J. Ahn & H. Freeman, 1984
"A program for automatic name placement"
Cartographica, 21(2&3):101-109

# A PROGRAM FOR AUTOMATIC NAME PLACEMENT

J. AHN and H. FREEMAN, *Image Processing Laboratory, Rensselaer Polytechnic Institute, Troy, New York, USA*

## INTRODUCTION

FEATURE NAME LABELS form an important part of a map for several reasons. The most obvious reason is that the labels tell us the names of the geographical entities represented. However, there are additional reasons. A label can give us an implicit understanding of the linear or area extent and orientation of a map feature, of the nominal class to which the geographical feature belongs, and of the relative size or importance of the feature. This is accomplished through the font, size, spacing and placement of the label. Finally, feature name labels serve also to enhance the aesthetic appearance of a map. It has been said that 'good form and placing of type makes a good map' (Imhof 1975).

The quality of a map depends strongly on the way it is annotated. However, despite all the advances made in map data processing in recent years, the annotation or name placement process has not been automated to any significant degree. This paper describes the development of a computer program which will automatically place names on a map – for point, line and area features – applying the rules used by cartographers. The process is to be fully automatic; however, an interactive editing system will be provided to make minor name placement adjustments or corrections in special cases (Hirsch 1980, Kelly 1980).

### *Basic Definitions*

The data represented in a map are usually of one of three types: – point data, line data, and area data. Point data refer to small, localized places on a map. Examples of point data are cities, villages, mountain peaks, ports, towns, mines, churches, and historic sites. For purposes of computer processing, point data can be represented by x, y-coordinate pairs.

Line data (more precisely, curve data) consist of geographical features that have linear or ribbon-like extent. Examples of line data include rivers, highways, canals, railroads, streets and ship courses. Line data can be described in a computer by the x, y-coordinate values of one or both endpoint(s) and a description of the curve that connects the endpoints. Of the several methods available for these purposes, one is use of a list of closely-spaced point coordinates along the curve; another is chain coding (Freeman 1974).

The third type of data on a map is area data, which represents a region on a map. Counties, states, lakes, oceans, and bays are examples of area data. Area data can be represented in a computer by describing their boundaries in a manner similar to the description of line data. (Note, however, that the boundaries may not always be clearly defined, as in the case of oceans.) Area data can also be represented by a quadtree (Samet 1982) or by tightly closed boundaries (TCBS) (Merrill 1973).

The three types of data can be generalized to one another, depending on the scale of the map. For example, a city, regarded as point data on a national map, may have to be treated as an area on a county map. The problem of map generalization is, however, not addressed in this paper.

### *General Principles*

According to Imhof (Imhof 1975) there are six general principles that should be followed in annotating a map. They are:

1 Names should be easily readable and easily locatable;
2 A name and the object to which it belongs should be easily recognizable;
3 Covering, overlapping, and concealment should be avoided;
4 Names should assist directly in revealing spatial situation, territorial extent, connections, importance, and differentiation of objects;
5 Type arrangement should reflect the classification and hierarchy of objects on the map;
6 Names should not be evenly dispersed nor be densely clustered.

The automatic name placement program developed here was designed to adhere to these principles.

## ANNOTATION RULES

In this section, the rules used by the automatic map annotation system are described. It should be pointed out that the rules given do not form a closed set; special rules may apply to special-purpose maps, and additional rules can be expected to be added as experience is gained with the system. For this reason it was regarded as important that the system be 'rule-based' in its structure and facilitate the addition, modification, or deletion of annotation rules. There is, of course, no single set of annotation rules that is accepted by all cartographers. There are national and regional preferences, some agencies or organizations have particular rules of their own, and finally there are individual style preferences. An automatic annotation system must have the flexibility to accept specific rule sets, corresponding to particular agency standards, particular application requirements, or individual style preferences.

### *Basic Rules*

The rules that apply to the labels of all types of map features are as follows:

1 A name should not overlap another name or a point feature. If a name does overlap a line feature (or a boundary of an area feature), the line, not the name, should be interrupted;
2 Names should not be evenly dispersed nor be densely clustered.


