DETAIL FILTERING IN GEOGRAPHIC INFORMATION VISUALIZATION

Laura Tarantino
Department of Electrical Engineering, University of L'Aquila, Italy

Keywords: fish-eye views, focus + context visualization, geographic maps, interaction and visual models, level-of-detail (LOD) models, map exploration, multiresolution, spatial databases, surface simplification, topological models

Contents

Summary

Geographic information systems (GIS) deal with storing, querying, manipulating, and displaying geographic information. In conventional cartography the database almost coincided with the map, thus implying an *a priori* detail filtering with a reduction of the original set of data, some of which were lost. This limitation has been overcome by modern GISs and by technological developments in automated data capturing tools, allowing the collection of huge volumes of geographic data representing areas of interest with extreme accuracy. As a consequence, manipulation and rendering of complex geographic datasets in real-time applications may require processing resources beyond the power of state of the art computers. The necessity of detail filtering does not disappear but is rather shifted to visualization and interaction tasks. The problem is particularly apparent when data have to be transmitted over networks (for example, on the World Wide Web), because of bandwidth limitation. A controlled data reduction is necessary for dynamically adapting the volume (and the precision) of data to the needs of specific interaction tasks. Surface simplification, level of detail and multiresolution

models, and topological models are possible approaches all aimed at detail filtering, sharing the goal of improving performance in data analysis and visualization. Multiresolution models appear particularly flexible and suitable to be applied in emerging information visualization strategies.

1. Introduction

To an ever-growing degree, information technology is based on interactive visual media, and visualization techniques are emerging as the primary support to decision-making tasks. Allied to this trend is the growing computer capacity to process and render heavy graphics in real time, along with the increasing technological capacity to make visible things that our eyes could not see unaided (for instance, satellite pictures). Human existence is more visual and visualized than ever, to the extent that the term "visualization," from referring to an internal construct of the mind, is nowadays used to denote an external artifact. A modern definition for "information visualization" is "the use of computer-supported, interactive visual representation of data to amplify cognition." Cognition enhancement is achieved when visual externalizations are designed so as to take advantage of the human processing system. Hence, visual interfaces constitute a crucial aspect of modern computer applications, and are capable of determining the success on the market of software products. This is particularly true when data handled by the application program are inherently graphic, as in GIS (see also *Interacting with GIS: From Paper Cartography to Virtual Environments).*

GIS deal with storing, querying, manipulating, and displaying geographic information. The core of a GIS is a spatial database management system handling the spatial and the thematic components of the database, describing what is present and where (for more details on this kind of management, see *Advanced Geographic Information Systems)*. More precisely, the *spatial component* encodes the geometric aspects of the physical objects under consideration (for example, location, shape, size), while the *thematic component* contains information about the non-geometric properties of the realm of interest. A GIS normally includes modules devoted to application-specific tasks, such as map production, spatial analysis, and data visualization. Though modern GIS functionality goes far beyond the mere automation of map production and map analysis (computers provide a means to interact with data in ways that were impossible with printed information in conventional mapping), undoubtedly cartography can be regarded as the starting point in the evolution of GIS.

For centuries geographers have collected geographic data and surveys to be recorded in pictorial forms by mapmakers and cartographers, and to then be used in a number of scientific, government, military, business, and administrative applications. However, while in conventional cartography the map was the database, in a GIS a map is just a particular view of the spatial database at a given time (see *Spatio-Temporal Information Systems*). Differences between conventional cartography and modern GIS refer to both the recorded data and the available data analysis. Constraining the database in a map implied an *a priori* detail filtering (with a great reduction of the original set of data, some of which were irremediably lost), for the production of a static document whose updating could be prohibitively expensive. Conversely, the growing computer data recording capability, paralleled by technological developments in automated data capturing tools, has allowed the collection of huge dynamic volumes of geographic data representing areas of interest with extreme accuracy. Spatial data models of greater and greater detail have come into use in several disciplines: in remote sensing, terrain data is acquired from satellite photographs; in computer vision, range data is captured using scanners; in volume visualization, isosurfaces are extracted from high-resolution datasets, just to mention a few examples potentially giving rise to extremely high volumes of spatial data. Consequently, manipulation and rendering of complex geographic datasets in real-time applications may require processing resources beyond the power of state of the art computers. The necessity of detail filtering, hence, did not disappear, but rather shifted to visualization and interaction tasks. The problem is particularly critical in distributed and World Wide Web applications, since the transmission bandwidth of networks is usually not sufficient to deal with the full resolution of geographic datasets (see *Web-Based Spatial Decision Support: Technical Foundation and Applications).*

