AUTOMATED INSETTING: AN EXPERT COMPONENT EMBEDDED IN THE CENSUS BUREAU'S MAP PRODUCTION SYSTEM

April A. Martinez
Geography Division
U. S. Bureau of the Census
Washington, DC 20233

ABSTRACT

The U.S. Bureau of the Census operates in a mass map-production environment, producing hundreds of thousands of unique, individual maps in batch mode. Given a very short window of time, interactive cartographic decision-making is impossible. Consequently, the Census Bureau has developed, and continues to enhance, "expert" components for its automated map production system. These expert components represent the quantification of collective cartographic decision-making processes. In use at this time are automated names placement, automated scaling, and automated insetting.

This paper discusses the development of one of the expert components: automated insetting. It describes how cartographers at the Census Bureau identify and quantify approaches for inset requirements, and examines how they developed the automated system once the approaches were identified and rules were established. The paper also addresses plans for future improvements of the existing system.

INTRODUCTION

Most cartographic design decisions are made based upon scientific and artistic rules of cartography. Cartographic design is a process of selection of scale, symbolization, and so forth. (Dent, 1985) The rules are continually applied while producing a map in an interactive automated environment. When, however, many thousands of maps are produced by automated means and constrained by limited time schedules, cartographic interaction is impossible and the work must be done in batch mode. It then becomes necessary to design an automated system in such a way as to imitate a cartographer's decisions, thereby developing an expert system.

The expert cartographic system should reflect the cartographer's decision-making process. The most important and difficult component of developing an expert system is the quantitative definition of the cartographer's reasoning process, particularly the implementation of aesthetic preferences.

At the Bureau of the Census, production of the maps for the 1990 census data collection is totally automated. Interactive cartographic decision-making is impossible. There are at least eight different types of maps, each having nationwide coverage, resulting in an estimated

2.5 million unique map sheets. Production of these maps began in April 1988 and will be completed by March 1991. This tight time frame prohibits cartographers from interactively designing the maps. To produce the maps in batch mode, the Bureau of the Census has developed "expert" components for its automated mapping system.

Background

One of the "expert" components developed for the automated map production system was insetting. It grew directly from the Census Bureau's first attempts to automatically determine an adequate map scale. The traditional scale determination rule for one type of field map was that the length of the smallest side of census blocks must be at least one inch to accommodate "map spotting."* The direct application of this rule in production consistently resulted in map scales too large for most of the geographic areas mapped, consequently yielding many more map sheets than were required, and than a manual system would have produced.

To reduce the number of map sheets, the Census Bureau decided to change the scale determination rule by forcing only 60 percent of census block sides to equal at least an inch. This change produced numerous isolated areas on the map for which the scale was too small and unsuitable for map spotting. Adequate map scale to allow enumerators to successfully complete their field assignments was required. A scale too small for specific areas on a map limits the use of the map. As a result automated insetting was developed.

In the present environment, the production system first determines the latitude and longitude limits of the geographic entity being mapped, followed by the scale of the map. Map scales are not standard. They are determined by the use of the map and the feature content of the geographic area within the window. (Martinez, 1987) The scale chosen is considered the best scale for the majority of the area within the window. This means there can be areas within the window for which the determined scale is too large or too small. Areas within the window for which the map scale is too small are selectively plotted on separate sheets at larger scales. These areas are selected by the map production system's insetting component.

DEVELOPMENT

Automated insetting attempts to emulate the steps a cartographer takes when deciding which geographic areas within a map to inset. Defining the decision-making process was the first phase of development. After reviewing numerous maps and selecting areas that required insetting, cartographers at the Census Bureau found that

* A map spot is a dot penciled with an associated number on an enumerator map by a census enumerator to show the location of a housing unit, multiunit structure, special place, or business establishment for field operations.

their decision to inset is based upon the answers to three questions:

1. What constitutes dense areas?
2. Which features are to be considered when determining density?
3. What determines the extent of the insetted area?

Density

On 1990 census maps used for field operations, the driving factor for insetting is "feature density." Because of the nature of field operations, most of the features in the Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) File must be shown on the maps. The maps are used for a variety of census operations. Some are used to depict political or statistical area boundaries, such as a county or census tract, while others are used for map spotting. One map type in particular must label and display every census block within a geographic entity. Because of map requirements cartographic features on these maps cannot be generalized.* Therefore, the major problem for an automated insetting system is defining the density of features on the map and insetting those areas whose feature density adversely affects the use of the map for a specific operation. To meet this requirement, the original programs written for the insetting system attempted to allow a computer to analyze the density of the features on a map similar to the way a cartographer views the map.

