Selection of Best Reference Objects in Object Localizations

Klaus-Peter Gapp
Cognitive Science Program, Dept. of Computer Science
University of Saarbrücken
P.O. Box 151150
D-66041 Saarbrücken, Germany
gapp@cs.uni-sb.de

Abstract

The quality of an object localization depends essentially on the adequate selection of a suitable reference. In most computational approaches developed so far only the distance between the located object and a potential reference object has been used as a decision criterion. However many other criteria have to be considered for a cognitively plausible selection of adequate reference points. In this paper we investigate how object and context dependent properties, like referentiality, visual salience, functional dependencies, or prior knowledge, influence the quality of a reference object. Each factor is quantitatively determined and scaled by relevance to a certain context. The scaling permits the necessary comparability of the different quality criteria. Finally, on the basis of these factors a computational model is presented which permits a context dependent determination of the best reference object in a particular spatial configuration.

Introduction

"Where is object A positioned?". The answer to spatial queries like this requires locating a known object A with reference to another (reference) object. The localization of a particular object is an often required procedure in many applications dealing with the domain *space*. The applications range from general spatial information systems, like geographic information systems or driver information systems, and systems using multi-modal instructions, to applications in Virtual Reality. Understanding the use of spatial references permits the development of better interfaces between systems and human users in the spatial domain.

Most algorithms developed so far in this context use very simple mechanisms for the determination of adequate reference objects (e.g., (Carsten & Janson 1985; Wazinski 1992). The distance from the located object (LO) to the reference object (RO) is often the only criterion used. However many other crucial factors which depend either directly on the objects or on context factors need also to be considered.

The most easily recalled attributes of a region are typically referred to as landmarks. They are used to denote distinguishing features of a route or a region (Lynch 1960; Appleyard 1969; Downs & Stea 1973; Siegel & White 1975). In the first case they support navigational decisions, whereas in the second case landmarks allow for the maintenance of general geographical orientation. Recognizability (Lynch 1960), use (Appleyard 1969), and cultural meaning (Moore 1979) have been emphasized as being the key factors for the relative landmark status of a place. The metaphor *cognitive map* is usually used to describe mental representations of environments, including landmarks, knowledge about the spatial relations between them, routes and metric survey information (Gärling & Golledge 1989). However the cognitive map seems to be systematically distorted and potentially contradictory (see e.g., (Stevens & Coupe 1978; Sadalla, Burroughs, & Staplin 1980; Tversky 1981; Holyoak & Mah 1982; Hirtle & Jonides 1985)), and thus not easily reconcilable in a map-like structure (Tversky 1993). This led Tversky to propose the term *cognitive collage* as often being a more appropriate metaphor for environmental knowledge than cognitive map.

Some landmarks function as spatial reference points, points that serve as the basis for the spatial location of other non-reference points. The concept of spatial reference points implies that the position of a large set of non-reference locations in a particular region is defined in terms of the position of a smaller set of reference locations (Sadalla, Burroughs, & Staplin 1980). Places, known as reference points, are relatively better known and serve to define the location of adjacent points. According to the findings of Rosch (1975) about the existence of asymmetries in similarity judgments between semantic reference points and non-reference points, Sadalla, Burroughs, and Staplin were able to measure asymmetries of cognitive distances between spatial reference points and non-reference points. For example, subjects were able to indicate the proximity of reference points faster than they could indicate the proximity of equidistant non-reference points, and that the direction of particular target locations is more quickly verified relative to reference locations, than relative to non-reference locations. An experimental cluster

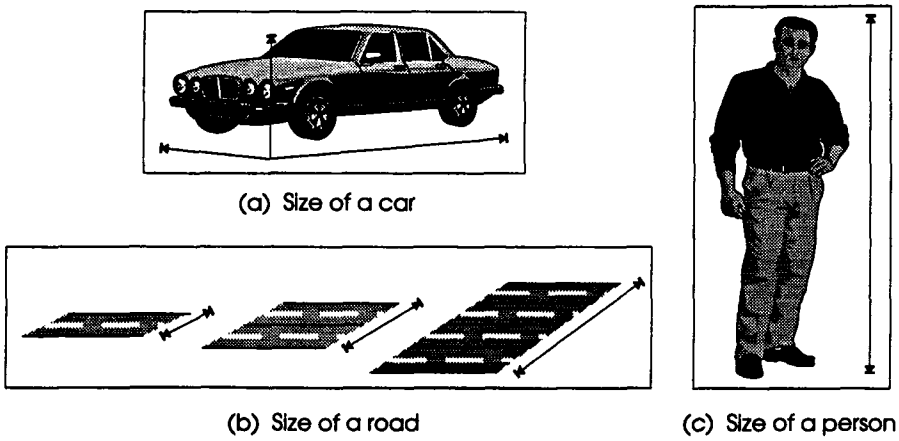

Figure 1: The size of an object

analysis of the six clusters size, familiarity, cultural value, form quality, marker salience, and visual salience showed that the attributes of locations designated as reference points are related to the scales of familiarity, visibility from a distance, domination of nearby places, and cultural importance. This gives us an initial idea of how to categorize the attributes which affect the quality of an object, or of a place to serve as a reference object.[1] The first category is related to features specific to the object or place, and the second is determined by additional context dependencies, like functional dependencies or the prior knowledge of the listener.

In this paper we first analyze the factors that govern the choice of reference landmarks in cognitive usage, and how they can be quantitatively measured. Finally, a domain independent model is developed for selecting suitable reference objects to describe a scene to a human user or to answer queries about spatial configurations.

Quality Criteria for Reference Objects

Object localization requires the establishment of a spatial relationship between an object to be located and one or two reference objects (Talmy 1983; Herskovits 1986). The localization task can be divided up into three parts: First, the object of interest, the LO, is identified, then an object suitable for use as a referent, the RO, is selected, and finally a linguistic description of the relationship between the objects from some specific point of view is specified. In this paper we will put emphasis on what makes an object a good reference object, and how this can be integrated into a computational theory.

[1]Note: *Reference object* in this context means all kinds of spatial objects or places which can be used to locate an object's position.

Talmy (1983) categorized the located or primary object, in contrast to the reference or secondary object, as having spatial variables that tend to be: More mobile, smaller, conceived as being geometrically simpler (often point-like), more salient, more recently on the scene, and more in awareness.

