AN EXPERT SYSTEM FOR DENSE-MAP NAME PLACEMENT

Jeffrey S. Doerschler
Hamilton Standard
Division of United Technologies Corp.
Windsor Locks, CT 06096

Herbert Freeman
CAIP Center
Rutgers University
Piscataway, NJ 08855-1390

ABSTRACT

A dense map consists of closely-spaced topographical and political features labeled by aesthetically placed names. Most existing systems designed to place names automatically are limited because they label lower-level points and lines, not higher-level features. To overcome these limitations, a fully automated rule-based cartography system was developed to label higher-level features on dense maps. The feature-based approach allows the cartographer to create independent spatial data that may partially define several features; to create different thematic maps from a single database; to define relationships among classes of features; and to determine how features will be represented and labeled.

INTRODUCTION

A map is a collection of topographical features within a geographical region. These features may be represented on the map by symbols, lines or boundaries. Names on the map identify and classify features and relate them to their physical counterparts.

An automatic map-production system must be able to access a digital map database, select the features that are to appear to be on a map, annotate features and produce maps in a human-readable form. Placement of names on these maps is one of the more complex and difficult processes to automate. The difficulty depends to a large extent on the density of the map. On sparse maps, there are many empty spaces where names may be placed. On dense maps features are close together, leaving little room for names.

Most previous attempts at automated name-placement have been limited to labeling point features. Yoeli (Yoeli 1972) accomplished this by dividing a map into a dense grid with each character being placed into a unique cell. Others (Kelly 1980, Hirsh 1982, Langran 1986) used different techniques to improve point feature name placement.

Name placement is much more complex when point, line, and area features need to be labeled. Several papers (Basoglu 1982, Ahn 1983, Freeman 1984) describe the development and operation of a program to label maps representing these three types of features.

Most existing name-placement systems represent and label points, lines, and areas differently; however, the complexity and variety of maps that each can represent and label is limited. Some of these systems are limited by the types of features that can be represented, by the number of features that can be represented by a single point or line, by the number of lines that can intersect, or by the ability to label features represented by different symbols or line types.

A feature-based system has been developed to demonstrate one approach to automated dense map name-placement. This system uses rules based upon those enumerated by Imhof (Imhof 1975) to determine the location, orientation, font, size, and slant of each character in feature names. After each rule is processed, the name-placement quality is measured to determine whether additional rules must be processed. If the quality of a name placement deteriorates after other names are placed, backtracking occurs and the name is repositioned.

THE DATA STRUCTURES

The major data structures used by the name-placement system include tightly-closed boundaries, k-d trees and map databases which contain topographical features. These databases are used during all phases of map production. The data comprising the features consist of spatial data, classifications, names and abbreviations, miscellaneous data, and character placement information as shown in Fig. 1.

The spatial data are points, lines, and boundaries which may describe the location, size, shape and extent of features. Many features share common points or lines. For example, a line may represent both a river and a political boundary. Spatial data is independent of feature data. It may, however, be used as part of the definition of the feature. All spatial data is stored in units of longitude and latitude so that the database will be independent of any map projection.

Each feature may have one name and an unlimited number of abbreviations, one of which will be placed on the map. Only non-standard abbreviations need to be entered, since the labeling algorithms can automatically abbreviate common words such as "N." for "North".

The classification of a feature (e.g. river, city or interstate highway) determines if the feature will be placed on a map, how it will be represented, and how it will be labeled. Other classification-dependent information may also be required to properly place names. For example, city populations may be needed to determine proper character sizes.

The last feature attribute in the map database is a list of character placements consisting of the characters in the feature name, the longitudes and latitudes of their locations, and the orientations, font, height, width, and slant of the characters. Character placements can only be determined after the map projection, style and size have been determined.

When a map is plotted, area features are often represented by solid regions drawn as a series of horizontal lines. To generate these lines, the boundary of the area is first converted to a tightly-closed boundary (Merrill 1973). Pairs of points on the boundary are grouped to form endpoints of horizontal lines. These horizontal lines or y-partitions may then be plotted or used by the name-placement algorithms.

