Robot Navigation and Manipulation Control Based-on Fuzzy Spatial Relation Analysis

Jiacheng Tan, Zhaojie Ju, Steve Hand, and Honghai Liu

Abstract

This paper addresses the problem of control and commanding robots in their work spaces. The problem is approached by deriving the image-space and object-space spatial relations. The relations are represented as fuzzy sets to capture the ambiguity inherent to the linguistic terms for the relations. Distances and object sizes and shades have been explored as the fuzzy qualifiers in conditioning the spatial relation reasoning. By using a simple syntax, simple natural language sentences are mapped to fuzzy queries for robot commanding. Implementation and experiment have verified the feasibility and demonstrated potentials of the approach.

Keywords: Robotics, fuzzy sets, spatial relations, image processing.

1. Introduction

Given the advances in computer science and artificial intelligence in the past few decades, it is still a huge challenge for robots or autonomous devices to survive unknown tasks in unknown environments. The problem is due to the inability of current robots in perceiving and understanding their environments and in identifying and organising their tasks. The situation is unlikely to improve greatly in the foreseeable future. For this reason, human operators are still indispensable in the control loops of critical robotic applications such as in space explorations, underwater servicing, landmine disposal, guided-missile targeting, and so on. In such missions, human operators, partially or completely, take over from the robots the responsibility of environment assessment and task planning, and the robots act on receiving the instructions from the operators. For this to happen seamlessly, communicating the intentions of the operators to the robots is a pivotal issue. In this paper, we discuss a fuzzy spatial relation-based technique for robot commanding. We attempt to find an efficient and natural way for human operators to interact with a mobile robot or a robot manipulator to guide them in task operations.

Commanding a robot, in particular a robot manipulator, to perform a task operation has long been an issue in robot control. The problem can be readily solved in industrial applications where the environment and the tasks can be completely modelled. The issue remains when a robot has to execute unknown tasks or negotiate unknown environments. To improve the performance of the robots in such applications, both command line-based operation and virtual environment-based manipulation have been studied [1, 2]. However, direct manipulation still prevails in practice. In this paper, this problem is approached by fuzzy spatial relation analysis. By mapping the workspace of a robot into fuzzy spatial relations and by borrowing the power of fuzzy reasoning, we are able to construct robot command in natural language or in a form close to it, which could ultimately make the voice-based robot operation control viable.

2. Related Work

Spatial reasoning is considered as the domain of spatial knowledge representation, in particular spatial relations between spatial entities and of reasoning on these entities and relations [3]. Humans effortlessly use such knowledge in everyday life. Spatial reasoning can be used to solve sophisticated space-related problems and plays a very important role in many areas of science and engineering, for example geographic information system (GIS) [4], image understanding and content-base retrieval [5], and robot navigation [6]. Rendell et al. [7] have proposed the region-connection calculus (RCC) for modeling the topological spatial relations and reasoning about the spaces and relations. As a generalization of the RCC, a 9-intersection model [8] was proposed to treat the set of the topological spaces and their relations in general. However, in real-world applications, the uses of the RCC theory have been hindered by the theory’s inability to cope with the uncertainty and imprecision of real world information. To make the RCC able to accommodate the imprecision, Schockaert et al. [9] presented a fuzzy version of RCC that interpreted spatial regions as fuzzy sets and connections as fuzzy relations.

As an efficient tool for modeling the vagueness in the linguistic definitions of regions and relations, the fuzzy
set theory has been well studied [10, 11, 12]. It is evident that the fuzzy representation of spatial regions and relations has provided an adequate framework for spatial knowledge representation and reasoning: it captures the imprecision inherent to the linguistic expressions of spatial regions and relations; it reduces the semantic gap between symbolic concepts and numerical information [13]. More theories and applications can be found in [14, 15] for GIS application, in [16, 17] for robotic applications, and in [18, 19, 20] for image processing and image interpretation.

3. Spatial Relations and Their Fuzzy Representations

The aim of spatial relation analysis is to derive the geometric and/or topological relations among a collection of objects. In this paper, we wish to explore such relations to establish an inference system in which the location of a single object can be uniquely defined and deduced, and subsequently the deduced location is used to instruct a robot or its end-effector to the location to accomplish navigation or manipulation tasks. In essence, by exploring the spatial relations we wish to establish a coordinate system equivalent to the Cartesian one but without recourse to the coordinates of real numbers (as far as specifying a location is concerned). The differences between the two are in that, one defines only a finite number of locations or regions by linguistic coordinates and a certain amount of impression is allowed, and the other defines an infinite number of locations by real numbers without any ambiguity.

