

From Regions to Transitions, From Transitions to Objects

Paulo Santos and Murray Shanahan

{p.santos, m.shanahan}@ic.ac.uk

Dep. of Electrical and Electronic Engineering,
Imperial College, London, SW7 2BT

Abstract

This paper describes the construction of a qualitative spatial reasoning system based on the sensor data of a mobile robot. The spatial knowledge of the robot is formalised in three sets of axioms. First of all, axioms for relations between pairs of spatial regions are presented. Assuming the distance between regions as a primitive function in the language, the main purpose of this initial axiom set is the classification of relations between images of objects (from the robot's vision system) according to their degree of displacement. Changes in the sensor data, due to the movement either of objects in the robot's environment or of the robot itself, are represented by transitions between the displacement relations. These transitions are formalised by the second set of axioms. The predicates defining the transitions between image relations are connected to possible interpretations for the sensor data in terms of object-observer relations, this issue is handled by the third set of axioms. These three axiom sets constitute three layers of logic-based image interpretation via abduction on transitions in the sensor data.

Introduction

Classical research in robotics concerns low-level tasks (e.g. sensory processing, manipulator design and control) leaving aside questions about high-level information processing such as reasoning about space, time, actions and states of other agents (Cox & Wilfong 1990)(Latombe 1991). Such issues have been addressed by the knowledge representation sub-field of Artificial Intelligence (Shanahan 1997)(Reiter 2002). Most knowledge representation (KR) theories, however, have been developed in isolation from empirical issues such as how the knowledge about the world is acquired and what are the physical mechanisms by which it is embodied in the agents.

The present paper describes a logic-based representation of the spatial knowledge constructed from the sensor data of a mobile robot. One of the main purposes of this theory is to bridge the gap between KR theories and practical robotics, equipping the robot with the basic machinery for deriving and manipulating information about physical objects (including the robot itself).

Briefly, this work proposes that incoming robot sensor data can be explained by hypothesising the existence and the dynamic relationships between physical objects, with respect to a (possibly moving) observer. This process recalls the sensor data assimilation as abduction proposed in (Shanahan 1996). This framework handles the issue of spatial occupancy, however, it does not take into account the relationship between the observer's viewpoint and the spatial objects. The main motivation for the present paper is to propose a spatial representation framework capable of coping with this issue. The notions of space and time (together) play a central role in the representational system described in this paper, an issue that has been generally overlooked by the knowledge representation community, as pointed out in (Vieu 1997).

Considering qualitative theories of space, particularly relevant to this work are the Region Connection Calculus (*RCC*) (Randell, Cui, & Cohn 1992)(Cohn *et al.* 1997) and the Region Occlusion Calculus (*ROC*) (Randell, Witkowski, & Shanahan 2001). From *RCC* this paper inherits the use of regions and connectivity relations in the construction of the spatial ontology, on the other hand, the way we deal with observer's viewpoint recalls *ROC*. The present framework, however, can be used to extend both *RCC* and *ROC* in the sense that it assumes sensor information in the foundations of the knowledge representation formalism.

Briefly, Region Connection Calculus is a many-sorted first-order axiomatisation of spatial relations based on a dyadic primitive relation of *connectivity* ($C/2$) between two regions. Assuming two regions x and y , the relation $C(x, y)$, read as " x is connected with y ", is true if and only if the closures of x and y have at least a point in common.

Assuming the $C/2$ relation, and that x , y and z are variables for spatial regions, some mereotopological dyadic relations can be defined on regions. They are, $P(x, y)$ (x is part of y), $O(x, y)$ (x overlaps y), $DR(x, y)$ (x is discrete from y), $PP(x, y)$ (x is a proper part of y), $Pi/2$ and $PPi/2$ (the inverses of $P/2$ and $PP/2$ respectively), $DC(x, y)$ (x is disconnected from y), $EQ(x, y)$ (x is equal to y), $PO(x, y)$ (x partially overlaps y), $EC(x, y)$ (x is externally connected with y), $TPP(x, y)$ (x is a tangential proper part of y), $NTPP(x, y)$ (x is a non-tangential proper part of y), and $TPPi/2$ and $NTPPi/2$ (the inverse relations of $TPP/2$ and $NTPP/2$ respectively).

Extending RCC, the Region Occlusion Calculus was designed to model spatial occlusion of arbitrary shaped objects, thus relating images of physical objects to the observer's viewpoint. The formal theory is defined on the degree of connectivity between regions representing two-dimensional images of bodies as seen from some viewpoint. In this paper we present a dynamic characterization of occlusion taking into account information from a mobile robot's sensors.