The region where an object can be located depends mainly on its size (Habel & Pribbenow 1988). This means that for large objects, like buildings, the search space for adequate reference objects is significantly larger than for smaller objects, such as a chair or table. Size, as well as many other factors, function as quality criteria in the selection process of adequate reference objects. Features directly related to the potential reference object are considered first, followed by factors related to the particular context.

Relevant Object Features

An important property of a reference object in a general localization task is its visual salience. Visual salience of an object depends on the interaction of basic features like size, shape and color, correlated to the corresponding attributes of the surrounding objects (Treisman 1988). Objects which are large in size with a salient shape and/or color are therefore preferred reference points.

Visual Salience With an object's size one usually refers to the length, width, and height of the object (cf. Figure 1a). However, for some objects, size refers only to one specific dimension, and this dimension varies between objects. For example the size of a road means its width (cf. Figure 1b), and the size of people, their height (cf. Figure 1c). It seems as if the side which varies most obviously, is usually used for size discrimination.

A phenomenon which is related to size perception, and which makes the exact determination of size more difficult, is that of illusory distortions. Classical il-

lusion configurations such as the Müller-Lyer illusion (Figure 2) give rise to strong misperceptions of size. In the Müller-Lyer illusion the arrowheads on the end of the lines make them appear longer or shorter, depending on the direction in which they point. In most current explanations of size distortions it is assumed that this illusion arises because the arrowheads delineate a set of spans either side of the line, which are either longer or shorter than the line itself.

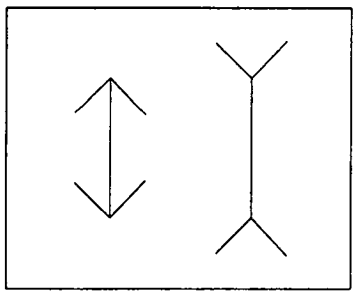

Figure 2: The Müller-Lyer illusion

In (Stuart, Bossomaier, & Johnson 1993) the argument was developed that size illusions are the consequence of interference effects in the size domain. Their empirical work showed that the size domain is coded in parallel, which is an important prerequisite for the assumption of sparse sampling of the size domain.

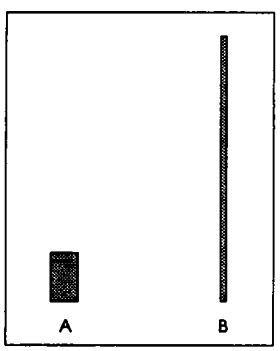

Figure 3: Object B stands out through its large vertical extension

In the context of reference quality, the size of different kinds of objects must be compared. It is therefore proposed, as a first approximation, to express each object's size through its visible horizontal and vertical extensions.[2] The relative size of an object compared to the surrounding objects is then determined by comparing the measured differences in each dimension. The result is that object B (e.g., a flagpole) presented in Figure 3 is more visually salient than object A because of the bigger difference in vertical size when compared to the difference in horizontal size. It remains to be seen which dimension (probably the vertical) has the bigger influence.

"What is smaller is preferably located with respect to what is larger" (Levelt, 1989, p 155), but there are exceptions: Size, as well as most other object features, is defined relative to other proximal objects. If in a certain context all potential candidates for a reference object have the same features of visual salience, except that one object is considerably smaller, then the smaller object might be more salient than the others. Consider for example the two spatial situations shown in Figure 4. In Figure 4a the answer to the question *"Where is the LO positioned?"* would be *"above R_4"*, whilst in Figure 4b the object R_1 would be the preferred reference object. In the first case R_4 is visually salient because it is the largest object. However, in the second case R_1 is more salient than the other objects, because its lower size distinguishes it from the surrounding objects.

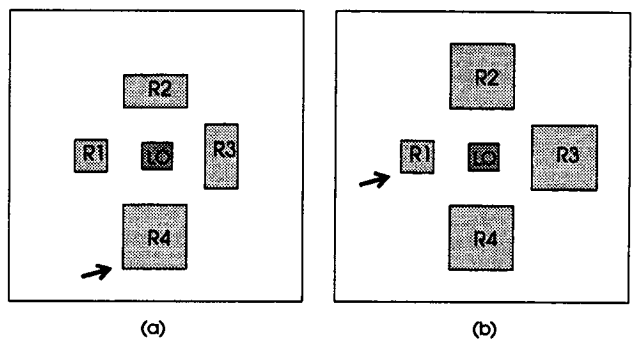

Figure 4: Size as a feature of relative visual salience

If objects are of nearly the same size, then the shape of the objects is another feature which helps to make them more distinguishable. The visual salience of a certain shape is again dependent on the shapes of the surrounding objects (cf. Figure 5a,b). Rotation can increase the salience of an object within a group of objects of the same shape (cf. Figure 5c). A car lying on its side is very salient in a parking area, if all the other cars are in normal position.

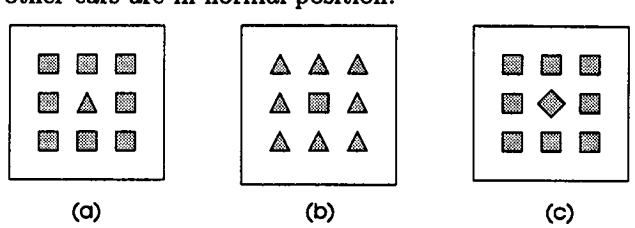

Figure 5: Shape as a feature of visual salience

A feature of objects which is easiest to perceive is *color*. Experimental evidence indicates color or, more exactly, color differences as being directly responsible for the conspicuousness of objects (Carter & Carter

[2]Note: In 3-dimensional space, the perceived objects are projected onto a 2-dimensional plane perpendicular to the line of sight.

1981). Again there is no absolute measure for the salience of a certain color, because, e.g., a light gray dot on a white background is not salient, whereas it is salient on a black background (cf. Figure 6).

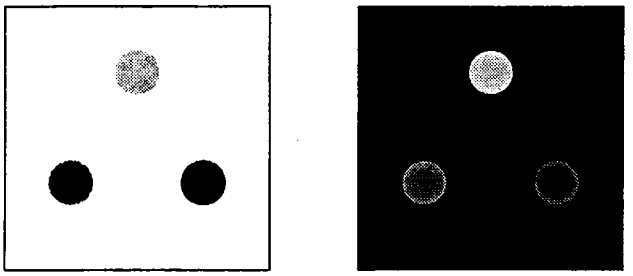

Figure 6: Visual salience: Color differences

Computer systems mainly use the additive, hardware-oriented RGB color theory. In the RGB model each color is divided up into a *red*, a *green*, and a *blue* value. However the use of the RGB color scale is unnatural for man machine communication, since its metrics do not represent color differences on a uniform scale, and colors are not organized in an intuitive manner.

