

Automated Detection and Delineation of Inset Maps

William G. Thompson

ABSTRACT: In order to create a useful map, the cartographer must select a scale at which the map reader can distinguish all features shown on the map and read all of the feature labels. However, the choice of a scale is also constrained by the size of the map's sheet, or in cases where the map can be tiled into multiple sheets, by the costs and inconvenience of printing and working with a large number of sheets. When the feature density pattern allows, one way to save paper and maintain legibility is to make the map at more than one scale: a small scale suitable for most of the map, while small areas of unacceptably dense features are shown separately on inset maps at larger scales. Creating inset maps requires the cartographer to make a series of complex, interrelated decisions. Are the feature density patterns such that a map sheet configuration employing inset maps is the most efficient? What is an appropriate scale for the main part of the map? How can areas where features are unacceptably dense be detected? How will the boundaries be drawn between the inset areas and the rest of the map? To employ an inset map capability in a non-interactive automated computer mapping system, the mapping software must answer these map design questions through algorithms that analyze feature density patterns and apply sound cartographic logic to issues of sheeting, scaling, and inseting. The Census Automated Map Production System (CAMPS), which was recently put into production by the U.S. Census Bureau to meet the mapping needs of the 2010 census, is a cartographically intelligent system that does this analysis and decision-making without any human intervention. This paper describes CAMPS' logic of scale selection, its discovery of clusters of dense features by means of the DBSCAN algorithm, the ways it delineates boundaries between inset areas and the rest of the map, how it determines the most efficient sheet configuration, and some of the difficulties and anomalies it encounters.

KEYWORDS: census maps, map scale, inset maps, feature density analysis, batch mapping

Introduction

In its many data gathering operations, its geographic information sharing programs, and its public data distribution, the U.S. Census Bureau is both a huge user and producer of maps. Since the advent of its digital TIGER database system, fully automated non-interactive map production has been a key part of the Census Bureau's activities (Trainor, 1990). Because the volume of mapping work and demanding production schedules preclude the possibility of any individual map-by-map design or editing, selection of map scale and sheet configuration must be made by software algorithms guided by parameter values. These algorithms apply a combination of cartographic principles, past experience, and analysis of the data to be mapped to create a product suited to the need and delivered on time.

Census Bureau map designs are highly project specific. Rather than creating a single scale, all purpose national map series, each map in a national census map series is individually focused on a subject area of interest—a legally or statistically defined geographic entity or a data collection work area. Each entity is mapped at whatever scale best fits the density of the features to be shown and the intended map use. If it is

necessary to use more than one sheet to create the map at the chosen scale, a configuration of multiple sheets will be created. The mapping system's performance is also crucial, as all the entities in the nation must be mapped repeatedly during the three years around the decennial census, and geographic updates gathered from a previous operation must be reflected on the maps made for the next operation just a few weeks or months later.

For the 2010 census, the Census Bureau has redesigned the MAF/TIGER database and along with it has written and put into production a new high volume batch mapping system, the Census Automated Map Production System (CAMPS). While its basic structure and processing flow is similar to the previous in-house census mapping systems, CAMPS employs some new strategies of cartographic analysis. This paper will discuss the theory and application of its algorithms that choose a map scale, detect areas suitable for inset maps, and delineate the inset maps.

Many census operations require paper maps created as digital files at census headquarters, plotted or printed at various locations around the country, to be used in the field or in a small office. The national scope of these projects magnifies the importance of using the minimum number of map sheets to map each entity, consistent with the map's legibility and purpose, in order to keep under control the costs of distribution and the burden of collating, storing, and using the thousands of maps created. Earlier census mapping systems have employed an automated inset map creation capability as the way to make the most efficient use of limited map sheet space. Muehrcke (1986) concurs that the advantages of mapping areas of varying density at different scales outweigh the disadvantages.

CAMPS draws its maps on two mutually exclusive canvasses. The parent canvas, tiled into multiple sheets if necessary, covers the entire entity of interest. Inset maps are clipped out of the parent canvas onto their canvas, leaving behind a blank area on the parent. Each inset is mapped on a separate sheet at its own larger scale appropriate to the features it contains. The complete sheet configuration of a CAMPS map consists of the parent sheets, the inset sheets, and an index sheet showing the outlines of the parent sheets and inset sheets for user orientation.

The methods for calculation of the complete sheet configuration for each CAMPS map have their origins in the early census batch mapping system described by Beard and Robbins (1990), and have evolved to match data formats and map design requirements. A four step process takes place at the start of each mapping job. First, a scale is selected for the parent map. Second, areas that meet a predefined feature density threshold are detected and designated as possible insets. Third, a boundary is drawn around each inset area designating the boundary between each inset map and the parent map. Finally, CAMPS tests other parent scales which lead to different sheet configurations to make sure it has selected the configuration with the fewest total sheets. The choice of a parent scale determines both the number of parent sheets and the size and number of the inset maps.

