MATCH CRITERIA FOR AUTOMATIC ALIGNMENT

Barbara Rosen and Alan Saalfeld
Bureau of the Census
Washington, DC 20233

ABSTRACT

Locally affine rubber-sheeting transformations are being used interactively at the Bureau of the Census to align maps. In trying to align two map representations of the same region, a person will use many visual clues to pair up features on the two maps. Although two individuals will not always agree on the point or line features to be matched, nevertheless, the criteria used to evaluate sameness or difference will probably be quite similar if not identical. The objective of this monograph is to classify the tests, checks, and clues one may employ to determine if a feature on one map is "probably identical" to a feature on another map.

Because exact matching is not always possible (for example, the two maps may be structurally or topologically different), one needs to develop relative or probabilistic measures of matching in order to achieve some degree of comparison of alternative pair-ups. Indeed, each individual's implicit assigning of some unspecified probabilities to possible pair-ups results in his matches being different from those of some other individual. While some "informed assignment of probabilities" and, hence, some arbitrariness is required, nonetheless, many rules may be given quite explicitly and without ambiguity. Those rules deal with relative likelihood or probability in situations and circumstances which are comparable. Without actually assigning an absolute numerical probability (such as 80%) to the event of a match occurring, one may determine that a particular match is more likely than another match because it satisfies all the criteria of the other potential match and then some additional criteria. Or perhaps the first potential match satisfies the same criteria to a higher degree based on some clear relative measures.

INTRODUCTION

Matches are made of corresponding map features: 0-cells of one map are matched to 0-cells of the other, 1-cells are matched to 1-cells, and 2-cells are matched to 2-cells. After a matched pair has been identified, the element in the moving map is repositioned directly on top of the corresponding element in the stationary map. Repositioning brings the moving map into better alignment with the stationary map; and additional matching features are more readily identified.

The approach described in this paper for matching map features through alignment is iterative. First the "most likely" matches are proposed for review by an operator. Then the approved matches are brought into exact alignment by a process of rubber-sheeting the moving map. Again "most likely" matches are proposed (some new candidates may appear due to the rubber-sheeting process) and reviewed by an operator. This alternating match/rubber-sheet process terminates when no new matches are proposed, usually after two or three iterations.

Next independent match criteria are weakened and less likely matches are proposed for review by the operator. At the same time, dependent match criteria (depending on prior matches) may be added to generate potential matches for operator review. Review is always necessary because even strong criteria may result in incorrect matches being proposed. Once again the selection of matched pairs by the operator is followed by a rubber-sheeting of the moving map, and the criteria are reapplied to the rubber-sheeted map until no new potential matches are proposed.

Match criteria may be weakened in small or large steps depending on the experienced reliability of the different criteria to produce accurate matches for the particular map type. For example, urban and rural areas exhibit different success rates with different match criteria. Match criteria may be independent of one another; and the operator may choose to weaken one criterion or another first.

The next section of this paper describes several candidates for match criteria and classifies them as discrete or continuous, topological or geometrical, local or semi-local, and independent or dependent. These criteria have been or are being implemented at the Bureau of the Census as part of a project to match and merge digital map files of the United States Geological Survey and the Bureau of the Census.

CLASSIFICATION OF MATCH CRITERIA

A map may be regarded as a topological object consisting solely of 0-cells, 1-cells, and 2-cells. The 1-skeleton of a map is simply the collection of 0-cells and 1-cells. The 1-skeleton of a map is a mathematical object called a graph. Graph theory studies combinatorial properties of sets of 0-cells and 1-cells or "points and lines," as they are more commonly called. If two maps are topologically the same, then their 1-skeletons will be the same (or isomorphic) graphs. If two maps are similar, their graphs will be similar. One may borrow terminology and results from graph theory to analyze similarity of maps. The graph of a map is also a network; and, hence, one may further borrow network terminology and concepts to analyze maps. A map is also a geometric object, and, as such, may be studied from a geometric point of view as well. Geometry includes the concepts of continuous distance and angle measure in addition to the topological and graph-theoretic notions of adjacency, incidence, orientation, and connectedness.

A 0-cell match is the most elementary match since a match of 1-cells engenders matches of the end point 0-cells of the 1-cells. Similarly, a 2-cell match implies 1-cell and 0-cell matches along the 2-cell boundary. The approach to feature matching in this paper is from smaller to larger: 0-cells are matched first; then 1-cell matches are recognized when corresponding segments have their respective end points already matched as 0-cells; and finally 2-cell pairings are identified as matches after their corresponding total boundaries have been matched. This paper will therefore focus first on 0-cell match criteria.