As the system is regarded as a type of coordinate systems, we first need to consider the space it will address and the reference or “origin” with respect to which spatial relations can be defined. In this paper, we consider the issue of using images as the visual guide for robot control. The workspace of the robot is considered as 2D scenes populated with objects. The depths of the objects, which can be estimated once their locations are established, are not considered in the discussion. We are going to use two reference systems, image space-based and object-based, and with each of them a set of spatial relations will be defined. It will become evident that the image space-based relations are necessary to guarantee any single location in the image is accessible and that the object-based relations, while not possessing the same attribute, can greatly improve the practicality of the reasoning system.

In addition to serving as a coordinate system, the spatial relations will be used to interface human operators with the robot, ideally, via natural language. As such, the spatial relations must be so defined that they are consistent with human’s conceptions of these relations and can tolerate the imprecision of the descriptions of these relations in natural language. This is where fuzzy theory comes into play.

A. Image-Space Spatial Relations

Given a space, a 2D and rectangular one in this case, it can be divided in an infinite number of ways. In defining the spatial regions and relations, we have paid attention to the following points:

- Geometrically, the division must be complete, by which we mean that no holes or gaps are left by the division;
- Conceptually, the division must be consistent with human conceptions of the spatial regions and relations;
- Algorithmically, the division scheme should allow the division be recursively applicable to the sub-regions. This is to ensure that every single location in the space is accessible.

These considerations naturally lead to us to divide the entire image into nine regions: a centre, a right, a left, a top, a bottom and four corners, as shown in Figure 1 (a). We name them purely for convenience of discussion.

With respect to the region centre, we define four primitive spatial relations: top, bottom, left, and right. We use them to describe the spatial relation of a point with respect the reference region, centre. So by relation right, we actually mean a point is on the right of the centre region. We avoid using the term “right of”, because we reserve it for naming an equivalent object-space relation. The relations have a second meaning: we use them to represent the set of all points on them a particular spatial relation holds. Therefore, there is a significant difference between the region right and the relation right. As we will see shortly, the relation right represents the whole region of the right hand side of the image. In the following discussions, when a relation is mentioned it normally refers to its second meaning.

![Figure 1. (a) An image is divided into nine regions: a centre, a right, a left, a top, a bottom and four corners; (b) The fuzzy membership of the relation right, where the total darkness represents 1.0 and white represents 0.0.](image)

Now, we consider the geometric meaning of these relations. In natural language, the meanings of these relations are rather vague. For example, a point anywhere within the shaded area of Figure 1(a) could be consid-
erred, to a more or less degree, to have a right relation with the centre region. This relation is certainly valid for all the points within the area with darker shade—the region we have named as right. However, for the points within the two corners the validity of the relation varies. If a point is very close to the top or bottom edges of the region right, the relation is almost certain. If a point is very close to the right edges of the top or bottom regions, the relation is almost invalid. Obviously, the distance between the point and the corresponding corner point of the centre region has no influence over the validity of the relation.

We model this variation in the validity of the relation as a simple linear function in \( \theta \), the angle between vertical line and the line joining the point \( P \) and the corner point of the centre region, as shown in Figure 1(a). Being such modeled, the relation right refers to the set of points that maintain a right relation with the region centre—the entire shaded region in figure 1(a). The points within this region can be adequately represented by a fuzzy set, with each point being assigned a fuzzy membership to reflect the validity of the relation at that point. Figure 1(b) show the fuzzy membership of each point of the region. So, the relations now have a third meaning: fuzzy sets.

The same argument applies to the relations top, bottom and left, and their membership can be similarly assigned. Figure 2 shows the membership functions for the four corners. With the relations being defined as fuzzy sets, the spatial relations of the four corners with respect to the centre region can be treated as the derived relations. For example, the relation of the top-right corner with respect to the centre region can be considered as the conjunction of the fuzzy sets top and right.