FIGURE 1. *Name placement for an area feature.*

*Area Feature Rules*

The rules that apply to the name of area features are as follows:

1 The label for an area feature should span the entire area and conform to the general shape of the feature, leaving about one and one-half letter spaces at both ends (Figure 1). However, if there is no significant difference between this placement and horizontal placement of an area name, then preference should be given to horizontal placement;
2 Non-horizontal-placed names should not be straight, but curved. The arcs should not be greater than 60 degrees;
3 A name that reads away from the horizontal is preferred over a name that reads toward the horizontal.

*Line Feature Rules*

The rules that apply to the name of line features are as follows:

1 The label for a line feature should conform to the curvature of the line;
2 Complicated and extreme curvatures should be avoided;
3 Line feature labels should not be spread out, but may be repeated at reasonable intervals along the line;
4 For horizontal line features, the names should be placed above the line. For vertical line features, there are two cases: If the line feature lies in the left half of the map, the name should be placed on the left side of the line to read upward. Otherwise, it should be placed on the right side of the line to read downward;
5 One should avoid placing a name near an endpoint of the line feature.

*Point Feature Rules*

The following govern the placement of point feature names:

1 The label for a point feature should be horizontal (usually east-west) and parallel to one of the map boundaries;
2 Point feature labels, like line feature labels, should not be spread out;
3 A point feature label should be close to the point feature to which it refers; though, some specified minimum distance must be maintained;
4 Since the English language has many more 'ascenders' than 'descenders', 'titles' (name labels that are above the point feature) are preferred to 'signatures' (name labels that are below the data item);
5 Although the best possible position of a point label is open to debate, Imhof recommends placement somewhat above and to the right of the point (Imhof 1975).

## THE ANNOTATION ALGORITHM

### *Basic Philosophy*

The basic requirement of any annotation system is that a name must unmistakably refer to the feature it designates and must not overlap point data. In order to achieve this goal, the following guideline was adopted: a name with a smaller degree of freedom is put in place before a name with a larger degree of freedom. This means that area features must be annotated first, since area names must be spread from one end to another, following the general shape of the area. When one considers that relatively large letters of an area name must be aligned with the other letters in the name while avoiding overlapping point data, it is easy to see that the degree of freedom for placing area names is relatively small. This is particularly true when the map is dense.

Point features have the next smallest degree of freedom since their names must be placed near the point to which the name refers. Line features can be labelled last since they have the greatest degree of freedom. Their names can be placed almost anywhere along the line; although, it is best to avoid placing a name too near the endpoints.

### *Area Feature Annotation*

An area name should span the area to which it refers. This makes it necessary first to find a shape description of the area. It was decided to use the so-called *skeleton* method for this purpose (Montanari 1969). The skeleton of an area is obtained as follows: the boundary layers of the area are peeled off one at a time, until the 'wavefront' meets another wavefront. When no more layers can be peeled off, the skeleton of the area will have been determined (Figure 2). Before finding the skeleton of an area, however, a polygonal approximation must be made to smooth out small bays and peninsulas. This is required to prevent small shape variations from having an undue effect on the overall shape description.

Once the skeleton of an area has been determined, the skeleton is divided into sections by identifying the skeleton points whose associated values are less than a threshold that depends on the typesize for the name to be placed. These points are the 'weak links' in the skeleton. All sections but the skeleton with the greatest associated area are deleted, and the skeleton path with the greatest associated area is determined. A straight, horizontal rectangle of fixed height and a curved

