Institute of Geography Online Paper Series: GEO-016

Visualisation of Settlements Over Large Changes In Scale

Omair Chaudhry & William Mackaness

Institute of Geography, School of GeoSciences,
University of Edinburgh,
Drummond St,
Edinburgh EH8 9XP
O.Chaudhry@sms.ed.ac.uk
William.mackaness@ed.ac.uk

Paper submitted for the ICA Workshop, a Coruna, Spain, July 2005

Visualisation of Settlements Over Large Changes In Scale

Omair Chaudhry & William Mackaness

Institute of Geography
School of GeoSciences,
University of Edinburgh
Institute of Geography
Drummond St
Edinburgh EH8 9XP
O.Chaudhry@sms.ed.ac.uk
William.mackaness@ed.ac.uk

Paper submitted for the ICA Workshop, a Coruna, Spain, July 2005

Keywords: Synoptic view, Model Generalisation, Buildings, Aggregation

Abstract:

This paper explores the application of map generalisation techniques in the creation of small scale mapping (1:250,000) directly from large scale data (1:1250 and 1:2500). Small scale mapping provides a synoptic overview. At 1:250,000 scale, it is not that there is less information contained in the map, rather that we see fundamentally different information and relationships – broad-brush information between higher order phenomenon. The evaluation criteria used at this scale are quite different from those used at large scale. This paper describes a methodology by which small scale mapping is derived directly from the large scale (without generating intermediate results) using simple generalisation techniques. The approach illustrated in this case study, creates city boundaries based on simple buffering of building features, and derivation of the single polygon surrounding clusters of buildings over a threshold size. The hull is simplified, small internal polygons removed. The algorithm was applied to large and small conurbations alike. The results obtained via this method were evaluated against paper maps. Initial results point to refinement of this technique.

1. Introduction:

It is only at small scales that we get a synoptic view of the world. That synoptic view is composed of highly characteristic forms – these forms representing 'ideas' or concepts, and their location. In the example (Figure 1) a mix of text and a highly generalised representation of the road network are used to delineate the concept 'city' – in this case London.

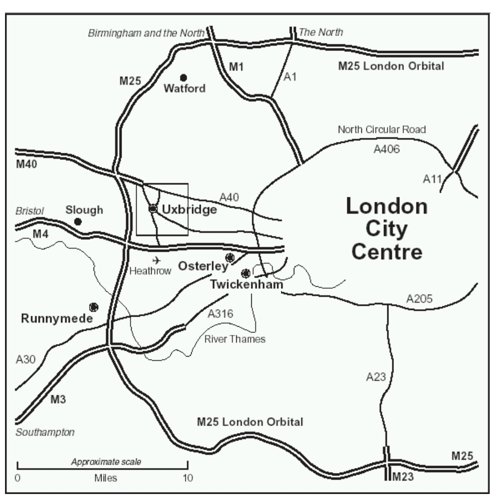


Figure 1: Small scale mapping at 1:1 300 000 (approx)

Examining Figure 1, we make a number of observations. First we see large scale classes subsumed by higher order forms (the dense cluster of buildings are replaced by the outline of the city). A different set of evaluation criteria are required - linked in some way to the success with which we have replaced large scale phenomenon with 'higher order' phenomenon associated with small scale, synoptic maps. There is less of a requirement in precision. The map conveys a topology completely different from that at the large scale (because this map is comprised a very different set of phenomena). The emphasis of the map has changed: the contextual information combines to emphasise the location of a set of places and a river, on the western side of London. We borrow specific properties of an entity, and symbolise it in a manner that both best enable its immediate identification and clearly separates it from other features. In Figure 1, one can see how symbolisation has been used both to capture the sinuous nature of the river Thames, and separate it from the hierarchical inter connectedness of the road/motorway network. In general what we see are a collection of caricatures of various concepts – exaggeration of a trait to reinforce the efficient (minimalist?) transmission of an idea.

Beyond the visual, (i.e. from a database perspective), information recorded at this scale can be used in spatial analysis – for example comparing spatial extents of UK cities, or in broad brush journey planning. In other words in support of 'scale dependent' analysis (Schneider, 2001 ; Levin, 1992). Finally we note that there is vagueness in the representation of these concepts. The boundary of London is inferred, the river and roads leading into London behaves like pieces of dropped cotton thread – but that the human eye combined with geographical experiences has no problem filling in the 'missing pieces'. Indeed we can see that the cartographer has taken full advantage of our conceptual model of 'city' in the creation of this map.

From a generalisation research perspective, this paper, reports on how these small scaled synoptic views can be generated from the large scale. In trying to generalise over large changes in scale, the research emphasises the importance of model generalisation – both in the creation and analysis of 'higher order' concepts, and as an essential precursor to cartographic generalisation. By model generalisation we mean database operations such as selection, aggregation and classification. From a pragmatic perspective it must be a solution that can cope with large volumes of data. The ideas are implemented for Generalisation of large urban buildings (Edinburgh) and for smaller settlements as well from Ordnance Survey (OS) MasterMap (1:1250 and 1:2500) to 1:250,000. The results are then evaluated against manually created OS Strategi (1:250,000) settlement layer (for more information on MasterMap and OS Strategi see (Ordnance, 2005b).

