THE AUTOMATIC LABELING OF GEOGRAPHIC MAPS - A PROBLEM IN COMPUTER AESTHETICS

Herbert Freeman
Rensselaer Polytechnic Institute, Troy, NY 12181

ABSTRACT

Although much has been accomplished in automating the production of geographic maps, one task, that of placing feature names on a map, has until recently defied all attempts at being automated. This paper describes the current state of a project to develop a system for automatically placing the names of area features, point features, and line features in a map. A so-called rule-based system is used which incorporates the body of knowledge employed by cartographers when performing this task. The system aims to achieve a fully annotated map in which (1) all names are unambiguously associated with the features to which they refer, (2) the necessary cartographic standards are satisfied, and (3) a high overall aesthetic quality is achieved.

RÉSUMÉ

Quoiqu'on ait déjà beaucoup accompli en automatisant la production des cartes géographiques, une tâche, celle d'arranger les noms caractéristiques sur une carte a jusqu'à présent défié tous les efforts d' automatisation. Cet article décrit l'état actuel d'un projet de développer un système pour placer les noms caractéristiques d'une région, d'un point, et d'une ligne sur une carte. Un système appelé "système basé sur les règles" est appliqué et incorpore les connaissances employées par les cartographes en faisant cette tâche. Le système aspire à exécuter une carte entièrement annotée dans laquelle (1) tous les noms sont associés sans ambigüité aux caractéristiques auxquelles ils appartiennent, (2) les règles cartographiques nécéssaires sont suivies, et (3) une qualité exceptionnelle est réalisée.

KEYWORDS: computer cartography, name placement, map annotation, region skeletons, knowledge-based systems.

INTRODUCTION

For more than two decades now there has been a concerted world-wide effort to automate the map production process. Enormous progress has been made in developing efficient systems for storing map data in large geographic data banks and for automating the process of producing maps of various kinds. Only one major task in this process has eluded automation, and that has been the one of placing the feature names in a map. To this date, this step is performed either literally by hand or manually through the use of an interactive graphics terminal. This paper describes the current state of a project dedicated to designing a system for fully automatic name placement.

A map is a medium for communicating spatially-structured information. As such it must conveniently render the information of interest, without ambiguity or uncertainty. Cartographers have over centuries developed the art of map making, establishing an extensive body of cartographic conventions. Any automatic system for name placement should obey the same conventions and approach the same standards of quality that one has come to expect from a map produced in the traditional manner. Most important, a map should present an aesthetically pleasing appearance - it must not look like an engineering drawing!

THE NAME PLACEMENT PROBLEM

The information that appears on a typical geographic map can be classified into three feature types: area features (e.g., counties, lakes, mountain ranges), point features (e.g., cities, villages, and mountain peaks), and line features (e.g., rivers, highways, and railroad lines). The process of labeling these features - that is, where and how to place their names - is different for each class, and follows well-established cartographic conventions.

Maps may contain a large variety of different kinds of information - political subdivision lines, population centers, highways, secondary roads, railroad networks, indications of land use, contour lines to indicate elevation, drainage lines, etc. When map data is stored in a computer data base, it is convenient to store these different kinds of information separately and treat them as overlays in assembling a particular map. This is illustrated in Fig. 1, where a representative set of such map overlays is shown. Thus if a map is requested that will contain, say, highways and secondary roads, contour lines, population centers of 5000 or larger, and all major rivers, one need merely select the appropriate overlays and combine them into a single hardcopy presentation.

These map overlay planes are normally stored in a size corresponding to geographic areas of 7.5-minute quadrangles. If the desired map is smaller than one quadrangle, then the data must be "windowed" out of the overlay planes. If the desired map is larger than a quadrangle or if it overlaps two or more quadrangles, a combination of "windowing" and "pasting" of the quadrangle overlay planes may be required to obtain the final map overlay planes.

Map name placement would be a much simpler matter if it could be done once for each overlay plane and then stored permanently with the overlay. Unfortunately this is not possible. When different overlay combinations are assembled, the names from one overlay may overlap those of another, or ambiguities of reference could occur. Making name placements that would be proper for an assembly that uses all planes, and then assigning the names to their respective overlay planes is also not practical. The placing of names is just as much influenced by the presence of other names and features as it is by their absence. Thus the proper placement of names for plane A would be different if the final map were or were not also to contain plane B.

