What Makes a Good Footprint?

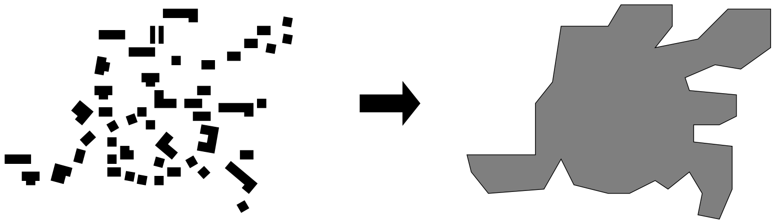
Antony Galton, University of Exeter

March 27, 2008

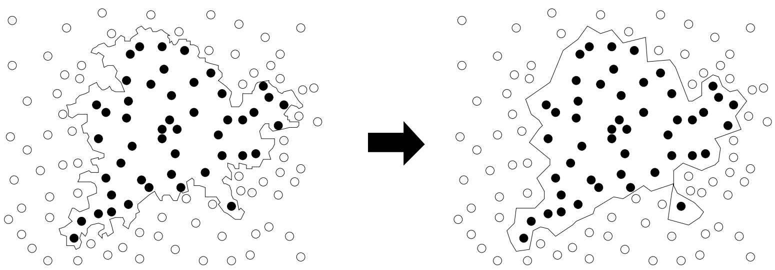
A problem of considerable interest in GIScience is that of characterising the region ('footprint') occupied by a set of discrete point-like elements ('dots'). The region is a 'higher-level' entity: it is not the location of the points as individuals, but of the collective which has those points as members.

Some instances of this problem are:

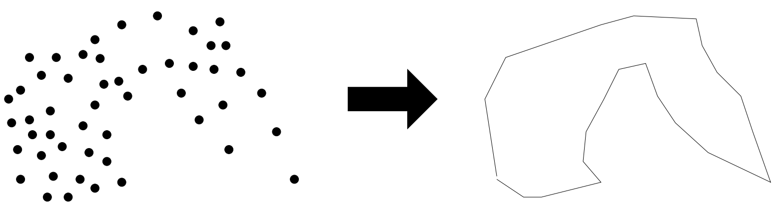
1. Map generalisation. A collection of buildings forms a settlement; in generalising the map we may wish to aggregate the set of points or small polygons representing the individual buildings to form a single polygon representing the settlement as an individual entity:



2. Region approximation [1, 2]. Given a discrete set of points, and for each point whether it lies inside or outside some region of interest, we want to draw a polygon to represent that region. This problem can appear in two guises: (1) the region in question is well-defined, but we either do not know, or do not wish to spare the computational resources to represent, its exact boundary; (2) the region is ill-defined ('vague'), and therefore we seek a 'reasonable' precisification of it:



3. Location of collectives [7]. Again, given a discrete set of points, but this time they represent individuals such as people in a crowd, or animals in a flock; we wish to represent the location of the crowd or flock as a region:



There are related applications in fields such as pattern recognition and computer vision, but there the goal is not typically to produce a crisp outline but to devise useful shape descriptors (relating to, e.g., convexity, elongation).

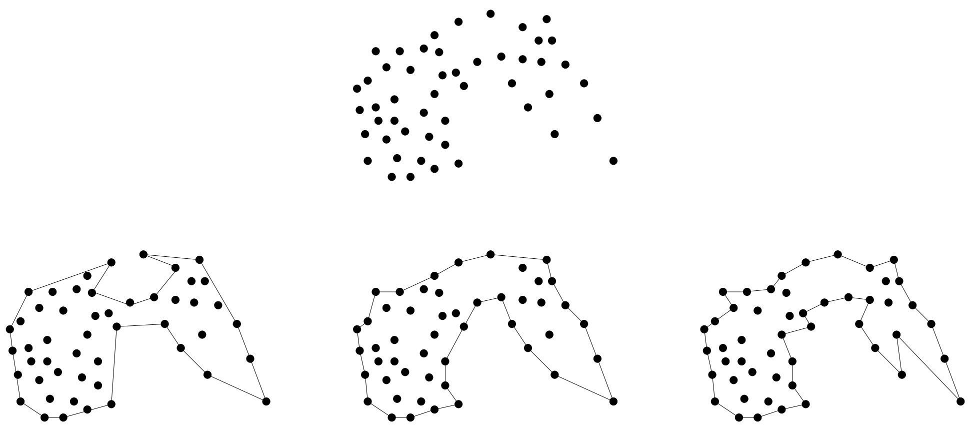
Different application contexts may have different requirements, and some of these were listed as evaluation criteria in [8]. They include: whether all the dots need to be included in the footprint, whether dots are allowed on the boundary of the footprint, whether the footprint must have various 'nice' topological properties (e.g., regular, connected, simple), and whether the outline of the footprint must be polygonal.