The first 0-cells to be matched should be special in some fashion. They should possess unusual, unique, or distinguishing characteristics. The 0-cells correspond to intersections of streets; and street intersections may be distinguished by the number and the directions of entering streets. On a map, a place such as "Seven Corners" is quite distinguishable from a typical north-south-east-west city block corner. Seven Corners is often chosen by a map reader in Northern Virginia as a reference point because of its uniqueness. If a map had several intersections with seven corners, however, then having seven corners would not be special and matching would not be so straightforward. This last example is an illustration of dependence—the notion that the likelihood of a match of a particular pair of 0-cells depends on the surrounding 0-cells and how they support or dispute the match.

Consider the following criteria or characteristics of 0-cells:

A. The position of the 0-cell on the map. When two maps are nearly aligned, the corresponding pairs of points should be in approximately the same position. After an initial alignment, pairs of points, one from each map, which form mutually nearest neighbors, become likely candidates for matching. These nearest neighbor pairs are the candidates tested to satisfy other matching criteria before distant pairs are checked. Position is a **geometric** and not a **topological** property because it changes when a topological transformation is applied. Position is a **continuous** variable and not **discrete.** Position of a feature of the rubber-sheeted moving map is **dependent** on the number and kinds of rubber-sheeting iterations performed. Position of a feature **depends** also on matches made of neighboring features. Points near to matched points are moved according to the way the matched points move. Therefore, position is a **semi-local** characteristic or criterion. Semi-local refers to the fact that behavior of nearby 0-cells affects the characteristic of the 0-cell under consideration.

B. The number of 1-cells attached to the 0-cell. This is called the **index** of the 0-cell and is a **discrete** measure because it can assume only positive integer values: 1,2,3, etc. If each of two maps have a unique 0-cell of index 7, then those 0-cells are a likely match. The uniqueness itself is a **dependent** property because it depends on other 0-cells' indices. Having index 7 is an **independent** property because it is not affected by other 0-cells' indices. Having index 7 is also a **local** property because it does not involve characteristics of neighboring 0-cells. Finally, having index 7 is a purely **topological** property and not a **geometric** property because it does not involve the direction or the lengths of the attached 1-cells nor does it involve the position of the original 0-cell.

The figure below illustrates different indices for 0-cells. A 0-cell with an index of 1 corresponds to a dead-end street; and a 0-cell with an index of 2 is called a shape point or inessential 0-cell.

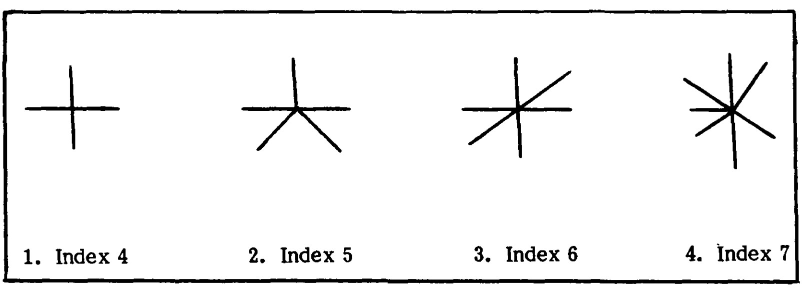

Figure 1. 0-cells of Index 4, 5, 6, and 7.

C. The directions of the 1-cells emanating from a 0-cell. Directions can vary **continuously** and may differ among 0-cells by arbitrarily small amounts. Tolerances and thresholds are required for comparing directions of emanating rays of two 0-cells. If angles are sufficiently similar, then the **local** 1-cell pattern or ray pattern should be regarded as a match. Alternatively one may divide the circle into sectors and thereby consider only a finite number of possible directions. This alternative approach will change direction measure to a discrete measure whose discrete value is the sector in which the 1-cell falls. An eight sector model of the eight principal directions is illustrated below in figure 2.

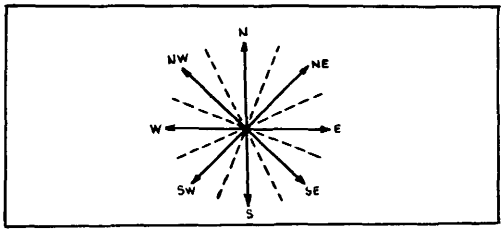

Figure 2. Eight Sectors and Their Defining Principal Directions.