K-d trees are used by the name-placement system to detect overlapping names and symbols. A k-d tree is a structure designed for efficient multi-key searches, including exact match, partial match, and region queries (Bentley 1975). K-d trees are formed for character and point locations. Longitudes and latitudes are used alternately on each level of the tree as keys. The leaves of the trees point to corresponding entries in the map database.

NAME-PLACEMENT SYSTEM

To create a map, a cartographer first selects the types of features to be drawn on the map. A set of rules is used to determine how each feature is to be labeled. After a name is placed, the aesthetic quality of the placement is measured, for example, by checking whether the name overlaps any existing feature or name. If the first position is unacceptable, another position is tried.

Data Files

The name-placement program uses a set of rules to determine the location of each character to be placed on a map. These rules are defined in data files to allow the cartographer both to specify the name-placement styles for each feature classification and to show relationships among these classifications. Other data files specify the order in which the names will be placed on the map, the sequences of rules to be used for each feature classification, how names are to be placed, and how their quality is to be evaluated. A block diagram of the name-placement data files is shown in Fig. 2.

The first six data files assign names to the placement algorithms. Each file contains lists of algorithm names and defines parameters required by the algorithms. Each algorithm

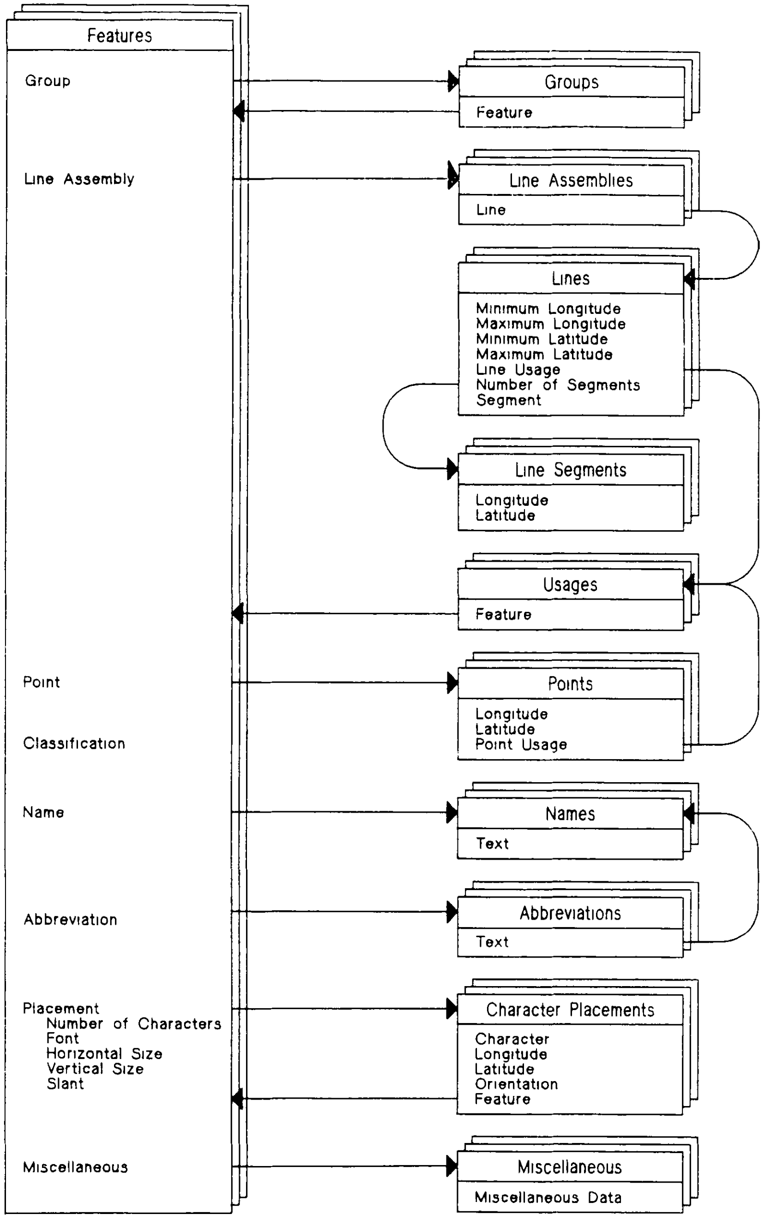

Fig. 1: Map Data Structures

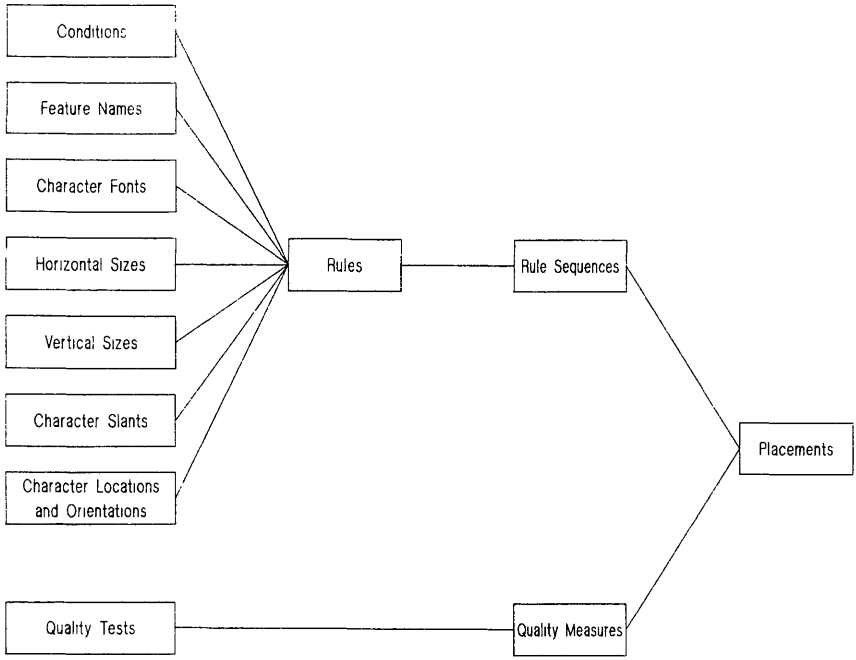

Fig. 2: Name-Placement Data Files

determines the value of one or two components of a name placement. A typical rule which uses these algorithms is the following:

```
Rule 101

    Letters
        Name_with_Capital_Letters_based_on_Population  25000
    Font
        Uniform_based_on_Population 1100 [Sans.1] 1101 [Sans.2] 8000
    Horizontal_Scale
        Population  8 1000 10 5000 12 8000 13 25000 15 100000 18
    Vertical_Scale
        Population  9 1000 10 5000 13 8000 14 25000 16 100000 18
    Slant
        Uniform 0
    Location
        Point_Upper_Right_Horizontal 5
```

In this example, rule 101 contains a list of name-placement components followed by the algorithm and its parameters which will be used to define the component. For example, the horizontal size of the characters will be defined by the algorithm "Population". The parameters to this algorithm are population intervals and character sizes in 0.1 mm units.

After a name is placed, one or more of the quality measurement algorithms are used to determine if the placement is acceptable. The names of these algorithms are defined in one of the data files whereas the set of quality measurement algorithms to be used for each feature classification and the importance of each algorithm are defined in the quality measurement data file.

If the quality of a name is unacceptable, the next rule must be tried. The rule sequence data file contains a list specifying the order in which the rules are to be tried. The last data file indicates the set of rule sequences and quality measures to be used for each feature classification.

Rule Processor

After the data files have been read, the name-placement rule processor determines the features to be labeled and the rules to be evaluated. The rule processor evaluates the first rule for each feature, possibly placing names on the map for every feature. The name placement quality will then be incrementally improved as additional rules are evaluated.