The paper is organized as follows: Section 2 gives a brief outline of our observations arising from a visual comparison between the input and the perceived output; Section 3 outlines the design of the algorithm and explains its main features (such as buffering, aggregation and simplification). Section 4 presents a case study, Section 5 evaluates the results followed by section 6, which outlines areas of further improvement. We conclude with a summary.

2.0 Observations:

Buildings from OS MasterMap (Figure 2a) were extracted and compared with the OS Strategi's settlement layer (Figure 2b). Comparison between the source map and target map provide interesting insight into the metamorphosis of phenomena into higher order phenomena. The authors are sensitive to the criticism of paper based comparisons nevertheless feel this to be a useful starting point showing how much the information content has changed with changing phenomena.

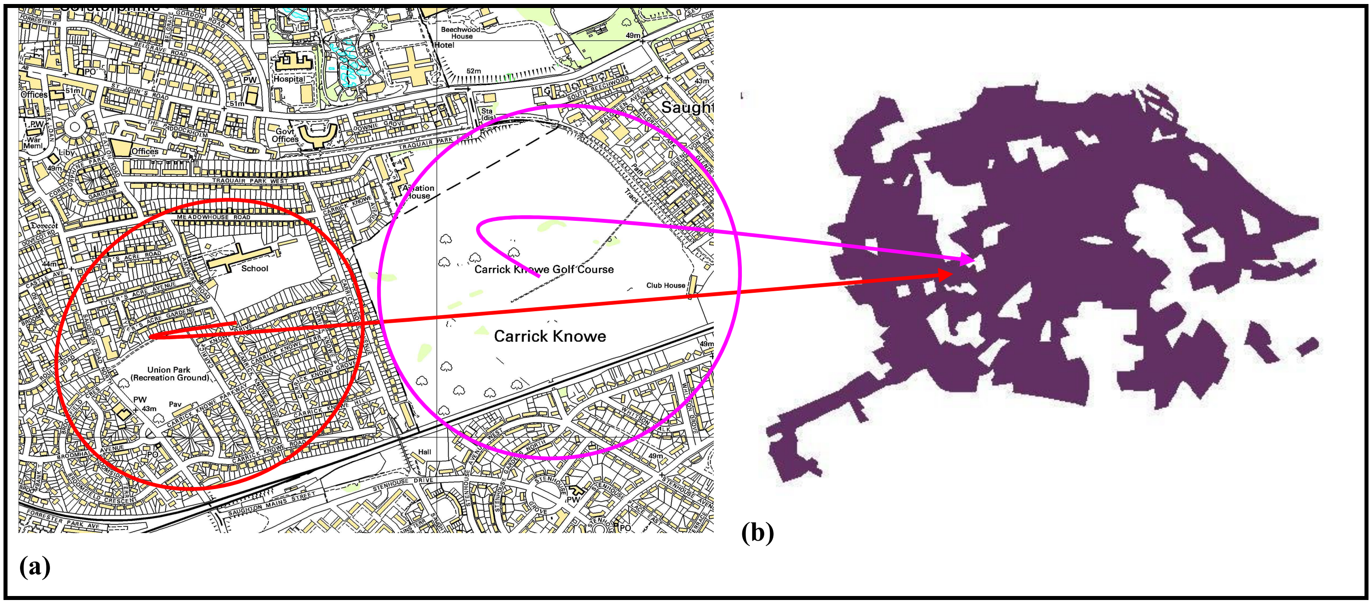


Figure 2: 2a. 1:10 000 scale map showing buildings and open spaces, and 2b: how they are represented at 1:250 000.

In Figure 2a, the emphasis is on detail, precise delineation and location. We can undertake spatial analysis at a large scale, between and 'within' different objects, and between classes of objects – measuring many properties - both Euclidean and topological. The eye is able to group objects and infer activities and process. More detailed comparison between the two figures elicits the following observations:

- Patterns, shapes, interactions and extents are discernible and measurable in both maps.
- There is not 'less information' in the synoptic view, merely *different* information.
- Large scale buildings are represented as individual polygons but at small scales they are amalgamated together with other phenomena into area object – city'.
- Whilst many of the same cartographic principles apply to both, in the synoptic view, locational precision and completeness is sacrificed in the interests of higher levels of abstraction – which facilitates ease of interpretation.
- That maps at such different scales enable different forms of analysis between fundamentally different types of entities to be undertaken (the link between task and scale of observation).

Our final observation relates to the continuum that connects these two maps together. Minsky observed that 'you cannot tell you are on an island by looking at the pebbles on a beach'. Yet aloft in a balloon, and those pebbles constitute a shoreline, which if found to be continuous and connected, would indeed reveal it to be an island. With respect to Figure 2, and in precisely the same manner, we could travel aloft, and discern the makings of a city – eventually making out its boundary. This research is premised on the idea that theoretically it should be possible to automatically create such a synoptic view 'directly' from the large scale database, though others have spoken of the challenges of transitioning these 'conceptual cusp' (Muller, 1991).

The question is whether this transition can be achieved through a set of incremental, (intermediate) results, or via a single step. These alternate approaches are presented figuratively in Figure 3. This research has deliberately chosen a single step approach in order to focus attention on how we can model fundamental changes in phenomena.

