Qualitatively Describing Objects Using Spatial Prepositions

Alicia Abella **John R. Kender**
Department of Computer Science
Columbia University
New York, NY 10027

Abstract

The objective in this paper is to present a framework for a system that describes objects in a qualitative fashion. A subset of spatial prepositions is chosen and an appropriate quantification is applied to each of them that capture their inherent qualitative properties. The quantifications use such object attributes as area, centers, and elongation properties. The familiar zeroth, first, and second order moments are used to characterize these attributes. This paper will detail how and why the particular quantifications were chosen. Since spatial prepositions are by their nature rather vague and dependent on context a technique for *fuzzifying* the definition of the spatial preposition is explained. Finally an example task is chosen to illustrate the appropriateness of the quantification techniques.

Introduction

The work presented in this paper is motivated by an interest in how spatial prepositions may be used to describe space and more interestingly, the spatial relationship among the objects that occupy that space. This work is not concerned with the natural language aspect of spatial prepositions. Given a particular environment and a particular task, where the task and environment may change, we wish for a framework that describes the elements in the environment. It is this framework that is of concern in this paper.

It is known that language meaning is very much dependent on context. An example of a context dependent use of the spatial preposition *next* as taken from [Landau and Jackendoff, accepted for publication] is *the bicycle next to the house.* We would normally not say *the house next to the bicycle.* This is the case because the house is larger in size and as such it serves as an anchor for those objects around it. The house in this example serves as a reference object, or in an environmental context, as a landmark. In the system presented in this paper either description is acceptable since the only concern is in the spatial arrangement of the objects irrespective of the size or the purpose of either of the two objects.

The treatment of objects in our chosen environment is a binary one. There is not a reference object, or landmark, because we wish to avoid choosing a reference object that would require the use of physical attributes such as color, size, or shape and focus solely on two objects' spatial relationship. If we think about the use of a preposition like *near* we realize that the requirement of a particular shape is not needed for its proper use. Landau and Jackendoff [Landau and Jackendoff, accepted for publication] have categorized spatial prepositions into those that describe volumes, surfaces, points, lines, and axial structure. They have pointed out that an object can be regarded as a "lump" or "blob" as far as most of the commonly used spatial prepositions are concerned. For example the preposition *in* or *inside* can regard an object as a blob as long as the blob has the capacity to surround. Likewise, *near* and *at* only require that the blob have some spatial extent. *Along* requires that an object be fairly linear and horizontal with respect to another.

The work presented in [Herskovits, 1986] covers the topic of spatial prepositions fairly extensively from a natural language perspective. The author only suggests the possibility of constructing a computational model for expressing spatial prepositions. The intent here is to demonstrate that a computational model can be constructed and that it indeed captures the vital properties sufficient for a succinct use of the chosen prepositions. We can encode the spatial prepositions fairly concisely because we are treating objects as "blobs" and because most of the properties characterized by these prepositions can be encoded using geometric properties such as alignment and distance. Other related works can be found in [Lindkvist, 1976; Talmy, 1983].

The following sections will provide the details of the encoding we have chosen and demonstrate them though the use of an example.

Notations and Definitions

The prepositions for which we have encoded are *near, far, inside, above, below, aligned, next.* We have defined a preposition as a predicate that maps k objects to true (T) or false (F); true if the k objects meet the requirements imposed by the preposition and false otherwise.

$$p : O^k \longrightarrow \{T, F\}$$

where p is a preposition and O^k is a k-tuple of objects. In this paper we will consider $k = 2$. Nevertheless, prepositions that involve three objects like *between* can also be represented, using a similar formalism.

Now that we have defined a preposition we need to define an object. Formally, each object is represented by a six element vector that depend on an object's area A, center (x_c, y_c), and inertia tensor $\begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix}$

It is important to scale the elements in this vector so that they have consistent units, in this case units of length, because we will use this vector in the fuzzification procedure described in section 4. Therefore, the k^{th} object is represented by a vector

$$\mu^k = (\sqrt{A^k}, x_c^k, y_c^k, \sqrt[4]{I_{xx}^k}, \sqrt[4]{I_{xy}^k}, \sqrt[4]{I_{yy}^k})$$

The pair of objects μ is represented by a 12-component vector

$$\mu = (\mu^1, \mu^2) \in \mathcal{R}^{12}$$

It is this scaled vector that we will be using in our future calculations.

The parameterization of objects presented above leads to the concept of a bounding box. A bounding box encloses the object using certain criteria. There are various ways in which to compute a bounding box for an object, one of which may be to find the maximum and minimum x and y values belonging to the object. The one we've chosen is defined through the values of ξ_x and ξ_y, that offer a measure of how much an object stretches in the x and y direction. See the Appendix for the derivation.