We now consider the region centre. Being used as the reference to define the primitive spatial relations, the region centre itself is not a relation. However, the space described as centre in natural language tends to be very small and is accompanied by imprecision. Within this region, the degree of a point being considered to belong in centre varies according to its distance from the centroid of the region. This characteristic of the region also calls for a fuzzy representation. In fact, we can view centre as a region consisting of the set of points that hold a geometric relation to the centroid of the region.

![Figure 2. The fuzzy membership functions for relations right, top, left and bottom.](image)

In this sense, centre is also a relation, but a special one because it has its own reference. The fuzzy definition of this relation is straightforward. It consists of the set of points each of which has its distance to the centroid of the region centre as its membership:

\[
\mu = 1 - \sqrt{2 \left( \frac{x-x_c}{w} \right)^2 + \left( \frac{y-y_c}{h} \right)^2 },
\]

with \(-w/2 \leq x-x_c \leq w/2\) and \(-h/2 \leq y-y_c \leq h/2\), where \((x_c, y_c)\) is the geometric centre, and \(w\) and \(h\) are the width and height of the region. This definition takes into account the aspect ratio of the region. Figure 3(a) shows the memberships of the points. Being such defined, relation centre becomes another primitive relation in the image space.

Obviously, we can define composite relations in terms of the primitive relations. The composite relations are necessary and useful: they guarantee that we can specify any position in the image space by applying the primitive relations recursively; they are necessary to express complex spatial relations. For example, the natural language description of the spatial relation of the shaded object in Figure 3(b) would look like “the object on the left at the bottom”. This description can be simply expressed as a composite relation left of bottom.

The image-space relations discussed in this section normally serve as the starting point of spatial relation reasoning, in particular, where a suitable reference object could not be found to determine the object-space relations.

B. Object-Space Spatial Relations

The object-space spatial relations have been well studied and found uses in many applications. The primitive relations, such as left of, right of, above, below and a few more have been identified and widely used for describing 2D spatial relations [21]. However, the actual definitions of these relations vary in practice. In this section, we discuss the fuzzy definitions of the object-space primitive relations and of the relations we have identified as being important for spatial reasoning.

In defining these primitive relations, one of the most important factors is to find an appropriate representation for objects of all possible shapes. A good representation should lead to unambiguous spatial relations and efficient evaluation of the fuzzy membership functions.
Therefore, the relation described by the four primitive relations already defined. Close in space and the relations among them cannot be described by the four primitive relations already defined. Therefore, the relation next to has a very limited scope; its uses are restricted to at the adjacency of the reference object where the relations left of, right of, above, or below fails. As illustrated in Figure 4(b), with the bounding box representation, the triangular object would be considered as closer to object A than it is to object B. Obviously, the implementation of the relation next to calls for an actual shape-based distance computation.

Semantically, the relation next to is very similar to the close to relation reported by other authors [13, 21], but with next to having a much narrower scope, which demands the actual object shapes being used in the evaluation of the membership functions. In contrast, the concept of “close to” can be more efficiently implemented with the bounding box representation, as will be discussed in the next section.

C. Some Fuzzy Qualifiers

So far, we have discussed a few spatial relations and expressed them as fuzzy sets. If we use these relations to specify or to query the location of an object, the results can hardly be unique, unless the scene is sparsely populated. With a cluttered scene, it is more likely that a collection of objects will fall under the same description or a collection of objects will be returned. As each object has a different membership with respect to different relations, one might attempt to use the numeric memberships as a kind of coordinate to differentiate the objects within the collection. However, doing so would defeat our purpose of devising these relations in the first place. To address this problem, we are going to introduce a few qualifiers.

Humans use object features to identify single objects when the spatial relations become ambiguous. Among these features are object shapes, textures, colors, sizes, and so on. Shapes are perhaps the most powerful and the most used feature in human object recognition. Unfortunately, mapping a general 2D shape to a linguistic symbol is technically difficult. Therefore, in this section, we will discuss some qualifiers that can be readily detected and easily described by linguistic symbols. They are: distance qualifiers, size qualifiers, and shade qualifiers.