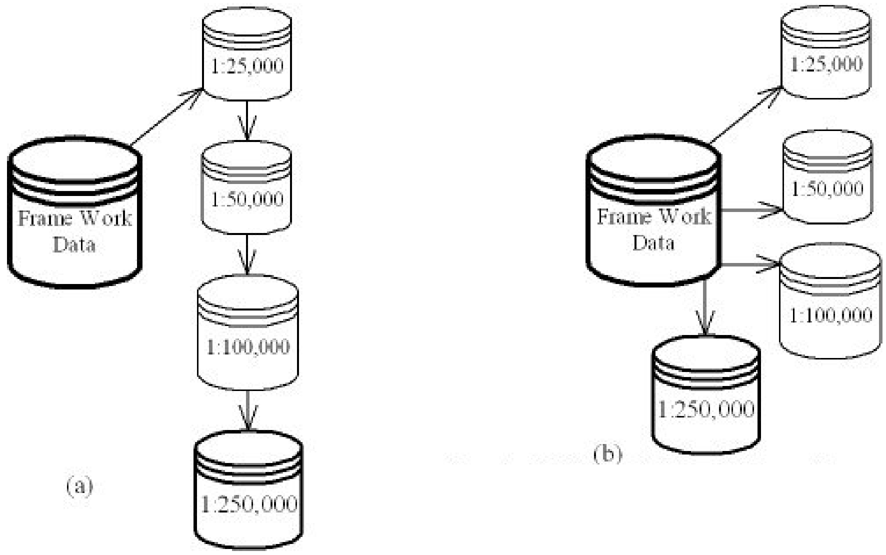


Figure 3: (a) A stepping stone approach to deriving small scale mapping, or b) a single step (leap!) approach.

3.0 Framework:

A lot of Generalisation algorithms for urban buildings have been proposed (Wang and Doihara, 2002; Basaraner and Selcuk, 2002; Chen et al., 2004) but they seldom extend over large changes in scale. In this research, the methodology entails selecting all building features, and aggregating them in a way that reveals large areas continuously covered by dense housing. Small, sparsely covered regions internal to these large areas (such as parks) must be managed in a suitable manner. Part of the challenge in developing this methodology is working with poorly attributed features. For example the database makes no distinction between city housing, and village housing. The overall design of the proposed system is summarised in Figure 4.

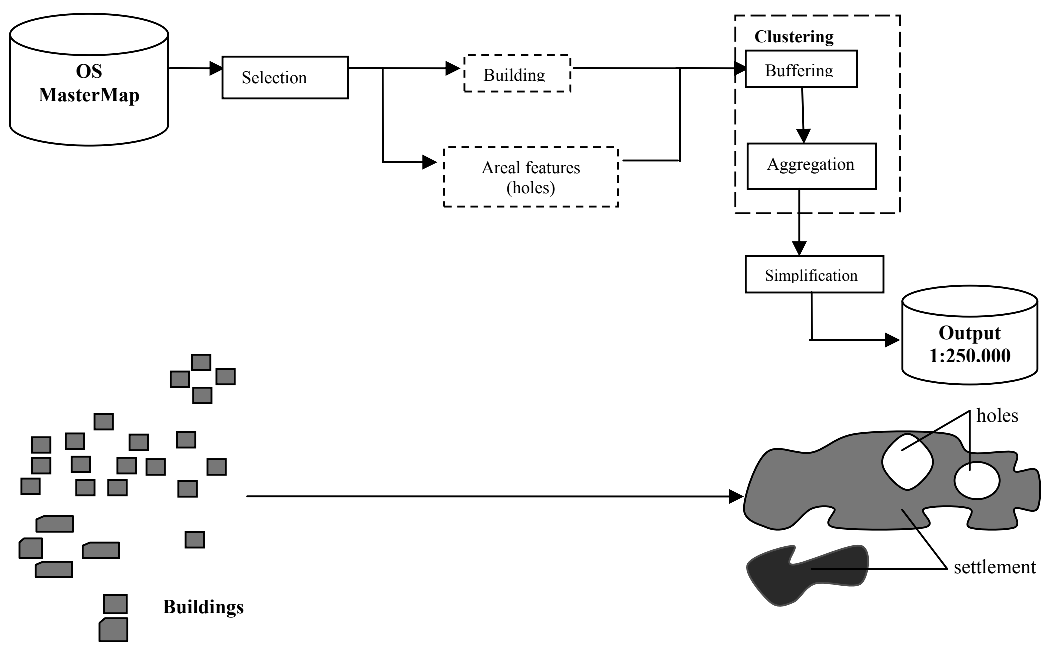


Figure 4: Overall Design

3.1 Buffering:

The first step is to extract all the buildings for the given area. This is achieved simply by using the attribute "Descriptive Group" and value "Buildings" in the MasterMap. The next step was to create clusters of buildings. The first step of the clustering was to create buffers around the individual buildings. A range of buffer sizes from 50 – 150 m were tried and through empirically testing it was determined that for large urban areas and for a target output of 1:250,000, 100 meters buffer was most suitable. It was of prime importance that the algorithm should be capable of handling large datasets which given the regional extent. So to ensure that we never run out of memory the buffers were stored back into Oracle. Although it has a negative effect on the performance due to I/O operations. Spatial indexing was used in Oracle to improve the efficiency. The resulting buffers had rounded corners. In order to remain as close as possible to the shape of a building the buffers were made angular (as per the illustration of Figure 5).