Two objects define a point in 12D space. A preposition p can be thought of as a set of points $U_p \in \mathcal{R}^{12}$ such that $U_p = \{\mu | p(\mu)\}$. The volume in this 12D space may be able to reveal some of the inherent properties associated with prepositions. In other words, examination of the space occupied by the various sets U_p may tell us something about the spatial prepositions. Vacancies in this 12D space may reveal why we do not have a word to describe certain spatial relationships among objects. The intersection and distances of volumes occupied by various spatial prepositions may reveal a correlation between various prepositions.

We say that objects O^1 and O^2 are in preposition p if $(\mu^1, \mu^2) \in U_p$. This "ideal" set is made up of pairs of object vectors that satisfy the constraints imposed by the preposition p. As we well know, prepositions are inherently vague in their descriptions, and their interpretation may vary from person to person. Because of this, it is important to add some *fuzzifying* agent to our ideal set. The fuzzifying technique is as defined through fuzzy set theory [Klir and Folger, 1988]. The theory of fuzzy sets is used to represent uncertainty, information, and complexity. The theory of classical sets[1] represents certainty. A classical set divides objects in the world into two categories: those that certainly belong to a set and those that certainly do not belong to a set. A fuzzy set, on the other hand, divides the world much more loosely, by introducing vagueness into the categorization process. This means that members of a set belong to that set to a greater or lesser degree than other members of the set. Mathematically, members of the set are assigned a membership grade value that indicates to what degree they belong to the set. This membership grade is usually a real number in the closed interval between 0 and 1. Therefore a member that has a membership grade closer to 1 belongs to the set to a greater degree than a member with a lower membership grade. Because of its properties fuzzy set theory can find application in fields that study how we assimilate information, recognize patterns [Abella, 1992], and simplify complex tasks. In our notation the fuzzified ideal set is defined through a membership function

$$f_{U_p}(\mu) \in [0, 1]$$

We also define a threshold value that depends on how much vagueness we allow before we decide that two objects are no longer describable with the given preposition:

$$f_{U_p}(\mu) \geq \theta_p$$

Computational Model of Spatial Prepositions

The quantification of prepositions entails representing objects through certain physical properties that can then serve as a basis for expressing prepositions. The physical properties we've chosen include object area, centers of mass, and elongation properties. These properties are calculated through the use of the zeroth, first, and second order moments. The basis for this choice of attributes is simplicity and familiarity. What ensues is a brief description of the various prepositions we've chosen to illustrate. Each preposition is defined through a set of inequalities. This results in sets U_p having nonzero measure (i.e. full dimensionality) in $\mathcal{R}^{12}$ which is necessary for the fuzzification procedure described in section 4.

NEAR

We've defined *near* so that objects' bounding boxes

[1] Referred to as "crisp" sets in fuzzy set theory.

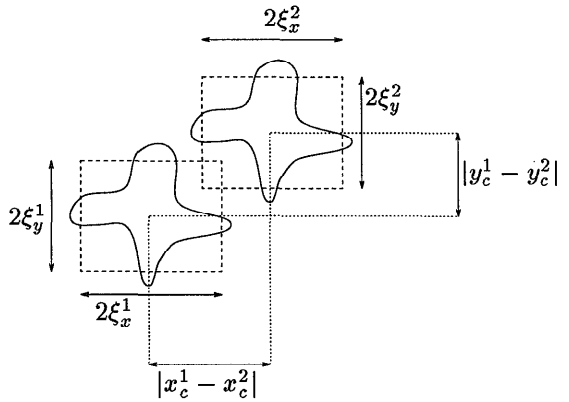

Figure 1: Two objects that are *near* each other

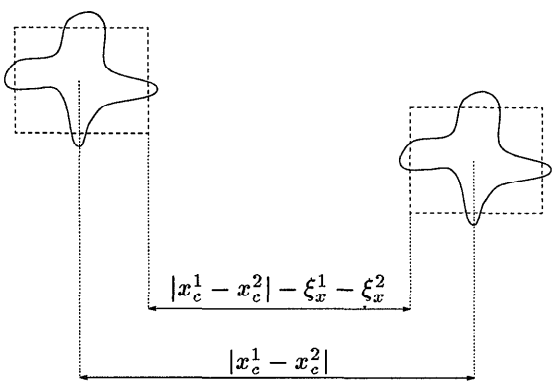

Figure 2: Two objects that are *far* from each other

have a non-empty intersection (see figure 1). Mathematically this is :

$$\xi_x^1 + \xi_x^2 > |x_c^1 - x_c^2| \text{ and } \xi_y^1 + \xi_y^2 > |y_c^1 - y_c^2|$$

