# CONFLATION: AUTOMATED MAP COMPILATION
## ----A VIDEO GAME APPROACH----

Maureen P. Lynch and Alan J. Saalfeld
Statistical Research Division
Bureau of the Census
Washington, DC 20233

## ABSTRACT

An interactive system for semi-automatic map matching/merging (conflation) is being developed at the Census Bureau. By means of that system an operator matches together features of two maps on the terminal screen. Several features of video games facilitate the operator's interaction with the screen image. Those features include movement simulation (rubber-sheeting), cross-hair targeting and pointing, color coding to identify status, flashing highlights, system prompts, system constraints to permit only "legitimate" moves, and a screen-clearing (partial image removal) strategy. Other game-like capabilities include preview and review of operations.

Full automation of map merging appears ever more feasible as testing continues with the prototype conflation system. Plans are underway to conflate United States Geological Survey maps and Census Bureau maps of all jointly mapped regions in the entire United States.

## THE SHORT HISTORY OF MAP CONFLATION

The word "conflation" (from the Latin **con flare** meaning "blow together") was coined for use with automated cartography in 1979 by the wife of James Corbett, the father of the Census Bureau's DIME files. "Conflation," many cartographers were surprised to discover, is a legitimate word traditionally used to describe the merging of two manuscripts into a third combined version. Most dictionary references to conflation mention the Bible, an all-time favorite text for conflating. Conflation of maps refers to a combining of two digital map files to produce a third map file which is "better" than each of the component source maps.

Much of the initial theorizing about the proper way to conflate maps centered on the subjective definition of the "best" features of each individual map and how to extract them and recombine them into a unique improved map. Many arguments arose about what constituted a match of map features; and counter-examples were presented to every conflation approach proposed. In the purely abstract and theoretical framework of philosophizing about conflation, too many alternatives precluded the possibility of a global theory; and, hence, for a long time there was no theory at all, only assorted ideas for handling specific cases. Map matching theory, or map conflation, really began developing in the course of a concrete experiment undertaken at the Census Bureau in 1984. Only when real digital files were compared could the experimenters begin to measure what would work and what would not work in terms of merging the two files into one.

The Census Bureau's interest in implementing some type of conflation system arose when the United States Geological Survey (USGS) agreed to provide laser-scanned line files from their 1:100,000 scale cartographic data base. The Census Bureau currently has a set of hand digitized maps covering approximately 5 percent of the land area of the United States and 60 percent of the country's population. The Census Bureau's Statistical Research Division began implementing some newly developed rubber-sheeting techniques on a high-resolution color graphics terminal to begin an experiment attempting to conflate the maps from the USGS and the Census Bureau.

As the experiment got underway, the research team decided to limit the scope of the file matching and merging task. Their goal was not going to be total conflation. Several factors entered into that decision, probably the greatest of which was that conflation in its broadest sense did not seem possible, whereas several subtasks undertaken did produce tangible and useful results. Moreover, the subtasks moved the Census Bureau closer to its goal of facilitating the GBF/DIME file updating/correction procedures. The final product of the current conflation effort will not be a unique third "better" map; however, it will be two "improved" map files which are linked in many ways, and which could be used to create one or more "better" maps at some point or to transfer feature-based information from one map to the other.

## SCOPE AND OBJECTIVES OF THE JOINT USGS/CENSUS PROJECT

Although total conflation (one new improved map) was not the objective, information transfer between files remained an important goal. The Geological Survey had made their digital files available to the Census Bureau with the understanding that the Bureau would tag the features in exchange for the use of the files. Tagging involved two levels of information transfer: tagging features by type, such as "class I roads", and tagging features by street name and house number ranges. Tagging by type is being accomplished by a highly interactive manual screen-based graphics system. Adding names and address ranges is more complex because of the quantity and quality of matches required. It requires a preliminary step of matching enormous numbers of linear features in a **consistent** and **comprehensive** manner to prevent duplicates and missed coverage. That preliminary matching step will be one of the products of the present limited conflation package being developed and used at the Census Bureau.