This article surveys existing approaches for the controlled reduction of spatial data that have to be visualized and managed during interaction tasks. A brief presentation of spatial data models will be the starting point of the discussion, since not only does the adoption of a particular model influence the type of data that can be used, but it also determines the kind of spatial analysis that can later be undertaken. The article then discusses methods for simplifying complex surfaces (representing, for example, terrains), level-of-detail and multiresolution approaches (based on the idea that resolution is dynamically adapted to application needs and users' interests), and topological models (aimed at filtering out metrics information while retaining information about the relative position of objects). The article finally presents the main features of a topological interaction model which supports visual interfaces characterized by details-on-demand data presentation and zoom-based navigation.

2. Interactive Database Applications

We have already remarked how spatial analysis (and the consequent necessary kind of detail filtering) depends on the adopted spatial data model. A "data model" provides a mathematical formalism including a notation for describing data, and a set of operations used to manipulate that data. In other words, a data model provides mechanisms for formally describing the reality of interest, and for interacting with it. A "database" consists of some collection of interrelated, integrated, shared, and persistent data, structured according to some data model. "Spatial databases," collecting spatially referenced data, separate the information into a "spatial component," encoding the geometric aspects of the physical objects under consideration (for example, location, shape, size), and a "thematic component," containing information about the non-geometric properties of the realm of interest.

2.1. Levels of Abstraction

Even if the only data that actually exist are at the physical level (organized as a collection of files), there are many level of abstractions between the computer, dealing with bits, and the final user, dealing with logical concepts. Traditionally, a three-tier architecture is used to separate the user applications from the physical database:

1. The *internal level* is concerned with the physical representation of data in terms of storage and access structures.
2. The *logical level* is concerned with the representation of data in terms of the *logical data model* provided by the database management system.
3. The *external level* provides external views, that is, portions of the logical database (often at the same level of abstraction as the logical database), that are of interest for individual users or classes of users.

In modern interactive database applications, two additional interdependent levels of abstraction on top of the above ones have to be considered, to formalize visual interaction with data:

- The *visual model*, concerning the way in which the final users perceive data, provides the mechanisms for displaying data and for visually interacting with them.
- The *interaction model*, intermediate between the visual and the logical models, bridges the semantic gaps between the two, acting as a formal counterpart of visual elements and primitives. The interaction model provides interaction structures (abstractions of the database, hiding details non necessary for specific interaction tasks), and primitives acting on such structures. Most of the simplification approaches discussed in this article pertain to this level of abstraction.

The design of any interactive application requires a series of interdependent stages, to define the different database schemata and interfaces between them. A crucial stage is the *conceptual design*, aimed at defining a *conceptual schema*, that is, a formal description of the reality of interest, independent of the technical environment constraints, and reflecting the requirements of the specific applications (therefore focusing on semantics of data and their interrelationships, along with expected usage). In the remainder of this section we briefly present the main approaches for the conceptual design of spatial data. For a complementary discussion on this issue, see *Conceptual Modeling of Geographic Applications*.

2.2. Conceptual Views of Spatial Data

The description of geographical phenomena requires a description of what is present and where it is. Several approaches can be used for the formalization of space and of spatial properties; the adoption of a particular model influences the type of data that can be used and the kind of spatial analysis that can later be undertaken. The two extremes in the range of conceptual approaches are:

- To perceive space as being occupied by entities which are described by their properties and mapped using a coordinate system, or
- To imagine that attributes of interest vary over the space as some continuous mathematical function or field.

The choice of approach generally depends on the scientific or technical application: the entity-based approach is appropriate for administrative contexts (for instance, cadastral applications, which view areas as a series of distinct units), while the continuous-field approach looks more suitable when the understanding of spatial processes in the natural environment is the aim of the user of the database.

2.2.1. Describing Entities

When viewing the space as populated with objects, the first step is recognizing the entities (town, rivers, streets, and so on); the next steps are listing their attributes, and defining their geographic properties (such as boundaries and locations). A basic spatial entity is further described in terms of three basic geographical data primitives, namely points, lines, and areas, which are the fundamental units of the vector data model.

The "vector data model" represents space as a series of discrete entity-defined point, line, or polygons units, geographically referenced by Cartesian coordinates. While points, lines, and polygons are supposed to be unchanging, it has to be observed that they refer to a particular resolution level: while the representation of some entity may remain unchanged across resolution levels (for example, in the case of cables, in utility-oriented applications), in general changing the scale implies a change in the representation (a town might be a point at country level, and a polygon at regional level).