When names are placed on a map, their quality is evaluated to provide feedback to the name-placement rule processor. If a placement is unacceptable, additional positions are tried. To determine if a name placement is acceptable, its quality is evaluated using the rules specified in the quality measurement data file.

Each feature classification has a quality threshold between 0.0 and 1.0. If the quality is greater than or equal to this threshold, the name placement is acceptable. If it is less than the threshold, the placement is unacceptable. The threshold is lowered each time a rule is tried so that a placement which is initially unacceptable may become acceptable if a better placement cannot be found easily. The amount by which the threshold is lowered is specified in the rule sequence data file.

An existing name placement may, at times, have to be moved to make room for another name. This condition, known as backtracking, is required whenever a new name placement causes the quality of an existing placement to become unacceptable. After a name is placed, the quality of nearby names which may have been affected is reevaluated. If a placement is unacceptable, existing names may be repositioned.

Placement

The components of a name placement include the characters to appear on a map and their locations, orientations, vertical sizes, horizontal sizes, slants, and fonts. When a name-placement rule is evaluated, several algorithms are executed, each defining one or more of these components.

The first name-placement component that should be defined is the name or abbreviation to be placed on the map. Next, several algorithms can be used to define the character font, horizontal size, vertical size, and slant. The algorithms may determine the value of these parameters based upon the feature classification, importance of the feature, amount of room on the map or by some other characteristic such as a population.

Locations and orientations are the most difficult name-placement components to define because of the large number of possible positions, restrictions, and interactions. These two components are defined by the same algorithms because they are interdependent. The location of a character is specified by its longitude and latitude; its orientation is specified by the number of degrees it has been rotated.

Algorithms which label point features may place names next to the feature in one of eight directions or shift the name horizontally or vertically until an acceptable position is found. Long names may be split into multiple lines before they are placed on the map. Depending upon its location, a multi-line name will be left-justified, right-justified, or centered. The multi-line algorithms are illustrated in Fig. 3.

Route numbers are names which are centered horizontally on a highway. These names may also be shifted to avoid overlapping other characters, symbols, or intersections. Route numbers should normally be placed in a county, state, federal, or interstate highway symbol. Since the highway symbol and route number need to be at the same location, both are positioned by the name-placement program. A symbol is placed into the map database as if it were a character.

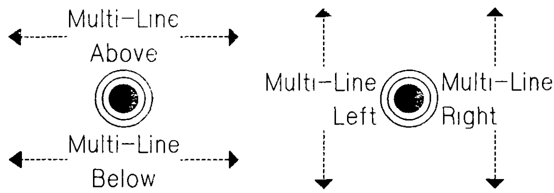

Fig. 3: Point Feature Multi-Line Name Placement

Other line features may be labeled by parallel names centered on or offset from the line. Once the name has been placed, it may be shifted to avoid overlapping existing names, existing features, map borders or sharp bends in the line. These names may also be shifted beyond the end of the line or repeated at predetermined intervals.

Many algorithms have been developed to place names within area features, along area boundaries, or adjacent to areas. A simple algorithm places a horizontal name in the middle of a feature. More complex ones will shift the name horizontally and vertically until an acceptable position is found. If the area feature is too small, the name can be placed next to the feature. Some of the same techniques used to place names next to symbols are also used to place names next to area features.

Other algorithms place names along the major skeleton of area features. A skeleton is the locus of points inside a region which is equidistant from the two or more closest distinct points on the boundary. The major skeleton is the longest continuous line in the skeleton as shown in Fig. 4. These algorithms may place a name in the center of the skeleton or shift the name along the skeleton until an acceptable position is found.

Finally, names of some area features, such as political subdivisions, should be placed along boundaries. These names will be repeated wherever a neighboring feature shares the boundary. Area feature boundary algorithms are illustrated in Fig. 5.

If none of the preferred name-placement algorithms produce an acceptable placement, additional steps can be taken. The size of the name could be reduced or the name could be eliminated from the map entirely.