One may also form four-sector or sixteen-sector models. A discrete sector-based measure, while easier to implement than a continuous measure of angle difference, may incorrectly classify as non-matches those intersection patterns whose rays fall near the same sector boundary, but not in the same sector. One may avoid this pitfall by testing for **either** an eight-sector match **or** a sixteen-sector match, since the sector boundaries of the two tests do not coincide.

In experiments with real maps at the Bureau of the Census, the intersection pattern was coded as an eight-bit binary number with each bit storing information on the existence (1) or non-existence (0) of a ray emanating in the corresponding sector. The eight bit positions corresponded to the following directions:

SW	W	NW	N	NE	E	SE	S
1	2	3	4	5	6	7	8

The binary numbers corresponding to the intersection patterns in Figure 1 are as follows:

1. (Index 4): 01010101.
2. (Index 5): 11010110.
3. (Index 6): 11011101.
4. (Index 7): 11111011.

The binary coding of intersection patterns as a single number proved to be an effective shortcut method for storing information. Matches were made with an excellent rate of success after an initial map alignment was performed and nearest neighbor pairs were compared.

Two drawbacks of the eight-sector binary representation are that only eight rays may be coded, and two rays in the same 45^{o} sector cannot both be coded. These limitations did not present any problems when working with real maps. Matches that may have been missed because of these constraints were detected by means of other tests.

The binary coding of the total intersection ray pattern as a single number has been descriptively named the **spider function** of the 0-cell. The spider function exhibits some rather useful properties when two 0-cells have similar, but not identical, intersection ray patterns:

(1) For the eight-sector spider function, counterclockwise rotation of the ray pattern of 45^{o} multiplies the spider function value by 2 modulo 255. If one 0-cell has spider function equal to twice the value of the spider function of another 0-cell (mod 255), then the street pattern of the first 0-cell may be a slight rotation of the pattern of the second 0-cell.

(2) If one ray pattern has an extra ray and everything else the same, then the spider functions differ by a power of 2.

(3) If two ray patterns have the same number of rays, and all ray pairs but one in the same sector; and if furthermore the one pair of rays which do not have matching sectors fall into neighboring sectors, then the spider functions will differ by a power of 2.

Properties (2) and (3) above suggest the likelihood of similar ray patterns if spider functions differ by a power of 2. Property (1) suggests another similar possibility when the quotient of spider functions is 2. When one is looking for clues for matching, one may perform various checks on the spider function in addition to testing for equality.

Dependent Match Criteria

The focus on individual 0-cell properties in the previous section could not avoid the discussion of interactions among 0-cells. The context and the map structure force many dependent relations on every 0-cell—an index value will be unique or not unique based on other index values. There is an intrinsic dependence resulting from underlying topological and geometric structure—there cannot be an odd number of 0-cells of odd index, for example. Another kind of dependence which should be examined for clues to matching, however, is not global (i.e., depending on the whole map), but rather semi-local, depending on neighbors of the 0-cell. Semi-local criteria have been proposed for study at the Bureau of the Census, but have not yet been tested to determine their efficacy. This section will therefore merely list some of the possibilities for semi-local matching criteria.

D. Match precipitation. Once a match has been made of a pair of 0-cells, the 1-cells emanating from those 0-cells lead to very likely candidates for additional 0-cell matches. The neighboring 0-cells, that is, those 0-cells which are a graph distance 1 away from the original 0-cells, may be tested with somewhat relaxed criteria for matches (such as one of the weaker spider function criteria). One may iterate this procedure, following the network out from a match in a spider-web-like manner until either everything is matched or one of the networks terminates or changes form. The figure below indicates the network expansion of matches resulting from the match on the unique 0-cell of index 5:

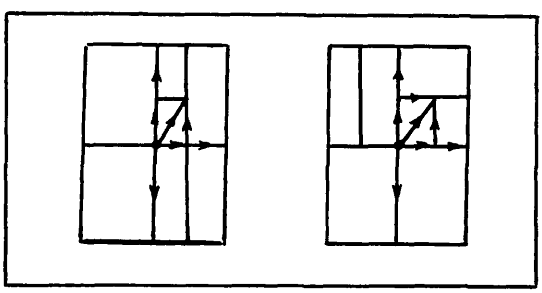

Figure 3. Matches Precipitated by an Index 5 Match.

Matches which can be made by following the network may be made in stages. Each stage involves consideration of all 0-cells at a graph distance 1 from current matches.