Even if feature names could be permanently placed for each overlay, there would still be a problem when the extent of the desired map is considered. This is illustrated in Fig. 2, where a map is to be windowed out of a larger region. Note how the area-feature name PLAINLAND is clipped, how the name for BIGTOWN is retained even though BIGTOWN itself lies outside of the desired map, how NEW-CITY is included but its name is lost, and how the river (a linear feature) will not be properly labeled. It is clear that name placement must be deferred until the selection of all the overlays has been made and the geographic extent of the desired map has been established. This means, of course, that name placement is a task that must be performed each time a particular map is to be produced. This does not hold for other map information, such as, for example, the encoded line data representing the contour lines. Once obtained and properly stored in the data base, this data would normally be subjected only to windowing and pasting operations, and possibly to a geometric transformation to satisfy a particular map projection requirement.

Because of the need to re-do name placement every time a map is to be produced that differs from a previous map either in extent or in data content, it is critically important to find a way for automating this one remaining manual task in the map production process. Unfortunately, until recently all attempts at automatic map name placement have failed at yielding an acceptable quality level. We shall here describe a new approach to the problem which has already yielded some promising results.

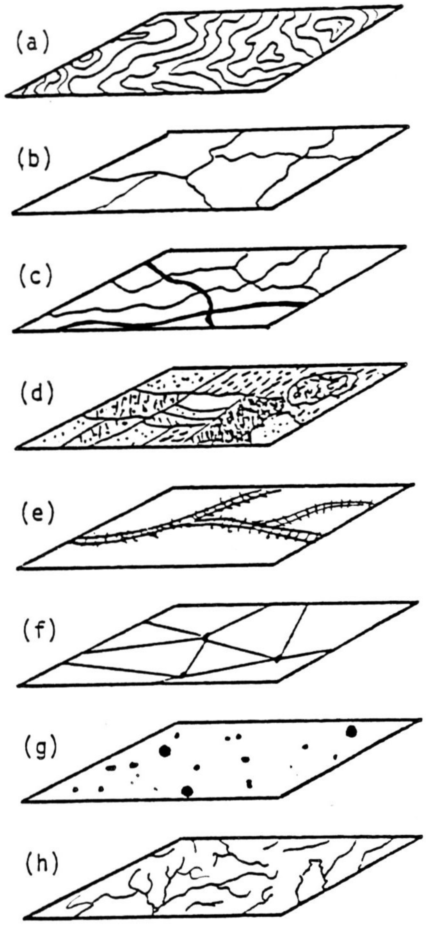

Fig. 1. A map as a coordinated set of feature overlays: (a) elevation contours, (b) political boundaries, (c) highways, (d) land usage, (e) railway network, (f) geodetic control data, (g) cities and towns, (h) drainage system.

NAME PLACEMENT RULES

Cartographers follow a fairly explicit set of rules in deciding where to place the names for the various kinds of features that appear on a map. These rules have been developed over many years. Some of them are considered basic and must be followed in virtually every instance.

Others are more flexible and may, in fact, only reflect relative preferences, and exceptions are always possible. In developing an automatic name placement system, the objective is, of course, to follow the same rules, and with the same degrees of priority and preference that one would expect a cartographer to use under the same circumstances. Fortunately, fairly good documentation exists as to what constitutes this body of name placement rules (1-3). The following summarizes some of the most important of these rules.

General Name Placement Considerations

1. A name must be unambiguously associated with the feature it is intended to label. It should be easy to read and locate.

2. The placement should assist in communicating spatial relationships, territorial extents, relative importance, and nature of a feature.

3. The placement should be aesthetically "non-disturbing" - it should be neither excessively clustered, exhibit an overly uniform distribution, nor present some clearly discernible (and distracting) pattern.

