CDMTCS
Research
Report
Series

Imprecise Reasoning about Geographic Information

Hans W. Guesgen
Department of Computer Science
University of Auckland

CDMTCS-058
September 1997

Centre for Discrete Mathematics and
Theoretical Computer Science

Abstract

This report summarizes preliminary work on a framework for imprecise reasoning about spatial information, in particular spatial information in geographic information systems. It is based on papers previously published by Jonathan Histed, Ute Lörch, David Poon, and the author.

Geographic information systems have gained an increasing interest over the recent years. However, their abilities are restricted in that they usually reason about precise quantitative information only, which means that they fail whenever exact matches cannot be found. They do not allow for any form of reasoning with imprecision.

In this report, we describe a way of incorporating imprecise qualitative spatial reasoning with quantitative reasoning in geographic information systems. In particular, we show how tessellation data models can be extended to allow for qualitative spatial reasoning. The idea is to associate qualitative information with fuzzy sets whose membership grades are computed by applying the concept of proximity.

In addition, we will show how images like geographic maps or satellite images can be analyzed by computing the distances between given reference colors and the colors that occur in the image, and how the results of this analysis can be used in the fuzzy spatial reasoner.

1 Introduction

Geographic information systems have been in use for nearly thirty years, but there has been little change in the functionality of the systems. The way in which they perform spatial reasoning, i.e., the extraction of new information from stored spatial data, has been quantitative in nature. On the other hand, human beings often prefer a qualitative analysis over a quantitative one, as this is more adequate in many cases from the cognitive point of view.

In the following, we will look at qualitative spatial reasoning in geographic information systems. We will develop a data model to support qualitative spatial reasoning based on constraints, and we will provide an example to illustrate our model.

A geographic information system is viewed here as a special case of a spatio-temporal information system, i.e., a set of procedures to store and manipulate spatially and temporally referenced data. The data used in a spatio-temporal information system may be of any scale. However, the most common systems use data at a geographic scale, i.e., they describe features of our world like roads, lakes, rivers, land, trees, etc.

Geographic information systems require data models (see, for example, [Laurini and Thompson, 1992] for an elaborate introduction to these models). These models vary from GIS to GIS. If the locations in the GIS need to be classified by the value of one of their attributes (e.g., temperature, type of vegetation, rainfall, etc.), then tessellation models are generally used. A tessellated data model is a model in which data is stored in tiles which cover a given area.

Classification is done by assigning values to each tessellation based on ranges defined by an operator. For example, a digitized map from a satellite may show areas of desert, bush, and water at a particular wavelength. By assigning water as 0–20%, vegetation as 21–65%, and desert as 66-100%, a false-color map can be produced which clearly outlines each area on the map.

The most common tessellated model is the raster model, in which an area is split into a matrix of squares. There are three main problems with the raster data model:

1. The size of each tessellation must be determined prior to the entry of data. Therefore it is not possible to change the resolution of the data once it is in the system.

2. The value assigned to a tessellation is assumed to apply evenly over the entire area of the tessellation. If a scale is selected that does not uniquely define the attribute represented in the tessellation, i.e., more than one attribute value may reside within a tessellation, then the value assigned to that tessellation will not accurately describe the contents of the tessellation.

3. The model requires a huge amount of storage space. For example, covering an area of one square kilometer to a resolution of five meters requires 40,000 squares of data. When a data set contains large areas of a single attribute value, many data squares are stored which contain the same data. This is obviously very redundant.

A way of solving most of these problems is to use a many-level tessellation (see Fig. 1) and to implement it by a quadtree. This approach has the advantage that the size of each square is determined solely by whether it has been reduced to cover a single attribute value, i.e., large

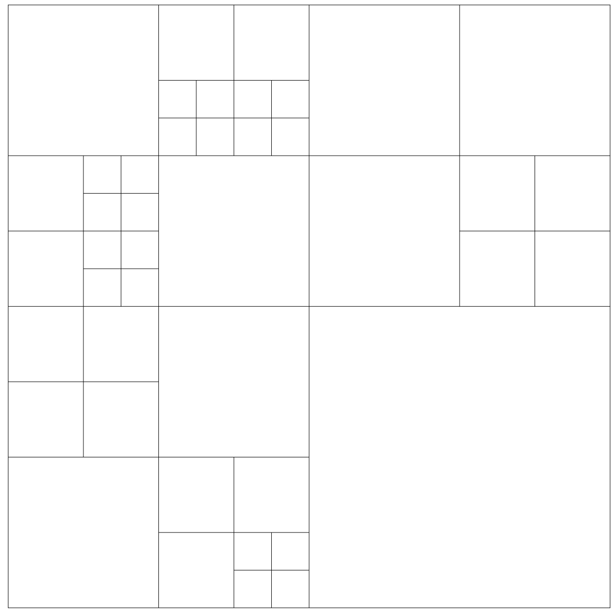


Figure 1: A many-level tessellation. Each area is divided into four smaller areas until each contains only one data value.

areas containing the same data are stored as one large data square, thus reducing storage space and removing redundancy.

