

Label and Attribute-Based Topographic Point Thinning

Paulo Raposo¹, Cynthia A. Brewer¹, Lawrence V. Stanislawski²

¹The Pennsylvania State University, University Park, PA
paulo.raposo@psu.edu, cbrewer@psu.edu

²The United States Geological Survey (USGS), Rolla, MO
lstan@usgs.gov

Abstract

Point features on topographic maps should be generalized through scale with respect to contextual, thematic information attributed to them. Also, because point features are typically presented with labels, their generalization should account for the area on the target map required to render the text. This paper presents an example automated workflow for the progressive thinning by selection of mountain summits. The workflow enriches the point features with a geomorphologic measure (prominence) and generalizes using this relative quantity as a ranking. The procedure uses scale-dependent rectangular tessellations to limit point density on the scaled target map such that each retained summit point has ample surrounding space within which to be labeled. The method is applicable to any thematic point map features which require uncluttered labels.

1. Introduction & Background

The computational geometry of point generalization is different from many other kinds of map features because a point has no dimensionality except that of its symbol. Consequently, point generalization is often performed in response to graphic collision or coalescence problems, or else to aggregate, typify or select/eliminate points in order to reduce the complexity and count of a distribution of points across the map. There is a need to meaningfully generalize thematic point data in many contexts, particularly because un-generalized data typically lead to incomprehensible maps suffering from the “push-pin problem”, where the symbols of many point features clutter a map and obscure both themselves and their base maps; to go further and label such a map is futile.

This research considers point generalization as the underpinning problem in map labeling in the United States Geological Survey (USGS) context. Many of the label data used across various feature types in USGS topographic maps are georeferenced by point vector data. Also, the USGS uses the Esri Maplex engine to automate run-time labeling decisions, including those of selection and placement. This research sets out to simplify the computational complexity of the labeling tasks Maplex is asked to perform across features and scales, by pre-selecting scale-based subsets of points and labels based on a pre-computed importance metric not already present in the data. This process simplification is expected to both improve the cartographic quality of label selection and placement decisions made by Maplex at run time, due to the reduction of less-significant points in competition for local label selection, and, by extension, speed up processing. Specifically, we present work on the scale-dependent pre-selection of summit points based on their computed prominences.

2. Point and Label Generalization

2.1 Point Generalization

As with any cartographic feature, point generalization frequently requires processes which transform both the spatial and attribute data, as well as the graphical representation of the map features, with the two stages referred to respectively as model and cartographic generalization. Edwardes, Burghardt and Weibel (2005) treat point generalization comprehensively, identifying the five generalization operators most meaningful to point data as *i*) selection, *ii*) aggregation, *iii*) simplification, *iv*) typification, and *v*) displacement. Operators *ii*) and *iii*) are understood to be processes of model generalization, operators *iv*) and *v*) are processes of cartographic generalization, and *i*), selection, is understood to be a process of either model or cartographic generalization, depending on whether it happens on a global or local level, respectively. They further address point generalization in an important use scenario, namely mobile mapping, where users typically access query-response thematic content on devices with small screens, exacerbating the “push-pin problem”. This use scenario is addressed by Bereuter and Weibel (2013), who also take into account the particular need for real-time generalization of thematic point data for dynamic query delivery to mobile users. Bereuter and Weibel (2010) also address the possibility of altering map space rather than point locations.

Computer scientists de Berg *et al.* (2004) consider the case of generalizing point maps in which global point distribution and local point density are important qualities for retention (e.g., dot density maps). Generalized points are regarded as a subset of the initial points, and their paper describes algorithms for approximating the initial distributions and densities using various heuristics. Subsets of points are also implied by “The Radical Law” (Töpfer & Pillewizer, 1966), whereby a reduced quantity of points is calculated given the initial count and the change in map scale. Other authors have suggested other means of determining appropriate generalized point distributions, including “map capacity” with respect to the number of features a map can support (Ratajski, 1967), as well as methods based on information theory entropy and capacity limits thereof (Bjorke, 2012).