In addition to providing the USGS with information, the Census Bureau also expects to benefit from using the USGS files. The Census Bureau wants to compare the two digital representations of the same area to find discrepancies that would permit the Bureau to focus on possible recent housing development or on map drawing errors of one file or the other. In addition, the Census Bureau wants to improve appearance of its maps. The rendering standards used to create the USGS files give USGS maps an appearance far superior to the Census Bureau's manually digitized maps. The Census Bureau wants to borrow that appearance if possible, to straighten the lines, smooth the curves, and thereby gain publication quality maps in place of the "stick maps" that are currently produced from DIME or DIME-like files.

The appearance enhancement of Census maps will be accomplished by rubber-sheeting Census maps to align them with the USGS maps. Rubber-sheeting involves moving some of the Census map's points to align exactly with corresponding points on the USGS map. The remaining Census map points are then moved in a proportional fashion so that the relative positions of all of the Census map's points are not changed. The Census map is shifted and stretched to fit the USGS map. More points are then aligned and the process of rubber-sheeting continues until all features that can be identified are identified.

During the rubber-sheeting of the Census map, the USGS map does not move. In some sense the USGS map behaves as the underlying "truth" to which we fit our map. The USGS map does not move because to distort it would destroy the straightness of lines and smoothness of the curves, which we have already recognized as important to preserve. One conceivably could move both maps some average distance to bring them together at points that can be matched; however, the matching of features would not be improved and both maps would lose their original shapes.

Another reason to leave the USGS coordinates alone and change only the coordinates of the Census map points is the Census Bureau's promise to **add** information to the USGS digital file and to change only errors of an agreed upon type and level on the USGS files.

The Census Bureau's more limited objectives for the conflation system may be summarized as follows:

1. Improve coordinates of the Census Bureau's Geographic Base File by exactly adopting the USGS coordinates wherever possible, which may include the addition of points that define shape.

2. Establish match flags and pointers between Census map files and USGS map files for all line and point features that can be paired on the two files.

3. Establish non-match flags for those line and points features which appear on one map and not the other.

4. Create maps which highlight matches and non-matches in a manner which facilitates secondary verification of controversial features.

The Census Bureau will create a new set of coordinates for all of the point features on its maps. The Census maps eventually may be drawn with either the new or the old coordinates. Additional information will be available at the end of the conflation process; no old information will be destroyed or overwritten.

Even with the new coordinates the Census maps will maintain their topological integrity. The whole point of triangulating and rubber-sheeting is not to alter the relative positions of map features. If a Census map requires correction of topological relations, that adjustment must occur elsewhere.

Rubber-sheeting of the Census maps is accomplished by computationally fast piecewise-linear local transformations defined on a triangulation of the map space. Coordinates of 0-cells (intersections) are recomputed and 1-cells (streets) are redrawn. Rubber-sheeting is applied iteratively to increase the number of map points aligned at each step. Each rubber-sheeting step leaves Census map features in the same relative positions but more closely aligned to the USGS map.

## EVOLUTION OF THE COMPUTERIZED ALIGNMENT SYSTEM

The iterative process of map alignment was initially designed to be highly manual because of the expected difficulty in deciding what to match. An operator was expected to apply complex decision rules to a large variety of situations. The initial design called for real-time operator interaction with a color graphics screen displaying the features of each of the maps. That design produced a video game type of situation. The similarity to video games was further exploited by adding system features which rewarded or reinforced correct choices by the operator via the screen display. The operator could be challenged to move quickly and accurately and could see visible results of moves made.

Recent experience with real maps has pointed to the feasibility of a more fully automated approach to matching, and such an approach is underway. The situations and decisions facing an operator are not as complex in general as was first imagined. The initial fears that a fully automated rubber-sheeting system would completely distort the Census map have been allayed. A stepwise iterative alignment procedure now matches a few points based on very strict criteria on the first pass, then relaxes match criteria with each successive iteration to match additional points. A first-pass match of a few points gives, in general, a very good initial alignment of the maps, and after an initial alignment of maps is accomplished, future distortions may be avoided entirely by simply not allowing large point movements.