In (Johansson 1949) it was stated that the basic attributes by which humans characterize and distinguish the appearances of colors are 1. hue, 2a. chromaticness 2b. saturation, 3a. lightness and 3b. clearness. According to the international lighting vocabulary (CIE 1987) these values are defined as follows:

Hue is an attribute of a visual sensation according to which an area appears to be similar to one of the perceived colors, red, yellow, green, and blue, or to a combination of two of them.

Chromaticness is an attribute of a visual sensation according to which the perceived color of an area appears to be more or less chromatic.

Saturation is the chromaticness and colorfulness of an area, judged in proportion to its brightness.

Brightness is the attribute of visual sensation according to which a given visual stimulus appears to emit more or less light.

Lightness is the attribute of visual sensation according to which a given visual stimulus appears to emit more or less light in proportions to that emitted by a similar area perceived as having a "white" stimulus.

Johansson proposed that three variables, one from each of the three categories, are necessary to define the appearance of colors unambiguously. In a 3-dimensional geometrical representation of colors, any of these three may be used as major axes. Colors representing constant amounts of these attributes, or constant differences in any of these attributes, in such a geometrical representation are simply illustrated by straight lines.

Systems which describe and classify color based on human color perception are called color appearance systems (Derefeldt 1991). They are defined by perceptual color coordinates or scales and by a uniform or equal visual spacing of colors according to these scales. Only a few color systems fulfill these criteria. These are the Munsell (Munsell, Sloan, & Godlove 1933), the DIN (Normen 1980), the NCS (Hård & Sivik 1981), and the OSA/UCS (Judd 1955) color systems. In many computational applications the CIELUV and the CIELAB color spaces are also used as approximations to color appearance systems[3]. Despite the fact that none of these systems can claim the advantage of representing all the basic appearance attributes of surface colors and their interrelations, the use of color appearance systems in computer applications always requires the measurement of the chromaticity coordinates of the monitor's RGB phosphors. However, in most cases approximations should be precise enough to account for the color when coping with the visual salience of objects.

The actual evaluation method for measuring color differences or contrasts between an object and its background is largely independent of the underlying color system. The problem of a cognitively adequate determination of color salience hasn't been fully solved yet. Therefore an approximation is proposed which works well under the assumption of single-colored objects without textures.

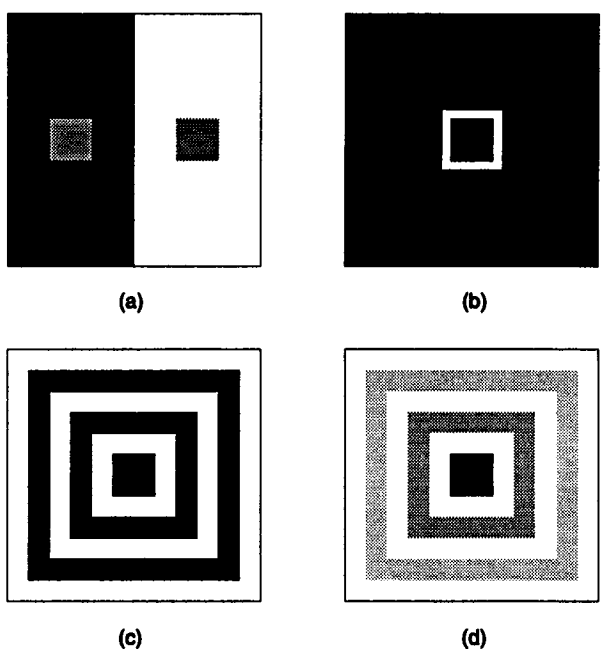

Figure 7: Color contrast

Color contrast has to be computed locally. For example, the small gray squares in Figure 7a are locally

[3]See Derefeldt (1991) and Billmeyer (1985) for an overview of color appearance systems.

salient but globally not salient; in the latter case the background color would also have to be gray. The size of the local background is directly dependent on the extension of the object considered. For example in Figure 7b, the area has to be chosen in such a way that parts of the white *and* the black background are focused. The same is true in Figure 7c: The small black square is not salient if one considers the local background. But in (d) the square is slightly salient because its black color differs from the white and gray colors.

The following algorithm accounts for all phenomena shown in Figures 7a-d. The color structure of an object's local background area is analyzed and related to the object's own color (cf. Figure 8a). The determination of the single color areas, taking into account the current perspective view, can be accomplished with a *z-buffer* or a *ray-tracing* algorithm, depending on the precision required (cf. (Foley *et al.* 1990; Glassner 1989)).

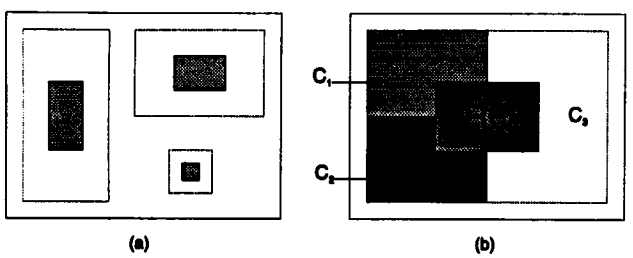

Figure 8: Evaluation of color contrast

The color contrast is then computed as follows: Let n different colors $C_i, i = 1, \ldots, n$, be determined in the local background area of the potential reference object and let $A(C_i)$ be the area occupied by the color C_i. Furthermore, let C_{Ref} be the reference color of the object and D_{cs} a function which measures the visually perceivable distance between two colors with reference to the color appearance system cs. The color contrast CC between an object O and its background is then defined as the sum of the single color differences weighted according to their occupied areas:

$$CC(O) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \frac{A(C_i) \cdot |D_{cs}(C_{Ref}, C_i)|}{\sum_{j=1}^{n} A(C_j)}$$

The algorithm is illustrated in Figure 8b. In the example, only differences in brightness, on a scale between 0 and 1, are used as color differences. The color coordinate of the potential reference object is $C_{Ref} = 0.5$. Three different background colors have to be taken into account $C_1 = 0.4$, $C_2 = 1.0$, and $C_3 = 0.0$. The proportions of occupied background areas are .25, .25, and .50 respectively. Hence the color contrast adds up to:

$$\begin{aligned} CC(RO) &\stackrel{\text{def}}{=} \sum_{i=1}^{3} \frac{A(C_i) \cdot |C_{Ref} -_{cs} C_i|}{\sum_{j=1}^{3} A(C_j)} \\ &= 0.25 \cdot |0.5 - 0.4| + 0.25 \cdot |0.5 - 1| \\ &\quad +0.5 \cdot |0.5 - 0.0| \\ &= 0.25 \cdot 0.1 + 0.25 \cdot 0.5 + 0.5 \cdot 0.5 \\ &= 0.4 \end{aligned}$$

The resulting contrast value of 0.4 corresponds to a moderately salient object with respect to its color. If the object color had been white (black) instead of gray, the contrast would have been 0.35 (0.65). These results support the validity of the proposed algorithm for the computation of local color contrasts independently of a specific color system.