Methods described in the literature typically produce a definite set of answers to these criteria. A common kind of algorithm will produce a topologically simple, regular, connected, polygonal footprint whose vertices are all in the set of dots and for which all or most of the remaining dots lie in the interior (apart from sporadic outliers permitted by some methods). Various methods for producing footprints that have been proposed may be found in [5, 3, 9, 10, 1, 2, 8, 11, 4].

These criteria all relate to the suitability of a method for some kind of application, but they do not help us to determine the quality of the results of any particular algorithm, and it is striking that in the literature very little is said about this. In the case of region approximation, there is a clear criterion of success: how closely does the footprint that is generated resemble the original region? But in other applications there is no 'original region' to compare the footprint with: the purpose of generating a footprint is to create a region to represent the spatial distribution of the dots, not to recover a pre-existing region from which the dots have been sampled.

In much of the existing literature, the focus is largely on computational criteria, particularly computational complexity. This is an evaluation of the algorithm itself, not of the results. Evaluation of the results typically amounts to little more than an assertion that the shape generated is "a good approximation to the perceived shape" of the dot pattern. What we find, therefore, is a focus on the properties of an algorithm and away from the properties of the results it generates.

I would like to redress the balance: to look seriously at the problem of how to evaluate a footprint that is proposed for a set of dots — and therefore how to evaluate an algorithm for producing such footprints. As an example, in the illustration below, which of the three shapes best captures the 'perceived shape' of the dot pattern?



Clearly this is not a question with a unique and well-defined answer, in general. None the less, we as humans do exhibit definite preferences: given a dot pattern and asked to draw a polygonal outline which captures the shape we perceive the dots as forming, people will readily produce an outline which to them seems satisfactory while at the same time acknowledging that there is no unique answer. And given a collection of such outlines, we regard it as a meaningful (if not always easy) task to rank them as better or worse representations of the shape formed by the dots.

I am currently interested in finding objective correlates of the perceptual/aesthetic criteria we employ in evaluating the quality of regions presented as footprints for dot patterns. For the sake of simplicity I am focussing on *polygonal hulls*, that is, simple polygons whose vertices are members of the dot pattern, where the members of the dot pattern which are not vertices all lie in the interior of the polygon.

My hypothesis is that for a polygonal hull to be 'good' footprint, it should achieve a balance between the conflicting goals of minimising the area (in order to cut out the empty areas devoid of points which we will often find in the convex hull) and minimising the perimeter (in order to produce a perceptually 'clean' outline without too many gratuitous sinuosities). These are clearly not the only factors that make for a good footprint, but it seems reasonable to suppose that they are important ones.

To test this hypothesis, I recently conducted a small informal pilot study (reported in detail in a paper currently under review [6]), in which I presented 13 subjects with a set of eight dot patterns and asked them to draw their preferred polygonal outline. The dot patterns I used were small enough for me to be able to compute *all* the possible polygonal hulls and to plot them on a scatter-graph relating area and perimeter. The hypothesis is that the hulls drawn by the subjects in the pilot study should all lie on or close to the Pareto front in the area-perimeter graph; and the results of the pilot study showed that this is indeed the case.

Here I illustrate the results for just one of the dot patterns used, shown in Figure 1. In Figure 2 are shown, on the left, the plot of perimeter against area for all possible polygonal hulls for this dot pattern (area is plotted along the horizontal axis, perimeter up the vertical, so the convex hull, which simultaneously maximizes area and minimizes perimeter, is at the extreme lower right), and on the right, the points on the Pareto front of this graph (shown as dots) and the points corresponding to the outlines drawn by the subjects (shown as circles). It will be seen that the latter are all on or close to the Pareto front, as predicted by the hypothesis.