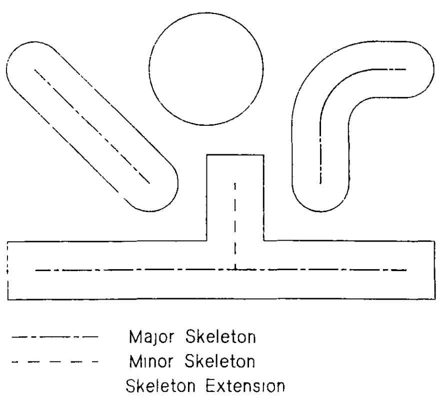

Fig. 4: Skeletons

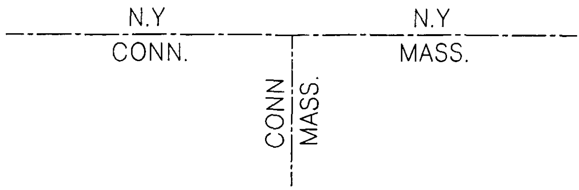

Fig. 5: Area Feature Boundary Name Placement

Quality

Once names have been placed on a map, their aesthetic quality must be evaluated to provide feedback to the name-placement rule processor. If this quality is above the current threshold, the placement is acceptable. If not, another placement rule must be tried.

The quality measurement algorithms determine the acceptability of a name placement. Some of the algorithms determine whether a name overlaps existing names or features or if the name is curved excessively, making it difficult to read. The quality measures also determine if a feature has been labeled or should be labeled. Others determine if a name is off the map or outside area feature boundaries. Some measures indicate a placement is poor if it is partially inside a region and partially outside. The algorithms are also used, for example, to prevent a route number from being placed over the intersection of two highways.

When evaluating the quality of a name placement, each character in the name must be examined to determine whether it overlaps other characters or symbols. Character and symbol locations, defined in terms of longitudes and latitudes, are placed into k-d trees. A region query is then performed to determine if a character or symbol already occupies the same location.

Performance

The time required to place names on two maps of similar density should be proportional to the number of names on the map, since the average number of rules used to place each name on these two maps should be similar. As the map becomes denser, more name-placement algorithms must be tried for each name. Therefore, the time required to place names will increase somewhat faster than the number of names on the map.

IMPLEMENTATION and RESULTS

The name-placement system described here was implemented on two different computer systems, a Prime 750 minicomputer and a Cromemco System Three microcomputer. A Hewlett Packard 7550 pen plotter was used to generate hard copies of the maps.

The 700 fully documented Fortran modules of the name-placement system were designed to be portable. All software was also designed using data abstraction and information hiding techniques.

Two maps were produced to illustrate the capabilities of the automatic name-placement system. The map in Fig. 6 is a 1:26,000 scale street map of Troy, New York. The map in Fig. 7 is a 1:1,160,000 scale regional map of Central New York State.

The Troy map illustrates dense-map name placement when most name-placement possibilities are very limited. Over 2,000 characters were placed on the map to label half of the 400 features. Streets are represented by parallel lines with their names or route numbers in the center. In dense areas, these names may extend beyond the end of the street. Town names are centered, whereas county names are placed along borders.

The regional map of central New York State shown in Fig. 7 has over 3,800 features consisting of 3,230 lines and 1,770 points. The lines are formed from over 28,000 individual segments. Approximately 18,000 characters have been placed on the map. This map is denser than the Troy map; however, there is more freedom to place names. This map clearly illustrates the capabilities and limitations of the dense-map name-placement system.

CONCLUSION

A rule-based system was developed to automate map production, The overall system creates map databases, places names, and plots maps of high aesthetic quality. All map data used by the name-placement system is based on features, not the graphics representing the features. This approach allows greater flexibility to define map styles and place names. If feature information were not present, it would be almost impossible to label a dense map adequately.

REFERENCES

Ahn, J. and H. Freeman. 1983, A program for automatic name placement: In *Proceedings Auto-Carto Six*, pages 444–453, Ottawa, Canada.

Basoglu, U. 1982, A new approach to automated name placement: In *Proceedings Auto-Carto V*, pages 103–112, Crystal City, Virginia.