2.2.2. Representing Continuous Fields

In the continuous-field approach, geographic space is represented in terms of continuous Cartesian coordinates in two or three dimensions. Though attributes (for example, temperature, air pressure, elevation) are usually assumed to vary smoothly and continuously over the space, the variation is generally too complex to be captured by simple mathematical functions. The usual approach is to discretize the geographical space into sets of single spatial units (such as square, triangular, or hexagonal cells), or into irregular triangles or polygons, tessellated to form geographic representations. The resulting tessellation is considered a reasonable approximation of reality at the level of resolution under consideration.

Meshes of irregular triangles have been used for a very long time in land surveying, as well as for diverse natural environment phenomena, for example, for dynamic modeling of wind fields, or surface water movement. A major advantage of irregular triangulations is that the density of the mesh can easily be adjusted to the degree with which the surface varies, thus favoring a variable resolution of the data. On the other hand, it is not uncommon that data capturing processes produce surface meshes with millions of polygons, which makes it necessary to simplify them.

The most common alternative to irregular triangulation is the grid cell representation of space (known as the "raster model"), which divides the surface into square cells, whose size is determined by the resolution required to represent the attribute variation for a certain purpose. Regular tessellations may also be nested to provide more spatial detail.

3. Displaying Maps at Different Levels of Detail

The characteristics of modern information-intensive applications push towards a controlled reduction of the handled data: the computing power and the transmission bandwidth of networks are, in fact, not sufficient to deal with the full resolution of spatial datasets (see *Spatial Data Quality*). The need arose for a trade-off between accuracy of representation and time/space constraints, particularly in real-time applications. The basic idea is to adapt the resolution dynamically to the needs of specific interaction tasks, possibly taking into account users' interest in displayed data (which might imply a variation in resolution over different areas of the entity).

To achieve this effect, the concept of "level of detail" (LOD) is introduced, sometimes summarized as "always use the best resolution you need – or you can afford – and never use more than that." Of course the principle has to be characterized to the needs of the specific application, interaction task, or visualization strategy. For example, in a flight simulator the LOD will be increasing with the distance from the viewpoint; in transmission over a network, the area of interest can be transmitted by increasingly finer levels of detail, allowing the client to interrupt the process when the desired LOD is reached; and in visualization techniques taking into account the user's degree of interest, the LOD will vary over the displayed area, enhancing accuracy in the regions of interest.

Despite application-specific differences, there is common ground among LOD models: all aim at defining sequences of gradually simplified representations of the same area that are structured either into a multilayer model or into a hierarchical model described by a tree. Appropriate transitions between levels of detail have to be defined by the interaction model (in other words, a LOD interaction model provides both a map structure and rules for traversing it). A gross categorization of LOD models singles out two classes, depending on whether representations at different levels of detail are stored explicitly, or implicitly defined to be derived at run-time:

- M*ulti-representation* models provide multiple representations of the same data at different resolution levels, and
- "True" *multiresolution* models generally provide an initial coarse map, plus the description of legal modifications to derive additional, finer maps from the initial one.

True multiresolution models generally perform better than multirepresentation models in terms of memory requirements and resolution adaptivity, particularly in the case of tessellated surfaces.

4. Multirepresentation and Multiresolution of Tessellated Surfaces

In the multirepresentation model for tessellated surfaces, a LOD representation consists of a collection of meshes of different sizes, each representing the object at a different resolution. Each mesh is defined independently of the others, and it is tagged with some range of detail. A multirepresentation may be built by the repeated application of a surface simplification algorithm with different parameters of simplification. For example, Figure 1 depicts the multirepresentation of a rectangular domain as a sequence of finer and finer meshes. The sequence can be regarded as the evolution of a mesh through a process of simplification starting from the mesh at the finest LOD. Our first step is therefore the study of surface simplification methods.