2 Quantitative Reasoning

There are various types of spatial reasoning used in standard geographic information systems, one of which is the creation of new maps from existing ones. We will illustrate this type of reasoning by using the city dump scenario as an example. In this example, the task is to find a suitable location for a new city dump given certain constraining factors that depend on the road network, waterways like rivers, lakes, and sea, classification of property into residential, commercial, industrial, rural, and parks, and various types of vegetation. In particular, there are the following constraints:

1. To limit the cost of transporting garbage and to maximize the useful life of the new dump, the city council has decided the following:
   (a) The dump must be within 500 meters of an existing road.
   (b) The dump must have an area of more than 1000 square meters.
2. Environmental legislation has further limited the possible locations for the new dump:
   (a) The dump must be at least 1000 meters from residential or commercial property.
   (b) The dump must be at least 500 meters from any water.
   (c) The dump must not be situated on land covered in native vegetation.

In existing geographic information systems, this type of reasoning uses the concept of map overlay. This is a method for merging different datasets to produce a final output. Various functions are available, including a buffer operator, which increases the size of an object by extending its boundary, and logical operators, such as $\neg$, $\wedge$, and $\vee$.

A possible solution to the city dump problem might be as follows:

1. Select all property which is residential or commercial.
2. Buffer the property selections by 1000 meters.
3. Use $\neg$ to select everything outside this area.
4. Perform similar actions for water.
5. Buffer the road data by 500 meters.
6. Select all vegetation that is not native.

At this point, we have a map showing all locations that are more than 1000 meters from residential or commercial property, a second one showing all locations that are more than 500 meters from water, a third map with all locations that are within 500 meters of a road, and a fourth one with all non-native vegetation. These maps are now overlaid to produce composite maps using the logical operators $\neg$, $\wedge$, and $\vee$. The property and vegetation maps are merged using $\wedge$ to produce a map showing all areas that are more than 1000 meters from residential or commercial property and that do not contain native vegetation. The water and road maps are merged to produce a map showing all areas within 500 meters of a road that are also more than 500 meters from water. Finally, the resulting two maps are merged to produce a map showing all possible locations for the city dump. A selection can then be made from all locations that are more than 1000 square meters in area.

3 Qualitative Reasoning

Quantitative spatial reasoning, as described in the previous section, delivers precise results, but is often too rigid and therefore not applicable to scenarios like the city dump scenario. The reason is that quantitative statements like *All locations that are more than 500 meters from water* and *All locations that are within 500 meters of a road* may eventually result in an empty map, as they restrict the search space too dramatically by excluding any areas, for example, which are 490 meters from water or 510 meters of a road. Such an area, however, might be the best choice available and therefore perfectly acceptable.

This problem can be solved by using qualitative spatial statements rather than quantitative ones. Instead of *All locations that are more than 500 meters from water*, we would put in the restriction *All locations that are far from water*. The system would then analyze the qualitative relation *far from* that is used in this restriction and would find the areas that best match this restriction and that are compatible with the other restrictions.

To achieve this goal, we interpret spatial relations among objects as restrictions on linguistic variables which represent spatial information about the objects.[1] Consider, for example, the position x of some object O in the city. A qualitative approach would specify x in terms of qualitative values like *near church*, *at harbor*, *downtown*, etc. This approach can be translated directly into an approach using linguistic variables.

Informally, a linguistic variable is a variable whose values are words or phrases in a natural or artificial language. The values of a linguistic variable are called linguistic values. For example, the position of O can be represented by a linguistic variable x whose linguistic values are from the domain $\{downtown, near\ church, at\ harbor, \ldots\}$. Using the notation from [Zadeh, 1975], we denote the domain of a linguistic variable as follows:

$$L(x) = downtown + near\ church + at\ harbor + \cdots$$

To express spatial information, we introduce restrictions on the values of the linguistic variables that represent these relations. For example, if O is either downtown or at the harbor, we restrict the value of x to $\{downtown, at\ harbor\}$ and denote this restriction as follows:

$$R(x) = downtown + at\ harbor$$

Spatial relations between objects can be represented by restrictions on composite linguistic variables. For example, the spatial relation between two objects O_1 and O_2 can be represented by introducing a binary composite variable (x_1, x_2), the values of which are from the domain $L(x_1, x_2) = L(x_1) \times L(x_2)$, and a restriction $R(x_1, x_2) \subseteq L(x_1) \times L(x_2)$ on the values of (x_1, x_2). In other terms, a spatial relation is a relation on linguistic variables representing spatial information.