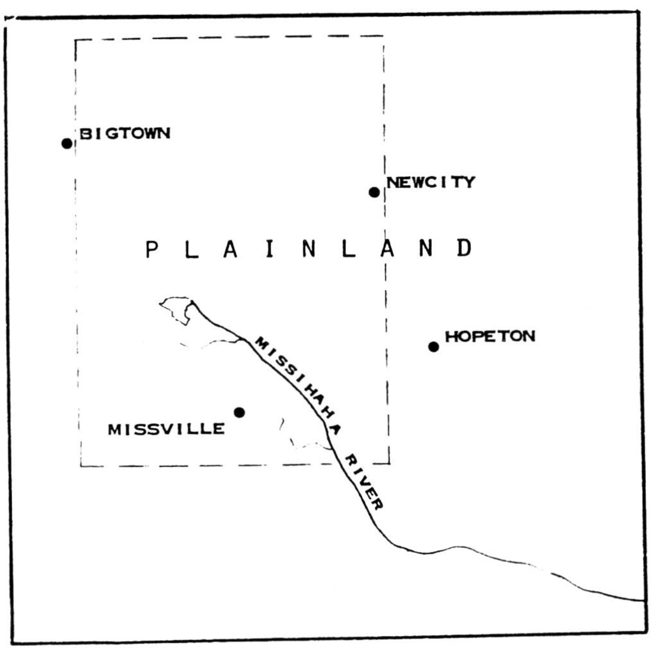

Fig. 2. Illustration of the necessity for deferring name placement until a map's extent has been selected.

4. Any overlap is to be avoided, as are also other placement situations that would tend to hide or conceal a name or a feature.

Rules for Area-Feature Names

1. The name for an area feature should span the entire area and conform to the general shape of the feature, with about one-and-one-half letter spaces at either end. The spanning of an area feature can be accomplished by judicious choice of font size and by spreading of the letters in the name.

2. Placement in conformance with area-feature shape is desirable; however, if there is little difference between this placement and horizontal placement, then preference is to be given to horizontal placement. "Horizontal" refers here to the constant-latitude lines (i.e., to the "parallels"), which will normally appear straight on large-scale maps (e.g., 1:24,000) and appear as circular arcs on small-scale maps (e.g., 1:1,000,000).

3. Names not placed horizontally should always be curved but with arcs that never exceed 60 degrees.

4. For non-horizontal name placement, it is desirable to have a name begin horizontally and then curve away from the horizontal rather than the opposite.

5. If an area feature exhibits a dominant boundary curve segment, then placement should try to conform (i.e., be parallel) to the curvature of this segment.

6. In choosing the name placement for an area feature, consideration must also be given to the placement of the names for neighboring area features to avoid aesthetic "clashes". (An example of an aesthetic clash will be shown later in this paper.)

Rules for Point-Feature Names

1. A point-feature name must be close to the feature to which it refers, and not so close to another point feature as to result in an ambiguity of reference.

2. The name for a point feature should be horizontal (i.e., parallel to the constant-latitude lines). Exceptions can be made in regions of extreme point-feature density and for point features bordering on large bodies of water.

3. Point features bordering on large bodies of water should normally have their names

placed "into the water" and be perpendicular to the shoreline.

4. Unlike area-feature names, point-feature names should never be spread out.

5. Point-feature names consisting of two or more words should as a first preference be placed in-line, though a two-line arrangement is permissible.

6. There is a slight preference for placing names above and to the right of a point feature. In general, placements above a point feature are to be given preference over placements below.

Fig. 3. Example of area-feature name placement. (The placement was done by the program described here, without human intervention.)

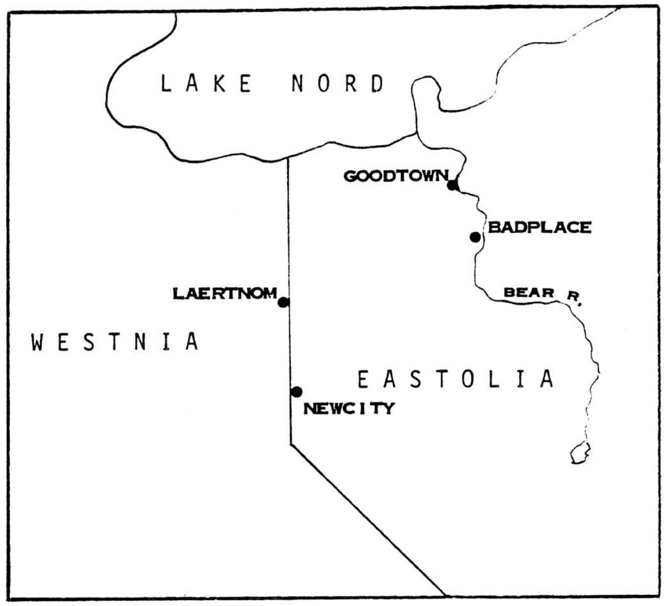

Fig. 4. The placement of point-feature names with due regard to nearby line features.

