Towards general theory of raster data generalization
Paulo Raposo & Timofey Samsonov

Introduction

Despite the raster model’s simplicity, deep conceptual ambiguities exist within it. Every cell covers a certain area in space, while an underlying continuous field phenomenon has variation throughout this area. A coded cell value can be determined only by using statistical techniques (e.g., mean, etc.). The same situation exists for classed rasters: cells may be crossed by several “classes”, and the coded value depends on the assignment method. Also, when the cell size (i.e., resolution), shape, or orientation is changed, the resulting values for a cell at a particular location also change. Thus, we have two ambiguities: the first is assignment ambiguity, concerned with the possibility of different calculation schemes for raster values, and the second is the Modifiable Areal Unit Problem (Openshaw 1984). Raster generalization must handle both problems, and involves one or both of two processes: a) recalculation of values for new cells, and b) transition to different, likely greater, cell sizes.

Certain important cartographic and GIS applications ask for a full-fledged system of raster generalization methods. These come primarily from thematic and special-purpose mapping and analysis (e.g., surface analysis, climatological visualization and analysis, etc.). For some geographic data the raster model is more natural than vector. Among these are DEMS, geophysical fields (e.g., temperature and currents), abstract statistical fields (e.g., probability surfaces), and virtually any other continuous phenomena.

Our study aims to inventory past and current research in raster generalization both in GIScience and related fields, and to offer principles and methods as context to evolving raster generalization systems.

Background

Several conceptual frameworks for raster generalization have been proposed (Peuquet 1979, Li et al. 2001). McMaster and Monmonier (1989) divide raster generalization tasks into four categories: a) structural generalization, b) numerical generalization, c) numerical categorization, and d) categorical generalization. They identify fundamental operators and evaluate a variety of techniques used for raster-mode generalization. Comparisons of raster- and vector-based generalization have also been of interest (Daley et al. 1997, Li & Openshaw 1992, Peter and Weibel 1999). A series of articles by Li et al. on mathematical morphology in rasters should also be mentioned (Li 1994, Li & Su 1996, Su & Li 1995, Su et al, 1997a, 1997b, 1998).

Perhaps the most researched topic in raster generalization is digital elevation models (DEM) processing. Early studies in DEM generalization were based upon kernel-based filtering operations (Loon 1978). Since filtering does not take phenomenon structure into account, adaptive methods using structural information,
typically in the form of skeletal structure lines, were developed (Weibel 1987, Zaksek and Podobnikar 2005). Fan et al. (2007) developed a four-stage method using low-pass, smoothing and threshold filters that are locally combined depending on the slope and curvature of the surface. Leonowicz et al. (2009) used low and high quartile filters for valleys and watersheds, and developed a curvature-based algorithm (Leonowicz et al. 2010). Jordan (2007) developed a method based on Strahler stream order; basins are derived for streams of certain hierarchical order and these are filled by triangulating watershed boundaries; Ai and Li (2010) use a similar methodology. Other interesting fields of research are based on Fourier analysis (Clarke 1988) and wavelet transformations (Wu 2000). Jenny et al. (2011) applied wavelet multiscale pyramids to DEMs. Effects of filtering on DEM characteristics are investigated by Stauffer et al. (2012). Also, some case studies were conducted in raster land cover generalization (Monmonier 1983, Johnsson 1996, Jaakkola 1997).

Within the field of computer vision, scale-space theory is a formalization of multiscale image (e.g., photograph) representation, frequently used to deal with feature detection operations such as facial detection. The general approach is to convolve images with a Gaussian kernel parameterized to scale, analogous to a GIS moving-window operation. “Any coarse scale representation can be computed from any fine-scale representation using a similar transformation as the transformation from the original image” (Lindeberg 2008), meaning that multiple, sequenced representations can be computed along a continuum of scale. We propose applying the principles of scale-space theory to geospatial rasters.

**Raster properties**

**Regular spatial sampling frequency:** An intrinsic property of rasters, this allows neighborhood-scope kernel operations to work across the whole dataset without anomalous biases (projection distortions remaining invariant).