In general, one can distinguish between noninteractive and interactive variables. Two variables x_1 and x_2 are noninteractive if the restriction on (x_1, x_2) is identical with the Cartesian product of the marginal restrictions on x_1 and x_2 [Zadeh, 1975]. For example, let x_1 and x_2 denote the positions of Ernie and Bert, respectively. Then a restriction like *Ernie and Bert are sitting at the table* would make x_1 and x_2 noninteractive, since Ernie's position at the table does not depend on Bert's position at the table (if we disregard the fact that usually Ernie can't sit where Bert is sitting). On the other hand, a restriction like *Ernie and Bert are sitting side by side at the table* would make x_1 and x_2 interactive, since Ernie's position at the table depends on Bert's position at the table (and vice versa).

Usually, spatial relations between objects are represented by restrictions on the values of interactive rather than noninteractive variables. For example, if we want to express the distance between two objects O_1 and O_2, then the relation on the composite variable (x_1, x_2), where x_1 and x_2 are the positions of O_1 and O_2, respectively, causes the variables x_1 and x_2 to interact with each other.

Linguistic variables provide us with a convenient means to express qualitative spatial relations. However, they alone aren't sufficient to integrate qualitative and quantitative spatial reasoning. Only when combined with fuzzy sets, they allow us to add quantitative aspects to the qualitative ones. The next section will discuss this issue.

[1] This step is just a shift in terminology rather than an introduction to a new reasoning method. The motivation for this step is the attempt to stay within the terminology used in fuzzy set theory.

4 Fuzzy Sets

A fuzzy subset $\tilde{R}$ of a domain D is a set of ordered pairs, $\langle d, \mu_{\tilde{R}}(d)\rangle$, where $d \in D$ and $\mu_{\tilde{R}} : D \rightarrow [0,1]$ is the membership function of $\tilde{R}$. In other words: Instead of specifying whether an element d belongs to a subset R of D or not, we assign a grade of membership to d.

The membership function replaces the characteristic function of a classical subset $R \subseteq D$, which maps the set D to $\{0,1\}$ and thereby indicating whether an element belongs to R (indicated by 1) or not (indicated by 0). If the range of $\mu_{\tilde{R}}$ is $\{0,1\}$, $\tilde{R}$ is nonfuzzy and $\mu_{\tilde{R}}(d)$ is identical with the characteristic function of a nonfuzzy set.

Fuzzy sets can be used to associate quantitative information with qualitative one. Consider, for example, a linguistic value like *downtown*. We can associate this qualitative value with a fuzzy set that characterizes for each coordinate on some given street map to which extend this coordinate represents some location downtown. Assuming that D represents the possible coordinate (usually a set of character–digit combinations), *downtown* may be represented by a fuzzy set such as the following:

$$\begin{aligned}\tilde{R} = &\langle M5, 1\rangle + \langle M4, 0.8\rangle + \langle M6, 0.8\rangle + \langle L5, 0.8\rangle + \\ &\langle N5, 0.8\rangle + \langle L4, 0.7\rangle + \langle L6, 0.7\rangle + \cdots\end{aligned}$$

In other words, each location on the city map is considered to be more or less downtown. If its membership grade equals 1, the location is definitely downtown. If it equals 0, then it isn't downtown at all.

If it doesn't cause any confusion, we denote a fuzzy set as follows:

$$\tilde{R} = \sum_{d \in D} \mu_{\tilde{R}}(d)\, d$$

In general, the fuzzy set corresponding to a spatial linguistic value may be a continuous rather than a countable or even finite set. For example, the spatial linguistic value *illuminated*, which specifies that an object is near some light source, may be associated with a fuzzy set $\tilde{R}$ in the domain of real numbers, $\mathbb{R}$. An element $d \in \mathbb{R}$ then indicates the distance of the object to the light source. If the distance is 0, then the object is definitely considered to be illuminated. The greater (the square of) the distance to the light source, the less we consider the object to be illuminated. Since $\tilde{R}$ is a continuous set, we denote it as follows, assuming that $\mu_{\tilde{R}}(d) = 1/(1+d^2)$:

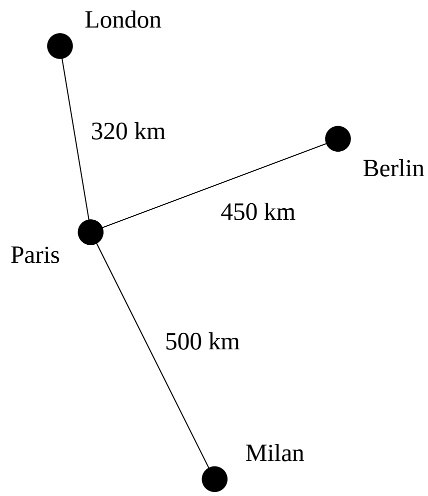
$$\tilde{R} = \int_0^\infty \left\langle d, \frac{1}{1+d^2} \right\rangle$$

In the following, we will discuss in more detail how to compute the membership grades for a qualitative spatial description. Since many qualitative descriptions use the concept of proximity, like in *close to a main road*, we will start with discussing the factors that influence the human perception of proximity.