Selecting a Parent Map Scale

The primary consideration in selecting a map scale is that all features crucial to the purpose of the map be drawn clearly and identified legibly. Therefore CAMPS bases its scale selection on the placement of text labels on important features drawn on the map. For area features, the label for each feature should fit inside the feature with some blank space around it. Each linear feature should be long enough that its corresponding label can be placed alongside it without overprinting

other features or labels. Point features must be far enough apart that an identifying label can be placed next to them without printing over another point.

To save processing time, CAMPS will not take into consideration all map features in its scale calculation. Instead it will base its determination on the features of a single key layer or feature type called the analysis layer. The analysis layer must be chosen carefully to result in a successful scale calculation. Because the features in the analysis layer represent all of the features on the map, their density pattern must be similar to the densities of the other feature types. The analysis features must be numerous enough to provide a good sample of the various sizes of all of the features shown. They also should be distributed somewhat evenly across the entire map. The examples in this discussion will be limited to the situation where the analysis layer consists of area features. Many CAMPS mapping projects use census blocks as the analysis layer, because census blocks are numerous and cover the entire landscape. Most importantly, census blocks conform to the pattern of streets, highways, streams, and other basic linear features, so their density closely corresponds to many other features shown on the map. Figure 1 shows the census blocks for a small Midwestern county, which will be used as the analysis layer in scale calculation for a map of this county.

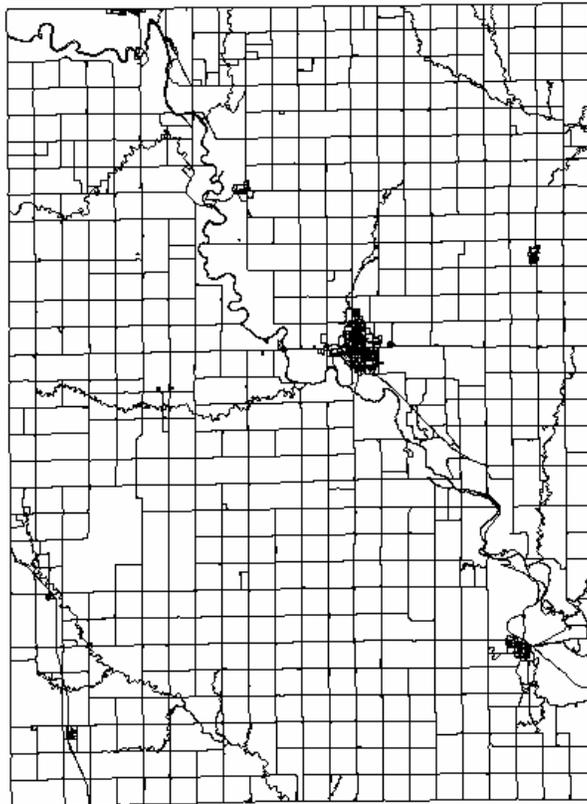


Figure 1. The analysis layer consisting of the census blocks of a Midwestern county.

The objective of scale calculation in CAMPS is to calculate an ideal scale for each of the features in the analysis layer and then select one of those scales as a satisfactory compromise scale for the entire map. The ideal scale for a feature is the one at which its identifying label can be placed on the map within the outline of that feature with an appropriate amount of empty space around it. Once again, in the interest of saving processing time, CAMPS assumes that all labels for the analysis features are the same

size, and uses this representative label for the calculation. Thus three parameters, set by the map designer before production time, drive the calculation. They are: L , the area of the representative label in square inches; C , the percentage of the feature, as it is shown on the page, that the label would cover if that feature were mapped by itself at its ideal scale; and P , how many (as a percentage) of the analysis features need to be mapped at their ideal scale or larger to have a workable compromise scale for the entire map. The ideal scale (S) for a single area feature is calculated by the formula

$$S = \sqrt{\frac{F \times (C/100)}{L}}$$

where F is the area of the feature on the ground, converted to square inches.

As an example, for a recent CAMPS mapping project the cartographer determined the size of the representative label to be .028 square inches, and at the ideal scale it should cover 70% of the area of the feature, and 80% of the analysis features should be mapped at their ideal scale or larger. CAMPS sorts all of the features in the analysis layer in descending order of size, then applies the ideal scale calculation to the feature at the 80th percentile of this list. Table 1 shows a part of a list of 100 analysis features. The entire map will be drawn at the ideal scale for feature 80.