2.2 Label Generalization

Several authors from cartography and other graphics fields have considered the sets of problems associated with map labeling, which typically surround the two tasks of label selection and label placement. Skupin (2000) notes that labels form a crucial consideration for map generalization, both because labels inhabit large areas of the product map, increasing its graphic complexity, and because their placement is a complex problem for automated systems (also: Shieber, Christensen, & Marks, 1995). Doddi *et al.* (1997) describe the label placement problem as consisting of finding a set of rectangles on the map wherein labels can be placed without graphic conflict. Of course, label placement is complicated further by every other graphic element on the map. Cecconi and Galanda (2002) describe a system of their design whereby labels are selected on-the-fly, according to a pre-determined importance attribute.

3. Summit Data and Enrichment for Generalization

The United States Geological Society (USGS) maintains the Geographic Names Information System (GNIS), which itself retains a database of some 2.2 million place names and toponyms across the United States, each attributed to a coordinate location. GNIS points are used in USGS

topographic map production as the point data source for numerous map feature types, natural and human. GNIS points are assigned an elevation value, but there is no other topographic information to describe natural features such as summits.

In order to generalize summits more meaningfully, we computed the relative relief (i.e., prominence) of each summit point using the method described by Chaudhry and Mackaness (2008; later adopted also by Sinha & Mark, 2010). A 10-meter digital elevation model (DEM) is used to create contour polygons at 15m intervals throughout the area of interest, as well as to sample the elevation of each plotted GNIS summit point. Prominence is defined as the difference between the summit elevation and the elevation of its lowest encircling (i.e., intersecting) contour polygon which does not also encircle a higher summit. (In order to detect and define summits, Chaudhry and Mackaness continue to derive further morphometric properties for topographic eminences, including partonomic range relationships, but we use only part of their methods because our summit points are already defined.)

4. Selection/Elimination by Scale-Referenced Tessellation

We present an example topographic automated workflow in which summits are enriched with a geomorphometric measure to facilitate their generalization by local selection. The work is thus an example of cartographic generalization, rather than model generalization, since the summit data themselves are not transformed.

The basic idea is to tessellate the map surface at any given map scale with rectangles, each of which defines the “neighborhood” within which one summit is permitted to be plotted and labeled. Rectangles are used over other possible tessellation shapes (e.g., hexagons) because label areas tend to be delimited by rectangles, given horizontal typing (Figure 1). By definition a tessellation possesses non-overlapping tesserae, ensuring that no piece of the map space is over-assigned as the neighborhood of more than one label. The approach is similar to that taken by Bereuter and Weibel (2013), who tessellate by quadtree, except that tesserae in this approach are defined by target scales and label “neighborhood” sizes, are not geometrically hierarchical, and are computed expediently, rather than forming a data storage structure.



Figure 1. Rectangles as a better approximation of label space than hexagons.

The rectangles are sized such that they are always 2 cm wide, and 1.236 cm tall at target map scale; these two numbers are (approximately) in the ratio ϕ (*phi*, approximately 1.618). This ratio was chosen because of its classical proportions, being taken here as a heuristic for the areal ratio a label of up to three lines long may take (Figure 2). A more data-driven approach is possible, if one were to analyze all labels under consideration, and take a mean or maximum width vs. height measure as the rectangle size.



Figure 2. Mountain names from our study areas in rectangles of ratio ϕ , with and without abbreviations.

Because the topographic mapmaking workflow of the USGS is entirely conducted with Esri's ArcGIS software, we have implemented our enrichment and generalization procedures within the same software, using the Python language to automate all steps.

Once the map space has been tessellated and the summit of most prominence calculated in each rectangle, the set of selected summits is flagged in the summits attribute table for visibility at the given map scale. This procedure allows a single summit feature class to be maintained, with a scale-dependent query-able attribute defining which summits are visible; this data structure complies with the standard mapmaking procedures currently in place at the USGS. The procedure progresses by a “ladder”, rather than “star”, approach (Stoter, 2005). Laddering ensures that as a user moves to progressively smaller map scales, the summits selected for viewing are always a subset of the summits selected at the previous viewing scale (i.e., summits never “blink” in or out across scales). This is particularly important to our approach since the scaling of rectangles according to map scale means that the resulting tessellations across scales are very unlikely to be hierarchical (i.e., “nested”). A star approach would likely mean that a given summit would have different intra-rectangle local neighbors across processed rectangle sizes among which to compare prominence; the result would be a discontinuous selection/omission sequence for that summit through the scale range processed.