5 Proximity Factors

There are numerous studies over a wide range of different data domains, including geographical space, of how humans make subjective judgements regarding distances. According to [Gahegan,

Map



Distance matrix

	Berlin	London	Milan	Paris
Berlin	0	?	?	450
London	?	0	?	320
Milan	?	?	0	500
Paris	450	320	500	0

Figure 2: Simple map and the corresponding Euclidean distance matrix showing the (absolute) distances between Berlin, London, Milan, and Paris.

1995], the human perception of closeness or proximity is influenced by the following:

1. In the absence of other objects, humans reason about proximity in a geometric fashion. Furthermore, the relationship between distance and proximity can be approximated by a simple linear relationship.
2. When other objects of the same type are introduced, proximity is judged in part by relative distance, i.e., the distance between a primary object and a reference object.
3. Distance is affected by the size of the area being considered, i.e., the frame of reference.

Proximity measures in spatial reasoning must behave in a way that follows the human perception of proximity. Otherwise, the result of the GIS is counterintuitive and therefore unreliable. In the following, we will address the factors that influence proximity in more detail.

Absolute Distance The simplest form of reasoning about proximity is based on the absolute (or physical) distance between objects. Absolute distance is the major factor that affects proximity. It can be defined by a symmetric Euclidean distance matrix, in which an entry $\delta(O_1, O_2)$ specifies the distance between an object O_1 with coordinates (x_1, y_1) and O_2 with coordinates (x_2, y_2):

$$\delta(O_1, O_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Figure 2 shows the Euclidean distance matrix for a simple map.

Relative Distance In the presence of other objects, the pattern of how the other objects are distributed should be considered in addition to the absolute distance. Taking the distribution from Figure 2, London and Milan are considered to be far apart, because Paris is closer to London than Milan. On the other hand, Milan would be considered to be close to London if Paris were absent from the map. Thus the perceived distance between Milan and London will depend on the presence of the other objects, in particular on the presence of Paris.

Frame of Reference The reference frame plays a significant role in comparing distances between objects. It defines the maximal distance, and therefore an upper bound on all distances under consideration. The maximal distance is given by the length of the diagonal of the boundary of the reference frame.

In Figure 2, Milan may be far from London in a European reference frame, but they are close to each other in a world reference frame. The scale of the reference frame influences the perceived distance. What might be considered as being close in one scale may be regarded as being far away in another scale.

Object Size Sometimes the object size will have an effect on proximity. The larger the size of an object, the closer the objects appear to be. For example, 100 kilometers is regarded as being far apart if we consider two houses, but as being close if we are looking at two cities.

Traveling Costs and Reachability Traveling costs can be expenses, effort, time, etc. Generally speaking, the lesser the traveling costs, the smaller the perceived distance between two objects. For example, it may take less time or money to travel from Paris to Milan, and therefore a traveler may regard Milan as being closer to Paris as London.

Reachability is related to traveling costs. If a city isn't reachable from another city, the costs for getting to that city are infinitely high, and therefore the perceived distance is huge. Or if the other city can't be reached easily because of a missing direct flight connection, then the perceived distance to that city is greater than to a similar city with a direct flight connection.

Traveling Distance In some sense traveling distance is related to reachability. If two objects are connected, then we can establish a traveling distance between the objects by the length of the path that we have to travel on to get from the one object to the other. The path can be a straight line, in which case the traveling distance is equal to the absolute distance. It is more likely, however, that the path is some curved line, which makes the traveling distance greater than the absolute distance.

Attractiveness of Objects Different types of objects involve different types of proximity. The perceived attractiveness of an object is a major factor that affects proximity. For example, *close to the park* may be 500 meters or less, whereas *close to the sewage ponds* may be 10 kilometers or less.