Distance computation can be efficiently done if two objects are sufficiently separated. Because they are sufficiently separated, the influence of their actual shapes over the distance computation becomes less influential; therefore the distance can be measured against their centroids. Of course, the computed distance between the centroids must be scaled afterwards in accordance with the actual shapes of the objects or their bounding boxes. To save the computing costs, we choose to scale the distance by the bounding boxes of the objects. It can be shown, in most cases, this choice is harmless.
We wish to classify the distances into five linguistic qualifiers: nearest, close, middle, far and farthest. The fact that these qualifiers are defined over the spatial order rather than the actual distances between the objects makes classifying the distances less useful. Therefore, instead of classifying the distance, we first sort the objects according to their distances to the reference object and map their orders in the sorted list into the interval of [0.0, 1.0], then the classification is done on this interval. Of the distance qualifiers, we want nearest and farthest to be crisp ones; because we always want a single object to be returned when using them. We always assign them the first and the last objects in the sorted list, or equivalently, the objects with memberships 1.0 and 0.0, respectively. For the three remaining fuzzy qualifiers, triangular membership functions are assigned to them, as shown in Figure 5.

We have treated the size qualifiers in the same way. We choose to use area as the measure for object size. The area of an object is calculated by counting the number of pixels of the corresponding image region. We also have five qualifiers for sizes: largest, large, medium, small and smallest. We could have added two more qualifiers to the set, for example, very large and very small, but it has proved to be difficult for humans to perceive and classify the minute differences in sizes. Again, we want the qualifiers largest and smallest to be crisp ones. Because area has similar characteristics to those of distance, the same membership functions are assigned to the size qualifiers.

Different from distance and area, the shade of an object refers to the fixed intensity values, for example, a shade of gray referring to an intensity value around 128. So the membership functions must be defined over the range of image intensity [0, 255] or its normalized equivalence [0.0, 1.0]. We noticed the fact that the visual perception of a shade is susceptible to the spatial configurations and the shades of surrounding objects [22]. However, without a quantitative analysis of the effect, it is difficult to account for this in this work.

Three qualifiers are defined over the range of the intensity: dark, gray and light gray. In addition, as in the cases of the distance and area qualifiers, we also have two relative qualifiers darkest and brightest. The mean of pixel intensities is used to measure the shade of an object. Considering that human’s visual discerning ability of shades is relatively low at the bright end and that bright objects are relatively rare in real work environment, we have defined the membership functions of the shade qualifiers as shown in Figure 6. The membership functions are effective when many dark objects are involved.

In this section, we have defined three sets of qualifiers, both crisp and fuzzy. Combined with the spatial relations, they form a set of tools for querying a particular object in images.

D. Querying Into Images

With the relations and the qualifiers in place, we can now address the problem of querying the locations of objects. In this section, we discuss how the spatial relations and the qualifiers can be used to form fuzzy queries for object locations.

While modeling the spatial relations, our motivations have been to bridge the gap between the linguistic symbols and the numeric values. We wish to use the linguistic symbols, rather than the numeric values, to express or to retrieve the locations of objects. Naturally, we would like to use the symbols in a natural way, i.e., embedding them in queries expressed in natural language.

Without indulging ourselves with natural language analysis, we have derived a simple syntax for query construction. The syntax takes the following form: <command>, <qualifiers>, <relation_1>, <qualifiers>, <relation_2>.

In robotic applications, the variable command could be move to, pick up, put down, and so on. In this paper, we use move to to illustrate the query structures.

The variable relation_1 takes zero or one argument. If the argument is non-empty, it must be an object-space relation along with the object it refers to. This implies that a relation involving more than one reference objects is not supported, as in the case of relation “between”, although there is no technical obstacles in implementing such relations or queries.

The variable qualifiers can take from zero to three arguments if the fuzzy qualifiers are used, one from each qualifier set, as in “medium and dark object”. But only one argument is permitted when a crisp qualifier is taken. Using two or more crisp qualifiers may lead to conflict in object attributes. For example, the darkest object is not necessarily the largest one.
The object detection unit processes the input image and detects objects. The outputs, the parameters of the objects, are passed on to the spatial relation evaluation unit where they are processed according to the requests received from the query processing unit.

The variable relation_2 can take zero or more image-space relations. Normally, it takes no more than two arguments, as in expressions like “on the left”, “at bottom right”, and so on, because composite relations involving more than two image-space relations are very hard to comprehend. If variables relation_1 and relation_2 take zero argument at the same time, the entire image will be used as the domains of the qualifiers.