The present work also builds on work reported in (Santos & Shanahan 2001), where an ontology based on depth, size and distance information is proposed for interpreting sensor data in order to address the problem of anchoring symbols to sensor data (Coradeschi & Saffiotti 2000). The qualitative spatial theory proposed in this paper, however, is based exclusively on information about the *displacement* of image relations. The new theory demonstrates how much can be achieved using distance information alone.

System Scheme

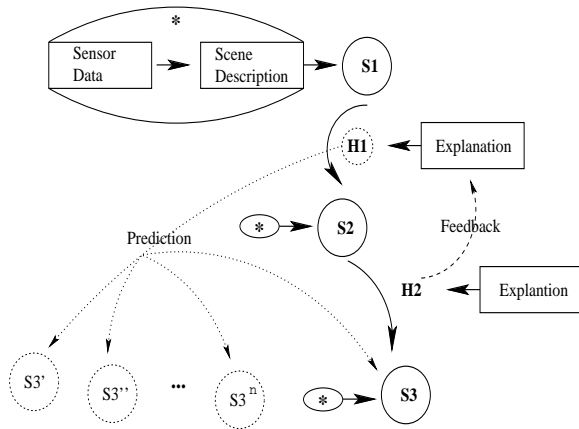


Figure 1: Preliminary scheme of the system.

Figure 1 illustrates the overall architecture of the system discussed in this paper. The first task in this system is the translation of snapshots of the world provided by the robot sensors into logic formulae, this process is represented by the scene description module in the diagram above. The scene descriptions are represented in the figure above by the labelled circles, $S1$, $S2$ and $S3$. Changes between two subsequent descriptions are explained by assuming hypotheses about the dynamic relations between an observer and the objects in the world. Examples of such hypotheses in the above diagram are $H1$ and $H2$ (respectively, the hypotheses explaining the transition from $S1$ to $S2$ and from $S2$ to $S3$). Both processes, scene description and explanation of changes, are handled in this work by abduction. In the former, abduction is used to assimilate static scenes into a logic theory about spatial regions, the next section presents this theory. In the second case, abduction is used to explain changes in the scene descriptions according to a theory about dynamic spatial predicates representing object-observer relations.

From a hypothesis, further sensor descriptions can be predicted. The prediction process is illustrated by the dotted lines in Figure 1, the dotted labelled circles represent the predicted scene descriptions. When a prediction is corroborated (i.e. when a scene description matches to a predicted scene) the new explanation hypothesis is used to modulate the confidence in the hypotheses about earlier changes. This feedback process is depicted by the dashed line in the figure above. Prediction in this work is assumed as a deductive process.

A brief discussion about how prediction and explanation interplay in the present framework is presented in a later section of this article.

For brevity, this paper assumes that all variables are universally quantified, unless explicitly stated.

A Spatial Logic Based on Regions

This section presents a many-sorted first-order axiomatisation of spatial relations assuming, initially, sorts for spatial regions and real numbers. Similarly to RCC (briefly introduced in the previous section), the axiomatic system presented below has spatial regions and the connectivity between them as fundamental concepts. However, this paper assumes the *distance* between pairs of regions as a primitive function with which the degree of connectivity is defined. Therefore, the relations between spatial regions are defined according to the degree of *displacement* (rather than connectivity) between them.

The reason for assuming distance as a primitive function for defining region relations resides in the fact that the relative distance between objects in a robot's environment (and between pairs of regions in images) is one of the features that can be extracted from the robot sensor data.

The concept of distance in this work should be understood as a qualitative notion of displacement, i.e., we are not interested in an accurate measure, but on how the distance between pairs of regions changes in time. Defining qualitative notions of distance, however, is not a straightforward task since the common sense concept of distance is context dependent (Lowe & Moryadas 1975). An initial work on qualitative notions of distance for artificial intelligence is presented in (Hernández, Clementini, & di Felice 1995).

For the purposes of this paper, however, we assume a distance function on pairs of spatial regions. This function can be intuitively understood as *the size of the shortest line connecting any two points in the two region boundaries*. In this work, assuming spatial regions x and y , the distance between x and y is represented by the function $dist(x, y)$, read as 'the distance between the regions x and y '.

With the $dist/2$ function, three dyadic relations on spatial regions are defined: $DC(x, y)$, standing for ' x is disconnected from y '; $EC(x, y)$, read as ' x is externally connected from y '; and, $Co(x, y)$, read as ' x is coalescent with y '. These relations, and the continuous transitions between them, are shown in Figure 2.

