Hypothesising Object Relations from Image Transitions

Paulo Santos and **Murray Shanahan**[1]

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.

1 Introduction

Much 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 [3][8]. Such issues have been addressed by the knowledge representation sub-field of Artificial Intelligence [17][13]. Knowledge representation (KR) theories, however, have largely been developed in isolation from empirical issues such as how knowledge about the world is acquired and what the physical mechanisms are by which it is embodied in the agents.

The present paper describes a logic-based formalism for representing knowledge about objects in space and their movement, and shows how to build up such knowledge 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 sensor data can be explained by hypothesising the existence of physical objects along with the dynamic relationships that hold between them, all with respect to a (possibly moving) viewpoint. The approach used recalls the abductive account of sensor data assimilation first proposed in [16]. However, while this earlier work deals with spatial occupancy, it does not deal with the question of the relationship between spatially-located objects and the viewpoint of an observer. One motivation for the present paper is to propose a spatial representation framework capable of coping with this issue. Moreover, as pointed out in [19], the knowledge representation community has produced very little work on formalisms that handle both space and time, a combination that plays a central role in the representational system described in this paper.

Two theories of qualitative spatial reasoning are particularly relevant to this work, namely the Region Connection Calculus (RCC) [11][1] and the Region Occlusion Calculus (ROC) [12]. 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 is reminiscent of ROC. The present framework, however, extends both RCC and ROC, in the sense that it assumes sensory information as the foundation of the knowledge representation formalism.

A brief overview of the RCC and ROC formalisms follows. RCC 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, ROC was designed to model the spatial occlusion of arbitrary shaped objects. The theory captures a set of spatial relations expressing object interposition that can hold between pairs of regions, each corresponding to the image of a body as seen from some viewpoint. The present paper builds on this by supplying a *dynamic* characterization of occlusion, based on the sort of information that is obtainable directly from the visual system of a mobile robot.

According to the present paper, the process of sensor data interpretation is composed of three sub-tasks. First, visual snapshots of the world are represented as 2D spatial regions. The relations between these regions are characterised in a logic-based language similar to RCC. This language is presented in Section 2. Second, from chronological sequences of these snapshots, a representation is formed of the transitions between them, in terms of dynamic predicates over images of objects. The way these predicates are mapped to transitions in the raw sensor data is described in Section 3.

Encoded in these sensor data transitions is information about the changing relationships between objects moving about in the world. So the third step in the interpretation process is to hypothesise sets of

[1] Dep. of Electrical and Electronic Engineering, Imperial College, London, SW7 2BT, email: {p.santos, m.shanahan}@ic.ac.uk

dynamic spatial relations between physical objects that can account for the transitions in the visual sensor data. The mapping between the predicates used to represent image transitions and those used to represent changing relationships between physical objects is described in Section 4. Section 5 presents a brief discussion of how to make inferences within this framework.

For brevity, the variables used in this paper are universally quantified unless explicitly mentioned.

2 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 is that an estimate of the relative distance between objects in a robot's environment (and between pairs of regions in images) can be extracted directly from the robot's sensor data, assuming the basic problems of image segmentation are overcome using off-the-shelf machine vision techniques.

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 [9]. Initial work on qualitative notions of distance for artificial intelligence is presented in [7].

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 length 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 1.

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, the relations DC, EC and Co are axiomatised by the formulae $(A1)$, $(A2)$ and $(A3)$.

$$(A1) \quad DC(x, y) \leftrightarrow (dist(x, y) > \delta)$$
$$(A2) \quad EC(x, y) \leftrightarrow (dist(x, y) \leq \delta) \wedge (dist(x, y) \neq 0)$$
$$(A3) \quad Co(x, y) \leftrightarrow dist(x, y) = 0$$

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 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.

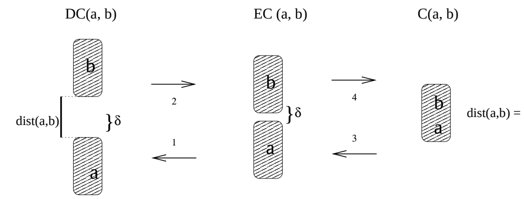

Figure 1. Relations on regions and the continuous transitions between them.

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

3 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 represents all information obtained by the robot sensors, which includes object reflections, occlusions, and sensor noise. A discussion of how these issues affect the logic-based interpretation of robot sensor data was initiated in [14]. However, further investigation should be conducted in order to properly solve these problems.