The title of this monograph reflects the initial idea of endowing the sytem with powerful manual tools for the operator. Experience suggests that manual intervention is not needed to the degree originally believed. The initial prototype system was almost entirely manual which tested the operator's skill and speed at making matches of point features of the two maps. Almost immediately it was realized that the speed of a purely manual operation would never accommodate even a small fraction of the large quantity of maps that must be processed. With sufficient machine guidance, however, the operator could be led to a significant number of error-free matches as experienced in our test cases. In later prototype systems, instead of leading the operator to error-free matches, the matches were made automatically, in batch, without operator intervention. To complete the analogy, the second prototype system was like the video game "Space Invaders" with an automatic pilot and an automatic "aim-and-fire" added to the operator's weapons. The operator merely presses the "start" button and sits back. Such an automatic system may take the fun out of the game by fighting all of the battles alone; nevertheless, it plays consistently and never loses!

## BASIC CHARACTERISTICS OF THE SYSTEM

The following properties of the map alignment/matching system were in the initial design. The system is:

(1) **Interactive.** A menu of commands allows the operator to make successive changes to the Census GBF map to improve its alignment with the USGS map and to select matching or non-matching features on the two maps.

(2) **Iterative.** By applying a fixed procedure repeatedly, the operator improves the number of match and non-match classifications at each stage.

(3) **Semi-automatic.** The machine applies topological and geometric tests after each iteration by the operator and then recommends, based on the results on those tests, additional matches or identifications. The operator then accepts or rejects all or some of the proposed automatic changes or makes additional manual changes.

(4) **Screen based.** "Screen based" means that the map image(s) on the screen are the principal (at times the only) data source used by the operator. The manipulation and selection of those images by the operator using highlighting, panning, zooming, and cursor positioning are the only interaction required. The computer programs take care of updating the data base accordingly. The programs also provide the tools for operating most efficiently and successfully with the screen image, including highlighting and prompting.

(5) **Overlay oriented.** This means that the matches depend on the ability of the operator and the system to position a feature of the

Census map directly above a feature of the USGS map; and, after such positioning is accomplished, the features remain coincident from that point on in the iterative procedure.

(6) **Full color coding.** Color is used to identify each of the map sources the node and segment matches and unmatchables, triangle areas, and feature types. By utilizing color, each of the distinct map components is made clearly distinguishable.

(7) **Simplistic design.** The overlay of maps are shown as simply as possible without any clutter for the easy identification of matching nodes. More details are available on request by the operator, including the highlighting of 0-cells or 1-cells, the display of previously matched or unmatched segments, and the display of statistical information that is available from the files.

(8) **Menu/Window driven.** Part of the display screen is reserved for the continuous display of available operator commands. (A tablet and mouse may be used instead to free more screen area.) A series of menus may be used to give the operator options for selecting a particular conflation process.

(9) **Subtractive reduction.** A procedure which tags and can (temporarily) remove from display items which have been identified as matches (or as non-matchable) facilitates the matching process. Such a procedure terminates allows the operator to "clear the screen" in order to focus attention on unresolved sections of the maps.

The following even more specialized capabilities of the system were added to improve efficiency and accuracy:

(10) **Subset selection and highlighting.** The operator must be able to find and see clearly the features on each map. Flashing displays and highly contrasting colors are employed to view such subsets as matched points or segments in a small neighborhood of a point, block boundaries, non-street features, nearest matched pair, and other subsets.

(11) **Separate image maintenance of each map with identifying links.** Each map must maintain its separate identity even as its features are transformed onto those of the other map. Image decomposition may be necessary to accomplish the identifications. Unmatching (correcting mistaken matches) is possible with this approach.

(12) **Nearest point subroutines to facilitate identifications.** Many of the search procedures used in later stages of matching are based on proximity of features. Effective local searching in specified subsets speeds nearest neighbor searches.

(13) **Geometric/topological editing to prompt likely matches.** A series of tests for similarity of map features assist the operator to make match/non-match selections. Test results are used to prompt the operator or to make a match decision that the operator need only accept or reject.

(14) **Topological edits to prevent unreasonable manual identification.** Unless specifically over-ridden, the system does not permit the operator to nail down two points which are grossly different in terms of their topological and geometric make-up.