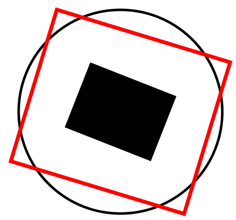


Figure 5: Buffering a polygon and Squaring it

3.2 Aggregation:

The next step is to aggregate the buffers and thus generate the boundary of the urban area. This required us to dissolve the overlapping buffers. Usually most dissolve functions in GIS are not efficient when working with large data set but here we used Oracle's Aggregate function which dissolved the buffers into one single 'multi-polygon'. The drawback is that all the buffers are dissolved into one single 'multi-polygon' (an example is given in Figure 6). This means that it is not possible to identify areas outside the urban boundary nor the holes inside it. To overcome this problem an algorithm was implemented that goes through the Oracle spatial object information array and identifies each inner-polygon which is part of the multi-polygon object. Thus we identify the outer polygons as separate entities in the database.

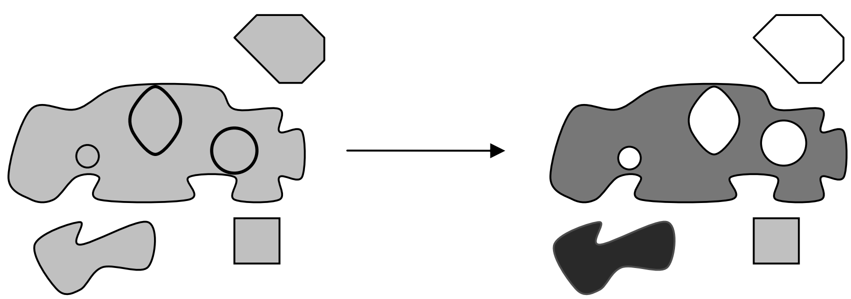


Figure 6: (a)All the buffers have been dissolved into one single polygon (same colour); (b) Polygons has been reclassified into different entities in the database

3.3 Selection and Simplification:

The next step is the classification and simplification of the polygons, for which we require a definition of what is large or small urban area – ie a built up area. According to OS 1:250,000 map specification (Ordnance, 2005a): '*An urban area shown on the Small Scale Mapping Intelligence (SSMI) 1:250 000 metric plot that measures over 1 sq km is captured as an large urban area.'* And *'An urban area equal to or greater than 0.01 sq km and less than 1 sq km is captured as a small urban area'.*

The area of each polygon is calculated and checked against this threshold. It is removed if it is less than this threshold (for example Figure 7). 'Holes' remain within the urban region reflecting low densities of buildings in parks and open spaces. In some instances these might be removed, amalgamated, or retained. It is time consuming to create the topology that would enable us to identify where these holes were. Instead we again made use of the Oracle spatial object specification, since a closed hole is stored as a linear ring in a multi-polygon geometry. This enabled us to identify the holes and calculated their areas. Using an area threshold value, smaller holes were removed and replaced by the urban area classification. This is pictorially presented in Figure 7.

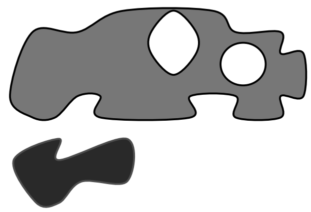


Figure 7: Removal of smaller areas and holes. In comparison to Figure 6b the small hole has been unified with the urban boundary. Small outer polygons have been eliminated.

To make the output more visually appropriate and to be in close resemblance to cartographers output. The output went through further simplification stage. All the polygons and holes were passed through series of simplification steps. Firstly an algorithm was implemented which reduced the number of vertices so that we do not end up with conflicts in the resultant polygon when we apply Douglas- Peucker (DP) algorithm on it. Douglas and Peucker (1973) algorithm was then applied using 1m as the threshold value. After that buffer of 25m was created inside the polygon and to simplify it further Douglas- Peuckar was applied again with 7m as threshold. This process was repeated for all the holes and areas in the output.

4. A Case Study:

We applied this methodology in the derivation of a synoptic view (1:250,000 map) directly from a large scale database (OS MasterMap). It was implemented and the results were evaluated against OS Strategi data set (1:250,000). The platform selected for the implementation was Java, SQLJ and Oracle 10g. The new Java Geometry API offers a wide range of functions to manipulate spatial objects in Oracle. Oracle 10g supports all the geometrical and topological functions defined by OGC as reported in (Oosterom et al., 2002). Taking advantage of the following functionality provided in Oracle (Oracle, 2005):

- **SDO_GEOM.SDO_BUFFER:** Returns a geometry object representing the buffer polygon around the given spatial object.
- **SDO_GEOM.SDO_ARC_DENSIFY:** Returns a geometry in which each circular arc in the input geometry is changed into an approximation of the circular arc consisting of straight lines.
- **SDO_AGGR_UNION:** Returns a geometry object that is the topological union (OR operation) of the specified geometry objects.
- **SDO_GEOM.SDO_AREA:** Returns the area of a two-dimensional polygon.
- **SDO_AGGR_CONVEXHULL:** Returns a geometry object that is the convex hull of the specified geometry objects.
- **SDO_UTIL.EXTRACT:** Returns the geometry that represents a specified element (and optionally a ring) of the input geometry.

The algorithm was applied to OS MasterMap of an urban area. Buildings were extracted from it as shown in the Figure 8. The number of buildings polygons were 138,616 (we wanted to ensure that the implementation was robust enough to handle large volumes of data).