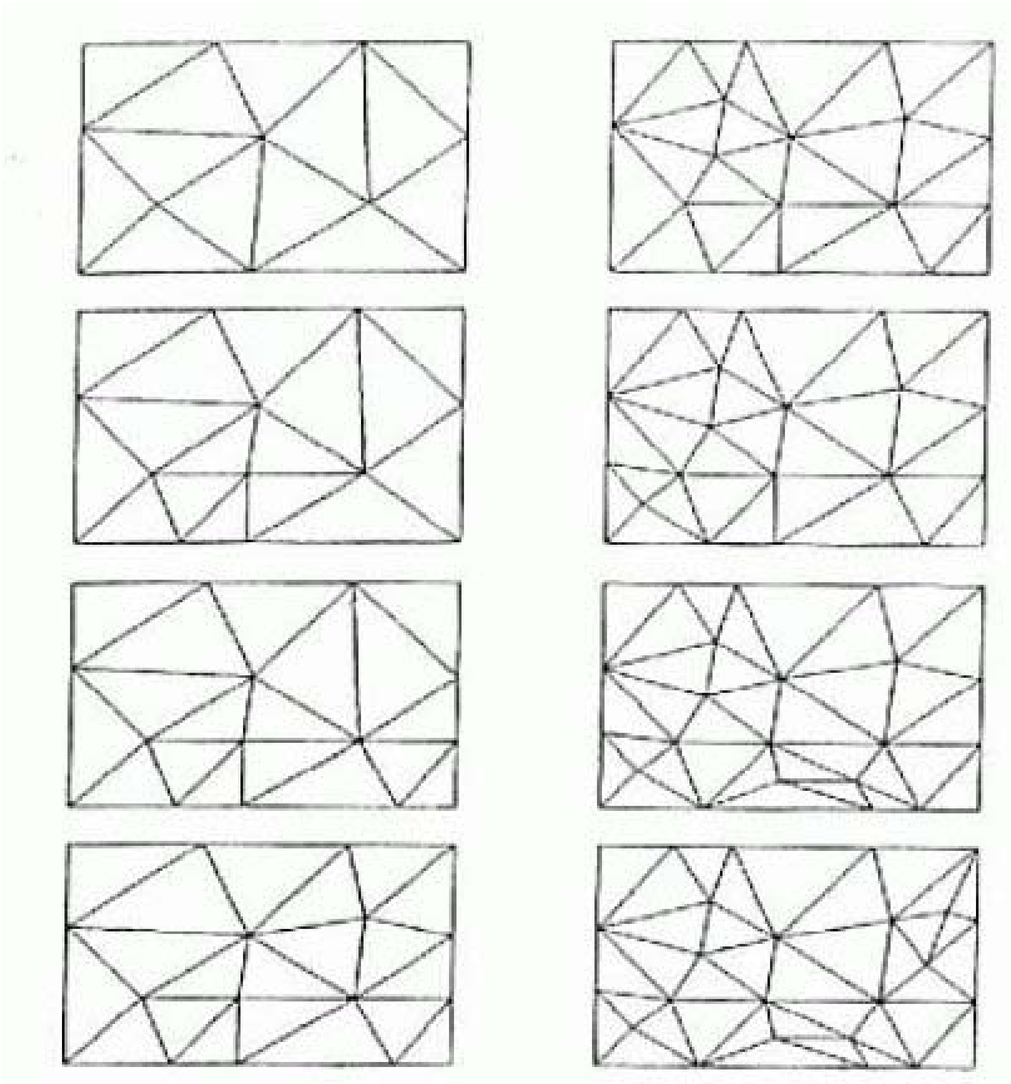

Figure 1. A multirepresentation of a rectangular domain. Source: Puppo (1998).

4.1. Surface Simplification

A variety of methods for simplifying curves and surfaces have been explored in a number of fields. There is common ground amongst methods for representing terrains developed in cartography, methods for approximating bivariate functions developed in computational geometry and approximation theory, and methods for approximating range data developed in computer vision. This is not surprising, since these disciplines share fundamentally the same technical problem, and cross-fertilization is therefore advocated. Algorithms for surface simplification may be categorized according to the class of surfaces on which they operate (that is, the object to be modeled):

- height field and parametric surfaces, or
- manifold surfaces.

The most general class of surfaces is that of non-manifold surfaces, which permit three or more triangles to share an edge, and arbitrary polygon intersections. Relatively few simplification algorithms can handle this degree of generality, and these are not discussed here. A brief overview of parametric and manifold surfaces follows.

-
-
-

Bibliography

Burrough P.A. and McDonnel R.A. (1998). *Principles of Geographical Information Systems*. Oxford: Oxford University Press. [This book presents a comprehensive introduction to the theory and practice of geographic information systems.]

Card S.K., Mackinlay J.D. and Shneiderman B. (eds.) (1999). *Readings in Information Visualization: Using Vision to Think*, San Francisco: Morgan Kaufmann. [This book is a collection of contributions on the various aspects of information visualization by several experts in the field. Some contributions are for deep specialists while some provide overviews and perspectives for a more general readership.]