(15) **Local screen rubber-sheeting and editing.** In conjunction with a zooming capability, the manual identification procedure should be able to operate locally. Changes to the image in the window (including accepting/rejecting the look-ahead automatic alignment) need not affect the rest of the image. Special software for windowing and subfile creation permits the operator to process the map locally.

(16) **Variable (adjustable) tolerance settings for automatic look-ahead alignment.** Points may be close in terms of distance or some topological measure; segments may be close in terms of direction. Because there are no a priori rules for determining how close is close enough and there are no summary results available on how close is typical or reasonable, variable tolerance settings permit different acceptance levels of matches as the process is iterated.

## ILLUSTRATIONS OF THE CONFLATION PROCESS

The technical details of the process are presented in several other papers included in the **AUTOCARTO 7 Proceedings**; and the mathematical theory behind the process is described in detail in those papers and in the references given in this paper's bibliography. This overview of the system will conclude with several illustrations of the stages of alignment. Colors are used on the screen display to differentiate line types and node types. The illustrations contained herein differentiate lines by type: solids, dots, and dashes; and differentiate points by different point symbols. All of the maps shown in the illustrations below are of Fort Myers, Florida.

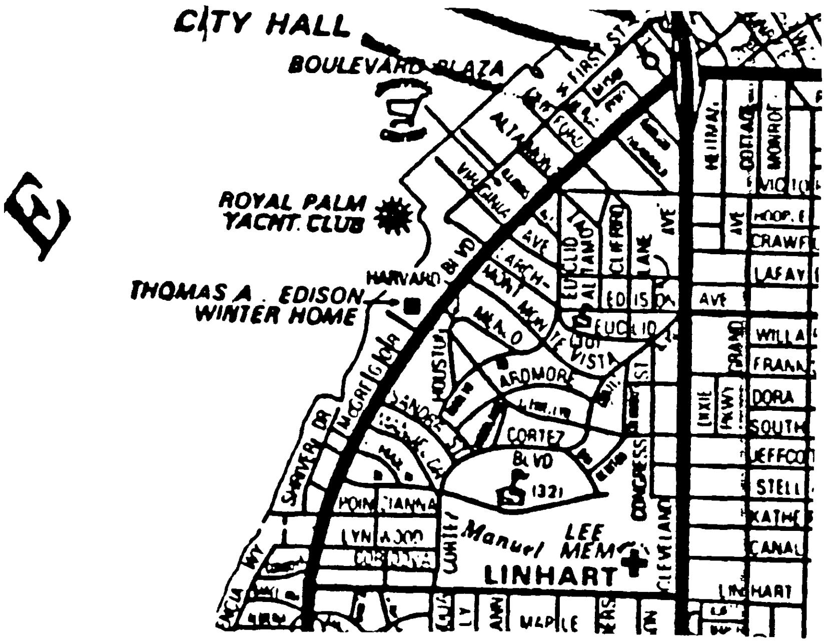


**Figure 1. Road Map of Area South of Caloosahatchee Bridge.**

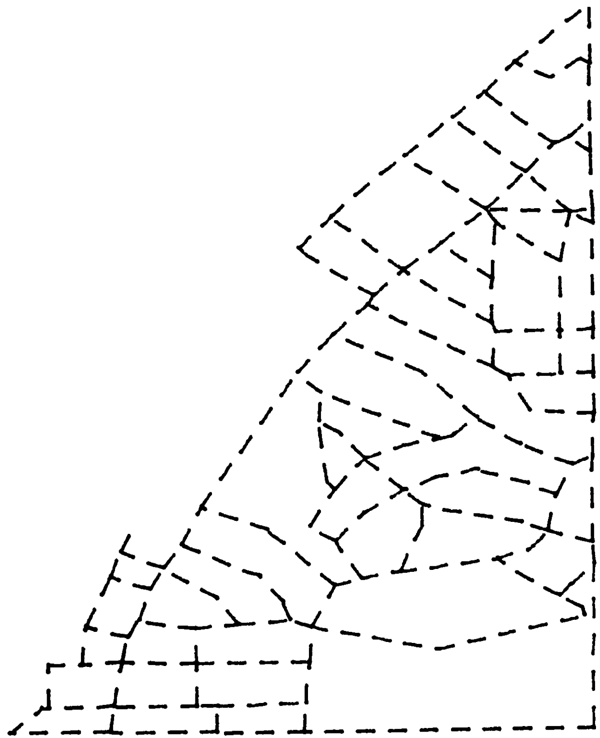

Figure 2. Long Dashes - USGS Map.