Bentley, J. L. 1975, Multidimensional binary search trees used for associative searching: *Communications of the ACM*, 18(9):509–517.

Doerschler, J. S. 1987, *A Rule-Based System for Dense-Map Name Placement.* Technical Report SR-005, CAIP Center, Rutgers, P.O. Box 1390, Piscataway, New Jersey 08855-1390.

Freeman, H. and J. Ahn. 1984, Autonap – an expert system for automatic map name placement: In *Proceedings International Symposium on Spatial Data Handling*, Zurich, Switzerland.

Hirsch, S. A. 1982, An algorithm for automatic name placement around point data: *The American Cartographer*, 9(1):5–17.

Imhof, E. 1975, Positioning names on maps: *The American Cartographer*, 2(2):128–144.

Kelly, P. C. 1980, *Automated Positioning of Feature Names on Maps.* Master's thesis, Department of Geography, State University of New York at Buffalo, Buffalo, New York.

Langran, G. E. and T. K. Poiker. 1986, Integration of name selection and name placement: In *Proceedings Second International Symposium on Spatial Data Handling*, pages 50–64, Seattle, Washington.

Merrill, R. D. 1973, Representation of contours and regions for efficient computer search: *Communications of the ACM*, 16(2):69–82.

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

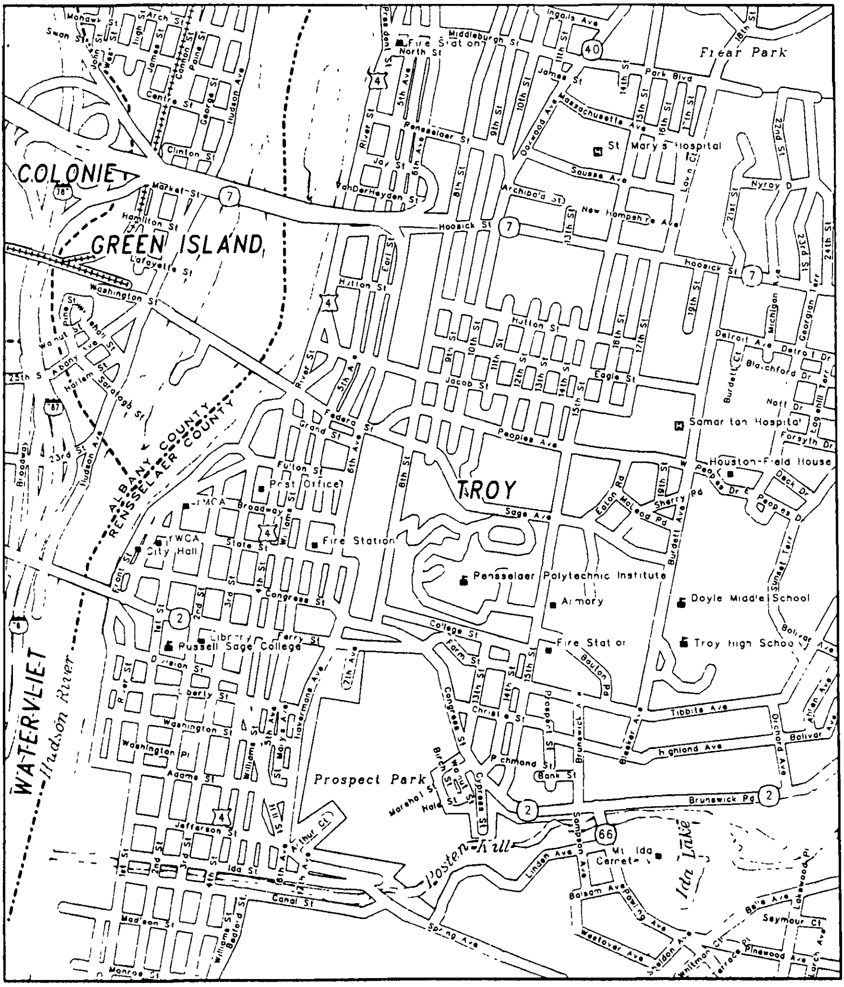

Fig. 6: Example of Automatic Name Placement (City Map)