The relations DC , EC and Co receive a special status in this work (amongst all of the possible relations between spatial regions) due to the fact that they can be distinguished via analyses on the sensor data.

Assuming the symbol δ as representing a pre-defined distance value, and x and y as variables for spatial regions, the relations DC , EC and Co are axiomatised by the formulae (A1), (A2) and (A3).

$$\begin{aligned} (A1) \quad & DC(x, y) \leftrightarrow (dist(x, y) > \delta) \\ (A2) \quad & EC(x, y) \leftrightarrow (dist(x, y) \leq \delta) \wedge \\ & (dist(x, y) \neq 0) \\ (A3) \quad & Co(x, y) \leftrightarrow dist(x, y) = 0 \end{aligned}$$

The distance δ is determined with respect to the application domain. For instance, in the domain of a mobile robot, assuming that the spatial regions in the calculus represent the regions of space occupied by physical bodies, δ can be assumed to be the size of the robot. Therefore the axiom (A2) can be understood as “two objects are *externally connected* if the distance between them constitutes an obstacle to the robot’s motion”. Thus, EC in this case can be used to define paths within a spatial planning system. Similar arguments apply for Co and DC .

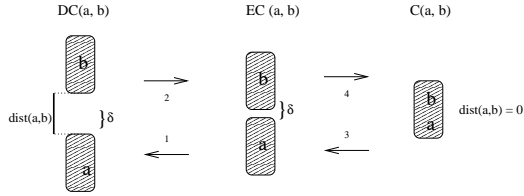


Figure 2: Relations on regions and the continuous transitions between them.

Transitions between spatial relations play a central role in this work. Next section describes the set of axioms (T1) to (T4) characterising the possible transitions between the relations above.

Interpreting Transitions

In this section the set of axioms (A1), (A2) and (A3) are extended in order to express the images of physical bodies with respect to viewpoints and the transitions between these images in time.

In order to represent transitions, the ontology for space described above is extended by assuming a sort for time points. New sorts for *viewpoints* and *visual objects* are also introduced in order to represent, respectively, the observer’s viewpoint and the objects noted by the sensors. It is worth pointing out that the *visual objects* sort represent all information noted by the robot sensors, which includes object reflexions, occlusions and sensor noise. A discussion of how these issues affect the logic-based interpretation of a robot sensor data was initiated in (Santos & Shanahan 2001). However, further investigation should be conducted in order to properly solve these problems.

In order to represent the images of *visual objects* as noted by a robot sensor, the image function $i/3$ is introduced (Randell, Witkowski, & Shanahan 2001). The function $i(x, \nu, t)$ is read as ‘the image of x as seen from ν at time t ’. I.e., $i/3$ represents a mapping from *visual objects*, *viewpoints* and *time points* to *spatial regions* in the sensor images.

Assuming the arguments of Co , DC and EC as being the output of the function $i/3$, the axioms (A1), (A2) and (A3) (as described in the previous section) can be included in the extended ontology.

In this language, the transitions between the displacement relations are represented by the dynamic predicates *receding/3*, *approaching/3*, *splitting/3* and *coalescing/3*. Assuming a and b as two distinct visual objects in the environment noted by the robot sensors, these predicates are intuitively defined below.

- *approaching*($i(a, \nu, t), i(b, \nu, t)$), read as ‘the image of a and b are *approaching* each other as noted from the viewpoint ν at time t ’;
- *receding*($i(a, \nu, t), i(b, \nu, t)$), read as ‘the images of a and b are *receding* from each other as noted from the viewpoint ν at time t ’;
- *coalescing*($i(a, \nu, t), i(b, \nu, t)$), ‘the images of a and b are *coalescing* as noted from the viewpoint ν at time t ’;
- *splitting*($i(a, \nu, t), i(b, \nu, t)$), ‘the images of a and b are *splitting* from each other as noted from the viewpoint ν at time t ’.

These predicates are axiomatised in the formulae (T1) to (T4). Axioms (T1) to (T4) are defined over time intervals $[t_1, t_2]$. In this paper $[t_1, t_2]$ is assumed to be shorter than the time for an object to suffer two consecutive changes in the world. The left hand side of (T1) to (T4) represents high-level hypotheses to explain the world at a time $t \in (t_1, t_2)$; thus, the assumption that $[t_1, t_2]$ is a short time interval guarantees that the interpretation at t is a possible explanation for one single change in the robot’s environment.