FAR

Far is not the complement of *near* as one may initially suspect. We may be faced with a case where an object is neither *near* or *far* from another object, but rather it is *somewhat near* or *somewhat far*. This notion of *somewhat* will be explained more fully when we introduce the concept of *fuzzifying* our "ideal" set. For now it suffices to say that *far* is defined so that the distance between two bounding boxes in either the x extent or the y extent is larger than the maximum length of the two objects in that same x or y extent (see figure 2). Mathematically,

$$|x_c^1 - x_c^2| - (\xi_x^1 + \xi_x^2) > 2\max(\xi_x^1, \xi_x^2) \text{ or}$$
$$|y_c^1 - y_c^2| - (\xi_y^1 + \xi_y^2) > 2\max(\xi_y^1, \xi_y^2)$$

INSIDE

Inside requires that the bounding box of one object

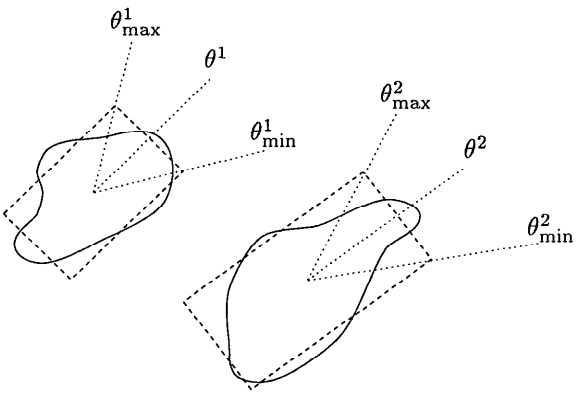

Figure 3: Definition of relevant angles for *aligned*

be completely embedded within the bounding box of another. Formally,

$$\xi_x^1 - \xi_x^2 > |x_c^1 - x_c^2| \text{ and } \xi_y^1 - \xi_y^2 > |y_c^1 - y_c^2|$$

ABOVE, BELOW

Above requires that the projections of bounding boxes on the x axis intersect and that the projections of bounding boxes on the y axis do not intersect. The mathematical relationship is

$$\xi_x^1 + \xi_x^2 > |x_c^1 - x_c^2| \text{ and } \xi_y^1 + \xi_y^2 < y_c^1 - y_c^2$$

Note that *above* is non-commutative. We define *below* similarly. As with *near* and *far*, *above* and *below* are mutually exclusive prepositions. However, not-*above* does not strictly imply *below*.

ALIGNED

The alignment[2] property is angular in nature, therefore its quantification involves inequalities between angles, rather than lengths as the previous prepositions had. For this purpose we define a different type of bounding box that is centered at the object's center of mass and oriented along the object's principal inertia axes with dimensions proportional to the object's maximum and minimum moments of inertia. θ, θ_{min} and θ_{max} are as shown in figure 3. With this in mind, the preposition *aligned* is defined as:

$$\max(\theta^1_{\min}, \theta^2_{\min}) < \theta^i < \min(\theta^1_{\max}, \theta^2_{\max}), \quad i = 1, 2$$

NEXT

We've defined *next* as a combination of the prepositions *near* and *aligned*. Therefore the definition for *next* is:

$$U_{next} = U_{near} \cap U_{aligned}$$

The preposition *next* is an example of a spatial preposition that is a combination of more elementary

[2] Although not a preposition from a language perspective we've adopted it as a spatial preposition.

prepositions. This hints at the possibility of a natural hierarchy of spatial prepositions. It also shows evidence of the possible partitioning of the 12D space mentioned previously.

The Fuzzification of Spatial Prepositions

This section describes why and how we fuzzify spatial prepositions. We need to fuzzify spatial prepositions because they are vague by their very nature; they depend on context and depend on an individual's perception of them with respect to an environment. For these reasons we need to allow for some leeway when deciding if two objects are related through a given preposition.