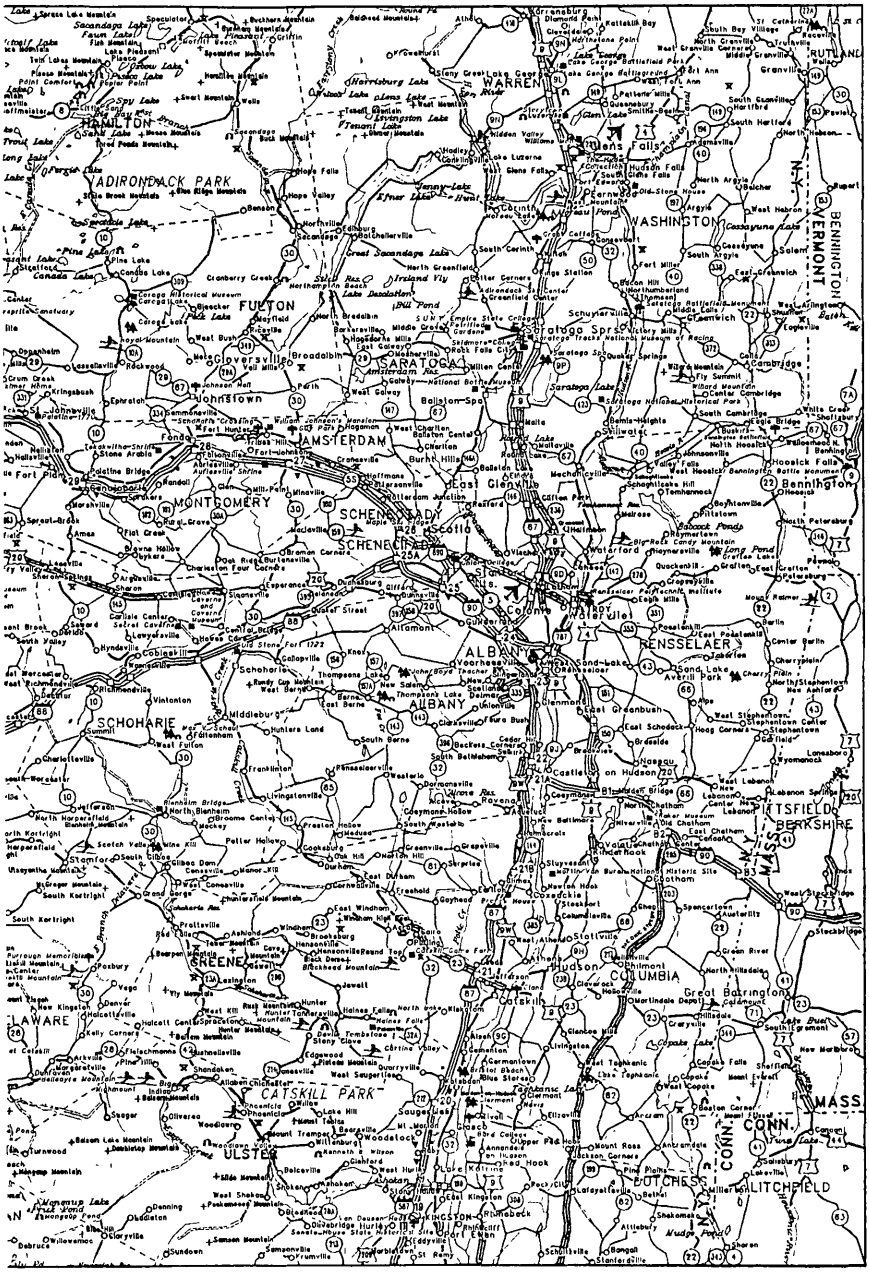

Fig. 7: Example of Automatic Name Placement (State Map)