Feature	Area(sq. meters)	F	C	L	S
75	10203	15814618	70	.028	19883
76	9995	15492219	70	.028	19680
77	9993	15489119	70	.028	19678
78	9102	14108071	70	.028	18780
79	8437	13077323	70	.028	18081
80	7836	12145775	70	.028	17425
81	7220	11190977	70	.028	16726
82	6965	10795728	70	.028	16428
83	6743	10451629	70	.028	16164
84	5988	9281381	70	.028	15232

Table 1: Features 75 through 84 of a list of 100 total features. According to the scaling parameters set for this example, the 80th feature is selected and its ideal scale of 1:17425 is chosen for the entire map.

The compromise scale for this map is 1:17425. This scale is ideal or larger for 80% of the analysis features, but the other 20% will be shown at smaller than their ideal size. If the constraints of map use allow, many of the 20% of features can still be labeled acceptably by placing the label outside the polygon and connecting the two with a leader line. However, if many of these small features are clustered together, the use of leader lines will result in label overprinting and confusion. The alternative is to make inset maps of these clusters. In most cases, this is more desirable than mapping at the ideal scale for the smallest analysis feature on the map, which will lead to an excessive number of sheets. The next section discusses how these clusters of small, dense features are detected.

Detecting Dense Feature Clusters Suitable for Insets

As with scale determination, the search for clusters of features that are unacceptably dense takes place only on features belonging to the analysis layer. The compromise map scale calculated above, 1:17425, now becomes the parent scale, the scale at which the parent map will be made and which will give the context for definition of unacceptably dense clusters. The analysis starts by creating a centroid to represent each analysis feature (Figure 2), and density is detected by moving a square window across the map sheet, centered in turn on each centroid, and counting the centroids that fall within the square. The cartographer defines inset-worthy clusters through two parameters: a window dimension, and the minimum number of points needed to fall within the window in order to form a dense cluster (i.e. density threshold). For instance, the cartographer might determine that all areas that have five or more points falling within a window on the map that measures one inch on each side is too dense and those areas must be mapped on inset maps.



Figure 2. A portion of the analysis layer shown in Figure 1. A centroid has been calculated for each of the analysis features

The building of dense clusters proceeds according to the DBSCAN algorithm (Ester et al., 1996). The example (Figure 3) uses a density window measuring one inch square on the map sheet at the parent scale and a density threshold of five points. One of the points is chosen arbitrarily and the square window is centered on it. If the density threshold is not met (Figure 3, point A), then the center point is marked as not belonging to a cluster and the test will move to the next point in the database. If the density threshold is met (Figure 3, point B), then a cluster is started that contains the points within the window.

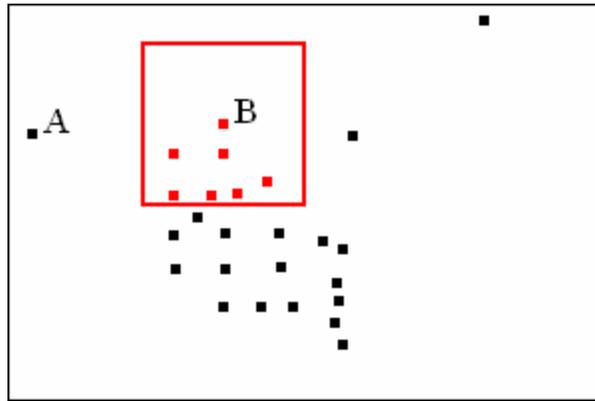


Figure 3. The red points form the initial part of a dense cluster.

Once a cluster is started, density testing proceeds through the other points belonging to the cluster. When the density threshold is met, new points within the window are added to the cluster (Figure 4).

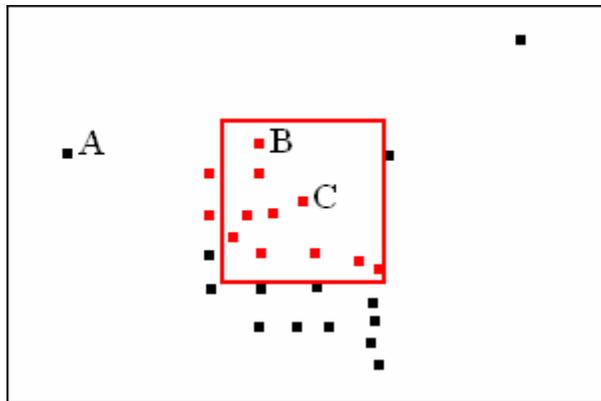


Figure 4. Five more points are added to the cluster when point C is tested.

If a point belonging to the cluster is tested, but the density threshold is not met, then no points are added to the cluster (Figure 5). Testing proceeds to any untested points remaining in the cluster.