More salient environmental features may have precedence over less salient ones (Taylor & Tversky 1992a). Experimental studies in (Mangold 1986) showed that color dominates size and shape in object identification tasks. Furthermore, size is more easily recognizable than shape. This ordering can be used for a weighted combination of the three features color, size, and shape, related to a single attribute of visual salience.

Mobility A property of an object which is independent of other objects is the distinction between permanently located and movable objects. In most cases, the preferred reference objects are objects with a stationary setting within a certain reference frame (cf. (Talmy 1983)). Even if a movable object is located at a particular place for a long time (e.g., a parked car), a permanent stationary reference object is more suitable.

Permanently located objects are preferred reference points because localizations using movable reference objects run the risk of being applicable only at a certain moment of time. However, there are exceptions, e.g., if the object to be localized is also moving. In such cases, objects moving in the same direction can be appropriate references. Hence it follows that mobility is an object-specific feature which in certain situation is influenced by aspects of the context.

To account for the mobility status of an object in a certain situation, four different *degrees of mobility* are proposed.

1. *permanent*

   This class contains objects (e.g., buildings, countries, oceans) which are not able to move by themselves and cannot in general be moved by external forces.

2. *immobile*

   Objects in this mobility class (e.g., pens, books, tables) are not able to move by themselves, but they can be moved by external forces.

3. potentially mobile
   These objects (e.g., parked cars, anchored ships) can move independently, but they are in a temporarily immobile state.
4. *mobile*
   Mobile objects (e.g., humans, animals, moving cars) are in motion at the time of localization. Hence they will occupy another position at the next moment.

The single classes describe the degrees of mobility of an object at a certain time. Therefore, at different times, objects can belong to different mobility classes. For instance, a brick belongs to Class 2. But its mobility grade changes to permanent, if it becomes part of a house. A car can change its class membership in the same way between the degrees mobile and potentially mobile depending on its current mobility status.

Frame of Reference A last point we want to address in the context of specific object features belongs to the coordination between perception and language required for producing and understanding spatial expressions. Perceptual cues about spatial relations between objects in the perceived environment and expressions that describe these relations must be mapped onto some mental representation of space in order for communication to occur (Miller & Johnson-Laird 1976). The interaction between perception and language requires the adoption of a frame of reference with respect to which spatial positions can be defined (Carlson-Radvansky & Irwin 1993). One distinguishes three different frames of reference: The viewer-centered or deictic, the subject-centered or intrinsic, and the environment-centered or extrinsic frame of reference (Fillmore 1975; Miller & Johnson-Laird 1976; Marr & Nishihara 1978; Talmy 1983; Levelt 1984; Pinker 1985; Retz-Schmidt 1988).

The question, which frame of reference has to be used for communication in a certain context, is not entirely solved yet (cf. (Schober 1995)). It was argued in (Miller & Johnson-Laird 1976) that the choice of the perspective is dependent on features of the configuration to be described. They proposed the preference of an intrinsic perspective where possible, i.e., if the reference object has an intrinsic front. If a deictic interpretation is intended when an intrinsic interpretation is possible, the speaker will usually add explicitly *"from my point of view"* or *"as I am looking at it"* (Miller & Johnson-Laird, 1976, p 398). In contrast, Levelt showed that the same spatial configuration can be described from both an intrinsic and a deictic perspective. However the speaker has to keep a chosen perspective constant for the whole description (Levelt 1982). In (Ehrich 1985) subjects were asked to describe the arrangement of furniture in a doll's house and to analyze the factors that determined the selection of reference objects and relations. In most descriptions a deictic perspective was used. Ehrich interpreted the dominance of the deictic perspective as the necessity of keeping a selected frame of reference constant. Therefore, it is much easier to describe a complete spatial configuration, such as a room, from a deictic perspective.

Following these results the intrinsic perspective is proposed to be used, if the reference object has an intrinsic front. However, if the task is to completely describe a particular environment with more than three objects, then the deictic perspective is preferred.

Context Dependencies

The situational context in which an object localization takes place has a great influence on the selected reference point. Although one often tends to see the basic object features like color, shape, and size as the crucial factors when deciding which reference object to choose, context dependencies are sometimes more decisive.

Referentiality Referentiality, for example, is crucial in the process of object localization. It is not possible to refer to an object if one cannot conclude from its visible parts its semantic interpretation, i.e., if the object cannot be unambiguously identified from a certain perspective. The necessity of referentiality was also verified in (Sadalla, Burroughs, & Staplin 1980), where *visibility from a distance* was rated as one of five important factors for reference points.

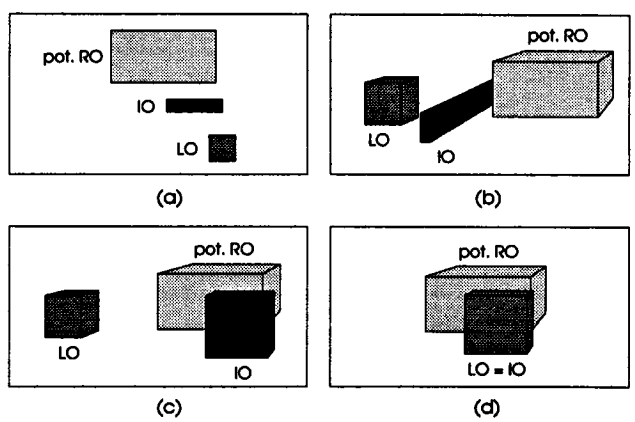

Figure 9: Intervening objects

Object identification is complicated if an object is not in its intrinsic position or if parts are occluded by other objects. Three different kinds of such *intervening objects* (*IO*) can be distinguished. An intervening object can appear between the located object and the potential reference object (cf. Figure 9a,b), or it can be located between the viewer and the reference object (cf. Figure 9c). Also the located object can itself function as an intervening object, if it is positioned between the viewer and the reference object (cf. Figure 9d). Depending on its dimensions, an object located between the located and the reference object functions like a barrier and consequently decreases the quality of a potential reference object (Gapp & Maaß 1994). The

same occurs if in 3-dimensional space from a particular point of view, a reference object is fully or partially occluded from another object. If the occlusion is too severe, e.g., referentiality of the potential reference object is no longer given, then either another object has to be selected for reference or, if possible, the point of view must be changed.