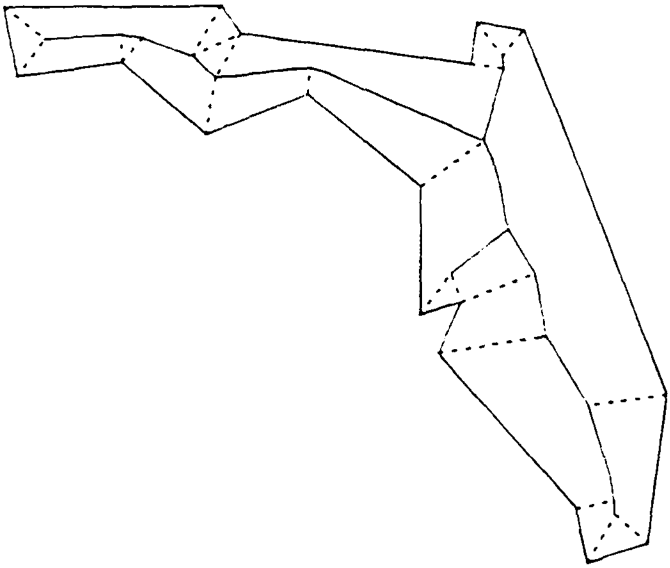

FIGURE 2. *Determination of the skeleton of an area feature.*

non-horizontal 'sausage' of fixed height are fitted to each non-disjoint subsection of this path. The heights of the rectangle and the 'sausage' depend on the typesize of the name. The 'sausage' or rectangle with the greatest area will be the area in which the name is placed.

If adjustments in the position of a name must be made because of conflicting data, they are made by either shifting the name perpendicular to the direction of the name, or altering the spacing between the letters. If a name is too large to fit into an area, it is placed outside the area, using the point-feature name placement algorithm.

### *Point Feature Annotation*

To determine the optimum point-feature name placement, a graph of possible name positions is created. A node in the graph represents a point name. Two nodes are connected with a branch in the graph if their valid name placement areas 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 by choosing any unprocessed node and processing all nodes connected to it, by traversing the connected component in a breadth-first manner. This is repeated until no unprocessed nodes remain in the graph.

Once all the connected components have been determined, each is processed separately, since a node in one connected component can not effect the name label of a node in another.

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 the grid cells that contain at least a part of the node's swept area. 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, that free-space block is removed from the list.

Using the free-space list and the possible-positions 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. The goal state is the state in which the names have been placed for all point features. The search algorithm used is a heuristic graph-searching algorithm similar to the well-known A* algorithm (Nilssen 1971). However, it differs from the A* algorithm in that the nodes are ordered such that the node with the smallest degree of freedom is checked first, regardless of the physical position of the nodes. Since a larger name in a densely-clustered area will have almost no free-space, its degree of freedom will be small and such names will tend to be placed first.

When 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 degrees of freedom for each node, and other internal information at the time the name labels are placed. Since the nodes are already sorted in the order of degrees of freedom, the amount of backtracking is relatively small in most cases.

### *Line Feature Annotation*

Since a line feature may have to be labelled in more than one place, it is divided into segments of fixed length. Each segment is then labelled independently by testing every possible position of the name label for overlaps and selecting the best non-overlapping position. Starting at a certain distance from the end of the segment, sections of the line long enough to contain the name are searched until all possible positions have been considered.

When checking for overlap among name label positions of a line feature and an existing name label or point data, it is only necessary to check the side of the line which should contain the name label. To determine which side should be checked, a straight line is fitted to the section of the line corresponding to the name label position. If the line is vertical (or close to it), the position of the line section relative to the entire map is checked. If the line section is in the left (right) half of the map, the name should be on the left (right) side of the line section. If the line is not vertical, the name should be above the line section. It is necessary to repeat this procedure for every line section since the line feature can change directions and the side of the line to be checked can change accordingly.

Two factors must be considered in deciding whether a particular position is a good position. The first factor, distance from the center of the line segment, is relatively simple to determine. The second factor relates to the curvature of the line at the particular name position. This is determined by computing the absolute deviation of points in the line from a straight line fitted to the line section. This value must be normalized by the number of points that constitute


FIGURE 3. *Example of automatic area-feature name placement.*

the line section before it can be combined with the distance from the center of the line segment to determine the 'desirability' of the name position.

### *Post-Processing Editor*

Although it is intended that the annotation system be completely automatic, it is clear that some provision must be included for interactive editing of the result. The purpose of such editing is to improve the appearance of the map, to correct data errors, and to correct possible mistakes made by the system.