Table 1 gives a few examples of the queries constructed by following the syntax. Expressed as imperative phrases, the queries can be easily parsed into composite relations in terms of the functions of the primitive relations. In complex cases, we may have to use the logic conjunctions of the simple queries. We will use the product \( T \)-norm to compute the membership of the conjunctions.

\[
T(\mu_1, \mu_2) = \mu_1 \cdot \mu_2
\]  

### 4. Implementation

To verify the relations we have to evaluate their uses in practical applications, a system that implements all the relations, qualifiers and other necessary components has been developed. This section gives the details of the system.

#### A. System Architecture

The system consists of, logically, three components: object detection, spatial relation evaluation, and query processing.

The object detection unit processes the input image and detects objects. The outputs, the parameters of the objects, are passed on to the spatial relation evaluation unit where they are processed according to the requests received from the query processing unit.

The output of the spatial relation evaluation could be one or more object locations. It could also be a position that the user wants to specify, as in the case of commanding a robot to go to a designated position. Upon receiving the output from the spatial relation evaluation unit, the object detection unit will highlight the object or position and perform a reprocessing of a local image region.

#### B. Object Detection

Object detection is a difficult task. Much of the reported work on spatial relation analysis has been conducted on images that have already been segmented, usually manually [5, 13]. Such a scheme can hardly work in robotic applications where object detection must be done on live images or video frames in real time.

In this paper, we integrate the object detection process with the spatial relation analysis. To simplify the problem, we have made no attempt at recognizing the objects or identifying their actual 3D parameters. The effects of perspective projection have also been ignored. We have not aimed at finding out every object in a scene in one attempt. In fact, it is extremely difficult to find a set of processing parameters to segment all the objects in an image when the contrast between the background and the foreground objects is small. Our aim of object detection is to find some anchoring objects in the scene with which spatial reasoning can start. Once started, if we wish, we can perform the detection process locally to bring out objects that have been missed in the previous processing passes.

In object detection, the image area corresponding to an object has been chosen as the critical feature for sifting out the candidate objects from the irrelevant items. The choice reflects the fact that a robot will only act upon an object if it is larger than a certain size and that small items are normally produced by background noises. It is worth pointing out that, measured in the number of pixels, the area of an object tends to be much smaller than its actual size.

The detection process consists of a few steps. The input image is first smoothed with a Gaussian filter to remove the irrelevant minute details. The optimal parameters of the filter depend on the content and quality of the input image. It is difficult to decide a set of parameters that will perform equally well in different regions of the image. The filtered image is then converted to a binary image by intensity thresholding. The value of the threshold is critical. A threshold can be found by analyzing the intensity histogram of the image, but to achieve the optimal value more sophisticated treatment is needed, which is beyond the scope of this paper. The binary image is then processed with morphological operations to remove speckles, sharp spikes and small holes. The left...
structures are considered as the potential scene objects. A labelling algorithm is then applied to give each object its own identity. The candidates are then filtered by their sizes to give the final scene objects.

Several geometric properties, such as the centroids, bounding boxes and convex hulls, of the corresponding image regions of the final scene objects are computed. As discussed in Section 3, the centroids and the bounding boxes are used for evaluation of the membership functions; the convex hulls are used for calculating the sizes and the intensity means of the objects. Other geometric properties such as orientations and extreme points can also be computed, but we have not explored their uses in this paper.

C. Spatial Relation Evaluation and Query Processing

The spatial relations and the qualifiers have been implemented as functions. The functions of the image-space relations and the size and shade qualifiers take an array of object indices as their inputs. The functions of the object-space relations and the distance qualifiers take one more input – the index of the reference object. The output of these functions is an array of indices of the objects along with their fuzzy memberships. The image-space relations also return the sub-regions that they define. The design ensures that the output of one function can be used as the input of another, which makes it easy to evaluate composite relations.

When a query for a location is received from the user, it is parsed into an ordered list of symbols of function names according to the query syntax. The symbols are then assembled into concatenated function calls. The system acts differently, depending upon the outputs of the function calls. If only one object is returned, the object is assumed to be the target object and marked as the current object in the image. The current object will be used as the reference object in the subsequent queries.

If the query is not about an object but about a space satisfying certain spatial constraints, then the space is marked. If multiple objects are returned, the object with the largest membership in the very last function call is chosen as the current object. If logic conjunctions are involved, the product T-norm will be evaluated.