Numerous factors were considered in defining areas that are considered to be "too dense." These were narrowed to eight factors quantifiable from data in the TIGER File. Eight algorithms for depicting density were developed and tested. The TIGER File structure is based upon entities known as 0-cells, 1-cells, and 2-cells. For a full description of these entities see referenced papers.** For the purpose of this paper, 0-cells can be considered as endpoints of a 1-cell, which in many cases are intersection points of line features. The 1-cells are lines connecting 0-cells, and can serve as bounding segments of a 2-cell. Two-cells are the smallest areas bounded by 1-cells. Aggregates of 2-cells make up higher-level census geography, beginning with census blocks. (Beard, Broome, and Martinez, 1986, 2)

The algorithms were:

1. One-cell, midpoint method. The coordinates of the endpoints of the 1-cells are added together and divided by two to get a midpoint.

2. One-cell, average of all points method. The coordinates of all the points along the 1-cell and the endpoints are added and the results divided by the count to get an average.

---

* "Generalized" in this sense refers to selectively eliminating specific types of cartographic features from a map to reduce clutter.

** See referenced papers by Broome, Kinnear, Boudriault, and McDowell for discussions on the TIGER System.

3. One-cell, endpoint method. The 0-cells are used.

4. Two-cell, envelope midpoint method. The maximum and minimum coordinates of the 2-cells are added and divided by two to get a midpoint.

5. Two-cell, weighted area centroid method. The area and geographic centroid of the 2-cell is determined and the centroid is assigned the value of the area.

6. Census block envelope midpoint method. The maximum and minimum coordinates for each census block are determined by aggregating the 2-cells that constitute the block. The sum is divided by two to determine the coordinates.

7. Census block 2-cell average centroid method. The 2-cell average centroid is derived by adding all the maximum and minimum coordinates of the 2-cells and dividing by the count to get an average centroid.

8. Census block weighted area centroid method. The area and geographic centroid of the block is determined and the centroid is assigned the value of the area.

A program was written to test each method. Production statistics such as processing time and computer cost were recorded for each algorithm. (Beard, Broome, and Martinez, 1986, 3)

After plotting the points from the intermediate files, four of the methods were retained for further research, because they were computationally the most efficient and observationally determined to best identify dense areas of features.* Of the eight methods, the four retained for determining feature density were:

1. The 1-cell midpoint;
2. the 1-cell average of all points;
3. the 2-cell, envelope midpoint; and
4. the block envelope midpoint method.

Further tests were made and analyzed for the ability to determine dense areas. The algorithm developed to examine density is based upon count of calculated points within a grid cell of a predetermined size. The size of the grid cell is related to the specific use of the map. For example, the size of the grid cell for a map that needs every polygon labeled with a block number that is approximately 0.24" wide and 0.08" tall, is 0.25" X 0.25"; this is smaller than the grid cell for a map that requires all linear features to be at least an inch long.

---

*Insetting based solely on the size of an area being labeled is an alternative method of insetting.

The count of occurrences within the grid cells for each method was stored in a large matrix. The resulting matrix of values was then smoothed by summing the counts of groups of nine grid cells and recording the total as the value for the center cell of the 3 x 3 group. This smoothing operation removes local irregularities due to the use of a single coordinate to represent a linear and/or areal feature (see figure 1). The matrix of smoothed grid values for each map was plotted and visually analyzed. By classing the grid values and shading the classes on the matrix, it became evident that every file clearly represented the general feature pattern of the map.

The next major step in developing the automated insetting system was to introduce cartographic expertise. This provided a way to interpret the grid cell values. Ten professional cartographers, with an average experience of five years in census map design, examined several plotted maps. They agreed that the need for an inset was related to the use of the map. Knowing the use of the maps, they were asked to mark on overlays the geographic extent of areas that required insets. In order to have a valid sample cartographers worked independently and did not discuss their results until after the test. In addition, none of the cartographers were shown the matrix of grid cell values created for each map.

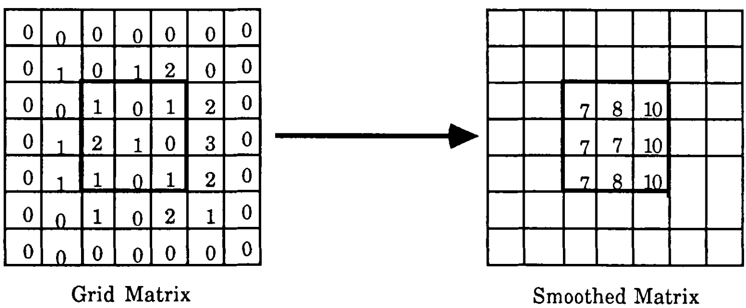

Figure 1. This is a portion of a map grid matrix. It shows the count of 1-cell midpoints that fall within the 0.25" grid cell. Summing the count of 3 x 3 grid cells and recording this total as the value of the center cell creates a smoothed matrix.