7. A point feature that lies a'jacent to a line feature (e.g., river, state boundary) should have its name placed on the same side as that where the feature is located.

Rules for Line-Feature Names

1. Line-feature names should reasonably conform to the curvature of the feature but should be placed at sections of the feature where the curvature is minimal.

2. Names may (and often must) be placed more than once along a particular line feature, with the number depending on the length of the feature and the need to achieve clear and unambiguous labeling.

3. Line-feature names should never be spread out.

4. For dominantly horizontal line features, the names should be placed above the feature. For dominantly vertical or other nonhorizontal line features, the names should be placed to the left of the feature and read upward in the left half of the map, and be placed to the right of the feature and read downward in the right half.

Some illustrations of the application of the foregoing rules are shown in Fig's. 3 through 5. Fig. 3 illustrates the manner in which an area-feature name spans the feature, conveying by its placement (as well as by the choice of font) a perception of the feature's importance and spatial extent. Fig. 4 illustrates the manner in which point-feature names should be placed relative to nearby line features: The city of Laertnom lies on the border between Eastolia and Westnia but belongs to the latter; hence its name must be placed in Westnia. The reverse is true for Newcity, another border town. The name for Badplace, a town on the west bank of the Bear River, is indeed badly placed. Fig. 5 illustrates correct placement of line feature names.

AREA FEATURE NAME PLACEMENT

Among the three classes of map features, the one for which name placement is most challenging is that for area features. The requirement that an area-feature name span the labeled area imposes severe demands on any automatic placement system. Because of the wide range of possible area shapes, there will necessarily be an enormous variety of name-placement configurations, and it will be most difficult to develop a specific set of rules to cover all the conditions that can arise. It is also easy to see that because of diversity of area-feature name placements, aesthetic considerations are likely to play a dominant role here.

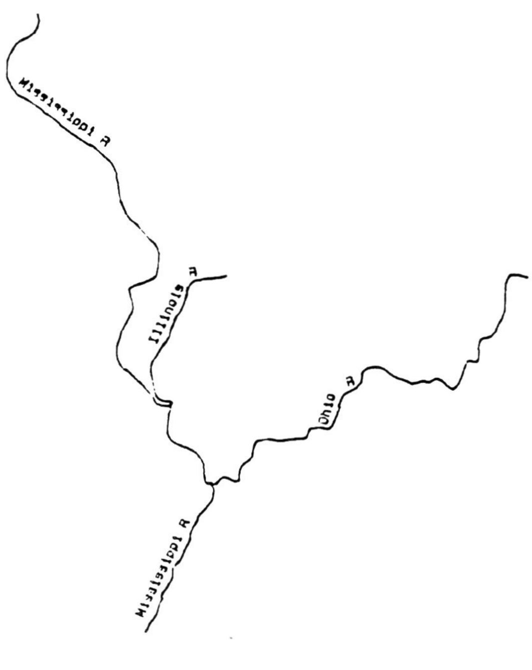

Fig. 5. The placement of line-feature names.

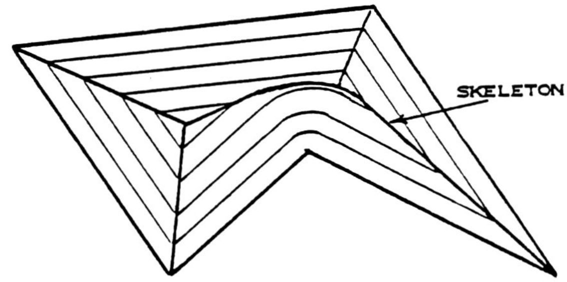

Fig. 6. Formation of the skeleton as the inward-shrinking boundary of a closed region.

Accordingly, since there is inherently a degree of competition for space among name placements for the three classes of features, it was felt necessary to give area-feature name placement priority over either of the others.

The first step in any area feature name placement process is that of describing the shape and then finding a shape-representing center line. Such a description is offered by the so-called medial-axis transform or skeleton technique (4,5). In this technique one begins with the area boundary and then lets it shrink uniformly toward the interior, layer by layer, until opposing parts of the boundary meet. The locus of points where this occurs is referred to as the area's skeleton.