Herein, the notions of *location* and a *viewpoint at a location* are assumed as interchangeable. In other words, we are not considering in this work that none direction of gaze in privileged and that at a certain location the observer could have a 360° view of the world. Distinguishing the notions of location and viewpoint in the ontology is a matter for further investigation.

The axioms (T1) to (T4) assume a primitive *between/3* in order to express an order on viewpoints. The statement *between*(x, y, z) is read as “ x lies in between y and z ”. In practice, the viewpoints of a mobile robot are chronologically ordered, i.e., if a robot has a viewpoint ν_1 at time t_1 and ν_2 at time t_2 , a viewpoint ν is considered in between ν_1 and ν_2 if the robot has ν at an instant that falls inside the interval $[t_1, t_2]$.

$$\begin{aligned} (T1) \quad & approaching(i(a, \nu, t), i(b, \nu, t)) \wedge \\ & (t_1 < t) \wedge (t < t_2) \rightarrow \\ & between(\nu, \nu_1, \nu_2) \\ & \wedge DC(i(a, \nu_1, t_1), i(b, \nu_1, t_1)) \wedge \\ & \neg Co(i(a, \nu_2, t_2), i(b, \nu_2, t_2)) \wedge \\ & (dist(i(a, \nu_1, t_1), i(b, \nu_1, t_1)) > \\ & dist(i(a, \nu_2, t_2), i(b, \nu_2, t_2))) \end{aligned}$$

Axiom (T1) expresses that if two images are *approaching* each other at a time point t then at some time point t_1 before t the images were *disconnected*; it is not the case that

the images of a and b were coalescing at t_2 , and the distance between them was larger than at a time instant t_2 after t . The condition that $i(a, \nu_2, t_2)$ and $i(b, \nu_2, t_2)$ are non coalescent at t_2 guarantees that *approaching/2* does not include *coalescing/2*, as described in axiom (T2).

$$(T2) \quad \text{coalescing}(i(a, \nu, t), i(b, \nu, t)) \wedge \\ (t_1 < t) \wedge (t < t_2) \longrightarrow \\ \text{between}(\nu, \nu_1, \nu_2) \wedge \\ (EC(i(a, \nu_1, t_1), i(b, \nu_1, t_1)) \vee \\ DC(i(a, \nu_1, t_1), i(b, \nu_1, t_1))) \wedge \\ Co(i(a, \nu_2, t_2), i(b, \nu_2, t_2)))$$

If two images are *coalescing* at a time instant t (as represented by (T2)) then they are *externally connected* at a time point t_1 before t and *coalescent* (*Co*) at a t_2 later than t .

$$(T3) \quad \text{splitting}(i(a, \nu, t), i(b, \nu, t)) \wedge \\ (t_1 < t) \wedge (t < t_2) \longrightarrow \\ \text{between}(\nu, \nu_1, \nu_2) \wedge \\ Co(i(a, \nu_1, t_1), i(b, \nu_1, t_1)) \wedge \\ (EC(i(a, \nu_2, t_2), i(b, \nu_2, t_2)) \vee \\ DC(i(a, \nu_1, t_2), i(b, \nu_1, t_2))))$$

Axiom (T3) expresses that if two images are *splitting* at a time instant t then they are *coalescent* at t_1 before t and *externally connected* at t_2 after t .

$$(T4) \quad \text{receding}(i(a, \nu, t), i(b, \nu, t)) \wedge \\ (t_1 < t) \wedge (t < t_2) \longrightarrow \\ \text{between}(\nu, \nu_1, \nu_2) \wedge \\ (DC(i(a, \nu_1, t_1), i(b, \nu_1, t_1)) \vee \\ EC(i(a, \nu_1, t_1), i(b, \nu_1, t_1))) \wedge \\ (dist(i(a, \nu_1, t_1), i(b, \nu_1, t_1)) \\ < dist(i(a, \nu_2, t_2), i(b, \nu_2, t_2))))$$

If two images are *receding* each other at a time point t , according to (T4), then at some time point t_1 before t the images were *disconnected* (or *externally connected*) and the distance between them was shorter than the distance at a time instant t_2 after t .