There is a lot of freedom in how we can fuzzify spatial prepositions, or equivalently, the "ideal" set, U_p. The idea we have adopted is to define the membership function $f_{U_p}(\mu)$ where $\mu \in \mathcal{R}^{12}$ as a function of a distance d between μ and U_p.

$$d = \min_{\mu' \in U_p} |\mu - \mu'|$$

Note that $d(\mu, U_p) = 0$ for $\mu \in U_p$. The distance d tells us by how much the defining preposition inequalities are *not* satisfied. Thus,

$f_{U_p}(\mu) = 1$ for $\mu \in U_p$

$f_{U_p}(\mu) \rightarrow 0$ as $d(\mu, U_p) \rightarrow \infty$

U_p is a multi-dimensional set defined by complex inequalities, for which computing d may be very burdensome. For this reason we resort to a Monte-Carlo simulation with a set of random points around μ that have given statistical properties. The experiments we've conducted use normally distributed random points with mean μ and covariance matrix diag $(\sigma^2, ..., \sigma^2)$. The exact form for f_{U_p} used is

$$f_{U_p} = \begin{cases} 1, & \mu \in U_p \\ \min(1, 2\frac{N'}{N}), & \mu \notin U_p \end{cases}$$

where N is the total number of random points in the Monte-Carlo simulation and N' is the number of points $\mu' \in U_p$. Note that the formulation of f_{U_p} ensures that f_{U_p} for μ very close to the boundary of U_p will have a value close to 1.

The following section will detail some experiments that use this fuzzification technique and put into effect the inequalities that define the given spatial prepositions.

Qualitative Description Experiments

We will use the image shown in figure 4 to illustrate several uses of the prepositions. Each object has been numbered to ease their reference. The image is read as a grey-scale pixel image. It is then thresholded to produce a binary image and objects are located using a sequential labelling algorithm [Horn, 1989]. Once the objects in the scene have been found, the attributes

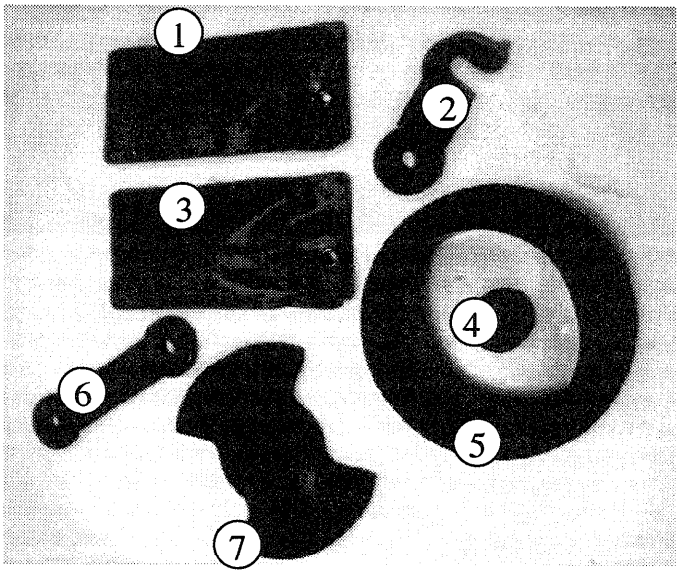

Figure 4: The experimental image

necessary for construction of the 12-dimensional vector are computed (e.g. the area of an object is the sum of all the pixels belonging to the object). Currently the system accepts a spatial preposition and displays all those objects that satisfy the preposition inequalities. The system also accepts as input two objects along with a preposition and it outputs how well those two objects meet the given preposition (the value of f_{U_p} for given σ). All intuitively obvious relations between objects are discovered by the system, e.g. objects 1 and 3 are *next* to each other, etc.

An interesting case, and one that demonstrates the effects of fuzzification is the case of supplying object 2 and object 6 along with the preposition *aligned*. With no fuzzification the system finds that 2 and 6 are not aligned. However, if we allow a certain amount of fuzzification with say $\sigma = 0.03$ the value of $f_{U_{aligned}}$ is 0.8. This value indicates that they may be sufficiently aligned to be regarded as such (which we actually see in the image!), depending on how much leeway we wish to allow. The dependency of $f_{U_{aligned}}$ on σ is shown in figure 5. From this graph we see that the value of the membership function significantly deteriorates for large values of σ. This simply means that the amount of induced uncertainty is so large that the objects cease to possess their original features (such as orientation in this case). This also indicates what the maximal acceptable value for σ should be. In this case, that is $\sigma < 0.1$.