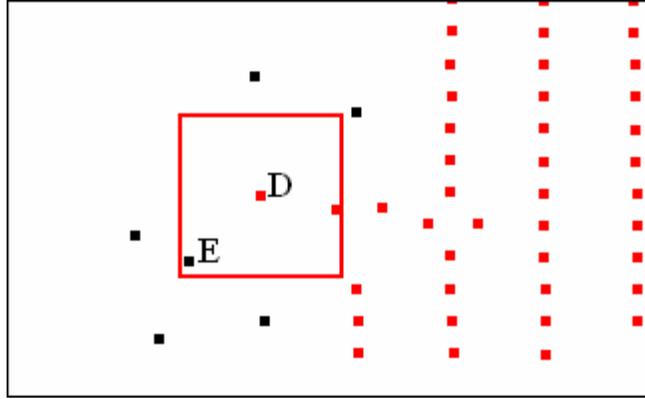


Figure 5. When point D is tested, the density threshold is not met, so point E is not added to the cluster.

When all the points belonging to the cluster have been tested and no more points can be added, the cluster point list is completed and saved, and the search goes to the next point in the database that has not yet been checked, trying to start a new cluster. In this procedure, each point serves as the window center only once.

Upon completion of density testing, each centroid will have been tagged with the name of the cluster it belongs to, or tagged as not belonging to a cluster. An inset area consists of all the analysis features whose centroids belong to a particular cluster. Figure 6 portrays two of the inset areas (clusters) that were detected in the example county.

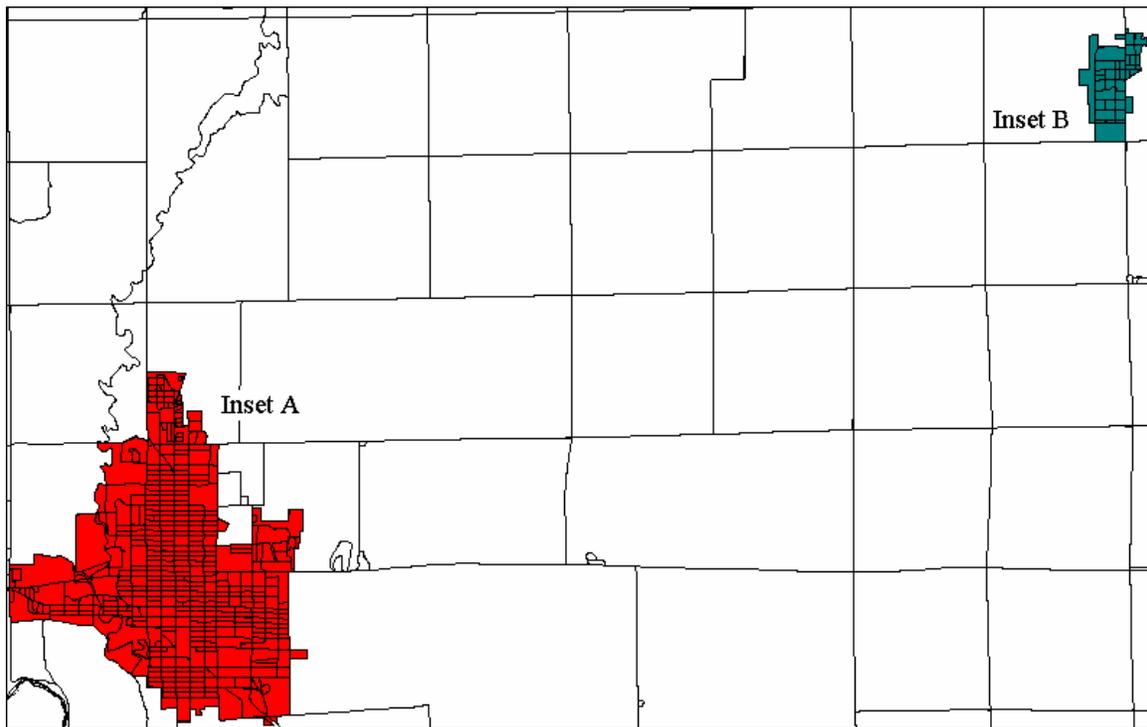


Figure 6. Two inset areas detected by density analysis and delineated by the bounds of the analysis features.

Once CAMPS has designated the analysis features belonging to each inset area, the next step is to draw the boundaries separating the insets from the parent map. CAMPS determines the shape of the inset maps in one of two ways. The first way is suitable when the map is used to do work on a feature-by-feature basis, and it would be confusing to split features between the parent map and the inset map, showing part of a feature at one

scale, and the rest of it at another scale. This approach delineates the bounds of the inset using the bounds of the analysis features, and is shown in Figure 6. The inset area takes on an irregular boundary in this situation, but it does consist of whole geographic features, so it is known as whole geography insetting. The other alternative is to draw a rectangle around the cluster of dense analysis features, as was done in Figure 7. Everything inside the rectangle becomes the inset area, and the area clipped from the parent map always has a simple rectangular shape. This is called rectangular insetting.