In order to represent the raw images of *visual objects* obtained by a robot's vision system, the image function $i/3$ is introduced [12]. The function $i(x, \nu, t)$ is read as 'the image of x as seem 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 to be 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 that a and b are two distinct visual objects observed by the robot's vision system, the previous 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 $(T4)$ below. For these axioms to be useful in an abductive setting, it is assumed that the interval $[t_1, t_2]$ is short enough to rule out the occurrence of multiple discontinuities between consecutive snapshots of the world.

The axioms make use of a notion of *location*, which is ontologically indistinguishable from that of a *viewpoint*. Accordingly, locations and viewpoints are assigned to the same sort. Intuitively, the idea of a viewpoint includes the direction of gaze as well as the location of the observer, but the direction of gaze plays no role in the present formalisation.

Axioms ($T1$) to ($T4$) also assume Tarski's primitive *betweeness* ($between/3$) in order to capture an ordering on viewpoints. The statement $between(x, y, z)$ is read as "x lies in between y and z", and it is intuitively defined as 'x, y, z are co-linear *and* each circle through y, z cuts both circles (y, yx) and (z, zx)' [20] (a more recent treatment of *betweeness* is presented in [4]). In fact, as $[t_1, t_2]$ is a short time interval, there is no loss in generality on assuming co-linearity of viewpoint pairs taken at t_1 and t_2.

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

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 $coalescing/2$ (axiom ($T2$)).

$$
\begin{aligned}
(T2) \quad & coalescing(i(a, \nu, t), i(b, \nu, t)) \longrightarrow \\
& \exists\, t_1\ t_2\ \nu_1\ \nu_2\ (t_1 < t) \wedge (t < t_2) \wedge between(\nu, \nu_1, \nu_2) \wedge \\
& \quad [EC(i(a, \nu_1, t_1), i(b, \nu_1, t_1)) \vee \\
& \quad\quad DC(i(a, \nu_1, t_1), i(b, \nu_1, t_1)] \\
& \quad\quad \wedge Co(i(a, \nu_2, t_2), i(b, \nu_2, t_2))
\end{aligned}
$$

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

$$
\begin{aligned}
(T3) \quad & splitting(i(a, \nu, t), i(b, \nu, t)) \longrightarrow \\
& \exists\, t_1\ t_2\ \nu_1\ \nu_2\ (t_1 < t) \wedge (t < t_2) \wedge between(\nu, \nu_1, \nu_2) \wedge \\
& \quad Co(i(a, \nu_1, t_1), i(b, \nu_1, t_1)) \wedge \\
& \quad\quad [EC(i(a, \nu_2, t_2), i(b, \nu_2, t_2)) \\
& \quad\quad \vee DC(i(a, \nu_1, t_2), i(b, \nu_1, t_2))]
\end{aligned}
$$

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* (or *disconnected*) at t_2 after t.

$$
\begin{aligned}
(T4) \quad & receding(i(a, \nu, t), i(b, \nu, t)) \longrightarrow \\
& \exists\, t_1\ t_2\ \nu_1\ \nu_2\ (t_1 < t) \wedge (t < t_2) \wedge between(\nu, \nu_1, \nu_2) \wedge \\
& \quad [EC(i(a, \nu_1, t_1), i(b, \nu_1, t_1)) \vee \\
& \quad\quad DC(i(a, \nu_1, t_1), i(b, \nu_1, t_1))] \\
& \quad\quad \wedge (dist(i(a, \nu_1, t_1), i(b, \nu_1, t_1)) \\
& \quad\quad\quad < dist(i(a, \nu_2, t_2), i(b, \nu_2, t_2)))
\end{aligned}
$$

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

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

$$
\begin{aligned}
(T5) \quad & static(i(a, \nu, t), i(b, \nu, t)) \longrightarrow \\
& \exists\, t_1\ t_2\ (t_1 < t) \wedge (t < t_2) \wedge \\
& \quad (dist(i(a, \nu, t_1), i(b, \nu, t_1)) = \\
& \quad\quad dist(i(a, \nu, t_2), i(b, \nu, t_2)))
\end{aligned}
$$

The above axioms are used in the following way. The robot's camera acquires a series of snapshots of its environment. The task of abduction is to find an explanation for the difference between pairs of consecutive snapshots, by hypothesising one of five dynamic relations between the 2D regions they depict. In practice, the interval between consecutive snapshots will be determined by the frame-rate of the robot's camera. As we will see in the next section, further abductive inference then takes place to explain these relations, in terms of the motion of 3D objects, including the robot itself.

4 From Transitions to Object Relations

In this section, the five possible dynamic relations between image regions presented above ($approaching$, $coalescing$, $splitting$, $receding$ and $static$) are related to the seven new relations on physical bodies described below.