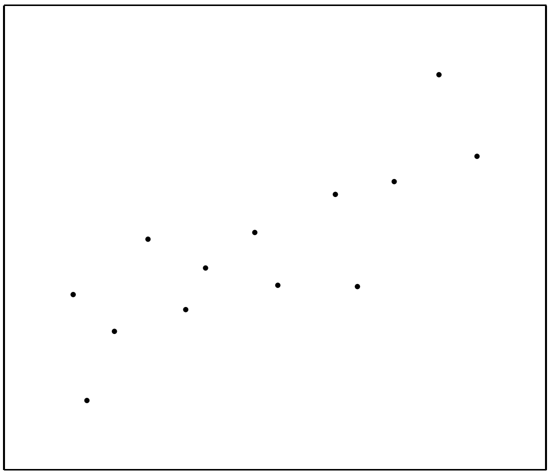


Figure 1: A dot pattern used in the pilot study

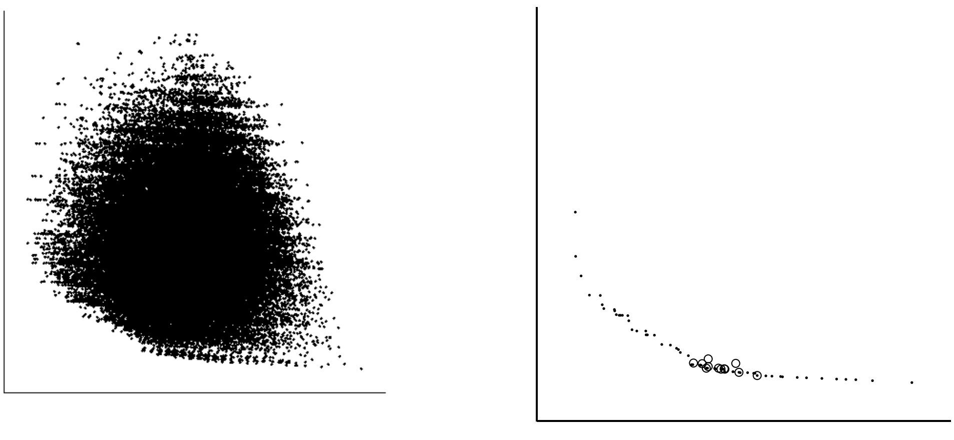


Figure 2: Left: Plot of perimeter against area for all polygonal hulls for the dot pattern in Figure 1. Right: The Pareto front (dots) and the hulls drawn by the subjects (circles).

There remain many further questions to answer. Given that the hulls chosen by the subjects lie close to the Pareto front, a further question is what determines their position along the front. In most of the patterns used in the Pilot study, the selected hulls were near the central portion of the front: for a typical dot pattern, the front rises steeply towards the left, meaning that once one has got the area down to a certain point, further reduction in area can only be achieved at the cost of a disproportionately large increase in perimeter. Selected hulls are not found in this region. It is also usually the case that hulls towards the right-hand end of the front, approaching the convex hull, are not favoured, except in the case of those dot patterns which look convex to us, e.g., a pattern obtained by filling a convex outline with dots at a more or less uniform density. Further research is needed to determine what other objective factors influence the subjective perceptual appropriateness of a given hull.

Once we have a good idea of what the main objective factors are, we will be in a position to undertake a principled evaluation of the various algorithms that have been proposed for generating footprints; and perhaps, going beyond this, to design new algorithms that are tailor-made to optimise with respect to those factors. In view of the multi-objective nature of the task, some form of evolutionary algorithm might be appropriate for this.

References

[1] H. Alani, C. B. Jones, and D. Tudhope. Voronoi-based region approximation for geographical information retrieval with gazetteers. *International Journal of Geographical Information Science*, 15(4):287–306, 2001.

[2] A. Arampatzis, M. van Kreveld, I. Reinbacher, C. B. Jones, S. Vaid, P. Clough, H. Joho, and M. Sanderson. Web-based delineation of imprecise regions. *Computers, Environment and Urban Systems*, 30:436–459, 2006.

[3] A. R. Chaudhuri, B. B. Chaudhuri, and S. K. Parui. A novel approach to computation of the shape of a dot pattern and extraction of its perceptual border. *Computer Vision and Image Understanding*, 68(3):257–275, 1997.

[4] M. Duckham, L. Kulik, M. Worboys, and A. Galton. Efficient generation of simple polygons for characterizing the shape of a set of points in the plane. *Pattern Recognition*, forthcoming. Accepted for publication, March 2008.

[5] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, IT-29(4):551–559, 1983.

[6] A. P. Galton. Pareto-optimality of cognitively preferred polygonal hulls for dot patterns. Submitted to Spatial Cognition 2008.

[7] A. P. Galton. Dynamic collectives and their collective dynamics. In D. M. Mark and A. G. Cohn, editors, *Spatial Information Theory*. Springer, 2005.

[8] A. P. Galton and M. Duckham. What is the region occupied by a set of points? In M. Raubal, H. J. Miller, A. U. Frank, and M. F. Goodchild, editors, *Geographic Information Science: Proceedings of the 4th International Conference, GIScience 2006*, pages 81–98. Springer, 2006.

[9] Gautam Garai and B. B. Chaudhuri. A split and merge procedure for polygonal border detection of dot pattern. *Image and Vision Computing*, 17:75–82, 1999.

[10] M. Melkemi and M. Djebali. Computing the shape of a planar points set. *Pattern Recognition*, 33:1423–1436, 2000.

[11] Adriano Moreira and Maribel Santos. Concave hull: a k-nearest neighbours approach for the computation of the region occupied by a set of points. In *Proceedings of the 2nd International Conference on Computer Graphics Theory and Applications (GRAPP2007), Barcelona, Spain, 8-11 March, 2007*, 2007.