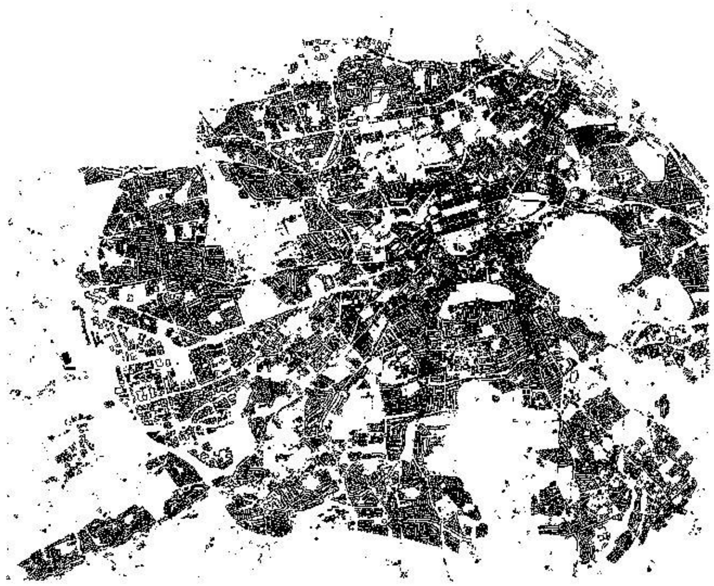
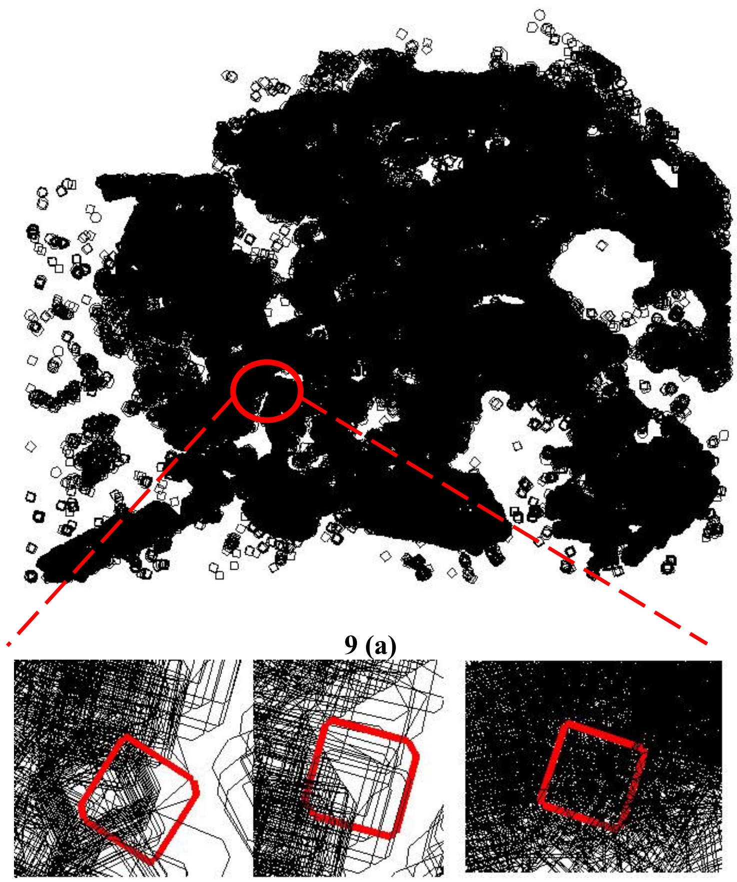


Figure 8:Buildings extracted from OS MasterMap of Edinburgh City

The next step was to create the buffers and modify their shape (Figure 9).



9 (a)

9(b)

Figure 9: (a) Showing buffers created around buildings of Figure 8. (b) Enlarged view showing the angular form of the buffers.

After buffering the next step was to generate the boundary of the urban area. This was done by dissolving the buffers using SDO_AGGR_UNION function the output obtained is shown in Figure 10. This function aggregates all the given polygons in the given table to one single multiploygon (Figure 10).

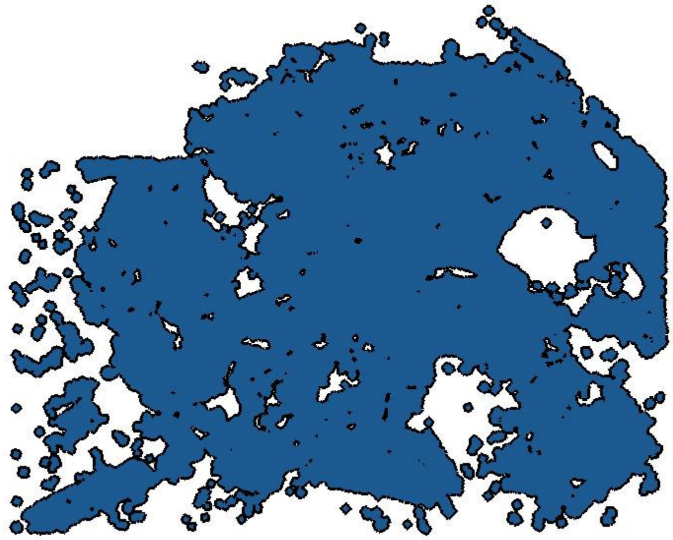


Figure 10: Initial output after the aggregation of buffers. All buffers have been dissolved into a single polygon.