1. $getting\_closer(a, b, \nu, t)$, read as 'objects a and b are *getting closer* to each other at time t from the viewpoint ν';
2. $ap\_getting\_closer(a, b, \nu, t)$, read as 'objects a and b are *apparently getting closer* to each other at time t due to motion of the observer';
3. $getting\_further(a, b, \nu, t)$, read as 'objects a and b are *getting further* from each other at time t from the viewpoint ν';
4. $ap\_getting\_further(a, b, \nu, t)$, read as 'objects a and b are *apparently getting further* from each other at time t due to motion of the observer ν';
5. $occluding(a, b, \nu, t)$, read as 'one of the objects a and b is *moving in front of* the other at time instant t from the viewpoint ν';
6. $touching(a, b, \nu, t)$, read as 'a and b are touching each other at time t as noted by ν';
7. $static(a, b, \nu, t)$, 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 $located/3$; $located(a, \nu, t)$ 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.

From here on, for brevity, we assume a constant — $robot$ — of the sort *visual object* that denotes the robot.

Based on the informal definitions above, the relationship between the predicates on images (described in Section 3) and the predicates on physical bodies is captured by the axioms ($IO\ 1$) to ($IO\ 4$) below.

$(IO\,1)$ $approaching(i(a,\nu,t), i(b,\nu,t)) \leftarrow$
$located(robot,\nu,t) \wedge [getting\_closer(a,b,\nu,t) \vee$
$ap\_getting\_closer(a,b,\nu,t)]$

$(IO\,2)$ $coalescing(i(a,\nu,t), i(b,\nu,t)) \leftarrow$
$located(robot,\nu,t) \wedge [occluding(a,b,\nu,t)$
$\vee touching(a,b,\nu,t)]$

$(IO\,3)$ $[splitting(i(a,\nu,t), i(b,\nu,t)) \vee$
$receding(i(a,\nu,t), i(b,\nu,t))] \leftarrow$
$located(robot,\nu,t) \wedge [getting\_further(a,b,\nu,t) \vee$
$ap\_getting\_further(a,b,\nu,t)]$

$(IO\,4)$ $static(i(a,\nu,t), i(b,\nu,t)) \leftarrow$
$located(robot,\nu,t) \wedge static(a,b,\nu,t)$

5 Inference

The purpose of inference within the framework presented in this paper is twofold: explanation of sensor data and prediction of their future configurations. Explanation is accomplished through abduction in a way similar to the proposed in [16]. Prediction, on the other hand, is handled by deduction. This duality between $abduction$ and $deduction$ was explored in [15] for temporal reasoning.

We briefly introduce the concept of abduction and relate it to the definitions presented in the previous sections. 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 (Section 2) comprises the set Γ. The background theory Σ is, then, assumed to be constituted by 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 in Section 3.

In order to clarify the concepts introduced in this paper and to give an idea of the sort of inference possible in this framework, we present the example below. This example assumes, as an abbreviation, that the ordering of time points is implicit in their own notation, 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 represent variables, while upper case letters are reserved for ground terms.

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., $\mathbf{u}$ is the skolemised version of u). Here, skolem functions are used to maintain the reference of variables from one inference step to the next.

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 2 and 3). For the sake of simplification, the framework developed in this paper is applied on cylindrical objects with added textures.

Consider the sequence of snapshots of the world in Figure 2. The first step of sensor data assimilation is the description of this sequence in terms of the displacement relations discussed in Section 2. The result of this task is exemplified by the formula (1) below (where O_1 and O_2 represent the two objects — rectangular areas —

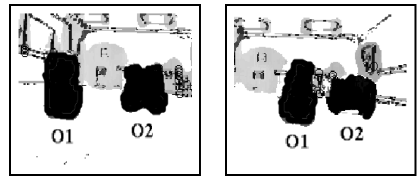

Figure 2. Depth maps at viewpoints V_1 and V_2.

in the scene; and, V_1 and V_2 represents the viewpoints where both pictures in Figure 2 were taken).

(1) $located(Robot_1, V_1, T_0) \wedge located(Robot_1, V_2, T_1) \wedge$
$DC(i(O_1, V_1, T_0), i(O_2, V_1, T_0)) \wedge DC(i(O_1, V_2, T_1),$
$i(O_2, V_2, T_1)) \wedge (dist(i(O_1, V_1, T_0), i(O_2, V_1, T_0)) >$
$dist(i(O_1, V_2, T_1), i(O_2, V_2, T_1)))$

Assuming formula (1) and the axioms described in Sections 2, 3 and 4, formulae (2) and (3) can be abducted as an interpretation of the sensor information in Figure 2.