O is closer to A than to B . . . . ☐
O is closer to B than to A . . . . ☐

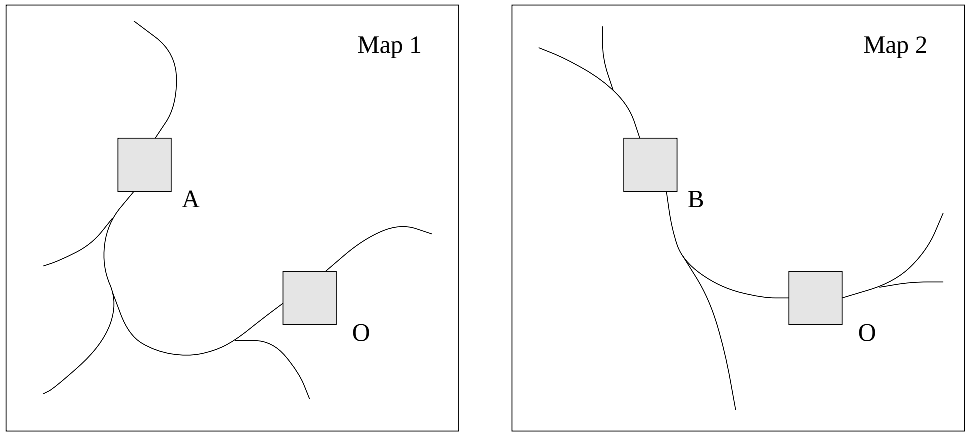


Figure 3: Sample question from a survey for the exploration of the human perception of proximity.

6 A Case Study

To find the most adequate measure of proximity for a GIS, we conducted a survey consisting of 10 questions like the one in Figure 3. Each of the questions referred to two maps, each map consisting of a primary object (O) and a reference object (A or B). The subjects were asked to determine which reference object is closer to the primary object.

Each question catered for a different proximity factor. The distance between O and A is identical with the distance between O and B in every question. If the proximity factor doesn't have any effect on the perception of proximity, then half the subjects should have chosen A as the answer, whereas the other half should have chosen B.

The survey was distributed locally through the internet and attracted 55 participants. Assuming that the subjects did the survey independently and that the probability of the outcome is constant across different subjects, then we can model the distribution of the number of subjects selecting answer A as a binomial distribution with parameters $n = 55$ and $p = 0.5$. Since n and p are large enough, we can use a normal approximation for this binomial distribution.

Although the calculations of critical standard derivations, critical values, etc. on the basis of a sample size of 55 may be inconclusive, they can at least be used as a first indication of the significance of certain proximity factors. In summary, these indications include the following:

- All factors have an effect on the human perception of proximity.
- The absolute distance is the most important proximity factor.
- As the problem reaches a certain degree of complexity, humans cannot reason about absolute distance any more.

7 Grades of Proximity

Since absolute distance seems to be the most important factor that affects the perception of proximity, we can apply the Euclidean distance between the primary object and a given reference object with coordinates (x_r, y_r) to calculate a grade of proximity. This grade can then be used to define a fuzzy set $\tilde{R}$ of reference objects that are in the proximity of the given primary object:

$$\tilde{R} = \int_{x,y \in [0,\infty)} \langle (x,y), \frac{1}{1 + (x - x_r)^2 + (y - y_r)^2} \rangle$$

Although this approach works well for queries related to one particular reference object, like *a place close to the town hall*, it becomes more complex if the spatial query refers to a class of objects, like *a place close to a waste dump*. In the latter case, the query doesn't specify a particular waste dump, but can refer to any waste dump available through the GIS, i.e, it uses a non-deterministic reference object. This type of query is commonly found in applications like resource planning and allocation.

The fuzzy set $\tilde{R}$ for a non-deterministic reference object is defined as the union of the fuzzy sets $\tilde{R}_i$ for the reference objects that belong to the class represented by the non-deterministic reference object:

$$\tilde{R} = \bigcup_{i=1}^{n} \tilde{R}_i$$

In general, computing this fuzzy set is a computationally expensive task. If we assume, however, that computing the union of fuzzy sets means maximizing the proximity grades, then it is sufficient to compute the proximity grade with respect to the closest reference object:

$$\tilde{R} = \int_{x,y \in [0,\infty)} \langle (x,y), \max_{i=1,\ldots,n} \frac{1}{1 + (x - x_{r_i})^2 + (y - y_{r_i})^2} \rangle$$

This means that finding the distance to a non-deterministic reference object is equivalent to searching for the closest reference object.

To find the closest reference object, we developed an algorithm which is related to the nearest neighbor searching algorithm introduced in [Samet, 1990]. Unlike Samet's algorithm, we don't dynamically update our search criteria but use a static criteria computed at the beginning of the search. The idea of the algorithm is to find a good candidate for the closest reference object and restrict the search to the circle that has the primary object at its center and the candidate reference object on its radius. If there are no other reference objects within the circle, then the candidate object is the closest reference object and its distance to the primary object determines the proximity grade of the primary object. Otherwise, we have to search for the closest reference object within the circle.