5. Experiment and Evaluation

In testing the system, we have used a grayscale image of a scene consisting of circuit boards, a power adaptor and other items, as shown in Figure 8.

After applying the processing procedure outlined in Section 4, we obtain the scene objects. For convenience of discussion, these objects are labeled with a number, as shown in Figure 9. It can be seen that, even given the high quality of the original image, misdetection still happens (Object 13). The objects with varying shades have lost their original shapes (Object 2, 3, 4 and 8). Given the defects, the detected image regions still provide reasonable indications about the locations of the objects.

With the image, we have experimented object retrieval using various queries. Note that the object labels are purely for convenience of discussion and have not been used in the object retrieval experiment.

A. Case 1
Suppose we want to retrieve the position of Object 11 to instruct a mobile robot or the end-effector of a robot manipulator to move to that position or to manipulate that object. This object has no distinguishing characteristic except for its shape. However, for the reasons we discussed previously, no shape-related qualifiers have been developed. We also do not have an object to refer to at this stage. However, by observation we know that the object is next to the largest object on the right hand side of the image (Object 12). Indeed, running the query “move to the object next to the largest on the right” retrieves its position, as shown in Figure 8.

B. Case 2
Suppose we want to retrieve Object 7 from the current
position (Object 11). This time, neither the object itself nor its neighbors have a distinguishing characteristic. Careful observation reveals that, comparing with the surrounding objects, Object 7 and 9 have a lighter shade. It appears that the query “move to the smallest dark gray object below the current object” would work. However, it does not. The query causes the execution of the following composite relation:

$$\text{smallest}(\text{dark_gray}(\text{below}(\text{current_object})))$$.

Because “dark gray” is a fuzzy qualifier, when the function returns, we will have a list of objects along with their membership of being “dark gray.” Table 2 compares the means of the actual pixel intensities and the memberships returned by the functions gray() and dark_gray(). In neither case, has Object 7 or 9 dominated the table. The result is partly due to the membership functions we have used (Figure 6), and partly due to the fact that the perception of shades is a very subjective matter.

Table 2. Comparison of gray() and dark_gray().

<table>
<thead>
<tr>
<th>Object No.</th>
<th>Mean of Intensity</th>
<th>Output by dark_gray()</th>
<th>Output by gray()</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.9182</td>
<td>0.9182</td>
<td>--</td>
</tr>
<tr>
<td>1</td>
<td>0.9038</td>
<td>0.9038</td>
<td>--</td>
</tr>
<tr>
<td>12</td>
<td>0.8880</td>
<td>0.8795</td>
<td>0.1205</td>
</tr>
<tr>
<td>6</td>
<td>0.8814</td>
<td>0.8142</td>
<td>0.1858</td>
</tr>
<tr>
<td>7</td>
<td>0.8582</td>
<td>0.5821</td>
<td>0.4179</td>
</tr>
<tr>
<td>9</td>
<td>0.8458</td>
<td>0.4576</td>
<td>0.5424</td>
</tr>
<tr>
<td>2</td>
<td>0.8243</td>
<td>0.2428</td>
<td>0.7572</td>
</tr>
<tr>
<td>4</td>
<td>0.7793</td>
<td>--</td>
<td>0.7926</td>
</tr>
</tbody>
</table>

In fact, this object can be retrieved by “move to the object next to the object below of AND left of object 11”. Table 3 gives the objects retrieved by functions below_of() and left_of() and the product T-norm of their conjunction. Another way, perhaps the most efficient way, of retrieving the object is using a composite image-space relation “move to the centre of the bottom”.

Table 3. below_of() and left_of() and their product T-norm.