The final step of label placement is automated in ArcGIS using the Maplex labeling engine, again a standard part of USGS mapmaking. The Maplex engine accounts for various graphic conflicts between labels and other features on the map. On a real topographic map, the engine would balance labeling requirements against roads, symbols, etc. A principle goal of the generalization described here is to generate subsets of labels for given map scales which reduce the spatial and graphic (i.e., computational) complexity with which the Maplex engine must deal at production time.

5. Example - West Virginia Appalachians

We demonstrate the workflow with an example set in National Hydrologic Dataset (NHD) hydrologic subbasin 02070001 in West Virginia, within the Appalachian Mountain range. This study area includes 174 summit locations and exemplifies the problem of over-crowding point symbols, each requiring a label. The problem of course becomes worse at smaller scales. Since it is impossible to label all summits at small scales, and because positional accuracy of summit location should not be compromised by displacement, the appropriate generalization operator is selection/elimination.

We selected subset summits for 1:24,000, as well as all scales from 1:50,000 to 1:1,000,000, laddering up by increments of 50,000. Computation of each summit's prominence was automated in ArcMap using minimal attribute table cursors and multiple Python tuple lists and dictionaries, and took about 5 minutes to run for all summits. The 21 ladder “rungs” computed

for scales from 1:24,000 to 1:1,000,000 took about 20 minutes in total to compute, again automated using minimal ArcGIS table cursors and multiple Python data structures. All computations were run on a desktop workstation (3.00GHz, 2-core processor, 4GB RAM). Samples of the workflow output are illustrated in Figures 3 and 4. Figure 3 provides descriptive statistics across the summits, as well as ranges of prominence values identifying selected summits for five scales. Figure 4 illustrates cartographic samples of the output.

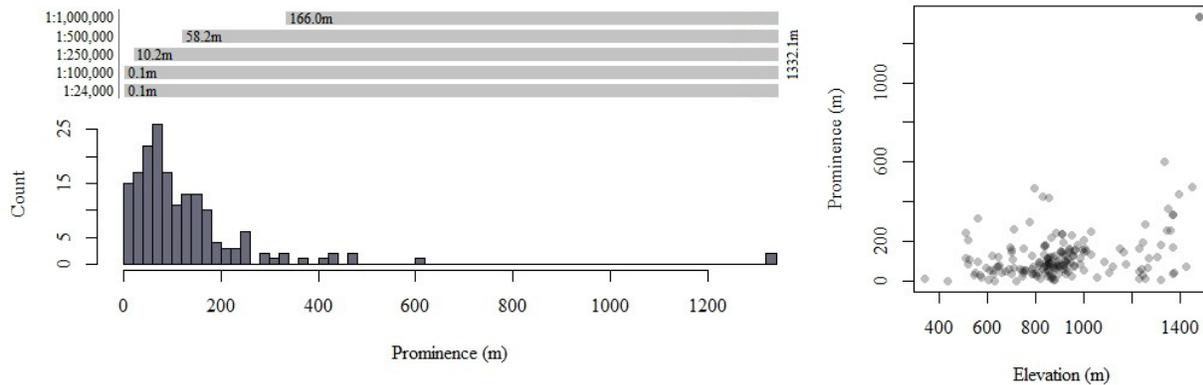


Figure 3. Descriptive statistics for the 174 summits in the study area. A histogram illustrates the distribution of prominence values among the summits. The horizontal bars above illustrate the ranges of prominence values among summits selected at five map scales, with numbers at left indicating the least prominence selected, up to the highest prominence selected (1332.1m). Summits are also plotted by elevation vs. prominence; analysis of distributions such as this one across various terrain types throughout the country may help to identify a rule associating appropriate prominence value thresholds to map scales.

6. Discussion and Conclusions

We feel the workflow described above effectively resolves the elemental problem of labeled summit point generalization for USGS topographic maps. The approach pairing thematic attribute data enrichment with tessellated local selection can be applied to other point and label geographic features. We note that this approach is only appropriate for map features that can be treated by selection/elimination operations driven by the hard constraint of available map space for feature labeling. While the label-driven selection process described here does not retain a sense of relative point density or spatial distribution among labeled points, these may still be retained as desired by also including unlabeled points. The fact that the tessellations we employ are driven not by the geometric distribution of the points, but by “neighborhoods” of map space required to place labels allows the method to address an important print/render-time problem in practical cartography.