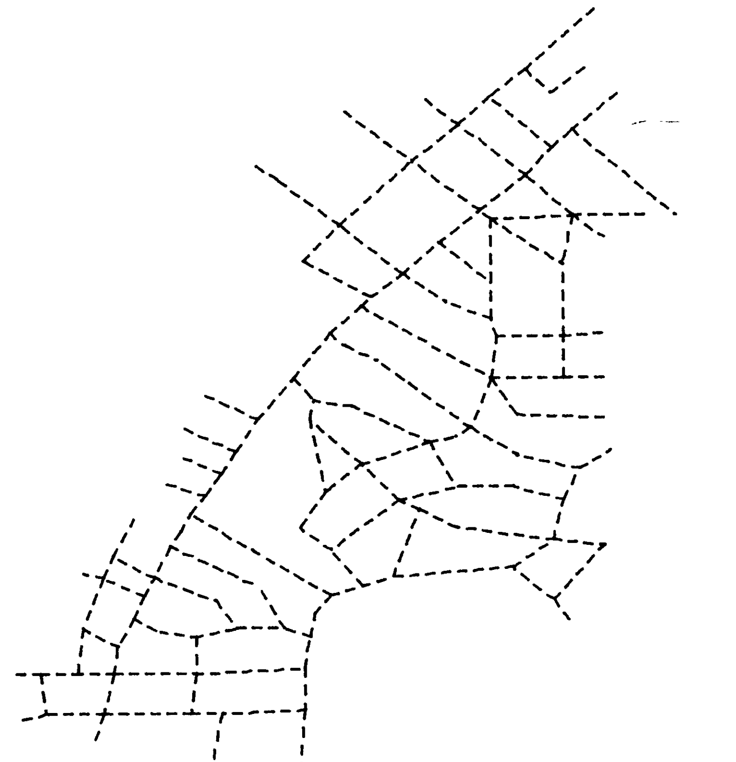

Figure 3. Short Dashes - Census Map of the Same Region.