To overcome the problem of one single 'multipolygon' the spatial object information was used to identify the inner rings (holes) and outer polygons. Using SDO_UTIL.EXTRACT function polygons were extracted from the above single multi-polygons. Then the simplification was performed by comparing the area of each hole and polygon against threshold value. SDO_GEOM.SDO_UNION function was used to merge these holes with the urban area. The smaller outer polygons were deleted. Smaller holes were further simplified by creating convex hull around them. Larger holes and the urban boundary were simplified by applying DP and buffering on them as described in section 3.3,the final output is shown in Figure 11.

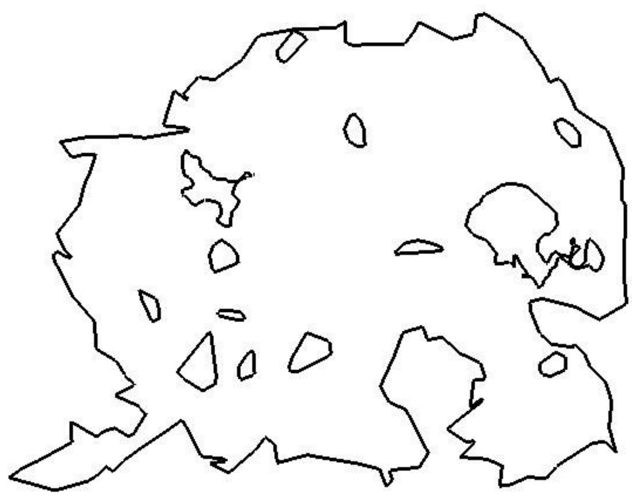
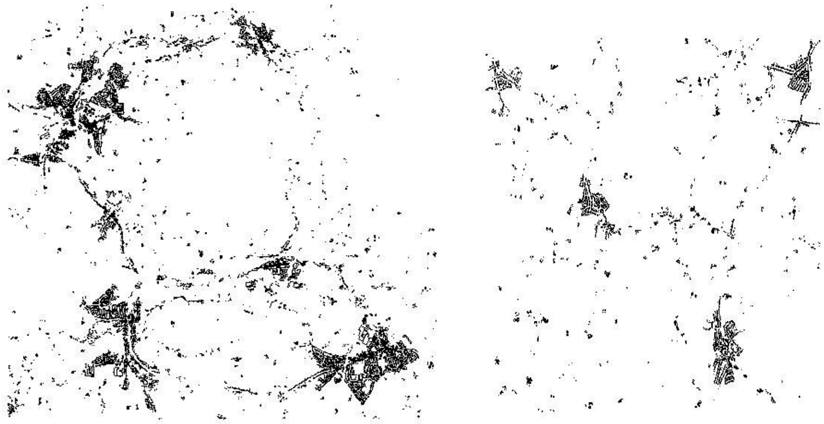


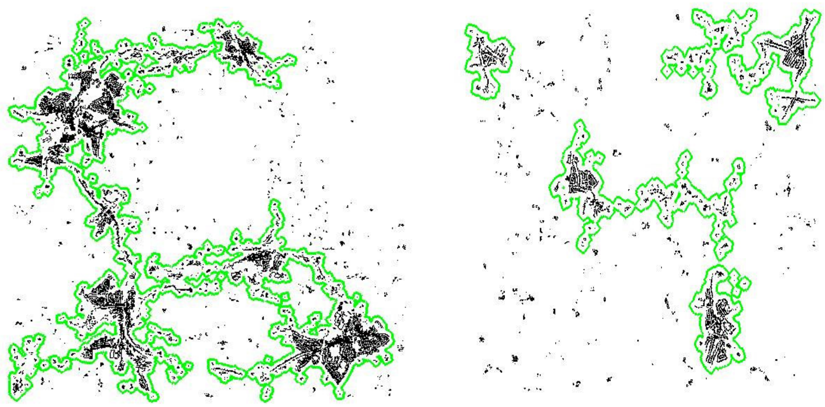
Figure 11: Output after being angluarized using a mixture of Douglas Peucker and Buffering

Although the focus here was on large urban areas but we tested the algorithm for smaller settlements and checked how robust it was. Also to check if same parameters can be applied on smaller settlements or not. If not what are parameters that need to be changed and why they need to be change them. Figure 12 shows the selected smaller settlement and Figure 13 shows corresponding outputs.



(a) (b)

Figure 12: Buildings extracted from MasterMap of a small settlement



(a) (b)

Figure 13: (a) Output obtained with 100m buffer size for input Figure12a. (b) Output obtained with 100m buffer size for input Figure12b

5. Evaluation:

Visual comparison was done between the output obtained (Figure 11) and OS Strategi settlement layer (Figure 14). Although the comparison with paper maps is a subjective issue (Mayer, 1998; Weibel and Dutton, 1999) it gives some indication of the success of the algorithm. As observed the algorithm has successfully maintained the overall structure which is same in both the maps. There are some localized differences (such as those highlighted by red circles in Figure 14). This might be because the cartographer has taken density as well as proximity into account whereas the algorithm proposed only takes proximity into account. Additional holes are present in OS Strategi layer and similarly some holes are missing because of the expanding nature of the city and the currency of Strategi data. We therefore argue that our output reflects a more current view.