Finally, if two images are static in t then the distance between them do not change from time point t_1 to t_2 as expressed by axiom (T5).

$$(T5) \quad \text{static}(i(a, \nu, t), i(b, \nu, t)) \wedge \\ (t_1 < t) \wedge (t < t_2) \longrightarrow \\ (dist(i(a, \nu, t_1), i(b, \nu, t_1)) = \\ dist(i(a, \nu, t_2), i(b, \nu, t_2)))$$

It is assumed, in axioms (T1) to (T5) above, that no information is available in between the end-points of every open time interval (t_1, t_2) . This assumption is plausible since the borders of the interval $(t_1$ and $t_2)$ represent the instants when information was acquired from the world by the robot sensors. Thus, the robot will have to *guess* what happened during the intervals given data about their borders. This hypothesised interpretation can be a wrong explanation for the sensor data but at least it can give a hint about the environment. Misleading hypothesis can be further updated (or revised) according to further perceived data about the development of the states of the world. This, however, is an issue for further investigation.

The next section explores the relationship between the dynamical predicates above and the physical bodies in the robot's environment.

From Transitions to Object Relations

In this section the predicates on images presented above (*approaching*, *coalescing*, *splitting*, *receding* and *static*) are understood as abstract definitions of the relations on physical bodies described below.

1. *getting_closer*(a, b, ν, t), read as 'objects a and b are *getting closer* to each other at time t as noted from the viewpoint ν ';
2. *ap-getting_closer*(a, b, ν, t), read as 'objects a and b are *apparently getting closer* to each other at time t due to motion of the observer';
3. *getting_further*(a, b, ν, t), read as 'objects a and b are *getting further* to each other at time t as noted from the viewpoint ν ';
4. *ap-getting_further*(a, b, ν, t), read as 'objects a and b are *apparently getting further* to each other at time t due to motion of the viewpoint ν ';
5. *occluding*(a, b, ν, t), read as 'one of the objects a and b is *occluding* the other at time instant t as noted by the viewpoint ν ';
6. *touching*(a, b, ν, t), read as ' a and b are *touching* each other at time t as noted by ν ';
7. *static*(a, b, ν, t), read as ' a and b are *static* at time t '.

For the purposes of rigorously presenting the connection between the previous set of relations and the abstract definitions defined in the previous section, herein we assume the predicate *located/3*; *located*(a, ν, t) represents the fact that a physical body a is located at ν at time t . Therefore, the remaining sections assume that the robot is equipped with a map with which it is able to locate itself in its environment. This simplification should be relaxed in future research, so that a similar framework to that described in this paper could be used in a robot map building process.

Herein, for brevity, we assume a physical body variable - *robot* - whenever a robot is referred to.

By assuming these definitions, the connections between the predicates on images (described the previous section) and the predicates on physical bodies are represented by the axioms (IO 1) to (IO 4) below.

$$(IO\ 1) \quad \text{located}(\text{robot}, \nu, t) \wedge \\ [\text{getting_closer}(a, b, \nu, t) \vee \\ \text{ap_getting_closer}(a, b, \nu, t)] \\ \longrightarrow \text{approaching}(i(a, \nu, t), i(b, \nu, t)) \\ (IO\ 2) \quad \text{located}(\text{robot}, \nu, t) \wedge \\ [\text{occluding}(a, b, \nu, t) \vee \\ \text{touching}(a, b, \nu, t)] \\ \longrightarrow \text{coalescing}(i(a, \nu, t), i(b, \nu, t))$$

- (IO 3) $located(robot, \nu, t) \wedge$
 $[getting_further(a, b, \nu, t)$
 $\vee ap_getting_further(a, b, \nu, t)]$
 $\rightarrow [splitting(i(a, \nu, t), i(b, \nu, t)) \vee$
 $receding(i(a, \nu, t), i(b, \nu, t))]$
- (IO 4) $located(robot, \nu, t) \wedge$
 $static(a, b, \nu, t) \rightarrow$
 $static(i(a, \nu, t), i(b, \nu, t))$

These axioms involve important issues concerning the relationship between the knowledge about objects and related knowledge about images. Issues such as the distinction of a moving viewpoint from moving objects, grounding symbols to sensor data, creating knowledge about the 3D world from under-constrained 2D images, etc. In order to solving these problems, the framework presented above has to consider (at least) information about the observer's motion, about the changes in the observer-relative positions of the background features and range data. This, however, warrants further investigation.

Inference

The purpose of inference within the framework presented in this paper is two fold: explanation of sensor data and prediction of their future configurations. Explanation is accomplished by assuming abduction as the inference rule in a similar way as proposed in (Shanahan 1996). Prediction, on the other hand, is handled by deduction. This duality between *abduction* and *deduction* was first explored in (Shanahan 1989) for temporal reasoning.