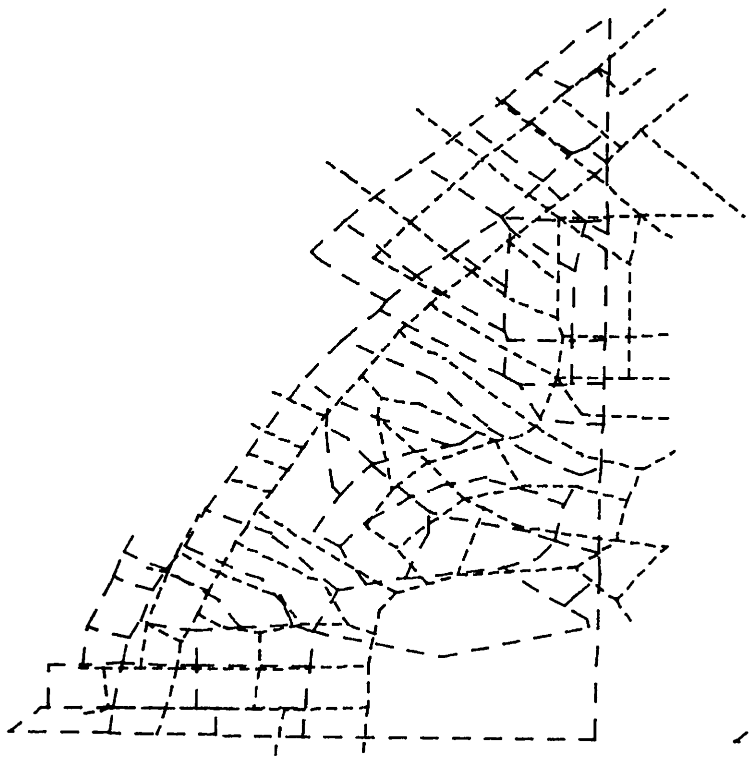

**Figure 4.** Overlay of USGS and Census Maps Based on Coordinate Transformations.

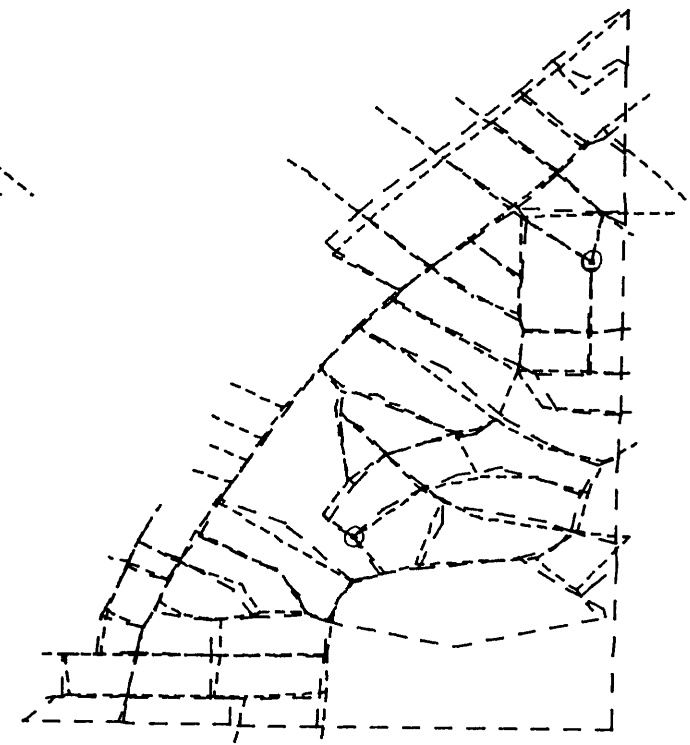

**Figure 5.** Rubber-sheeted Alignment After Two Pairs of Points Identified.

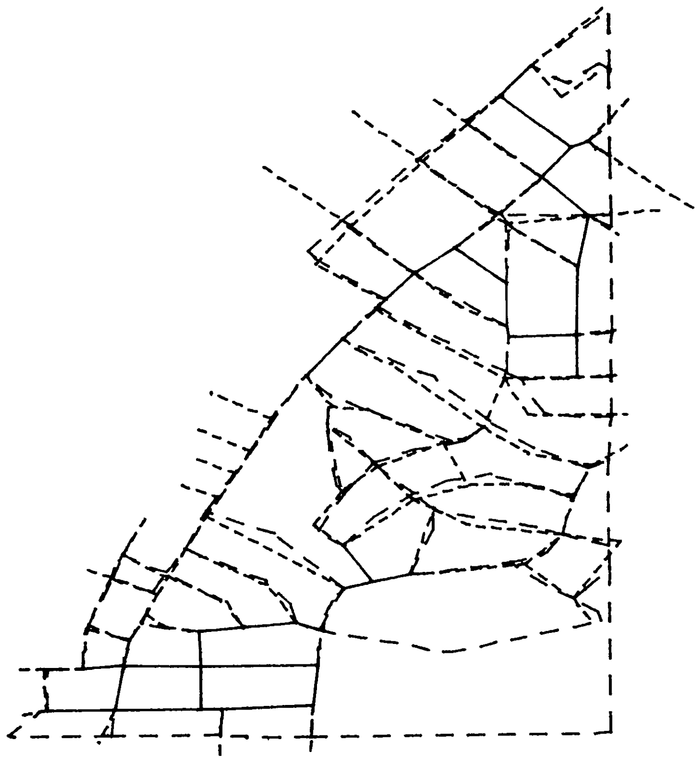

Figure 6. Rubber-sheeted Alignment After 36 Pairs of Points Identified.
Solid line indicates resulting street matches.