A comparison of the overlays revealed a close match between the choice made by the cartographers. The variances averaged less than one-fourth inch at map scale. This variance was considered acceptable for the intended purpose. So the inset overlays that the cartographers produced were registered to each other, and "average" inset overlays were produced.

Rules for density

The "average" inset overlays were registered to the plotted matrices of grid values of the same maps to see if any relationship could be established. A visual examination revealed that each of the four matrices appeared to have a specific breakpoint value; that is, a value within a grid cell that, if surpassed, was deemed too dense on the cartographers' average inset overlays. This density (breakpoint) value varies for each map type by:

1. The method used to determine density;
2. the grid cell size; and
3. the purpose of the map.

With known breakpoint values, the major obstacle of automated insetting was overcome; the system was able to determine feature density.

In production, the system creates, then scans, a grid cell matrix for cells with values higher than the breakpoint value. If higher values are found, the system checks systematically around the cell for adjacent cells that have values higher than that defined for an operation. If an adjacent cell is found, the system moves to this grid cell position and again checks for an adjacent cell, continuously counting the number of cells and storing the maximum and minimum grid cell position values. Finally, the system returns to the starting cell position. The system stores the results as a possible inset window only if it finds at least three cells adjacent to each other.

<u>Rules for qualify</u>

A map can have numerous insets based on feature density. However, some of these insets may be unnecessary depending on the census operation using the map. A major concern for field maps is the number of map sheets; if there is any way to reduce the number of sheets without affecting the quality of the census operation or the usability of the map, then an attempt is made to reduce the number of sheets. One way to reduce the number of sheets is by eliminating insets. Therefore, before a possible inset window determined by density alone can become an inset, it must qualify based on rules of inset qualification.

The rules of inset qualification were and are being determined through discussions with cartographers pertaining to the map information on which they base their judgement for eliminating insets. Some of the rules defined are general and are applied to every map generated in production. For example, some maps show features in fringe area outside the geographic area being mapped; however, it is unnecessary to have insets of dense areas within the fringe, so one of the first rules of qualification is that an inset must be within the geographic area being mapped (see figure 2).

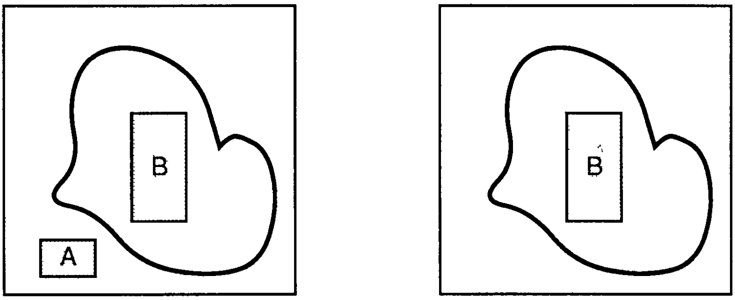

Figure 2. Example of an inset qualification rule: Only keep insets within the entity being mapped. Inset "A" is removed because it is outside the entity boundary.

Other rules for inset qualification are specific to the map use. A rule for the map that labels every polygon with a block number is that a possible inset window qualifies as an actual inset only if at least five whole census blocks are within it. (It was reasoned that block numbers can be arrowed from adjacent blocks into clusters of fewer than five blocks without significantly hindering readability.) In another operation, the map is used by officials to verify and update the location of their government's political boundaries; consequently, for that map type only insets with political boundaries would qualify as actual insets.

The rules of qualification are the most powerful tool within the automated insetting system. By defining the rules of inset qualification based upon anticipated map use, the cartographer affects the entire design of the map and the effectiveness of the census operation for which the map is intended.

<u>Rule for nearness</u>
A cartographer manually creating a map visually determines which areas on a map are too dense at the desired map scale. Some of the areas are close enough to other potential areas to be grouped together in one inset. The cartographer also visually determines the best geographic extent of an inset. The insetted area can be either an amorphous shape or a rectilinear window. The cartographer knows where to stop expanding the limits of the inset in either case.

In the automated system, the rule of nearness was defined by again having cartographers review numerous maps for which they determined a specific distance at map scale that constituted nearness on a map. The distance was further redefined during test production when it was decided that the system was combining too many insets, thereby creating large, multisheet insets.

The limits -- or the latitude and longitude windows -- of the actual insets were the maximum and minimum limits of the grid cells making up each inset. However, it was found that an inset's limits many times were coincident with part of an incorporated place. As a result, the inset limits were expanded by one-fourth of the nearness distance in an attempt to inset the entire incorporated place. This results in a geographic area portrayed entirely at the larger inset scale.