Assuming that we have stored the information about the reference objects in a quadtree as suggested above, the algorithm recursively proceeds down the tree until it finds the leaf node that corresponds to the location of the primary object. The reference object found in the leaf node becomes the candidate object (see Figure 4). If the leaf node that corresponds to the primary object is empty, we have to enlarge the search area. Every quadtree has the property that at least two of the siblings of an empty node are not empty. This means that the quadrant represented by the parent of an empty node must contain at least two reference objects (see Figure 5). Extending the search circle such that the parent quadrant is contained in the circle

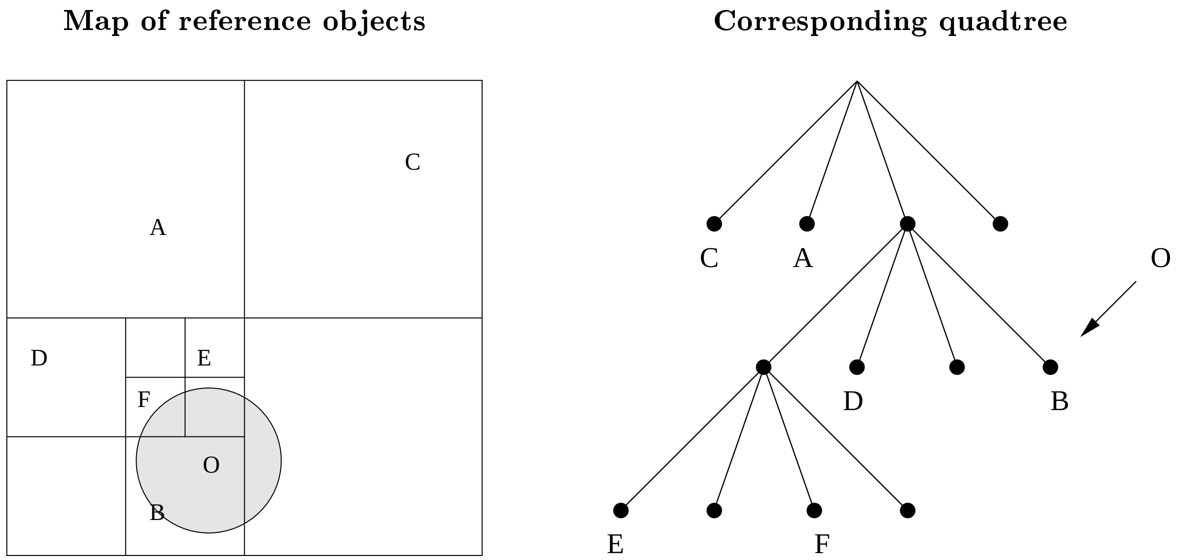


Figure 4: Finding a candidate reference object in a quadtree. In the case shown here, the primary objects corresponds to a quadrant that contains a reference object. The circle given by the primary object and the reference object limits the search space.

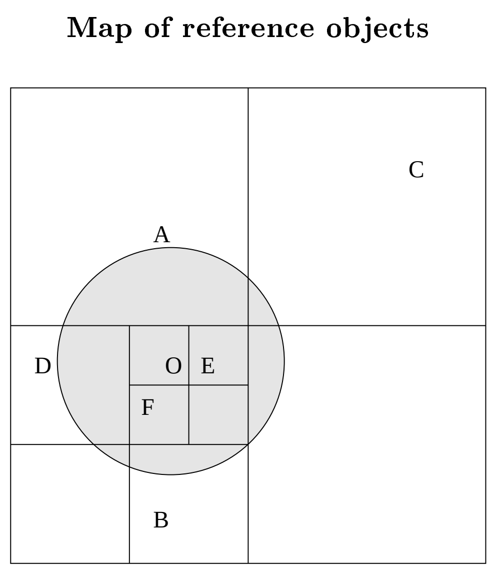


Figure 5: Extended search space for finding a candidate reference object. The primary object corresponds to an empty quadrant, so the search space is extended to include the parent quadrant.

therefore guarantees finding a suitable candidate object.

In summary, the whole algorithm can be divided into the following steps:

1. Compute a search circle that is as small as possible but guarantees containing the closest reference object.
2. Compile a list of the locations of all reference objects in the search circle.
3. Calculate the distance between the primary object and each of the reference objects in the search circle.
4. Select the reference object with the smallest distance.

It can be shown that the algorithm has a linear worst time complexity, due to the fact that the depth of the quadtree is $O(\log \sqrt{m})$, where m is the maximal number of reference objects.

8 Fuzzifying the Distances between Colors

The algorithm introduced in the previous section uses geographic maps as input. So far, we have always assumed that the information to produce these maps is already available in electronic form. This assumption is certainly valid for a large amount of geographic information. On the other hand, there is also a significant proportion of information that isn't yet available online, like old maps. Producing an online version of such maps is often not straightforward, as colors may have faded, lines may have become invisible, etc. In the following, we will develop a framework for converting such 'fuzzy' maps into a representation that can be incorporated with the fuzzy spatial reasoner. The main idea of our approach is to determine a set of reference colors, each color corresponding to an object of a particular type.

Usually, there is some standard as to how certain areas of a map are colored. For example, blue areas on a map most often represent areas of water, whereas green areas stand for rural land. A brute-force approach to analyzing color maps would be to scan in the map and look for certain colors in the resulting file. However, this approach fails in most situations, as colors differ slightly from map to map.