In each case the intervening object occludes parts of the reference object and therefore affects its visual salience. The following method is proposed to account for intervening objects: Depending on the location of the intervening object, either from the located object's center of gravity or the point of view, the intervening object is projected onto the reference object (cf. Figure 10a,b). Depending on the degree of occlusion, the hidden parts of the reference object influence the computation of an object's visual salience (e.g., size), as well as its degree of referentiality.

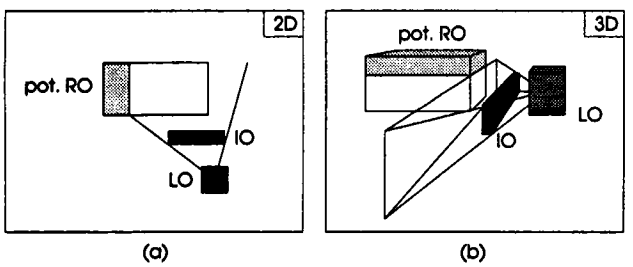

Figure 10: Projection of the intervening object onto the reference object

However, it is too simple to make the referentiality of an object only dependent on its degree of visibility. It is, rather, the visibility of an object's prominent parts, which allow for its identification, that defines referentiality. For example, if only the spire of a church is visible, the partially occluded object is still easily perceivable as a being church. However, if only the spire is occluded, the identification of the object as being a church might be much more difficult, although in the latter case the hidden part of the object is smaller than it was in the former.

The minimal set of an object's prominent parts which need to be visible for an unambiguous identification of the object can not yet be computed automatically and need to be marked for each object by hand. It is therefore proposed to use the degree of visibility of an object as a first approximation for its referentiality. In the examples shown in Figure 10a,b this corresponds to a referentiality factor of about 25% in the 2-dimensional and about 33% in the 3-dimensional case.

Distance The probability of the occurrence of intervening objects depends on the distance between the reference and the located object respective to the point of view. Since distance itself is a decisive criteria for reference objects, the closer a potential reference object is situated to the located object, the better or more precise the localization. This "nearness condition" of objects need not be explicitly mentioned in a localizational phrase, like *"the tree is in front of the church"*. On the contrary, "nearness" is the unmarked case, the default, unless one has evidence to the opposite (Herskovits 1985). If in a particular localization the reference point is further away than usual, distance information, like *500 feet away*, is usually added.

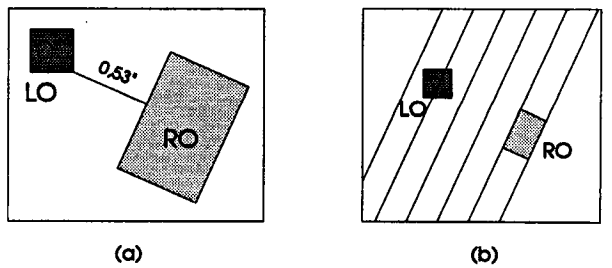

Figure 11: Two methods of distance measurement

The distance between two objects is usually defined as the minimal distance between the borders of the objects (cf. Figure 11a). To account for the asymmetrical distance judgments between ordinary objects and reference objects (Sadalla, Burroughs, & Staplin 1980) a new metric is necessary which, e.g., measures the distance relative to the reference object's extension (cf. Figure 11b). However, the size of an object is already a feature that is used in the determination of a best reference object; thus using such a new metric would consider the object's size twice. It is therefore proposed to measure the distance between the located object and a potential reference point using the shortest distance between the objects. The asymmetrical distance effect is then taken into account in combination with the separate feature "size".

Functional Dependencies Items that are related are more likely to be remembered together (Tulving 1962). Spatial distance as a relation between objects is one factor that leads to a grouping in memory (Stevens & Coupe 1978; Hirtle & Jonides 1985; McNamara 1986; Taylor & Tversky 1992b). Non-spatial organizations may also be used, for example, in remembering items together that are related by function rather than by spatial proximity. In (Hirtle & Jonides 1985) it was found that people tended to group commercial buildings with other commercial buildings, and university buildings with other university buildings, despite the fact that the buildings were spatially interspersed. According to this, in Figure 12 it is more likely that people will choose the monitor as a reference object for the computer than the chest. The semantic relationship between the monitor and the computer and their spatial nearness will lead to the formation of a spatial cluster and to a priming effect (Hirtle & Heidorn 1993). However it is important to note that the priming from non-spatial associations occurs indepen-

dently from a priming by spatial relations (McNamara & LeSeur 1989).

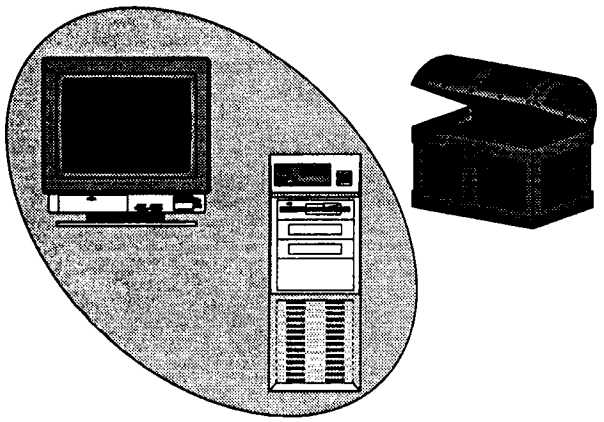

Figure 12: Functional dependencies

Prior Knowledge Grouping and clustering of objects by non-spatial associations is partly dependent on the prior knowledge of the user. The prior knowledge itself can be decisive in the selection process of a reference object. Imagine, for example, a scene with a swimming pool, a soccer field, and a school building as presented in Figure 13. If a pupil is asked where the swimming pool is, he will probably choose the school as the reference. However, a soccer player might tend to refer to the swimming pool as being close to the soccer field.