The skeleton-extraction process is illustrated in Fig. 6. The skeleton of the area corresponding to the state of Idaho in Fig. 7(a) is given in Fig. 7(b). The reader will note that only the dominant component of the skeleton has been retained here; all smaller branches have been eliminated. A smooth circular arc is then fitted to the dominant skeleton component to serve as the centerline for placing the area-feature name. This is illustrated in Fig. 7(c). The technique appears to work fairly well, and it has been applied in all the examples of automatic area-feature name placement shown here.

POINT-FEATURE NAME PLACEMENT

Point feature names must be placed so that the distance between a point and its name is no more than some small distance, fixed between relatively narrow limits (6-9). This forms a permissible placement region (the "placement domain") about the point feature, as shown in Fig. 8. The primary placement positions in the domain are numbered in descending order of desirability.

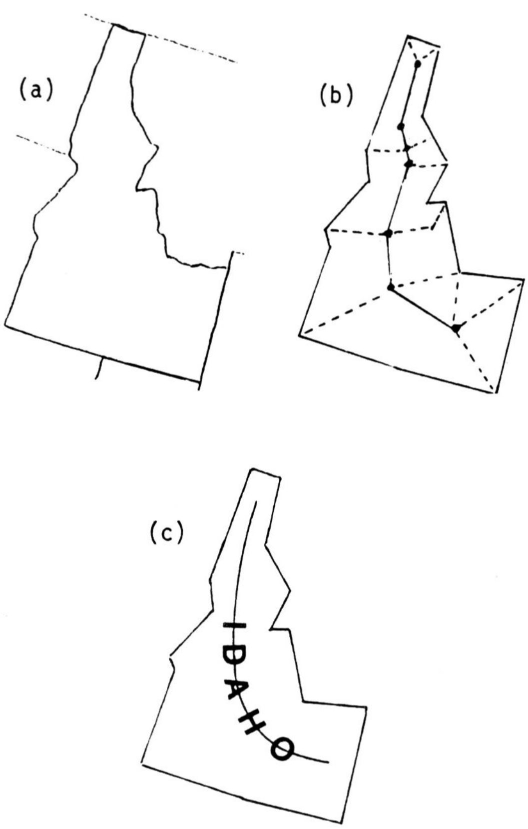

Fig. 7. A region (a), its dominant skeleton component (b), and its name placed on the circular arc fitted to the skeleton (c).

To determine the optimum point-feature name placement, a graph of possible name positions is created. For example, Fig. 9 shows the domains of all possible name positions for a set of 17 point features. A node in the graph represents a point feature. Two nodes are connected with a branch in the graph if their placement domains overlap. To avoid comparing every node against every other node, nodes are sorted in order of increasing Y values. Then only the nodes that fall within a fixed Y range need to be compared against each other.

Once a graph has been constructed, it is divided into connected components. Only those point features whose nodes belong to the same connected component need be considered in searching for a placement configuration that avoids overlap. For each node in a connected component, a list of free-space blocks is constructed by checking the positions of area names and neighboring point features. This is done by checking grid cells that contain at least a part of a node's placement domain. If an area name or a point feature falls within the existing free-space block, that free-space block is split by removing the area overlapped by the area name or the point feature. If the resulting free-space block is too small to contain the point name, the free-space block is removed from the list.

Using the free-space list and the placement domain list, a state-space search is carried out to place the point-feature names. The initial state is the state in which no name has been placed for any point feature in the particular connected component. The goal state is the state in which the names have been placed for all point features. The search algorithm used is a modified version of the familiar A* algorithm (10).

If it becomes impossible to place a name, the algorithm backtracks, removing the names already placed, to place them at different positions. Backtracking is helped by means of update records, which record the changes in the free-space blocks, the degree of freedom remaining for each node, and other internal information at the time the name labels are placed. In extreme cases, the algorithm may "give up" and request the deletion of one or more point features, the use of smaller fonts, or the use of a larger map format.

Observe that for large-font area-feature names, there is no objection to having a point-feature name be placed in the space between the characters of the area-feature name, as shown in Fig. 10. To facilitate such placement, large-font area-feature names have ther characters individually encased in rectangles. Only the character rectangles are then regarded as occupied space when searching for space available for point-feature name placement.