**Topology & neighborhood:** Every cell in a raster is clearly spatially related to its neighbors (4 or 8, depending on what kind of connectivity is used, see ‘connectivity paradox’ explained by Duff et al. 1973)—not always true for vector data. This allows intuitive analysis of every cell’s neighborhood, and thus calculation of various surface derivatives such as slope.

**Projection distortions:** Rasters contain all the distortions of their map projection. Usually raster processing does not treat these distortions, but in case of rasters covering large extents these distortions are inevitable. A possible solution is to partition the raster, reproject each partition independently to reduce distortion (e.g., convert to UTM), process each, and finally mosaic together. Another approach may be to use a variable kernel shape/size, such that the kernel varies in accordance to the local projection distortion to compensate for it. This latter case is developed in our case study.

**Constraints for raster generalization**

We suggest that the goal is to produce generalized rasters that maintain and/or optimize their information content through successive generalizations. This can be
conceived of as a matter of avoiding or minimizing error in the generalizations. This error can be of several types, including spatial and thematic (attribute) error (Luneta et al. 1991), each of which can be quantified and statistically compared. A directly related approach is to consider the pixel entropy of successive rasters (Bjorke 1996; Knöpfli 1983; Shannon 1948).

With increasing degrees of generalization we cannot indefinitely maintain or optimize the entropy or error of a raster; at some scale and resolution the pixels will be so coarse, or the variation between pixel values so minimal, as to make this impossible. We hypothesize that that this point of failure, similar to Ratajski’s (1967) points de généralization, is deterministically different for every input raster. If this is true, identifying that point would be very useful, since it would describe the range of scales over which a certain image is analytically reliable.

Some further possible constraints for raster generalization:

1. Geometry constraints (shape, area)
   - Horizontal boundary shapes should be preserved continuously
   - Areas and value distributions should be preserved continuously

2. Value constraints (statistics)
   - General statistics (mean, range, etc.) should be preserved continuously

In addition to novel innovative pixel operations, generalization processes may take the form of sequences of established raster processes (e.g., filters, pan-sharpening, opening and closing [“expand” and “shrink”]).

**Strategies of generalization**

Classed rasters have inherent “shapes” coded into them, and generalization of these may borrow techniques from vector generalization by operating on their boundaries. Continuous rasters require different approaches, probably involving calculus techniques on surface derivatives (e.g., Tobler 1979), as well as pattern-recognition methods that detect structures in the surface using edge detection algorithms and measures of spatial autocorrelation in pixel values.

**Unexplored topics**

McMaster & Monmonier (1989) offer a framework of four classes into which raster generalization procedures may fall. While many elementary, pixel-level operators are well-documented among varying disciplines such as GIScience, computer graphics, and medical imaging, there remain unexplored topics relevant to cartographers and spatial analysts. The issue of map projection distortion, in particular, is important whenever the scale of a map is small enough that Earth’s curvature cannot be ignored. Such an issue would be encountered by an analysis seeking to identify large-scale phenomena (e.g., a hurricane) from small-scale phenomena (e.g., wind speed recorded in 1 meter pixels). In this paper we examine the consequences of ignoring the variable latitudinal scaling in the Mercator projection across a large north-south extent in raster low-pass filtering; we offer a solution using variable-size kernels parameterized to areal distortion.
Further topics to address include the information content of raster images, and the evaluation of generalized rasters with regard to differences in information, or, equivalently, error. The Modifiable Areal Unit Problem remains an important, but largely unresolved issue in spatial analysis and representation since the work of Openshaw (1984); while not limited pixels, it is centrally important in any kind of structural raster generalization. Procedural issues yet to be explored include the effects of varied sequences of operators applied to rasters, and the effects of star vs. ladder approaches (Stoter 2005) with respect to particular operators.

**Case study: Raster processing with variable kernel shape**

Both local and global raster processing operations are significantly affected by projection distortions — a fact frequently ignored. An example consequence of this is how slopes in polar areas are heavily flattened using a DEM in the Mercator projection. In contrast, raster processing is motivated not by pixel re-projection, but by altering the properties of the modeled field or set of objects; when a cartographer performs DEM smoothing, she probably wants to smooth every terrain feature to an equal degree, and not have features at high latitudes generalized less than those at low ones.