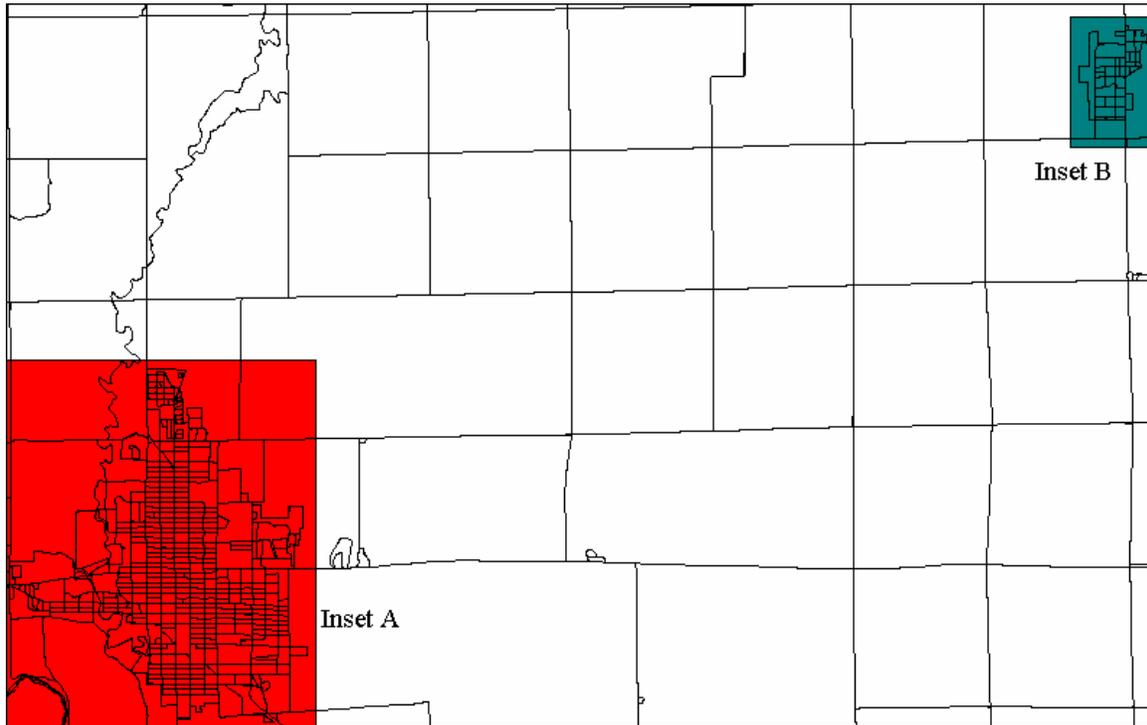


Figure 7. The same inset areas delineated by rectangles.

The county displayed in Figures 6 and 7 obviously has two distinct levels of feature density and delineating the inset areas is fairly simple. In many other cases, the differences between dense and non-dense areas are less clear cut, resulting in wildly irregular cluster shapes, clusters that touch or nearly touch, and clusters with holes or exclaves. Therefore, in order to improve the final delineation of the inset areas, CAMPS must perform some further processing on the dense clusters.

Merging and Simplifying Dense Clusters

Figure 8 shows a more complex set of analysis features where features assigned to dense clusters are shaded in gray. Application of the DBSCAN algorithm resulted in ten separate clusters in a relatively small area, some having holes, and some having quite irregular shapes. Rather than use ten separate sheets for the ten insets, some of the clusters that touch or are near each other could easily be combined for mapping on one sheet. Where there is a hole in a cluster the hole should become a part of the inset that surrounds it rather than being mapped as an island on the parent sheet. Simplifying the inset shape (if it is not rectangular) makes a more aesthetically pleasing map as well as simplifies the relationship between the parent and inset maps. CAMPS' merging and simplification routines differ slightly according to whether the insets are to be whole geography based or rectangle based.



Figure 8. Analysis features belonging to ten different dense clusters, all shaded with gray.