<table>
<thead>
<tr>
<th>Object No.</th>
<th>below_of()</th>
<th>left_of()</th>
<th>Product T-norm</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.8551</td>
<td>0.1454</td>
<td>0.1243</td>
</tr>
<tr>
<td>9</td>
<td>0.8368</td>
<td>0.1637</td>
<td>0.1370</td>
</tr>
<tr>
<td>12</td>
<td>0.7387</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>7</td>
<td>0.5907</td>
<td>0.4098</td>
<td>0.2421</td>
</tr>
<tr>
<td>6</td>
<td>0.4554</td>
<td>0.5451</td>
<td>0.2482</td>
</tr>
<tr>
<td>1</td>
<td>0.2027</td>
<td>0.7978</td>
<td>0.1617</td>
</tr>
<tr>
<td>4</td>
<td>0.1295</td>
<td>0.8710</td>
<td>0.1128</td>
</tr>
<tr>
<td>2</td>
<td>0.0340</td>
<td>0.9665</td>
<td>0.0329</td>
</tr>
<tr>
<td>8</td>
<td>--</td>
<td>1.0000</td>
<td>--</td>
</tr>
<tr>
<td>3</td>
<td>--</td>
<td>0.9756</td>
<td>--</td>
</tr>
<tr>
<td>5</td>
<td>--</td>
<td>0.2244</td>
<td>--</td>
</tr>
</tbody>
</table>

C. Evaluation and Observations

We cannot exhaust the ways of retrieving objects in a scene even as simple as the one we have used. However, by experimenting with different queries on different objects, we have arrived at some conclusions. The spatial relations and the qualifiers have provided a set of convenient tools for locating objects in a scene. They are flexible and powerful and often provide several alternative ways to access the same object. We have noticed that the image-space relations are indispensable. In many cases, for example when objects with similar sizes and shades are evenly distributed across the image, they provide the only access to an object. We have also noticed that, in querying an object, accurate user observation of the spatial relations is important for efficient retrieval. The shades of objects can be an efficient constraint when strong contrasts exist among the objects, but they become less reliable when the range of the contrasts is small.

6. Future Work

In this paper, we have addressed the problem of using spatial relations to command robots in their work spaces. The results of implementation and experiment have verified the feasibility and demonstrated the potential of the approach. The work can be improved in several aspects. The first is on the design of the fuzzy membership functions of the spatial relations and the qualifiers, for example, by training the membership functions of the shade qualifiers with some sample data. Secondly, more complex relations such as “between” and “group” and more object features such as colors and orientations could be integrated into the framework. A useful extension of the work would be to investigate the problem in 3D space where the issues of perspective shortening and object occlusion dominate.

References


JiaCheng Tan received BEng in mechanical engineering in 1983, Jilin University of Technology, Changchun, China, MSc in mechatronics in 1989, Xidian University, Xi’an, China, and Ph.D. in computer graphics in 2001, De Montfort University, U.K. He had been a visiting researcher in robotics in the Department of Electronic Engineering, University of Salford, Salford, U.K., 1996-1997, and a research fellow in artificial intelligence at the Open University, U.K. He had been a visiting researcher in robotics in the Department of Electronic Engineering, University of Salford, Salford, U.K., 1996-1997, and a research fellow in artificial intelligence at the Open University, U.K., 2000-2002. He joined the School of Computing, University of Portsmouth, U.K. in 2002. His research interests are in the areas of computer vision, image-base modeling, computer graphics and robotics.

Zhaojie Ju received the B.S. in automatic control and the M.S. in intelligent robotics both from Huazhong University of Science and Technology, China, in 2005 and 2007 respectively, and the Ph.D. degree in intelligent robotics at the University of Portsmouth, U.K., in 2010. Dr Ju is currently a postdoctoral research associate in the intelligent systems and biomedical robotics group at the University of Portsmouth. He has published over twenty refereed international journal and conference papers including one Best Paper Award and one Best Competition Paper Award. His current research interests include computational intelligence, pattern recognition, human-computer interaction and cognitive robotics.

Steve Hand was a researcher in applied entomology at Southampton University before becoming a lecturer at the University of Portsmouth specialized in expert systems and object orientated techniques. He subsequently became the head of the School of Computing and Mathematics, and is currently the head of the School of Creative Technologies.
Honghai Liu received the PhD degree in intelligent robotics from Kings College, University of London, London, UK, in 2003. He is currently a Professor of intelligent systems with the University of Portsmouth, Portsmouth, UK. He has previously held research appointments with the University of London and the University of Aberdeen, Aberdeen, UK, and appointed as the project leader in the large-scale industrial control and system integration. His research interests include approximate computation, pattern recognition, intelligent video analytics, cognitive robotics, and their applications, with an emphasis on the approaches of intelligent connection of perception with actions using contextual information. Dr. Liu is a Fellow of the Institution of Engineering and Technology.