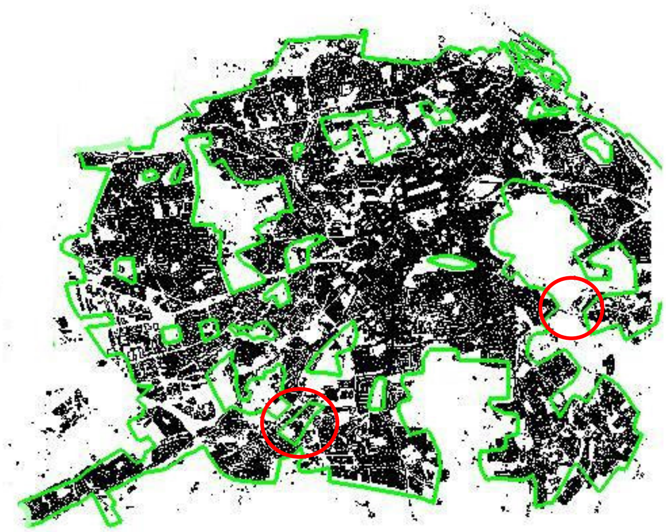


Figure 14:OS Strategi (1:250,000) settlement layer. Red circles highlight differences

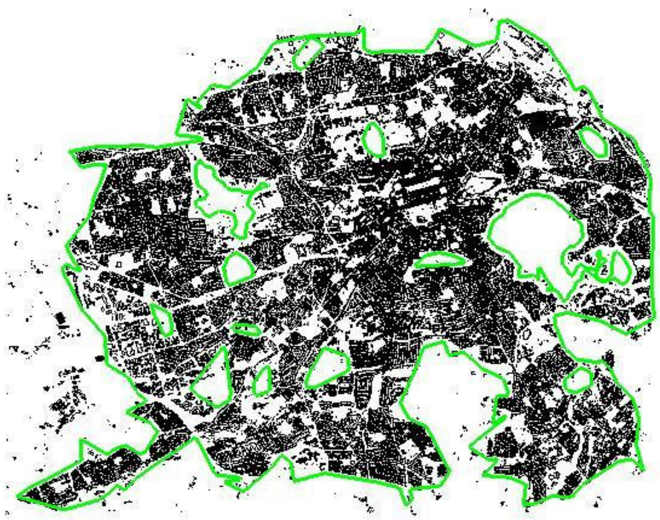
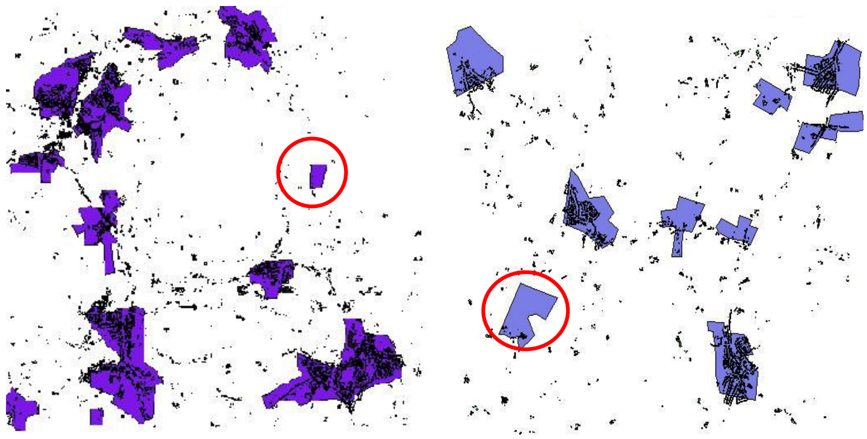


Figure 15: Output using the approach presented in paper along with input buildings

Figure 16 shows the OS Strategi output for the smaller settlement shown in Figure 12 a and 12b (overlaid on top of OS Strategi). Certainly there is a difference between our output (Figure 13) and OS's result. The main reason for this seems to be that the size of the buffer is too large for the smaller settlements so many buildings that should have be eliminated because of low density are kept in the output. Another reason is the same as in the case of urban area i.e. the OS Strategi result is not current so its not updated with the latest developments as shown in Figure 16. The next section outline further area of development and how to deal with the problems mentioned here.



(a) (b)

Figure 16: (a)OS Strategi (purple block) for the selected smaller settlement in Figure 12a (b) OS Strategi (blue block) for the selected smaller settlement in Figure 12b. The red circles show the problem areas where the density of buildings low yet in the OS Strategi database they are recorded as 'built up'

6. Further Work:

Further work will look into how the localized problems such as those highlighted in Figure 14 can be managed. And also to deal with the problems of the smaller settlements as shown in Figure 16. One thing that can be inferred from the results is that cartographers have not only proximity but also the density of the buildings whereas the algorithm proposed here only takes proximity into account. Further work will look at how density can be modelled and how dense buildings can be removed before the clustering starts. Road topology can be used to calculate the density of the buildings within each block (Wang and Doihara, 2002). Another approach is to use a grid of a specific cell size and lay it on top of the input area and measure the density within each cell (Wang et al., 2002).

It is apparent that we need to use different buffer sizes for large and small urban areas. We tested this on input shown in Figure 12b with buffer size of 50m and obtained the result shown in Figure 17.