The convex hull is a useful construct for simplifying complex boundaries like those delineating CAMPS whole geography insets (Li, 2007). Applied recursively, the construction of convex hulls can also be used to merge the smaller inset areas. CAMPS begins by drawing a convex hull around each dense cluster of analysis features (Figure 9). When any of the convex hulls touch or overlap, those clusters will be merged to form a new cluster consisting of all the features that lie entirely within those convex hulls. Then new convex hulls will be drawn around the newly merged clusters (Figure 10). These in turn will be checked for touching or overlaps, and merged as before. The process continues in the same fashion until a set of convex hulls is created where none overlap (Figures 11, 12). At this point any islands discontinuous from the rest of their cluster will be examined. Those that do not meet the density threshold will be dropped from the inset area. Those that do meet the threshold will make up a new separate inset. The final, merged inset area consists of all the analysis features that are completely contained within the final convex hull.



Figure 9. Convex hulls drawn around the original ten dense clusters.



Figure 10. The first round of merged clusters with new convex hulls drawn around them.

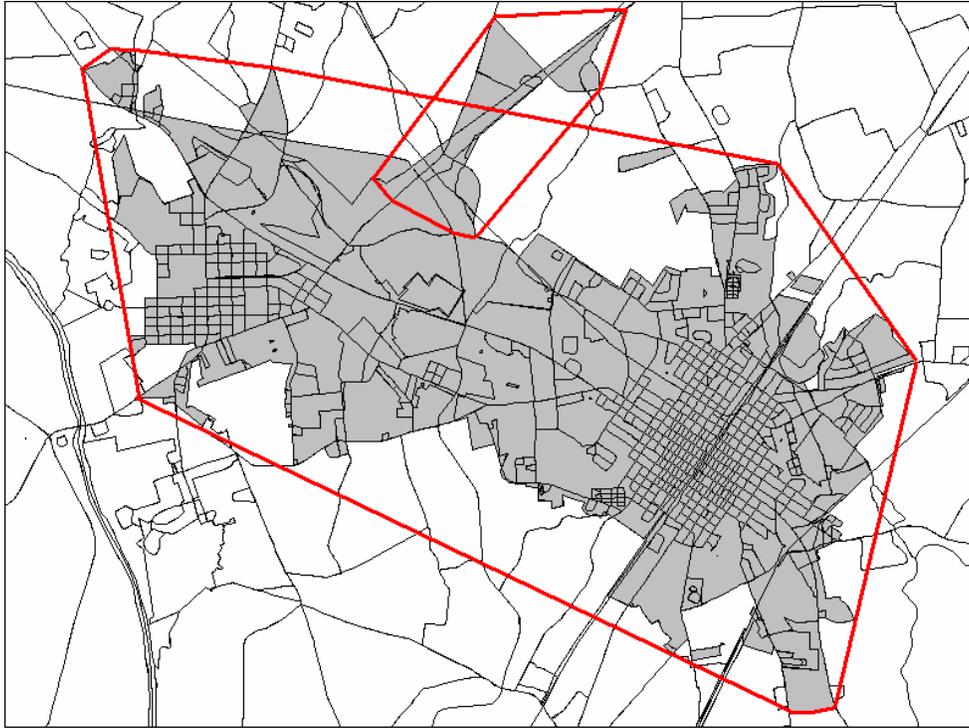


Figure 11. Merged clusters after the second round with the resulting convex hulls.