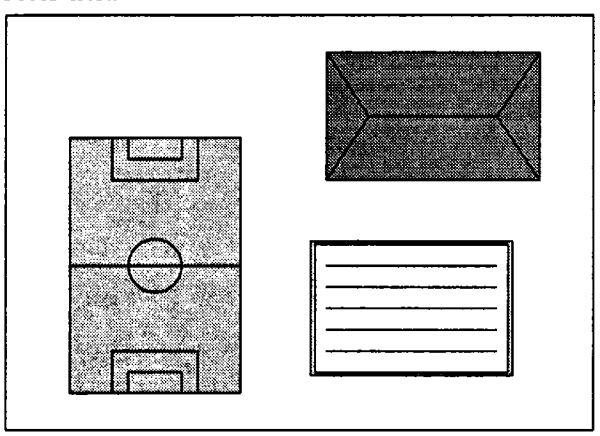

Figure 13: Prior knowledge: The swimming pool is close to the school/soccer field

Prior knowledge influences the localization task in two ways: On the one hand, the chance of an object being selected as a reference object only increases if the object is known to the person asked for its location. On the other hand, objects are also preferred as reference objects if they are more likely to be known to the person who asked for the object's location. In the latter case stereotypes can be defined, which help in a particular situation to classify to which degree an object might be known or not.

Previously Mentioned Objects A last factor we want to address appears if more than one localization has to be carried out, e.g., in a route description task. In such a situation, objects which have been mentioned as reference objects in previous localizations might be better reference objects than unknown new objects. The importance of the factor "previously mentioned" increases if the localizations are done by inspection of a mental map, which means that the user has no visual image of the scene.

Determining the Best Reference Object

In this section we focus on how a reference object in a particular situation can be determined on the basis of the criteria discussed above. The importance of these criteria varies as a function of the prevailing situative context. If, for example, the objects in question are directly visually perceivable, features like referentiality or visual salience are of high priority. On the other hand, if a localization is generated without visual contact, other factors like prior knowledge or already mentioned objects are more important. The following ordered enumeration gives an initial idea of which features are more important than others depending on the context:

<u>Visually (non-)perceivable context</u>

1. (1.) Distance
2. (8.) Referentiality
3. (2.) Mobility
4. (4.) Functional dependencies
5. (6.) Visual salience
   (color, size, shape)
6. (7.) Intervening objects
7. (3.) Previously mentioned objects
8. (9.) Frame of reference
9. (5.) Prior knowledge

However, the order is not strict, and it is intended only to an initial idea of which factors are of higher importance than others. A decrease in influence from one step to the next is not necessary. An exact specification of the priorities requires extensive empirical investigations.

The definition of a procedure which determines the best reference object in a particular situation is a complex task, because the criteria differ in nature, they are object- and context-dependent, and the values from different factors are not directly comparable.

The input of the algorithm is a set of potential reference object candidates which are proximal to the object to be localized. The size of this proximal region depends on the extensions of the located object. They can be efficiently retrieved by using a spatial indexing method, like quad-trees (Finkel & Bentley 1974) or range-trees (Lueker 1978).

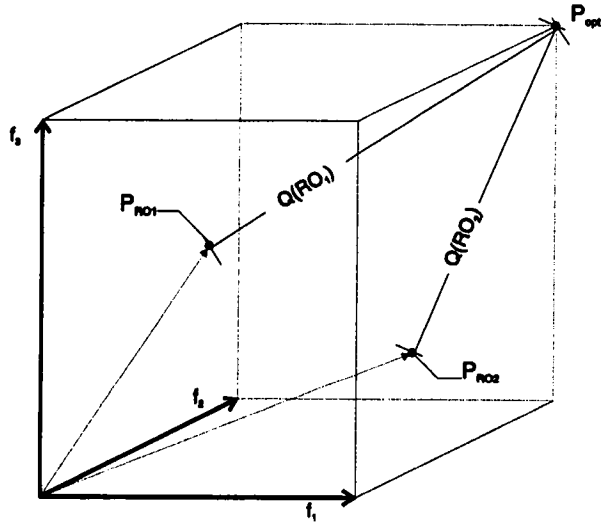

Figure 14: Example: Combination of three features

The following procedure is used for deciding which object should be selected as the best reference object: Each object receives a graded evaluation in the interval $\mathcal{I} = [0..1] \subseteq \mathbb{R}$ for each ordering feature, e.g., 0.0 for a mobile, 0.3 for a potential mobile, 0.6 for a immobile, and 1.0 for a permanently located object. Distance can be scaled by intervals and the complete range from 0 to 1 could be used for visual salience. The same is performed for each context dependent factor. If a reference object candidate was previously mentioned, the graduated factor "previously mentioned" for this object depends on the number of objects, located objects as well as reference objects, mentioned since its last occurrence. Functional dependencies are rated in relation to their intensity and prior knowledge is evaluated by the probability that a particular object is known to a listener.

Assume that n factors are used for the evaluation of an object's quality as a reference object. This results in a n-digit feature vector $f = (f_1, \ldots, f_n)$ in the feature space $\mathcal{F} = \mathcal{I}^n = [0..1]^n$. Depending on a particular situation, context factors can change the importance of one or more dimensions, e.g., if the user has only a mental image of the scene and no direct visual access. The influence of these situational aspects is accounted for by the possibility of weighting the affected dimensions using a context dependent n-dimensional scaling function SC_{cxt}:

$$SC_{cxt} : \mathcal{F} \rightarrow \mathbb{R}^n;$$

$$(f_1, \ldots, f_n) \mapsto (sc_1(f_1), \ldots, sc_n(f_n))$$

with

$$sc_i(f_i) : \mathcal{I} \rightarrow \mathbb{R}, \quad 1 \leq i \leq n$$

The scaling is done in correspondence to the context dependent priority list given above. Additional constraints can be used to modify or even eliminate a feature in a specific localization task.

The optimal reference object can now be defined as the $n - x$ dimensional subspace $\mathcal{S}^{n-x} \subseteq \mathcal{F}$, $1 \leq x \leq n$, in which x is the sum of factors defined by an interval. This means, that if an optimal reference object is defined by only one single value for each factor f_i, then hence $x = n$ and $\mathcal{S}^{n-x} = \mathcal{S}^0$ denotes a point in $\mathcal{F}$ (cf. Figure 14).