Unmatchable criteria as well as match criteria can be defined. The spider function can be applied there as well. The process of matches precipitated through a network may terminate when the sets surrounding the matched set (that is, all those 0-cells at a graph distance 1 from the matched set) are unmatchable.

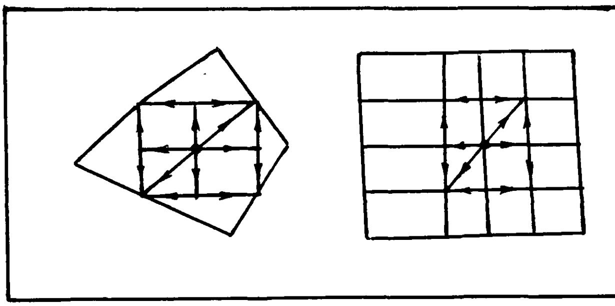

Figure 4. Networks Matched up to Surrounding Unmatchable Set.

E. Triangle or window equivalence (Positional dependence). In the course of the rubber-sheeting process, all of the 0-cells are assigned to enclosing triangles in successively finer triangulations of the space under consideration. On other occasions, quadrant windows are used to partition the space and consequently the 0-cells. When the three 0-cell vertices of a triangle are matched to three image vertices, the triangle interior is assumed to match the interior of image triangle. Naturally it is reasonable to search for 0-cell matches within corresponding triangles. Similarly, after reasonable alignment of the maps, windows should contain corresponding pairs of 0-cells; and it is again reasonable to limit the search for matching pairs to rectangular areas which correspond.

Matching criteria which compare only 0-cells in the corresponding triangles or rectangles are expected to miss a small number of correct matches; however, the search time saved will be considerable in the early stages of matching; and other non-positional criteria may be used later to detect any out-of-triangle or out-of-rectangle correct matches.

If matches are made which fall out of corresponding triangles, the rubber-sheeting may distort the space by producing folds (not one-to-one, but many-to-one affine functions). It is important to detect possible out-of-triangle matches in order to modify the locally affine rubber-sheeting map to recover its one-to-one character.

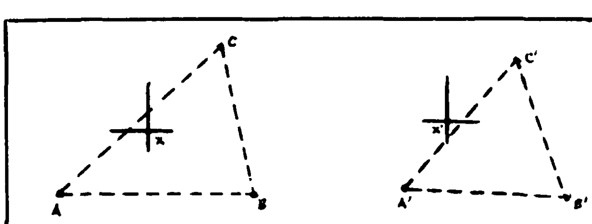

Figure 5. Potentially Dangerous Out-of-Triangle Match.

Another globally dependent match criterion is the following:

F. Minimum total "distance" measure. In order to match features on two maps *en masse*, one may define various "distance" measures (which may include infinite "distance" between known unmatchable pairs) for every possible pair of features, one from each map. James Fagan of the Bureau of the Census has begun experimenting with different distance measures as input to the classic Assignment Problem of linear programming. The solution to the map-match Assignment Problem is a pairing of similar cells which minimizes a measure of combined distance between paired cells. The method uses a single combination measure and its output proposed pairs may be tested by further applying other matching criteria. As experimentation continues, other procedures which are shown to discriminate further among the proposed matches may be incorporated into the defining distance measure. The Assignment Problem is an optimization procedure which produces a unique solution based on the distance measure chosen. More experimentation with distance measures will help assess the value of this approach. The technique may be modified to produce successful matches of only a few of the features as a first alignment routine.

CONCLUSION

Several matching criteria for automatic alignment of maps have been proposed and are being tested at the Bureau of the Census as part of a major development program to match and merge map files from the United States Geological Survey with the Bureau of the Census' own digital files. Combinations of the criteria already tested have produced useful iterative matching routines; and additional criteria are scheduled to be tested and eventually used in conjunction with the working routines to improve their efficiency and success rates. Many areas of mathematics have lent themselves to the development of match criteria, including topology, graph and network theory, geometry, statistics, and optimization.

BIBLIOGRAPHY

Corbett, J., 1979, Topological Principles in Cartography, Bureau of the Census Technical Paper 48, Washington, DC.

Harary, F., 1972, Graph Theory, Addison Wesley, Reading, MA.

Lynch, M.P., and A.J. Saalfeld, 1985, "Conflation: Automated Map Compilation—A Video Game Approach," **AUTOCARTO 7 Proceedings.**

Minieka, E., 1978, Optimization Algorithms for Networks and Graphs, Marcel Dekker, Inc., New York, NY.