## CURRENT STATUS AND DISCUSSION

Currently, the area and point annotation algorithms have been implemented on a PRIME-750 computer with a program written in RATFOR, a FORTRAN preproces-


FIGURE 4. *Example of point-feature name placement.*

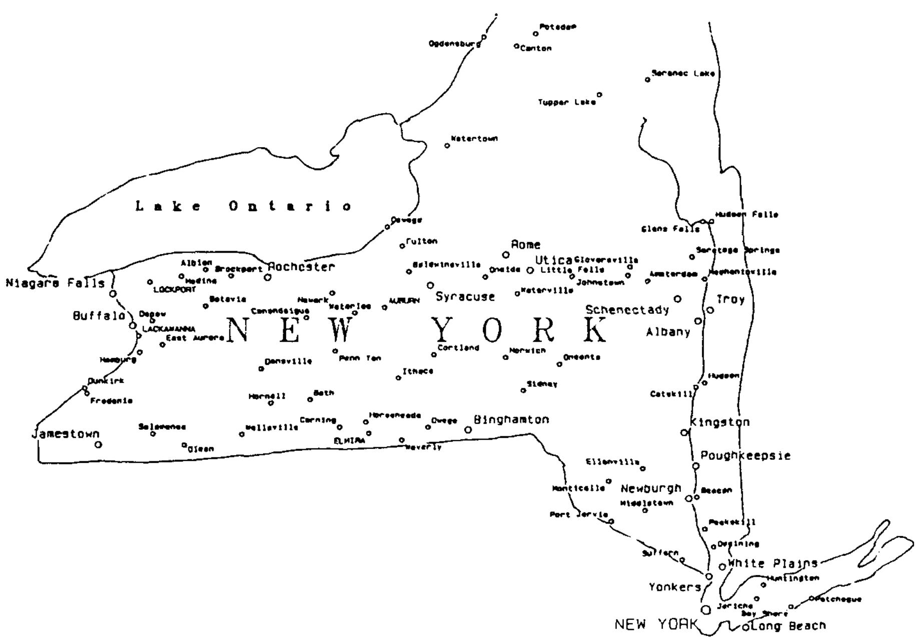


FIGURE 5. *Example of combined area-feature and point-feature name placement.*

sor. Results obtained with the area annotation algorithm are shown for the western part of the United States in Figure 3. An example of point-feature name placement is shown in Figure 4. A map labelled with the point and area annotation algorithms working together is shown in Figure 5.

Although it will be some time before an annotation algorithm will be able to compete in quality against an experienced cartographer, some of the results to date are impressive. A few of the names in Figure 3 are, however, clearly not in the best position. This is most notable for Arizona, for which the name runs in a direction opposite to those of the adjacent states. This is due to the fact that the skeleton considers only the local properties of the area features. A more sophisticated algorithm, which includes a set of rules involving interaction with the names of neighboring areas, has been outlined and will be implemented in the next version of the program.

## ACKNOWLEDGEMENT

The research reported here was supported by the National Science Foundation, Computer Engineering Program, under Grant ECS 81-11619.

## REFERENCES

FREEMAN, H. 1974. Computer processing of line-drawing images, *Computing Surveys*, vol. 6: 1.

HIRSCH, S.A. 1980. Algorithms for automatic name placement of point data, MS thesis, Dept. of Geography, SUNY at Buffalo, Buffalo, NY.

IMHOF, E. 1975. Positioning names on maps, *The American Cartographer*, vol. 2: 2.

KELLY, P.C. 1980. Automated positioning of feature names on maps, MS thesis, Dept. of Geography, SUNY at Buffalo, Buffalo, NY.

MERRILL, R.D. 1973. Representation of contours and regions for efficient computer search, CACM, vol. 16: 2.

MONTANARI, U. 1969. Continuous skeletons from digitized images, JACM, vol. 16: 4.

NILSSEN, N.J. 1971. *Problem-solving methods in artificial intelligence*, McGraw-Hill, New York.

SAMET, H. 1982. Neighbor-finding technique for images represented by quadtrees, CGIP, vol. 18: 1.