The context, specific weighting of the single dimensions, makes the factors comparable with each other. Each reference object RO_i is represented as a n-dimensional point $P_{RO_i} \in \mathcal{F}$, according to its evaluation of the n quality factors. A scale for the quality of a reference object can now be defined as the shortest distance between P_{RO_i} and $\mathcal{S}^{n-x}$. If $\mathcal{S}^{n-x}$ represents a point $P_{opt} = (f_1^{opt}, \ldots, f_n^{opt}) \in \mathcal{F}$ then the quality of a potential reference object $Q(RO_i)$ can be expressed as:

$$Q(RO_i) := |\overline{P_{RO_i} \; P_{opt}}| = \sqrt{\sum_{i=1}^{n} (SC_i(f_i) - f_i^{opt})^2}$$

Hence the best reference object RO_i is defined as the object with the minimal "distance" $Q(RO_i)$. If more than one object have the same distance, direct comparisons on the basis of the applicable spatial relations can be used for a final decision (cf. (Gapp 1994; 1995a; 1995b)).

This quality measure for reference objects represents only one possibility of combining the single feature dimensions. The special characteristic of linear combinations is that weak evaluations of features affect the final result disproportionately. For instance, if an object O receives the evaluations 0.5 and 1.0 in a 2-dimensional feature space, its quality to serve as a reference would be $Q(O) = 0.5$. The determination of the weighted average represents another method for a quality estimation which doesn't overvalue weak evaluations. The weighting is necessary to reflect the additional scaling of the single feature dimensions. In the example above, the use of the average would result in a quality measure of $Q(O) = 0.75$. It remains to be shown using empirical investigations which method represents a cognitively adequate quality measure for the determination of best reference objects.

The algorithm presented is not dependent on a certain number of features. The more features used, the better the performance of the algorithm. The performance also depends directly on the quality of the calculation of the individual factors and on their weighting. The individual features of an object can be computed in parallel, however dependencies between some of them can force the reconsideration of already independently computed factors. A parallel processing of all potential candidates is also encouraged. On the one hand this ensures efficient processing, and on the other hand it permits the use of *anytime* architecture

(cf. (Boddy 1991)) for the system, which allows the generation of answer at "any" time, with increasing quality the longer the computation lasts (resource dependent processing).

Discussion

A cognitively plausible description of an object's location in natural language requires the selection of an adequate reference object. The factors which determine the suitability of an object to serve as a reference object in a particular situation vary and depend both on the object itself and on the context. Problems are related to the calculation of single factors, which are partly dependent on each other, and to their importance to the quality of a potential reference object, which also differs with the context. An exact definition through the use of empirical studies seems to be difficult, given the high complexity of the problem. However, the computational model proposed above permits the computation of an approximated solution, which can be refined by using more factors, modified weights, and a more precise calculation of the single factors. The use of neural nets or genetic algorithms, instead of the distance measure for analyzing an object's reference qualities, also seem to be promising alternatives.

Conclusion

A cognitively plausible localization of an object requires a context dependent selection of an adequate reference object, which permits the optimal determination of the objects position. We investigated which object and context dependent factors are essential to fulfill this complex task. It turned out that factors which are directly dependent from the potential reference object, like visual salience, have to be defined relative to surrounding objects. On the basis of these investigations a theory for the computation of the best reference object was developed. The approach accounts for the different priorities of the single features dependent on the current context and the set of factors considered can be restricted or enhanced without changing the algorithm. An implementation can benefit in performance through the full parallel design of the algorithm. Even anytime architecture can easily be applied.

Acknowledgments

The author would like to thank Bernd Andes, Anselm Blocher, and Amitabha Mukerjee for valuable discussion. The clarity of this paper was improved thanks to comments from Marc Gurmann, Anthony Jameson, and the two anonymous reviewers.

References

Appleyard, D. 1969. Why buildings are known. *Environment and Behaviour* 1:131–156.

Billmeyer, F. W. J. 1985. AIC annotated bibliography on color order systems. Mimeoform Service, rear, 4805 Prince George's Avenue, Bletsville, MD 20705.

Boddy, M. 1991. Anytime problem solving using dynamic programming. In *Proc. of AAAI-91*, 738–743.

Carlson-Radvansky, L. A., and Irwin, D. E. 1993. Frames of reference in vision and language: Where is above? *Cognition* 46:223–244.

Carsten, I., and Janson, T. 1985. Verfahren zur evaluierung räumlicher Präpositionen anhand geometrischer Szenenbeschreibungen. Master's thesis, FB Informatik, Univ. Hamburg.

Carter, R. C., and Carter, E. C. 1981. Color and conspicousness. *Optical Society of America* 71:723–729.

CIE. 1987. International lightning vocabulary. Publication CIE 17.4, Bureau Central de la Commission Electrotechnique Internationale, Genève, Suisse.

Derefeldt, G. 1991. Colour appearance systems. In Gouras, P., ed., *The Perception of Colour*. Beccles and London, UK: The Macmillan Press. chapter 13, 218–256.

Downs, R. M., and Stea, D. 1973. Cognitive maps and spatial behaviour: Process and products. In Downs, R. M., and Stea, D., eds., *Image and Environment. Cognitive Mapping and Spatial Behaviour*. Chicago: Aldine. 8–26.

Ehrich, V. 1985. Zur Linguistik und Psycholinguistik der sekundären Raumdeixis. In Schweizer, H., ed., *Sprache und Raum*. Stuttgart: Metzler. 130–161.

Fillmore, C. J. 1975. *Santa Cruz Lectures on Deixis*. Bloomington, IN: Indiana University Linguisic Club. only reproduced, first published in 1971.

Finkel, R. A., and Bentley, J. L. 1974. Quad trees: A data structure for retrieval on composite keys. *Acta Informatica* 4:1–9.

Foley, J. D.; van Dam, A.; Feiner, S. K.; and Hughes, J. F. 1990. *Computer Graphics. Principles and Practice*. Reading, MA: Addison-Wesley, second edition.

Gapp, K.-P., and Maaß, W. 1994. Spatial layout identification and incremental descriptions. In *Proc. of the AAAI-94 Workshop on Integration of NL and Vision Processing*, 145–152.

Gapp, K.-P. 1994. Basic meanings of spatial relations: Computation and evaluation in 3d space. In *Proc. of AAAI-94*, 1393–1398.