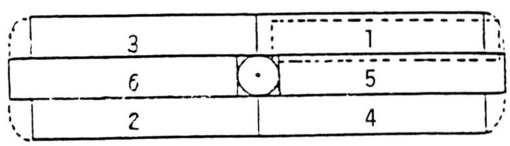

Fig. 8. Point-feature name placement possibilities, numbered in descending order of preference.

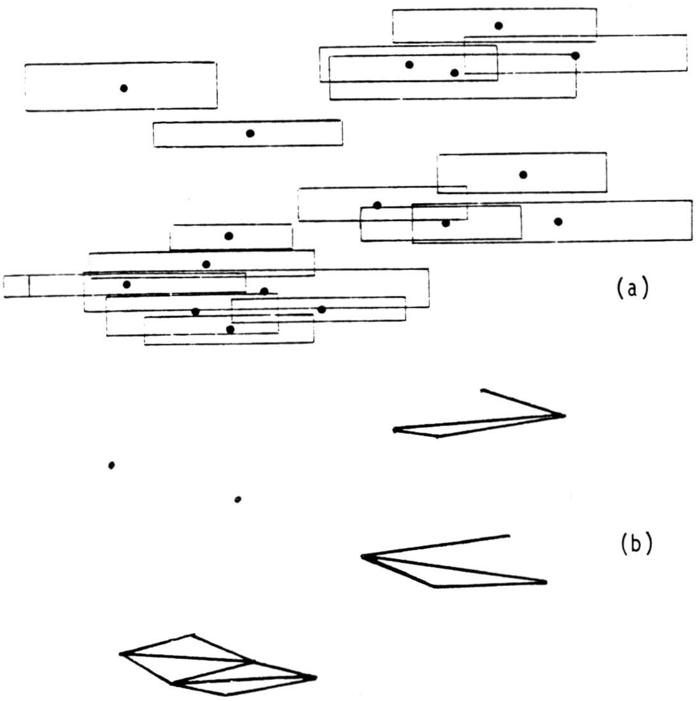

Fig. 9. The placement domains for a set of 17 point features (a), and the associated placement graphs for the connected components (b).

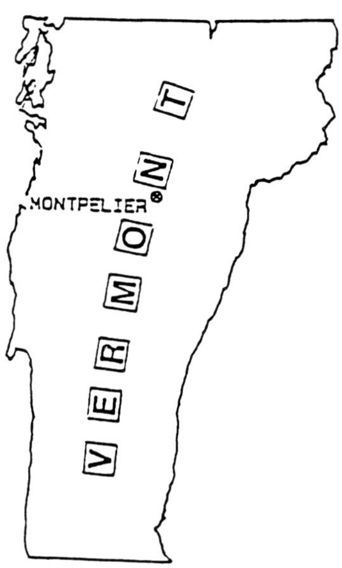

Fig. 10. Permissible interlacing of point-feature names with large-font area-feature names.

LINE-FEATURE NAME PLACEMENT

Line features are the only features that frequently require multiple name placements. Fortunately, they also exhibit the greatest degree of freedom for having their names placed. To facilitate multiple name placement for a line feature, the feature is divided into segments of fixed length. Each segment then has a name placed independently by testing all possible positions of the name for overlaps and selecting the best non-overlapping position. Starting at a certain distance from the end of a segment, a simple linear search is made over the extent of the feature segment until all possible positions have been considered.

When checking the possible placement locations of a line-feature name for overlap with other names already placed or with point features, it is only necessary to check the side of the line feature on which the name is to be placed. To determine which side is to be checked, a straight line is fitted to the section of the line corresponding to the intended name location. If the line is vertical (or nearly so), the position of the line section in the map is determined. If the section is in the left half of the map, the name should be to the left of the feature and read upward; if in the right half, it should be to the right and read downward. If the line-feature section is more horizontal than vertical, then the name should be placed above the feature.

IMPLEMENTATION

A name placement system, called AUTONAP, was developed during the past two years. The system was designed as a rule-based system (9,11), to facilitate extension and growth of the software. All the rules mentioned earlier were explicitly entered into the rule base, as well as secondary rules which refer to the relative priority of one rule over another in case of conflict. Although rule-based systems tend normally to be written in LISP, it was decided to use Fortran-77 instead because of the higher performance and the more ready portability of this language. The present implementation consists of approximately 12,000 Fortran statements.