Further research may identify minimum summit prominence values that are deemed appropriate at given map scales over the terrain types of the United States. For instance, a lower prominence value may not be relevant at a smaller scale for a mountain foothill area. This effort may also help control point density patterns for better contextual representation of terrain variations through summit point labeling.

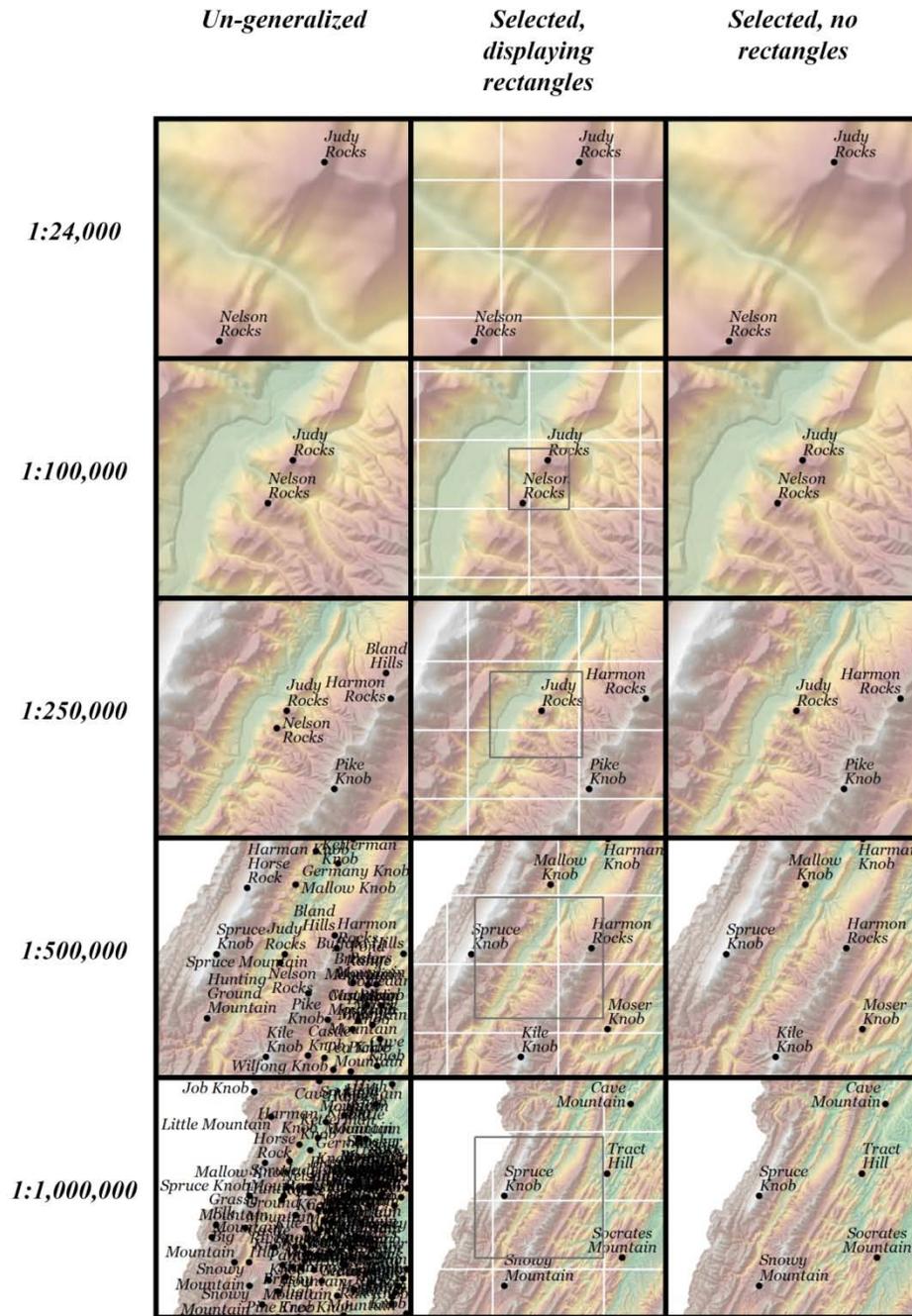


Figure 4. Workflow output for selected scales, depicted with and without the rectangles driving point and label selections. Grey outlines in the middle map tiles illustrate the extent of the above, larger scale tile. Map tiles are rendered with a hypsometric tint and terrain shading for context.