Gapp, K.-P. 1995a. Angle, distance, shape, and their relationship to projective relations. In Moore, J. D., and Lehman, J. F., eds., *Proc. of the 17th Annual Conference of the Cognitive Science Society*, 112–117. Mahwah, NJ: Cognitive Science Society.

Gapp, K.-P. 1995b. Processing spatial relations in object localization tasks. In *Proc. of the AAAI Fall Symposium on "Computational Models for Integrating Language and Vision"*.

Gärling, T., and Golledge, R. G. 1989. Environmental perception and cognition. In Zube, E. H., and Moore, G. T., eds., *Advances in environment, behaviour, and design*. New York: Plenum. 203–236, Vol. 2.

Glassner, A. S. 1989. *An Introduction to Ray Tracing*. London - San Diego - New York: Academic Press.

Habel, C., and Pribbenow, S. 1988. Gebietskonstituierende Prozesse. LILOG-Report 18, IBM Deutschland GmbH, Stuttgart.

Hård, A., and Sivik, L. 1981. Ncs-natural color system: A swedish standard for color notation. *Color Res. Appl.* 6:129–138.

Herskovits, A. 1985. Semantics and pragmatics of locative expressions. *Cognitive Science* 9(3):341–378.

Herskovits, A. 1986. *Language and Spatial Cognition. An Interdisciplinary Study of the Prepositions in English*. Cambridge, London: Cambridge University Press.

Hirtle, S. C., and Heidorn, P. B. 1993. The structure of cognitive maps: Representation and processes. In Gärling, T., and Golledge, R. G., eds., *Behaviour and Environment: Psychological and Geographical Approaches*. Elsevier Science Publishers. chapter 7, 177–189.

Hirtle, S. C., and Jonides, J. 1985. Evidence of hierachies in cognitive maps. *Memory and Cognition* 13(208-217).

Holyoak, K. J., and Mah, W. A. 1982. Cognitive reference points in judgement of symbolic magnitude. *Cognitive Psychology* 14:328–352.

Johansson, T. 1949. Characteristic properties of colour and colour combinations. *Revue d'Optique Theorique et Instrumentale* 28:241–246.

Judd, D. 1955. Progress report by optical society of america committee on uniform color scales (note). *J. Opt. Soc. Am.* 45:673–676. Progress Report No. 1.

Levelt, W. J. M. 1982. Cognitive styles in the use of spatial direction terms. In Jarvella, R. J., and Klein, W., eds., *Speech, Place, and Action*. Chichester: Wiley. 251–268.

Levelt, W. J. M. 1984. Some perceptual limitations on talking about space. In van Doorn, A. J.; van de Grind, W. A.; and Koenderink, J. J., eds., *Limits in perception*. Utrecht: VNU Science Press. 328–358.

Levelt, W. J. M. 1989. *Speaking — From Intention to Articulation*. Cambridge, MA: MIT Press.

Lueker, G. S. 1978. A data structure for orthogonal range queries. In *Proc. of the 19th IEEE Symposium on Foundations of Computer Science*, 28–34.

Lynch, K. 1960. *The Image of the City*. Cambridge, MA: MIT Press.

Mangold, R. 1986. *Sensorische Faktoren beim Verstehen überspezifizierter Objektbenennungen*, volume 185. Frankfurt am Main: Lang.

Marr, D., and Nishihara, H. K. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. In *Proc. of the Royal Society of London*, 269–294.

McNamara, T. P., and LeSeur, L. L. 1989. Mental representations of spatial and nonspatial relations. *The Quarterly Journal of Experimental Psychology* 15:211–227.

McNamara, T. P. 1986. Mental representations of spatial relations. *Cognitive Psychology* 18:87–121.

Miller, G. A., and Johnson-Laird, P. N. 1976. *Language and Perception*. Cambridge: Cambridge University Press.

Moore, G. T. 1979. Knowing about environmental knowing. *Environment and Behaviour* 11:33–70.

Munsell, A.; Sloan, L.; and Godlove, I. 1933. Neutral value scales. *J.Opt. Soc. Am.* 23:394–411.

Normen, D. 1980. *DIN 6164. Teil 1: Din Farbenkarte. System der DIN Farbenkarte f"ur den 2° Normalbeobachter*. Berlin, Germany: Beuth-Verlag.

Pinker, S. 1985. Visual cognition: An introduction. In Pinker, S., ed., *Visual Cognition*. Cambridge, MA: MIT Press. 1–63.

Retz-Schmidt, G. 1988. Various views on spatial prepositions. *AI Magazine* 9(2):95–105.

Rosch, E. 1975. Cognitive reference points. *Cognitive Psychology* 7:532–547.

Sadalla, E.; Burroughs, W. J.; and Staplin, L. 1980. Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory* 6(5):516–528.

Schober, M. F. 1995. Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about location? *Discourse Processes* 20:219–247.

Siegel, A. W., and White, S. H. 1975. The development of spatial representation of large-scale environments. In Reese, W., ed., *Advances in Child Developement and Behaviour*. New York: Academic Press.

Stevens, A., and Coupe, P. 1978. Distortions in judged spatial relations. *Cognitive Psychology* 10:422–437.

Stuart, G. W.; Bossomaier, T. R.; and Johnson, S. 1993. Preattentive processing of object size: implications for theories of size perception. *Perception* 22:1176–1193.

Talmy, L. 1983. How language structures space. In Pick, H., and Acredolo, L., eds., *Spatial Orientation: Theory, Research and Application*. New York, London: Plenum. 225–282.

Taylor, H., and Tversky, B. 1992a. Descriptions and depictions of environments. *Memory & Cognition* 20(5):483–496.

Taylor, H. J., and Tversky, B. 1992b. Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language* 31:261–292.

Treisman, A. 1988. Feature and objects: The fourteenth bertlett memorial lecture. *The Quarterly Journal of Experimental Psychology* 40a(2):201–237.

Tulving, E. 1962. Subjective organization in free recall of "unrelated" words. *Psychological Review* 69:344–354.

Tversky, B. 1981. Distortions in cognitive maps. *Cognitive Psychology* 13:407–433.

Tversky, B. 1993. Cognitive maps, cognitive collages, and spatial mental models. In Frank, A. U., and Campari, I., eds., *Proc. of the 1 st int. Conference on Spatial Information Theory*, 14–24. Elba, Italy: Lecture Notes in Computer Science 716.

Wazinski, P. 1992. Generating spatial descriptions for cross-modal references. In *3rd Conference on Applied Natural Language Processing, Trento, Italy*, 56–63. Association for Computer Linguistics.