Getting the system to a performance level at which it avoided all overlap and satisfied the basic cartographic requirements for name placement was relatively straightforward. The greatest difficulty was encountered in achieving results that were aesthetically acceptable. Numerous new rules had to be introduced (primarily affecting the placement of area-feature names) to improve the aesthetic performance of the system. An illustration of one of the difficulties encountered is shown in Fig. 11. The name placements for California and Nevada are by themselves quite acceptable. However, when both are placed in the same map, they are found to "clash", creating an unpleasant visual effect, as is readily apparent by referring to Fig. 12. This led to the introduction of new rules which permit the placement of an area-feature name to be affected by the placement of neighboring area-feature names.

An example of a map produced with the system is shown in Fig. 13. This example is representative of the current performance capabilities of the system.

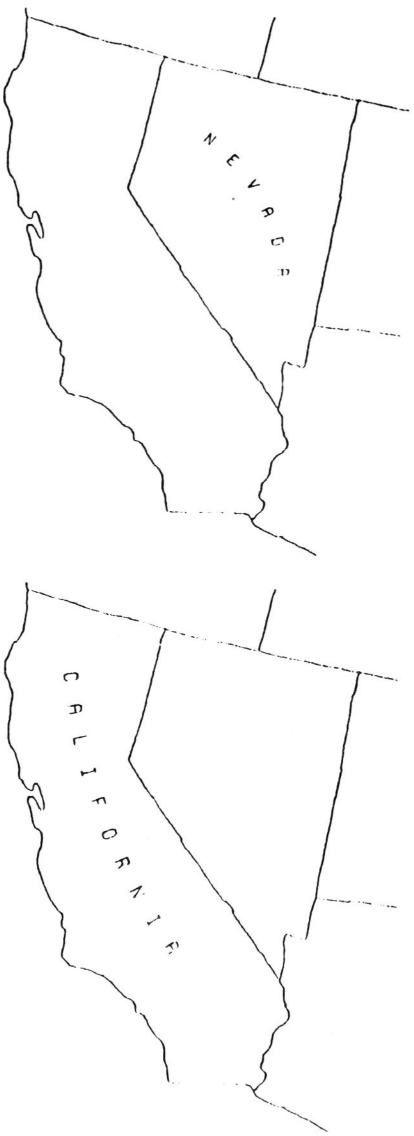

Fig. 11. Two examples of area-feature name placement.

Fig. 12. Aesthetically unsatisfactory area-feature name placement ("clashing") of the two names of Fig. 11 when these are placed in close vicinity.

ACKNOWLEDGEMENT

The writer wishes to acknowledge the contributions of John Ahn and Vinciane Lacroix, both former graduate students in the RPI Image Processing Laboratory, to the research reported here. The work was supported by the National Science Foundation, Computer Engineering Program, under Grant ECS84-07900.

REFERENCES

1. Yoeli, P., "The logic of automated map lettering", The Cartographic Journal, 9, (2), December 1972, 99-108.

2. Keates, J.S., Cartographic Design and Production, Longman, London, 1973.

3. Imhof, E., "Positioning names on maps", The American Cartographer, 2, (2), 1975, 128-144.

4. Blum, H., "A transformation for extracting new descriptors of shape", in Models for the Perception of Speech and Visual Form, ed. W. Wathen-Dunn, MIT Press, Cambridge, Mass., 1967, 362.

5. Montanari, U., "Continuous skeletons from digitized images", Jour. ACM, 16, (4), October 1969, 534-549.

6. Hirsch, S.A., "Algorithms for automatic name placement of point data", Master's thesis, Dept. of Geography, State University of New York, Buffalo, NY, December 1980.

7. Kelly, P.C., "Automated positioning of feature names on maps", Master's thesis, Dept. of Geography, State University of New York, Buffalo, NY, December 1980.

8. Ahn, J., and Freeman, H., "A program for automatic name placement", Proc. AUTO-CARTO SIX, Ottawa, Canada, October 1983.

9. Freeman, H., and Ahn, J., "AUTONAP - An expert system for automatic name placement", Proc. Int'l. Symposium on Spatial Data Handling", Zurich, Switzerland, August 1984.