So instead of searching for a particular color, we look for colors in a particular range, like all areas from light to deep blue. In other words, we are looking for a color that is close to a given reference color, which means that we need to define distances between colors. The smaller the distance of some color to the reference color, the more we consider this color to be the same as the reference color. By mapping distances to fuzzy values such that a distance of 0 is mapped to a fuzzy value of 1, we are able to define a fuzzy set for the reference color.

There are several different formats for image files, many of them using the RGB color space to specify the color of a particular pixel in the image. In the RGB color space, the color of a pixel is split into its red, green, and blue component, similar to how the color is generated on a CRT tube. Most visible colors can be generated that way. Our first attempt was to use the RGB color space to compute distances between colors. Let $c_1 = (r_1, g_1, b_1)$ and $c_2 = (r_2, g_2, b_2)$ be two colors, then the distance between c_1 and c_2 would be defined as follows:

$$\Delta(c_1, c_2) = \sqrt{(r_2 - r_1)^2 + (g_2 - g_1)^2 + (b_2 - b_1)^2}$$

Each color in the RGB color space can be associated with a fuzzy set for that color. Since the largest distance in the RGB cube is $\sqrt{3}$, we define this fuzzy set as follows:

$$\tilde{R}_c = \{\langle c', \mu(c')\rangle \mid \mu(c') = (\sqrt{3} - \Delta(c', c))/\sqrt{3}\}$$

Although this looks reasonable at first sight, it has a major drawback: it is not consistent with how humans perceive color differences. Given two colors, c_1 and c_1', and two other colors, c_2 and c_2', then c_1 and c_1' might look more different than c_2 and c_2', even if $\Delta(c_1, c_1')$ is equal to $\Delta(c_2, c_2')$. A way to solve this problem is to switch to a color space that—although not as well known as the RGB color space—reflects human color perception more accurately. This color space is the CIE Luv color space [Foley *et al.*, 1990].

Coordinates in the RGB color space can be transformed into coordinates in the CIE Luv color space by first mapping them to coordinates (x, y, z) in the CIE space and then mapping the result to coordinates (L, u, v) in the CIE Luv space. In particular, the transformations are as follows. Given a color (r, g, b) in the RGB color space, a transformation matrix M can be used to calculate the corresponding color (x, y, z) in the CIE space:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = M \begin{pmatrix} r \\ g \\ b \end{pmatrix} \qquad \text{with} \qquad M = \begin{pmatrix} x_r & x_g & x_b \\ y_r & y_g & y_b \\ z_r & z_g & z_b \end{pmatrix}$$

where x_r, x_g, and x_b are the weights applied to the monitor's RGB colors to find x, and so on.

After obtaining the coordinates of the color in the CIE space, we are able to calculate the coordinates (L, u, v) of the color in the CIE Luv space by applying the following formulas:

$$L = 116\,\sqrt[3]{y/y_n} - 16$$

$$u = 13\,L\,(u' - u'_n)$$

$$v = 13\,L\,(v' - v'_n)$$

$$u' = \tfrac{4x}{x+15y+3z} \qquad v' = \tfrac{9y}{x+15y+3z}$$

$$u'_n = \tfrac{4x_n}{x_n+15y_n+3z_n} \qquad v'_n = \tfrac{9y_n}{x_n+15y_n+3z_n}$$

where (x_n, y_n, z_n) are the coordinates of the color that is defined as white.

The new coordinates are in a color space that is perceptually uniform, which means that the distance between, for example, two reds is comparable with the distance between two greens, etc. In other words, given a reference color for a certain area like water, we can now associate with each point on the map a fuzzy value that indicates the degree to which the area is considered to be water. This fuzzy value can then be used in the fuzzy spatial reasoner to deduce new information from the given geographic information in the following way. Instead of classifying a priori each area as being of a certain type, we analyze the color of the area and compute fuzzy values by comparing that color to reference colors for water areas, roads, rural land, etc. These fuzzy values can be combined with the fuzzy values obtained from the qualitative spatial relations, yielding fuzzy values that indicate the degree to which a certain area satisfies the set of initially given restrictions.

9 Conclusion

Qualitative reasoning is reasoning in terms of linguistic values, whereas quantitative reasoning is reasoning based on numerical values such as measurements. Both qualitative and quantitative reasoning are used by humans to deduce new information from given one. However, it is believed that there is a preference towards qualitative reasoning, and that often some sort of translation takes place from quantitative to qualitative information:

> *Humans often convert precise quantitative information into qualitative values (or categories) to gain insight into the truth meaning of the data.* [Booth, 1989]

In this report, we introduced a scheme of representing qualitative spatial information by associating qualitative relations with fuzzy sets. We argued that there are several factors that influence the perception of closeness (or proximity), but that the absolute distance is the main factor. Based on the concept of absolute distance, we introduced an algorithm for computing proximity grades. In particular, we focussed on the problem of computing proximity grades for non-deterministic reference objects. The algorithm has been implemented in Java and is part of the 'playground' on which we explore new concepts and algorithms for reasoning about space.