We briefly introduce the concept of abduction and relate it to the definitions presented in the previous sections¹. Abduction is the process of explaining a set of sentences Γ by finding a set of formulae Δ such that, given a background theory Σ , Γ is a logical consequence of $\Sigma \cup \Delta$. In order to avoid trivial explanations, a set of predicates is distinguished (the *abducible predicates*) such that every acceptable explanation must contain only these predicates.

Assuming the framework proposed in the previous sections, the description of sensor data in terms of displacement relations comprises the set Γ . The background theory Σ is, then, assumed to be composed of the set of axioms (T1) to (T5) and (IO1) to (IO4). Finally, the abducibles are considered to be the abstract predicates, *receding/3*, *approaching/3*, *splitting/3* and *coalescing/3*, defined above.

In order to clarify the concepts introduced in this paper and to give a first approach to inference in this framework, we present the example below. This example is a sketch of the inference procedures in the present framework. A complete discussion of these procedures is a matter for further investigation.

In this section we assume, as abbreviation, that the order of time points is implicit in their own representation, i.e., $t_i < t_j$ if and only if $i < j$ (for time points t_i and t_j , and integers i and j) and that the viewpoint ν_i is related to the time point t_i . Moreover, lower case roman letters are used to

¹For brevity an explanation of deduction is omitted in this paper.

represent ground terms, while upper case letters are reserved for variables.

Skolemisation is also implicit in this example, lower case bold letters are used to represent the skolem functions of their non-bold counterpart variables (i.e., U is the skolemised version of U). Here, skolem functions are used to keep the reference of variables in one step of inference throughout its further steps.

For the sake of brevity, we omit in the example below details about how to make inferences about the location of the robot in its environment.

The example below assumes depth maps taken from the viewpoint of a robot navigating through an office-like environment (Figures 3 and 4). For the sake of simplification, the framework developed in this paper is applied on cylindrical objects with added textures.

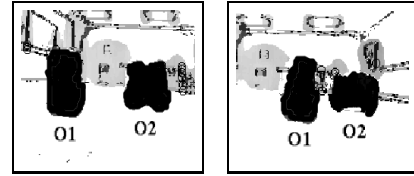


Figure 3: Depth maps at viewpoints ν_0 and ν_1 .

Considering the sequence of snapshots of the world in Figure 3, the first step of sensor data assimilation is the description of this sequence in terms of the displacement relations. The result of this task is exemplified in the formula (1) below (where o_1 and o_2 represent the two objects — rectangular areas — in the scene; and, ν_0 and ν_1 represents the viewpoints where both pictures in Figure 3 were taken).

- (1) $located(robot_1, \nu_0, t_0) \wedge located(robot_1, \nu_1, t_1) \wedge$
 $DC(i(o_1, \nu_0, t_0), i(o_2, \nu_0, t_0)) \wedge DC(i(o_1, \nu_1, t_1),$
 $i(o_2, \nu_1, t_1)) \wedge (dist(i(o_1, \nu_0, t_0), i(o_2, \nu_0, t_0)) >$
 $dist(i(o_1, \nu_1, t_1), i(o_2, \nu_1, t_1)))$

Assuming formula (1) and the axioms described in the previous sections, formulae (2) and (3) can be inferred as an interpretation of the sensor information in Figure 3.

- (2) $\exists V T located(robot_1, V, T) \wedge between(V, \nu_0, \nu_1)$
 $\wedge approaching(i(o_1, V, T), i(o_2, V, T)) \wedge t_0 < T$
 $\wedge T < t_1$ from (1) and axiom (T1);
- (3) $ap_getting_closer(o_1, o_2, V, T)$, from (2), axiom (IO1) and the assumption of object immovably;

Formula (3) is a hypothesis on the state of the objects in the world to explain the given sensor data.

From formulae (1), (2) and (3) we would like to derive a set of expectations (predictions) about the future possible sensor data and their interpretations in terms of object relations. One possible set of predictions is comprised by the

formulae in the **Predictions I** set.

Predictions I:

$$\left(\begin{array}{l} (I.1) \quad \exists W_1 U_1 \text{ located}(\text{robot}_1, W_1, U_1) \wedge \\ \quad \text{between}(W_1, \nu_1, \nu_2) \wedge \\ \quad EC(i(o_1, W_1, U_1), i(o_2, W_1, U_1)) \\ \quad \wedge t_1 < U_1 \\ (I.2) \quad \exists W_2 U_2 \text{ located}(\text{robot}_1, W_2, U_2) \wedge \\ \quad \text{coalescing}(i(o_1, W_2, U_2), i(o_2, W_2, U_2)) \\ \quad \wedge U_1 < U_2 \text{ from (I.1) and axiom (T2);} \\ (I.3) \quad \text{occluding}(o_1, o_2, \mathbf{W}_2, \mathbf{U}_2); \\ \quad \text{from (I.2), axiom (IO2) and} \\ \quad \text{the assumption of immovably of objects;} \end{array} \right)$$

Formula (I.3) is a hypothesis about the future relationship between objects o_1 , o_2 and the observer.