Cicerone S., Frigioni D. and Tarantino, L. (2001), *Exploration of Geographic Databases: Supporting a Focus+Context Interaction Style*, Technical Report Rep. R.01-62, Dept. of Electrical Engineering, University of L'Aquila, to appear in *Journal of Applied and System Studies*, Cambridge: Cambridge International Science Publishing. [This is an article presenting a topology-based LOD interaction model for focus+context visualization of geographic maps. It constitutes the basis of the description in Section 6.]

Heckbert P.S. and Garland M. (1997). *Survey of Polygonal Surface Simplification Algorithms*, Technical Report, Carnagie Mellon University, Dept. of Computer Science. [This is a survey of methods for simplifying and approximating curves and polygonal surfaces.]

Kuijpers B., Paredaens J. and Van den Bussche J. (1995). Lossless representation of topological spatial data. *Advances in Spatial Databases*, Proceedings of the Fourth International Symposium SSD'95 (eds. M.J.Egenhofer and J.R. Herring), Lecture Notes in Computer Science 951, Springer, pp. 1–13. [This paper introduces the definition of PLA structure described in this article.]

Kuper G., Libkin L. and Paredaens J. (eds.) (2000). *Constraint Databases*, Springer. [This book presents essential principles and applications of constraints databases. Part II is devoted to spatial and temporal data.]

Puppo E. (1998). Variable resolution triangulations. *Computational Geometry* **11**(3–4), 219–238. [This paper introduces multitriangulation, the general framework for multiresolution meshes discussed in Section 4.2.1. It is the source for Figures 1, 2, and 3 in this article.]

Puppo E. and Dettori G. (1995). Towards a formal model for multiresolution spatial maps. *Advances in Spatial Databases,* Proceedings of the Fourth International Symposium SSD'95 (eds. M.J.Egenhofer and J.R. Herring), Lecture Notes in Computer Science 951, Springer, pp. 152–169. [This paper presents an entity-based multiresolution model for plane maps composed by points, lines and regions.].

Puppo E. and Scopigno S. (1997), *Simplification, LOD and Multiresolution: Principles and Applications*, Tutorial Notes PS97 TN4 EUROGRAPHICS'97, **16**(3) (eds. D. Fellner and L. Szirmay-Kalos), Eurographics Association, Oxford: Blackwell. [These tutorial notes provide an introduction and discussion of simplification methods, level of detail models, and multiresolution models for surface meshes, along with their applications. Necessary theoretical background material is also included making the notes self-contained.]

Stell J.G. and Worboys M.F (1999), Generalizing graphs using amalgamation and selection, *Advances in Spatial Databases,* Proceedings of the Sixth International Symposium SSD'99 (eds.R.H. Guting, D. Papadias and Lochovsky), Lecture Notes in Computer Science 1651, Springer, pp. 19–32. [This paper presents a multiresolution model for network data structures, a simpler case than fully two-dimensional data structures, applicable, for example, in utility applications.]

Biographical Sketch

Laura Tarantino is Associate Professor of Computer Architectures at the Department of Electrical Engineering of the University of L'Aquila, Italy. After receiving the Laurea degree in Electronic Engineering from the University of Rome "La Sapienza" in 1986, from 1987 to 1990 she was Research Assistant at the Department of Computer Science of the same university, and lecturer at the University of L'Aquila. In October 1990 she joined the Department of Electrical Engineering of the University of L'Aquila, as a Researcher and Assistant Professor in Computer Science, and has been Associate Professor since 1998. She has been also a consultant to major Italian computer science industries. Her primary research activity is in human–computer interaction with databases, with focus on the definition of information visualization formalisms, on the formalization of interaction models bridging the gap between the underlying formal data model and the visual model, and on user interface development systems for database applications. She has also worked on object-oriented databases, with a particular interest in the definition of query languages based on the integration between logic programming and the object-oriented paradigms. In these areas she has published over 80 scientific papers in international journals, books and conference proceedings. She is a referee for several international journals and has served on the organizing committees and program committees of several international conferences and workshops. She was the Organizing Committee Chair in the 2000 Edition of the Italian National Conference on Advanced Systems for Databases. She was Program Co-Chair for the 1998 Edition of the International Working Conference on Advanced Visual Interfaces (AVI 1998), Program Chair for AVI 2000, and is on the AVI Steering Committee.