Beyond that, we outlined a system for fuzzy spatial reasoning. The system takes an image (geographic map, satellite image, etc.) and analyzes the colors in that image by first transforming the colors into a perceptually uniform color space and then associating fuzzy values with the colors. These fuzzy values are combined with the fuzzy values that are obtained from analyzing spatial relation like *close to* or *far from*. As a result, we get a fuzzy classifications of all points in the image.

The work described in this report is more or less closely related to other work on data models for geographic information systems [Frank, 1992; Goodchild, 1992; Gupta *et al.*, 1991; Herring, 1991; Herring, 1992; Raper and Maguire, 1992].

From the viewpoint of AI, the work described here is related to work on qualitative spatial reasoning in the form of reasoning about spatial relations. This has been addressed in a number of papers.

Both Freksa [Freksa, 1990] and Hernández [Hernández, 1991] introduced an extension of Allen's temporal logic [Allen, 1983] to spatial reasoning. The basic idea of these papers is to describe the relationship between two objects by one or several relations from a set of thirteen basic relations.

A similar approach is described in [Mukerjee and Joe, 1990], where objects of a two-dimensional world are characterized by the directions in which the objects are moving and by associating with the objects trajectories along which they are moving.

10 Acknowledgement

Many thanks to Jonathan Histed, Ute Lörch, and David Poon, who worked on various aspects of what is described in this report. We are grateful for the support we received from the University of Auckland Research Committee under grants A18/XXXXX/62090/F3414025, A18/XXXXX/62090/F3414052, and A18/XXXXX/62090/F3414065.

References

[Allen, 1983] J.F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26:832–843, 1983.

[Booth, 1989] L.C. Booth. *Qualitative Reasoning.* Morgan Kaufmann, San Mateo, California, 1989.

[Foley *et al.*, 1990] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes. *Computer Graphics: Principles and Practice.* Addison Wesley, Readings, Massachusetts, second edition, 1990.

[Frank, 1992] A.U. Frank. Spatial concepts, geometric data models, and geometric data structures. *Computers and Geosciences*, 18:409–417, 1992.

[Freksa, 1990] C. Freksa. Qualitative spatial reasoning. In *Proc. Workshop RAUM*, pages 21–36, Koblenz, Germany, 1990.

[Gahegan, 1995] M. Gahegan. Proximity operators for qualitative spatial reasoning. In A.U. Frank and W. Kuhn, editors, *Spatial Information Theory: A Theoretical Basis for GIS*, Lecture Notes in Computer Science 988. Springer, Berlin, Germany, 1995.

[Goodchild, 1992] M.F. Goodchild. Geographic data modeling. *Computers and Geosciences*, 18:401–408, 1992.

[Gupta *et al.*, 1991] A. Gupta, T.E. Weymouth, and R. Jain. An extended object-oriented data model for large image bases. In O. Günther and H.-J. Schek, editors, *Advances in Spatial Databases*, Lecture Notes in Computer Science 525, pages 45–61. Springer, Berlin, Germany, 1991.

[Hernández, 1991] D. Hernández. Relative representation of spatial knowledge: The 2-D case. In D.M. Mark and A.U. Frank, editors, *Cognitive and Linguistic Aspects of Geographic Space*, pages 373–385. Kluwer, Dordrecht, The Netherlands, 1991.

[Herring, 1991] J.R. Herring. The mathematical modeling of spatial and non-spatial information in geographic information systems. In D.M. Mark and A.U. Frank, editors, *Cognitive and Linguistic Aspects of Geographic Space*, pages 313–350. Kluwer, Dordrecht, The Netherlands, 1991.

[Herring, 1992] J.R. Herring. TIGRIS: A data model for an object-oriented geographic information system. *Computers and Geosciences*, 18:443–452, 1992.

[Laurini and Thompson, 1992] R. Laurini and D. Thompson. *Fundamentals of Spatial Information Systems.* Academic Press, London, England, 1992.

[Mukerjee and Joe, 1990] A. Mukerjee and G. Joe. A qualitative model for space. In *Proc. AAAI-90*, pages 721–727, Boston, Massachusetts, 1990.

[Raper and Maguire, 1992] J.F. Raper and D.J. Maguire. Design models and functionality in GIS. *Computers and Geosciences*, 18:387–394, 1992.

[Samet, 1990] H. Samet. *The Design and Analysis of Spatial Data Structures.* Addison-Wesley, Reading, Massachusetts, 1990.

[Zadeh, 1975] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning—I. *Information Sciences*, 8:199–249, 1975.