These predictions assume that the observer continued its motion in the same direction as taken when the snapshots of Figure 3 were obtained. In the general case, however, many possible predictions can be supposed. Further research should a mechanism for handling the predictions according to assumptions about the motions of objects and the actions of the observer.

Figure 4 shows the images from the robot's camera at viewpoints ν_2 , ν_3 and ν_4 .

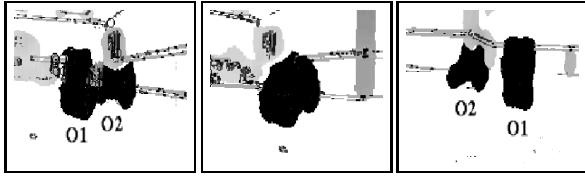


Figure 4: Depth maps at viewpoints ν_2 , ν_3 and ν_4 .

The sensor data noted at ν_2 is described by formulae (4). Formulae (5) and (6) follow from (4) and the axioms.

- (4) $\text{located}(\text{robot}_1, \nu_2, t_2) \wedge$
 $(\text{dist}(i(o_1, \nu_2, t_2), i(o_2, \nu_2, t_2)) \leq \delta)$ from sensor data;
- (5) $EC(i(o_1, \nu_2, t_2), i(o_2, \nu_2, t_2))$ from (4) and axiom (A2);
- (6) $\exists V T \text{ located}(\text{robot}_1, V, T) \wedge \text{between}(V, \nu_2, \nu_3)$
 $\wedge \text{approaching}(i(o_1, V, T), i(o_2, V, T)) \wedge t_1 < T \wedge$
 $T < t_2 \wedge \nu_1 < V$ from (4), (5) and axiom (T1);

Similarly to the first set of predictions, **Predictions II** hypothesises about the future possible sensor data and about the relationship between objects o_1 and o_2 , and the observer.

Predictions II:

$$\left(\begin{array}{l} (II.1) \quad \exists W_1 U_1 \text{ located}(\text{robot}_1, W_1, U_1) \wedge \\ \quad \text{between}(W_1, \nu_2, \nu_4) \wedge \\ \quad Co(i(o_1, W_1, U_1), i(o_2, W_1, U_1)) \wedge \\ \quad t_2 < U_1 \\ (II.2) \quad \exists W_2 U_2 \text{ located}(\text{robot}_1, W_2, U_2) \wedge \\ \quad \text{splitting}(i(o_1, W_2, U_2), i(o_2, W_2, U_2)) \\ (II.3) \quad \text{ap_getting_further}(o_1, o_2, \mathbf{W}_2, \mathbf{U}_2) \\ \quad \text{from (II.1) and axiom (IO3);} \end{array} \right)$$

- (7) $\text{located}(\text{robot}_1, \nu_3, t_4) \wedge$
 $(\text{dist}(i(o_1, \nu_3, t_4), i(o_2, \nu_3, t_4)) = 0)$ from sensor data;
- (8) $Co(i(o_1, \nu_3, t_4), i(o_2, \nu_3, t_4))$ from (6) and axiom (A3);
- (9) $\exists V T \text{ located}(\text{robot}_1, V, T) \wedge \text{between}(V, \nu_3, \nu_4)$
 $\wedge \text{coalescing}(i(o_1, V, T), i(o_2, V, T)) \wedge t_3 < T \wedge$
 $T < t_4$ from (7) and axiom (T2);

Formula (9), derived from the axioms and the descriptions of the images, confirms the prediction (I.2) and, consequently, (I.3).

Discussion

This paper described three sets of axioms constituting a logic-based hierarchy for scene interpretation. The first layer of this hierarchy, constituted by the axioms (A1), (A2) and (A3), formalises relations between pairs of spatial regions assuming a distance function as primitive. The purpose of this first set of axioms is to classify, in terms of displacement relations, images of the objects in space as noted by a mobile robot's sensors. Transitions between these relations in a sequence of sensor data were, then, axiomatised by the second set of axioms ((T1) to (T5)), defining the second layer of the image interpretation system.