To treat this issue we offer an approach using variable kernel shapes that are morphed at each pixel according to the local Tissot ellipse of distortion. In this case it is important that the initial kernel size should be defined not in pixels, but in raster projection units (meters).

The following algorithm is applied:

1. Define initial shape of the kernel (commonly rectangular or ellipse) and its sizes in both $X$ and $Y$ directions

2. Calculate the extent of the raster in geographical units (degrees).

3. Sample raster area by the control points, which are equally spaced in degrees. Sampling distance is defined by user or can be calculated as a function of raster geographical extent and resolution.

4. Calculate the parameters of distortion ellipse at each point using projection equations. These parameters are $m$ (meridian scale), $n$ (parallel scale) and $\theta$ (local angle between parallel and meridian).

5. Using distortion ellipse parameters, define the local matrix of affine transformation.

6. Transform initial kernel shape and rasterize it. Round the size of the kernel to the odd number if needed.

7. For each pixel in the initial raster find the closest control point and assign its number to the pixel.

8. Process the whole raster using kernels from assigned control points.

We selected a combination of mean filtering and the Mercator projection for preliminary testing of the approach. The Mercator projection’s linear distortions are calculated simply as $m = n = 1/\cos(B)$ at every point, where $B$ is the latitude, and $\theta =$
Thus the kernel need only be scaled according to value of \( m \). Control points were sampled each degree by latitude.

The source raster was a 5 km resolution DEM covering the European part of Russia. First, we processed the DEM using a 35×35 km (7×7 pixels) square kernel in a traditional, non-variable-kernel approach. Then we applied our variable kernel approach, with size varying from 7×7 pixels at 30° to 67×67 at 84°, always covering approximately the same 35×35 km area. Results of filtering show that our approach produces more geographically sound results than simple filtering. Compare the Kola peninsula in the northern part of the map and the Caucasus in the southern part (Figure 1). Enlarged terrain features at high latitudes appear to be generalized much less then southern territories by simple filtering. In contrast, application of a variable kernel shape smoothes features of the same geographical area, which can be important if the smoothing is performed for geographical analysis applications.

The difference becomes more evident when resulting and initial rasters are re-projected into Albers Equal Area projection (50° central meridian, Figure 2). In this case it can be seen that non-variable generalization is inconsistent throughout the area, having very detailed polar regions and smoothed subtropics. And in the case of the variable kernel shape, the generalization is the same over the whole area (particularly from south to north), showing that this map introduces little projection distortions. Figure 3 illustrates the same behavior using contour lines; a greater spatial frequency and complexity of contour lines derived from the non-variable-kernel smoothing indicates that it has failed to generalize northern latitudes to the same degree as more southerly ones in terms of generalizing the features similar in size.
Distortions should also be taken into account for calculation of global raster statistics. While in the case of raster minimum and maximum all pixels should be considered, the mean and standard deviation characteristics should better be taken from a distortion-aware approach. In this case the field of areal distortion can be calculated, then normalized and used as probability density function where higher density is in the areas with smaller area distortions. Then random sampling with a given density of points can be used to extract the values and calculate mean and standard deviation of raster values.
Conclusions

We have illustrated that the generalization of rasters is a nuanced procedure that requires attention to several spatial and statistical sources of error. Most notable among these are the effects of projection distortion across large areas, which make supposedly uniform pixels in fact not uniform, and the Modifiable Areal Unit Problem, which confounds statistical aggregation and analysis, such as seen in raster coarsening and resampling. We have begun to address the issues raised by implementing a method using variably-sized kernels which are calibrated using the areal error equations of the map projection used. Further work will address statistical aggregation techniques and methods of measuring structure and information in continuous raster surfaces to enable characterization and evaluation of generalized images.

Literature Cited


Peuquet D.J. Raster processing: an alternative approach to automated cartographic data handling, American Cartographer, vol. 6, 1979, pp. 129—139;