10. Nilsson, N., Problem-Solving Methods in Artificial Intelligence, McGraw-Hill, New York, 1971.

11. Hayes-Roth, F., Waterman, D., and Lenat, D., Building Expert Systems, Addison-Wesley, Reading, Mass., 1983.

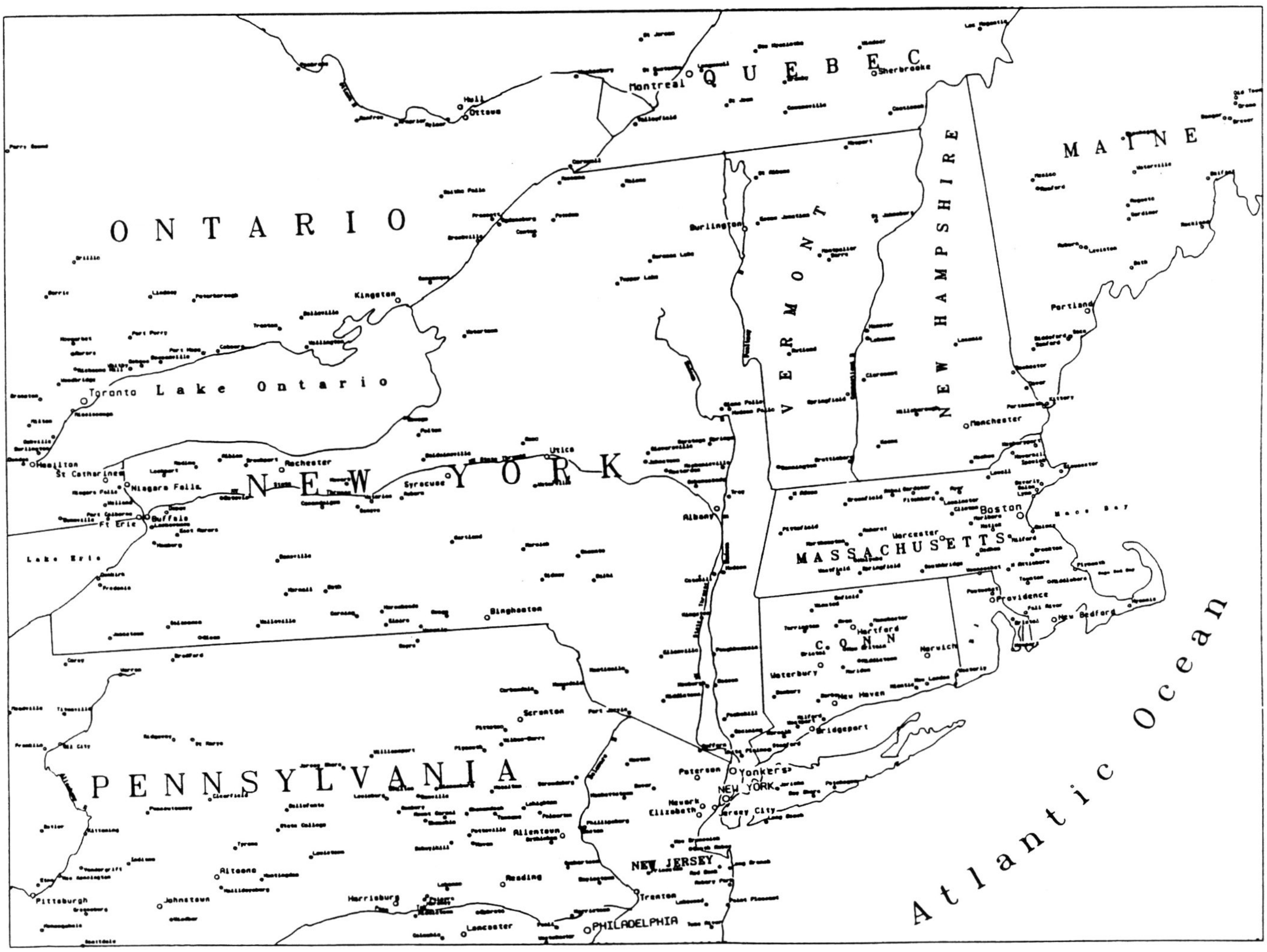

Fig. 13. Example of a map produced by AUTONAP. All name placement was accomplished fully automatically.