Another interesting case is that of supplying object 2 and object 6 along with the preposition *near* or *far*. Neither satisfies the inequalities precisely. However, if we again, allow for fuzzification, we get a most interesting result, as shown in figure 6. We observe that

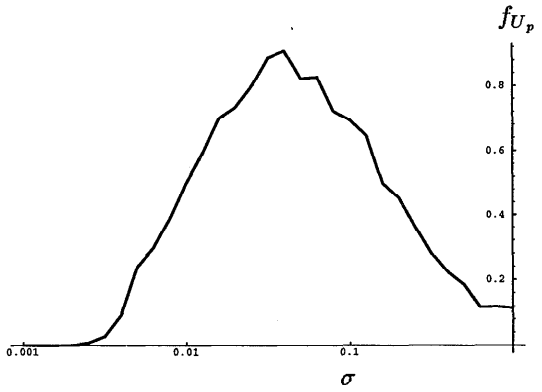

Figure 5: The dependency of $f_{U_{aligned}}(2,6)$ as a function of σ

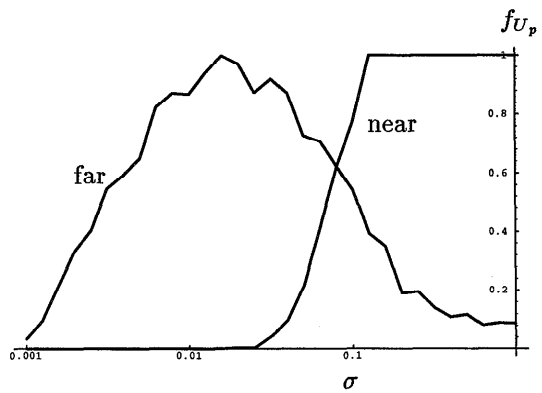

Figure 6: The dependency of $f_{U_{near}}(2,6)$ and $f_{U_{far}}(2,6)$ as a function of σ

although we can not say for certain that object 2 and object 6 are either *near* or *far*, we can say that they are *somewhat near* or *somewhat far*. How we decide which of the two to use can be seen in figure 6. If we examine the slopes of the two curves we see that for small values of σ the slope for *far* is steeper than that for *near*. Therefore it would seem more appropriate to say that 2 is *somewhat far* from 6 as opposed to 2 is *somewhat near* to 6.

Conclusion

The intent of this paper was to establish a computational model for characterizing spatial prepositions for use in describing objects. A quantification was established and demonstrated through the use of an example. A framework to deal with the inherent vagueness of prepositions was also introduced with the use of a fuzzification technique.

An extension of this work would be one in which a user could conduct a dialogue with the system, capable of understanding as well as generating scene descriptions. In other words, we may wish to describe a particular object with as few descriptions as possible through the feedback from the system. The goal would be to home in on the object we are truly referring to through repeatedly supplying additional prepositions to those objects that were singled out after previous inquires. An experiment using this technique may reveal that people naturally describe spatial arrangements in a series of descriptions, rather than once and for all. It may also demonstrate inadequacies in the vocabulary or complexity of a scene. We may also discover that certain environments require that we adopt prepositions that do not exist in the English language for describing a particular sort of spatial arrangement.

References

Abella, A. 1992. Extracting geometric shapes from a set of points. In *Image Understanding Workshop.*

Herskovits, A. 1986. *Language and spatial cognition: An interdisciplinary study of the prepositions in English.* Cambridge University Press.

Horn, Berthold K.P. 1989. *Robot Vision.* The MIT Press.

Klir, G. J. and Folger, T. A. 1988. *Fuzzy Sets, Uncertainty, and Information.* Prentice Hall.

Landau, B. and Jackendoff, R. ation. "What" and "Where" in spatial language and spatial cognition. *BBS.*

Lindkvist, K. 1976. *Comprehensive study of conceptions of locality in which English prepositions occur.* Almqvist & Wiksell International.

Talmy, L. 1983. How language structures space. In *Spatial orientation: Theory, research, and application.* Plenum Press.

Definition of ξ_x and ξ_y

We have used the following two equations to define how much an object stretches in the x and y directions.

$$\xi_x = 2\max\{\sqrt{\frac{I_{\max}}{A}}|\cos\theta|, \sqrt{\frac{I_{\min}}{A}}|\sin\theta|,\}$$

$$\xi_y = 2\max\{\sqrt{\frac{I_{\max}}{A}}|\sin\theta|, \sqrt{\frac{I_{\min}}{A}}|\cos\theta|,\}$$

The following is the derivation of the above formulas. The maximal moment of inertia is given by the formula

$$I_{\max} = \int\int_A u^2 dudv = k\xi_u^2 A$$

where u and v are axes of maximal and minimal moment of inertia respectively, A is an object's area and ξ_u is an elongation parameter that conveys information regarding how much an object "stretches" along the axis u. Constant k is chosen so that in the case of a circle with radius r we have $\xi_u = r$. Simple calculation gives $k = 2$, and formulas for ξ_x and ξ_y are obtained by projecting ξ_u and ξ_v onto axes x and y.