The second layer of the hierarchy aims the classification of transitions in the sensor data by means of *abstract* predicates (the left-hand side of axioms (IO1) to (IO4)). These predicates were, then, rewritten into possible explanations for the sensor data transitions in terms of object-observer relations. The last set of axioms ((IO1) to (IO4)) characterises this process, which constitutes the final layer of the hierarchy.

The purpose of the qualitative spatial theory presented in this paper is to demonstrate how much can be achieved using distance information alone. This simple theory composes the foundations of a more complex, practical, system.

The use of abstract predicates in section *From Transitions to Object Relations* recalls the idea of abstract reasoning (Console & Dupre 1994)(Giunchiglia & Walsh 1989). Abstract reasoning frameworks have concentrated mainly on using abstractions to provide general proofs in automated theorem proving in order to guide proofs in the ground space (Giunchiglia 1990). In the present paper, however, abstraction is used to give a general interpretation of an ordered pair of sensor data description. In this sense, the main purpose of using abstract definitions is to overlook the ambiguities in the sensor data, keeping every plausible interpretation of a scenario inside a more general abstract concept. Axioms (IO1) to (IO4) define the abstract predicates in terms of the more specific equally-plausible hypotheses to explain particular transitions. Therefore, not only can abstraction interleave *planning and execution* (as proposed in (Nourbakhsh 1998)) but also it can interleave sensor data interpretation and planning. The development of this issue is a problem for future research.

Another important subject for further development of this research is the exploration of the framework for perception incorporating feedback and expectation, as proposed in

(Shanahan 2002), into the sensor data interpretation process described in this paper.

Acknowledgments Thanks to David Randell and to the two anonymous reviewers for the invaluable suggestions and comments. This work has been partially supported by CAPES.

References

- Cohn, A. G.; Bennet, B.; Gooday, J.; and Gotts, N. 1997. Representing and reasoning with qualitative spatial relations about regions. In Stock, O., ed., *Spatial and Temporal Reasoning*. Kluwer Academic Publishers. 97 – 134.
- Console, L., and Dupre, D. T. 1994. Abductive reasoning with abstraction axioms. In *LNCS*, volume 810. Springer-Verlag.
- Coradeschi, S., and Saffiotti, A. 2000. Anchoring symbols to sensor data: preliminary report. In *Proc. of the 17th National Conference on Artificial Intelligence (AAAI-2000)*, 129–135.
- Cox, I., and Wilfong, G. 1990. *Autonomous Robot Vehicles*. Springer-Verlag.
- Giunchiglia, F., and Walsh, T. 1989. Abstract theorem proving. In *Proceedings of the 11th IJCAI*, 372–377.
- Giunchiglia, F. 1990. The use of abstraction in automatic inference. DAI research paper 454, University of Edinburgh.
- Hernández, D.; Clementini, E.; and di Felice, P. 1995. Qualitative distances. In *Spatial Information Theory*, volume 988 of *LNCS*. Springer-Verlag. 45 – 57.
- Latombe, J. C. 1991. *Robot Motion Planning*. Kluwer Academic Publishers.
- Lowe, J. C., and Moryadas, S. 1975. *The Geography of Movement*. Houghton Mifflin.
- Nourbakhsh, I. R. 1998. Using abstraction to interleave planning and execution. In *Proceeding of the third biannual world automation congress*.
- Randell, D.; Cui, Z.; and Cohn, A. 1992. A spatial logic based on regions and connection. In *Proc. of the KR*, 165–176.
- Randell, D.; Witkowski, M.; and Shanahan, M. 2001. From images to bodies: modeling and exploiting spatial occlusion and motion parallax. In *Proc. of the IJCAI*.
- Reiter, R. 2002. *Knowledge in Action*. The MIT Press.
- Santos, P., and Shanahan, M. 2001. From stereoscopic vision to symbolic representation. In *Working notes of the AAAI Fall Symposium on Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems*.
- Shanahan, M. 1989. Prediction is deduction but explanation is abduction. In *Proc. of the IJCAI*.
- Shanahan, M. 1996. Noise and the common sense informatic situation. In *Proceedings AAAI 96*, 1098–1103.
- Shanahan, M. 1997. *Solving the frame problem*. The MIT press.
- Shanahan, M. 2002. A logical account of perception incorporating feedback and expectation. In *Proc. of the KR*.
- Vieu, L. 1997. Spatial representation and reasoning in AI. In Stock, O., ed., *Spatial and Temporal Reasoning*. Kluwer.