is used, data such as number of stories, façade form or fenestration, cannot be derived. Through recently acquired funding from the Engineering and Physical Sciences Research Council, these various aspects are now being explored.

4. Conclusions

A simple pattern matching algorithm has been embedded in a GIS system to analyse the plan form of housing stock. The trial system was able to successfully identify, without further intervention, the age of a high proportion (60%) of dwellings in a sample of 2000. Pre-1919 housing was most readily identified with 69% of that sample being correctly identified. Housing constructed within the era 1965-1980 was the least well recognised, with only 12% being correctly identified.

Higher success rates will be required in a working system. It is considered that improvements in the success rate will require the use of more advanced pattern matching methods and algorithms and the use of supporting information, such as distance from the road centre or roof form (as may be determined from satellite imagery for instance). These aspects are currently being investigated.

5. Acknowledgements

The authors would like to thank the Engineering and Physical Sciences Research Council for funding the furthered development of this work and for the development of the original Energy and Environment Prediction (EEP) model.

6. References

BARR, L.S., 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: Planning and Design, volume 31(3) May, pages 397 – 418

BELONGIE, S., MALIK, J., and PUZICHA, J. 2002, Shape matching and object recognition using shape contexts, , IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 24, Issue: 4, p 509-522

JAIN, A.K., and DUBES, R.C., 1988, Algorithms For Clustering Data, Prentice-Hall.

JONES, P., WILLIAMS, J.L., and LANNON, S., 2000, Planning for a Sustainable City: an Energy and Environmental Prediction Model, Journal of Environmental Planning and Management, vol. 43, no. 6, pp. 855-872(18)

Biography

Mr S. Lannon is a Senior Research Associate of the Welsh School of Architecture, whose research specialities include GIS, database integration and programming. He is co-author of the Energy and Environment Prediction (EEP) model and was part of the Housing and Neighbourhoods and Health (HANAH) project.

Mr D.K. Alexander is a Senior Research Fellow and has developed a number of computer based building energy and environment models, as well as managing the Centre‘s physical scale modelling facilities

Professor P.J.Jones is Head of Department of the Welsh School of Architecture and is director of the Centre of Research in the Built Environment (CRiBE). Professor Jones has been active in research on energy and sustainability in the built environment for more than twenty years.