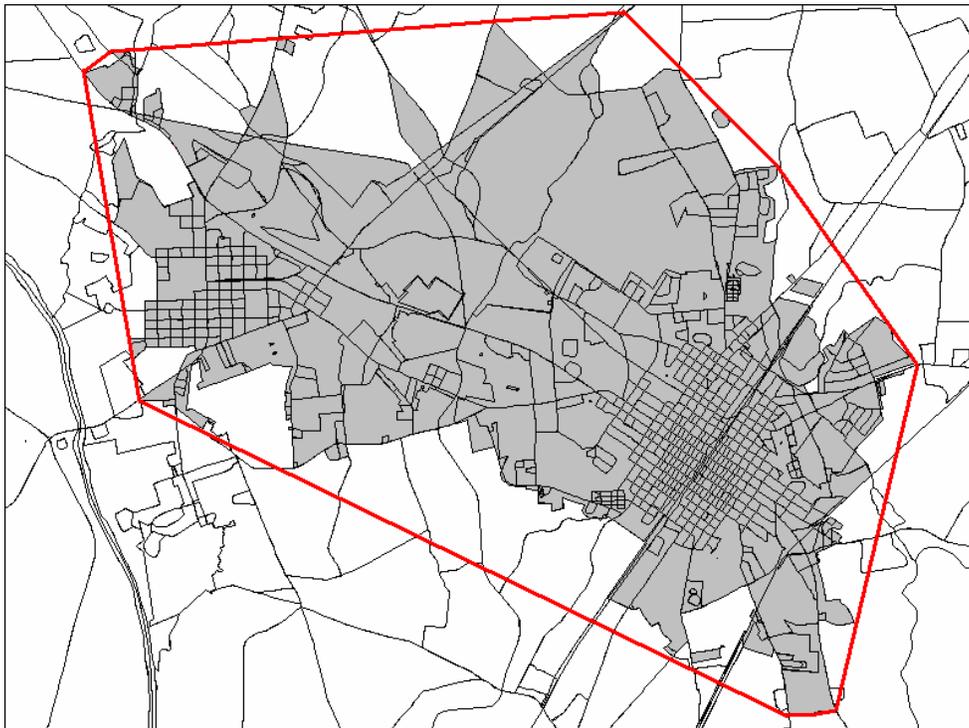


Figure 12. The final merged inset area.

CAMPS proceeds along similar lines if the insets are rectangular. The rectangular boxes delineating the insets are checked for overlaps (Figure 13). If any overlaps are detected, a new box is created from the minimum and maximum extents both vertically and

horizontally of the overlapping boxes. This procedure also continues repeatedly until no more overlaps are detected (Figure 14).

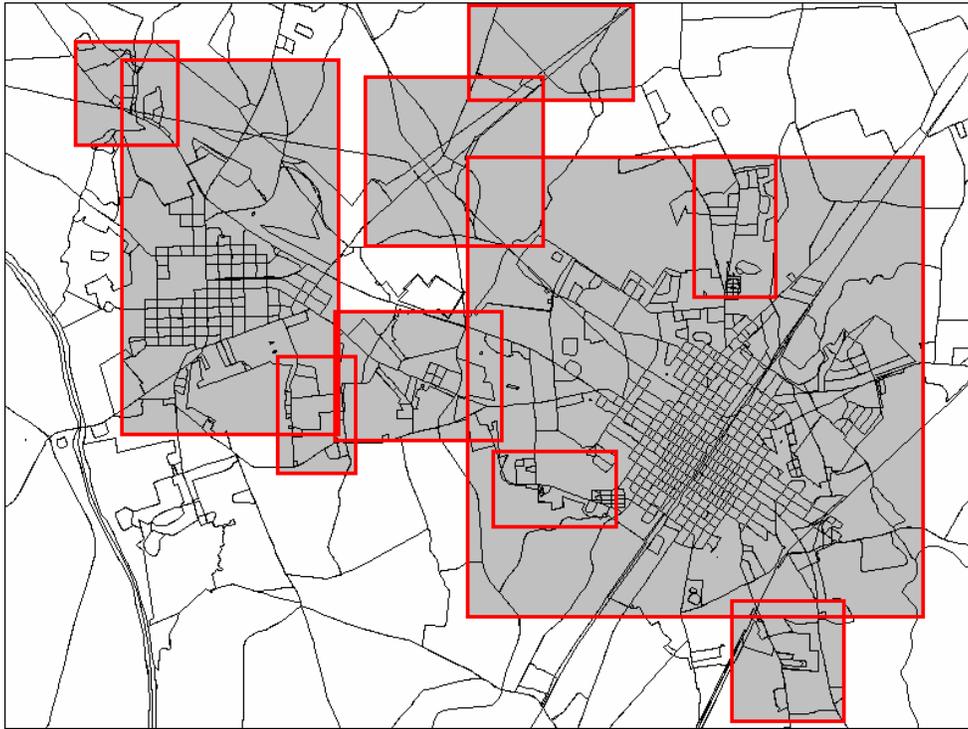


Figure 13. The originally determined rectangular inset areas.



Figure 14. The final merged rectangular inset area.

Finding the Best Combination of Parent and Inset Sheets

If the overall goal of parent and inset sheet configuration is to map the entire entity legibly on the fewest sheets, a secondary goal is also to keep the number of inset sheets to a minimum. Each time an inset is created, it makes the map harder to use by requiring the user to refer to two sheets to understand a single area: the parent sheet to get the context, and the inset sheet to get the detail. When CAMPS calculates a parent scale and creates a parent map and inset sheets based on that scale, it creates an acceptably legible map of the entity, but is there another parent scale at which the sheet configuration results are better? If a slightly larger parent scale were selected, the result might be fewer inset sheets and the same or fewer total sheets (parent plus inset). Therefore, CAMPS' automated inset detection and sheet configuration will try more than one parent scale.