Finally, a longer term goal is to build a knowledge base of multiscale cartographic representations, such as summit prominence estimates associated to terrain feature extents, that can be used in ontology design patterns to enable the automated identification, extraction, and

visualization of complex topographic features through a semantic interface. This work will follow cognition-based extraction concepts demonstrated by Sinha and Mark (2010) with possible implementation through ontology design patterns presented by Varanka and Jerris (2010).

Acknowledgements

Research funding provided by the U.S. Geological Survey's Center of Excellence for Geospatial Information Science (USGS-CEGIS) through the Department of Interior Cooperative Ecosystems Studies Unit (CESU). The work of Paulo Raposo and Dr. Brewer is funded by grant # 06HQAG0131, "Symbol and Label Design Solutions for Electronic Topographic Mapping for *The National Map* of the United States". Barbara Buttenfield is a collaborator on the larger project encompassing this work and we appreciate her advice. We also appreciate the support of E. Lynn Usery, Director of CEGIS, as well as the work of Chelsea Hanchett and Andy Stauffer. The Gould Center in the Department of Geography at Penn State supported our work with computing and facilities.

References

- Bereuter, P., & Weibel, R. (2010). *Generalisation of point data for mobile devices: A problem-oriented approach*. Paper presented at the 13th Workshop on Progress in Generalisation and Multiple Representation, Zurich, Switzerland.
- Bereuter, P., & Weibel, R. (2013). Real-time Generalization of Point Data in Mobile and Web Mapping Using Quadrees. *Cartography and Geographic Information Science, advance online publication*. doi: <http://dx.doi.org/10.1080/15230406.2013.779779>
- Bjorke, J. T. (2012). Exploration of Information: Theoretic Arguments for the Limited Amount of Information in a Map. *Cartography and Geographic Information Science, 39*(2), 88-97.
- Cecconi, A., & Galanda, M. (2002). Adaptive Zooming in Web Cartography. *Computer Graphics Forum, 21*(4), 787-799.
- Chaudhry, O. Z., & Mackaness, W. A. (2008). Creating Mountains out of Mole Hills: Automatic Identification of Hills and Ranges Using Morphometric Analysis. *Transactions in GIS, 12*(5), 567-589.
- de Berg, M., Bose, P., Cheong, O., & Morin, P. (2004). On simplifying dot maps. *Computational Geometry, 27*(1), 43-62.
- Doddi, S., Marathe, M. V., Mirzaian, A., Moret, B. M. E., & Zhu, B. (1997). *Map labeling and its generalizations*. Paper presented at the Eighth Annual Association for Computing Machinery (ACM) - Society for Industrial and Applied Mathematics Symposium on Discrete Algorithms (SODA), Philadelphia, PA, USA. <http://dl.acm.org/citation.cfm?id=314250>
- Edwardes, A., Burghardt, D., & Weibel, R. (2005). Portrayal and Generalisation of Point Maps for Mobile Information Services. In L. Meng, T. Reichenbacher & A. Zipf (Eds.), *Map-based mobile services: theories, methods and implementations* (pp. 11-28). Berlin, Heidelberg: Springer.
- Ratajski, L. (1967). Phénomène des points de généralisation. In K. Kirschbaum & K. H. Meine (Eds.), *International Yearbook of Cartography* (Vol. 7, pp. 143-152). Bonn-Bad: Godesberg.

- Shieber, S. M., Christensen, J., & Marks, J. (1995). An empirical study of algorithms for point feature label placement. *Transactions on Graphics, 14*(3).
- Sinha, G., & Mark, D. M. (2010). Cognition-Based Extraction and Modelling of Topographic Eminences. *Cartographica, 45*(2), 105-112.
- Skupin, A. (2000). *From metaphor to method: cartographic perspectives on information visualization*. Paper presented at the IEEE Symposium on Information Visualization (InfoVis) 2000.
- Stoter, J. E. (2005). *Generalisation Within NMA's in the 21st Century*. Paper presented at the International Cartographic Conference, A Coruña, Spain.
- Töpfer, F., & Pillewizer, W. (1966). The Principles of Selection. *The Cartographic Journal, 3*(1), 10-16.
- Varanka, D., & Jerris, T. J. (2010). *Complex Topographic Feature Ontology Patterns*. Paper presented at the AutoCarto Conference, Orlando, Florida.