In production, once the system determines if all the inset windows qualified based on the inset qualification rules, it eliminates insets that fall within another inset. The system also expands windows of insets that overlap each other. The result is one larger inset window (see figure 3). Finally, the system checks if multiple insets are close enough to each other to merge to become one larger inset.

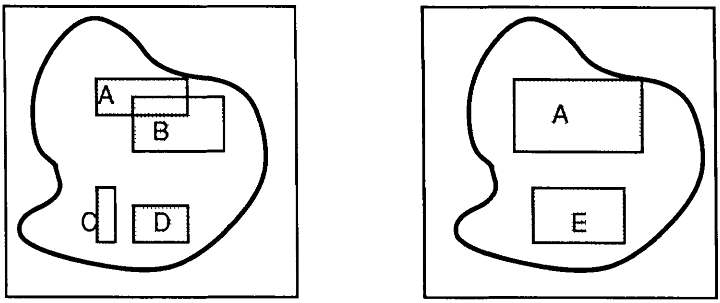

Figure 3. Example of overlapping insets and insets near each other merged to make larger insets.

ENHANCEMENTS

An automated cartographic mapping program used in batch production needs an intelligent insetting system. Insetting by feature density is the most common method used at the Bureau of the Census. The rules are defined by map use, which varies by map type. The major rules for insetting are first to define the density breakpoint value for the map based on grid cell size and algorithm used to determine feature density. Second the inset must qualify based on the rules of inset qualification. Finally, the extent of the inset must be determined.

Modifications to the rules and algorithms are applied as new and unanticipated circumstances arise. New density values for census operations have been defined and rules have been added to inset qualification. Most enhancements to the system are in the rules for inset qualification.

At this time, the Bureau of the Census is beginning to develop a publication mapping system. One area of research involves the feasibility of having the automated insetting system inset entire incorporated places if any portion of an incorporated place falls within an inset. This could eliminate the need to create new plot files solely for the purpose of creating maps for each incorporated place; the inset would serve as a larger-scale inset area within a county map as well as an incorporated place map. However, many issues need to be addressed on the overall affect of this concept. New cartographic rules would be required for the system. For example:

> How does the system handle urban sprawl? Should the system inset a network of features that is not dense simply because it is within the limits of an incorporated place?; if not, where does one stop expanding the inset? How does one handle corporate corridors? How is the number of sheets affected?
>
> What happens if more than one place falls within an inset window? Does the system make two insets, or does it put both incorporated places on one inset map? Should it make an inset of the incorporated place that covers the biggest portion of the original inset?
>
> Will the whole map become an inset?

These are only a few of the questions that are being considered, as additional rules for the system are being defined and algorithms developed.

CONCLUSION

Map production for the 1990 census maps began in spring of 1988. Automated insetting has been successfully implemented for every operation to date. In batch production the insetting component first determines possible insets by density. Once these have been selected, the system then checks if the inset qualifies, and finally it merges insets that are within or near each other.

The system has grown rapidly from its original intent; however, the basic concept of a system defining feature density and insetting areas with areas of high feature density is still the same. As the insetting system continues to be used, new rules are defined and added to the component cartographically enhancing the quality of the insets defined by the system. Future developments of automated mapping systems for batch production must include inset determination if for no other reason then to tell the system when a given map has areas too dense for effective use.

REFERENCES

Beard, C., Broome, F., Martinez, A. 1987, Automated Map Inset Determination: Proceedings Auto Carto 8, pp. 466 - 470.

Boudriault, G. 1987, Topology in the Tiger File: Proceedings Auto Carto 8, pp. 258 - 263.

Broome, F. 1984, Tiger Preliminary Design and Structure Overview: The Core of the Geographic Support System for 1990: presented at the 1984 Annual Meeting of Association of American Geographers.

Dent, B. 1985, Principles of Thematic Map Design: Addison-Wesley, p. 22.

Hammond, R., McCullagh, P. 1978, Quantitative Techniques in Geography: An Introduction: Oxford University Press, pp. 36 - 40.

Kinnear, C. 1987, The TIGER Structure: Proceedings, Auto Carto 8, pp. 249 - 257.

McDowell, T., Meixler, D., Rosenson, P., Davis, B., 1987, Maintenance of Geographic Structure Files at the Bureau of the Census: Proceedings Auto Carto 8, pp. 264 - 269.

Martinez, A. 1987, Applications of Expert Rules in Automated Cartography", presented at the 1987 Applied Geography Conference, pp. 249 - 257.