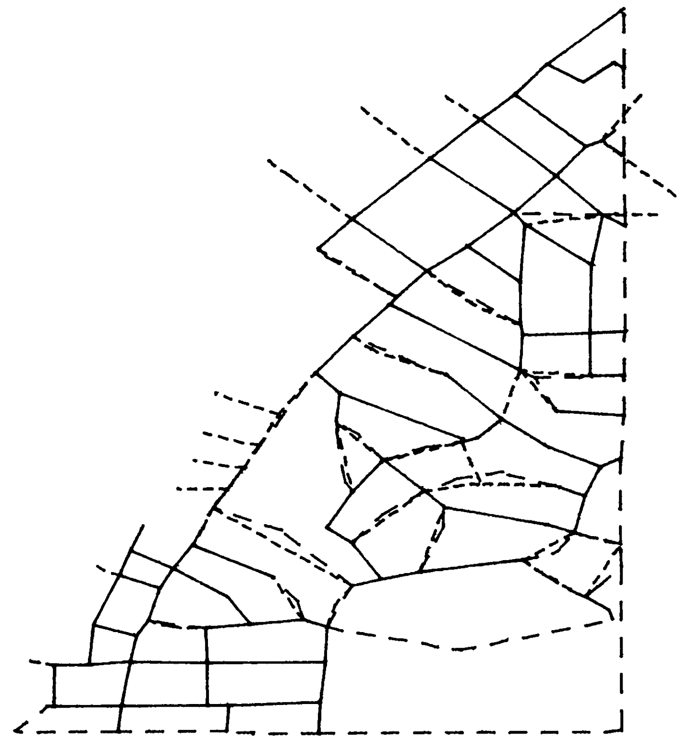

Figure 7. Rubber-sheeted Alignment After Full Processing.

## ONGOING RESEARCH AND DEVELOPMENT IN CONFLATION

The principal focus of research in conflation at this time at the Bureau of the Census is in the area of map file storage and manipulation. The large data sets involved with digitized map files must be organized to interact with local rubber-sheeting routines of the prototype system. Data structures will be made compatible with windowing features so that appropriate data subsets are automatically selected when working windows are chosen by the operators. In principle, the entire map, whatever its size, will be processed as a whole, although local routines will be applied to chosen subsets while the remainder of the map is simply "fixed by the identity transformation."

The working prototype programs utilize arrays to store point and line information; and the size limits of those storage methods are currently 400 to 500 points and 400 to 500 line segments per map. It has been observed that a screen image of approximately 250 line segments and 150 intersection points per map approaches the limits of managability both in terms of the limits of screen resolution and operator visual ability to discriminate features without a great deal of magnification.

Because the expected work unit will be a managable screen image of approximately 400 total line and point features per map, the existing rubber-sheeting routines in the prototype package will not require rewriting for processing larger maps. Additional refinements of the triangulation procedures will be necessary to interact with local editing and local windowing, and those refinements are part of the current research plan.

## BIBLIOGRAPHY


Gillman, D., 1985, "Triangulations for Rubber Sheeting," **AUTOCARTO 7 Proceedings.**

Griffin, P., and M. White, 1985, "Piecewise Linear Rubber-sheet Map Transformations," to appear in October issue of The American Cartographer.

Keates, J.S., 1973, Cartographic Design and Production, New York, NY, Halsted Press.

Murch, G.M., 1984, "Physiological Principles for the Effective Use of Color," IEEE Computer Graphics and Applications, 4(11), pp. 49-55.

Rosen, B., and A. Saalfeld, "Matching Criteria for Automatic Alignment," **AUTOCARTO 7 Proceedings.**

Saalfeld, A., 1985, "A Fast Rubber-Sheeting Transformation Using Simplicial Coordinates," to appear in the October issue of The American Cartographer.

U.S. Geological Survey/Bureau of the Census, 1983, "Memo of Understanding for the Development of a National 1:100,000 Scale Digital Cartographic Data Base," Washington, DC.

White, M., 1981, "The Theory of Geographical Data Conflation," internal Census Bureau document.