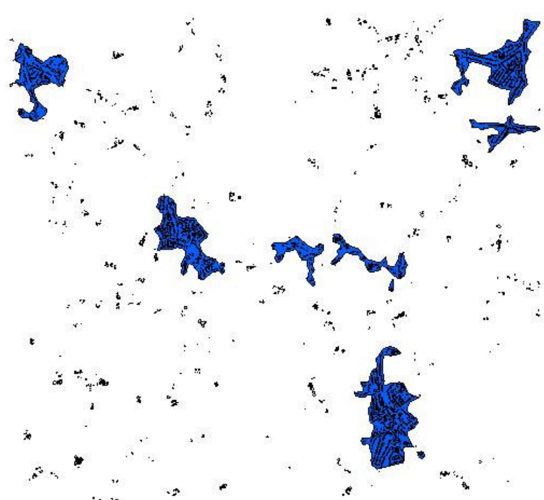


Figure 17: Output with Buffer size of 50m on input (Figure 12b)

From the results shown above it is apparent that there is a correlation between the buffer size and the areal extent of a settlement. One approach to deal with the issue of different buffer sizes is to first run the program with buffer size of 100m. After the identification of the large urban area for the remaining buildings we rerun the program with a smaller buffer size.

7. Conclusion:

The challenge of deriving maps over large changes in scale is not new (Watson, 1970; Smaalen, 2003). But in this paper we argue that the creation of the synoptic view – the representation of concepts - requires a more focused development of model generalisation techniques. We argue that this work is relevant to spatial analysis and exploratory data analysis – that generalisation is fundamental to the revelation of different patterns and interdependencies between geographic phenomenon.

The paper proposed a methodology by which a synoptic view can be created. The paper also looked into the spatial functions provided by Oracle 10g and a case study was carried out to generalize the urban buildings and generate boundary of the urban settlement which compared quite favourably with the manually created urban settlement layer of 1:250,000 map. Further work will look at dealing with smaller settlements and mixtures of small and large settlements. Also it will look into density calculations of the buildings which appears to be an important precursor to clustering.

Acknowledgements:

We gratefully acknowledge support and funding from the Ordnance Survey. Many thanks to Patrick Revell and Nicolas Regnauld for their support and ideas.

References:

Basaraner, M. and Selcuk, M. 2002, 'Strategies for Generalisation of Buildings and Built-up Areas in Topographic Geospatial Databases', in *International Symposium on Geographic Information Systems*, pp. 23-26.

Chen, J., Li, Z., Yan, H. and Ai, T. 2004, 'Automated Building Generalisation based on Urban morphology and Gestalt theory', *International Journal of Geographical Information Science,* vol. 18, pp. 513-534.

Douglas, D. and Peucker, T. 1973, 'Algorithms for the Reduction of the Number of Points required to represent a Digitised Line or its Caricature', *The Canadian Cartographer,* vol. 10, no. 2, pp. 112-122.

Levin, S. A. 1992, 'The Problem of Pattern and Scale in Ecology', *Ecology,* vol. 73, no. 6, pp. 1943-1965.

Mayer, H. 1998, '*Three Dimensional Generalisation of Buildings based on Scale-Spaces'*,Report, Chair for Photogrammetry and Remote Sensing, Technische Universität München.

Muller, J. C. 1991, 'Generalisation of Spatial Databases', In D. J. Maguire, M. Goodchild and D. Rhind, (eds.), *Geographical Information Systems*, London, Longman Scientific, pp. 457-475.

Oosterom, P. J. M., Quak, C. W. and T.P.M., T. 2002, 'Testing current DBMS products with real spatial data', *In Proceedings UDMS 2002*, Prague, Czech Republic.

Oracle 2005, 'Spatial User's Guide and Reference 10g Release 1 (10.1)', Available: [http://www.oracle-10g-buch.de/oracle_10g_documentation/appdev.101/b10826/toc.htm] (15th Feb 2005).

Ordnance, S. 2005a, '1:250,000 specification', *Ordnance Survey internal document*.

Ordnance, S. 2005b, 'Ordnance Survey Home Page', Available: [www.ordnancesurvey.co.uk] (15 Feb 2005).

Schneider, C. 2001, 'The Rise of the Concept of Scale in Ecology', *BioScience,* vol. 51, no. 7, pp. 545-553.

Smaalen, J. W. N. v. 2003, 'Automated Aggregation of Geographic Objects: A New Approach to the Conceptual Generalisation of Geographic Databases', *Doctoral Dissertation, Wageningen University, The Netherlands*.

Wang, P. and Doihara, T. 2002, 'Automatic Generalisation Of Roads And Buildings', in *ISPRS Commission IV*, Ottawa, Canada.

Wang, P., Doihara, T. and Lu, W. 2002, 'Spatial Generalisation: An Adaptive Lattice Model Based on Spatial Resolution', in *ISPRS Commission IV*, Ottawa, Canada.

Watson, W. 1970, 'A study of generalisation of a small scale map series', *International Yearbook of International Cartography,* vol. 16, pp. 24-32.

Weibel, R. and Dutton, G. 1999, 'Generalising Spatial Data and Dealing with Multiple Representations', In P. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind, (eds.), *Geographical Information Systems*, vol. and Sons, John Wiley, New York, pp. 125-156.