CAMPS works iteratively through a table of standard parent map scales set up for the mapping project (Table 2). The first step is to use the text placement based parent scale calculation described above to determine the starting point in the standard scales table. In the example, the seed parent scale calculated was 1:17425, so CAMPS chooses the next largest scale, 1:16000, from the standard scales table as the actual initial parent scale, and does the inset detection and delineation using that initial parent scale. The first sheet configuration is saved and CAMPS goes to the next largest scale in the standard scales table and makes it the parent scale for the second sheet configuration. Inset detection and delineation is repeated for the second parent scale and the configuration is saved, then compared to the first configuration. If the second configuration results in a higher number of total sheets, then iterative configuration is stopped and the first configuration is used to make the maps. If the second configuration results in fewer or the same total sheets, a third configuration will be prepared using the next largest parent scale, and that configuration will then be compared to the second. The process continues in this manner until a configuration with more total sheets than the previous configuration is made. At that point it terminates and the next to last configuration is used. If the process reaches the largest parent scale on the table, that configuration is used. Throughout the iterations, two trends are operating that will tend to arrive at the best combination of parent and inset sheets. First, each step tries a larger parent scale, which will tend to reduce the number of inset sheets. Second, the process stops where the total number of sheets is at a minimum. For the example in Table 2, iteration ends after the third configuration, and the second configuration is used to make the map.

Configuration	Parent Scale	Parent Sheets	Inset Sheets	Total Sheets
	200000			
	100000			
	48000			
	24000			
1	16000	10	25	35
2	12000	14	15	29
3	8000	20	10	30
	4000			
	2000			
	1000			

Table 2: A table of standard parent scales showing the results of three iterations of sheet configuration. Configuration 2 results in the fewest total sheets.

Discussion

The CAMPS automated scale calculation, inset detection, and inset delineation algorithms described in this paper have now been used to produce a variety of multiple sheet maps of many types of geographic entities—counties, county subdivisions, large and small incorporated cities, American Indian reservations, and several types of census data collection units. In general they are effective in producing useful and legible maps. However, nationwide map production for a variety of entities encounters many difficult and unusual geographic situations.

Problems often arise when the analysis layer does not truly represent the nature of the density pattern for the entire map, as when the analysis layer contains only a few features or the features have an extremely large range in size. For instance, an entity to be mapped might contain only ten analysis features, three of which are small polygons about the size of city blocks, while the other seven are all very large, each perhaps many square miles in size. The ideal scale of the second or third smallest feature, the one most likely chosen as starting parent scale, is far too large for the entire map, resulting in an excess number of sheets. Even the iterative sheet configuration process will not ameliorate this problem because the starting point for the parent scale is already too large and at further iterations it can only get larger. The problem is to be able to detect these situations, so that the parent scale calculation can be short circuited to start the iterations at a smaller parent scale and probably produce a better map.

At other times, there may be an important layer whose features are more dense than the analysis layer features, such as a collection of short street features where the analysis census blocks are much larger. It may be worthwhile to do the initial parent scale analysis on more than one type of feature.

Map users have occasionally found difficulty in using maps with whole geography insets. The highly irregular shapes that sometimes result when insets are formed by clustering together analysis features can be hard to relate to the rest of the map. Making the inset area bounds conform where possible to a compact, pre-defined geographic entity such as a block group, census tract, or even an incorporated city should help orient the reader.

This paper has described only the use of area features as the basis for scale calculation and inset delineation. At this time CAMPS can also do these operations based on point features, such as housing unit locations. Dense clusters are built from the point features themselves in the same way they are built from the polygon centroids. Point-based insets can only be delineated with rectangles that bound the dense areas. There is a danger that

these insets, magnified enough to show the points clearly, will not include within their bounds other types of features that might provide orientation.

As the number of map projects successfully completed with CAMPS grows, so does the storehouse of knowledge that can be used to evaluate the cartographic intelligence of these automated map decision making techniques. In fact, the process eventually may be enhanced by data from past projects to produce better and faster results. CAMPS developers continue to modify and shape the system in response to problems with current mapping projects and needs of future projects, building mapping software that more closely imitates the ingenuity and resourcefulness of the human cartographer.

REFERENCES

- Beard, C., and Robbins, A. (1990) Scale Determination and Inset Selection within a Totally Automated Map Production System. *Cartography and Geographic Information Systems*, 17, 1, pp. 57-68.
- Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996) A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*. Portland, OR.
- Li, Z. (2007) *Algorithmic Foundation of Multi-Scale Spatial Representation*. Boca Raton: CRC Press, pp. 184-188.
- Muehrcke, P. (1986) *Map Use Reading, Analysis, and Interpretation*, 2ed. Madison: JP Publications, pp. 60-61.
- Trainor, T. (1990) Fully Automated Cartography: A Major Transition at the Census Bureau. *Cartography and Geographic Information Systems*, 17, 1, pp. 27-38.

William G. Thompson, Computer Mapping Specialist, Geography Division, U.S. Census Bureau. Washington, DC 20233. E-mail: <william.g.thompson@census.gov>.