(2) $\exists\ v\ \ t\ located(Robot_1, v, t) \wedge between(v, V_1, V_2) \wedge$
$approaching(i(O_1, v, t), i(O_2, v, t)) \wedge T_0 < t \wedge$
$t < T_1$ from (1) and axiom $(T1)$;

(3) $ap\_getting\_closer(O_1, O_2, \mathbf{v}, \mathbf{t})$, from (2), axiom (IO1) and the assumption of object immovably.;

Formula (3) is a hypothesis about 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:

$(I.1)$ $\exists\ w_1\ u_1\ located(Robot_1, w_1, u_1) \wedge$
$between(w_1, V_2, V_3) \wedge$
$EC(i(O_1, w_1, u_1), i(O_2, w_1, u_1)) \wedge T_1 < u_1$

$(I.2)$ $\exists\ w_2\ u_2\ located(Robot_1, w_2, u_2) \wedge$
$coalescing(i(O_1, w_2, u_2), i(O_2, w_2, u_2)) \wedge$
$\mathbf{u_1} < u_2$ from (I.1) and axiom (T2).;

$(I.3)$ $occluding(O_1, O_2, \mathbf{w_2}, \mathbf{u_2})$;
from $(I.2)$, axiom $(IO2)$ and
the assumption of immovably of objects;

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 after the snapshots of Figure 2 were obtained. In the general case, however, many possible predictions can be deduced. Further research is required to design a set of criteria for ranking competing hypotheses, and a method for handling these possible hypotheses within the inference.

Figure 3 shows the images from the robot's camera at viewpoints V_3, V_4 and V_5.

The sensor data noted at V_3 is described by formulae (4). Formulae (5) and (6) follow from (4) and the axioms.

(4) $located(Robot_1, V_3, T_2) \wedge$
$(dist(i(O_1, V_3, T_2), i(O_2, V_3, T_2)) \leq \delta)$ from sensor data;

(5) $EC(i(O_1, V_3, T_2), i(O_2, V_3, T_2))$ from (4) and axiom $(A2)$;

(6) $\exists\ v\ t\ located(Robot_1, v, t) \wedge between(v, V_3, V_4) \wedge$
$approaching(i(O_1, v, t), i(O_2, v, t)) \wedge T_1 < t \wedge$
$t < T_2$ from (4), (5) and axiom $(T1)$;

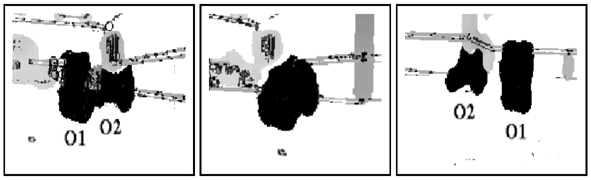

Figure 3. Depth maps at viewpoints V_3, V_4 and V_5.

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:

$$\left(\begin{array}{ll} (II.1) & \exists\, w_1\ u_1\ located(Robot_1, w_1, u_1) \wedge \\ & \quad between(w_1, V_3, V_5) \wedge \\ & \quad Co(i(O_1, w_1, u_1), i(O_2, w_1, u_1)) \wedge \\ & \quad T_2 < u_1 \\ (II.2) & \exists\, w_2\ u_2\ located(Robot_1, w_2, u_2) \wedge \\ & \quad splitting(i(O_1, w_2, u_2), i(O_2, w_2, u_2)); \\ (II.3) & ap\_getting\_further(O_1, O_2, \mathbf{w_2}, \mathbf{u_2}) \\ & \quad \text{from } (II.1) \text{ and axiom } (IO3); \end{array}\right)$$

$$\begin{array}{ll} (7) & located(Robot_1, V_4, T_4) \wedge \\ & \quad (dist(i(O_1, V_4, T_4), i(O_2, V_4, T_4)) = 0) \text{ from sensor data;} \\ (8) & Co(i(O_1, V_4, T_4), i(O_2, V_4, T_4)) \text{ from (6) and axiom (A3);} \\ (9) & \exists\, v\ t\ located(Robot_1, v, t) \wedge between(v, V_4, V_5) \wedge \\ & \quad coalescing(i(O_1, v, t), i(O_2, v, t)) \wedge T_3 < t \wedge t < t_4 \\ & \quad \text{from (7) and axiom (T2);} \end{array}$$

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

In this paper, depth information about the scene was ignored. Further research will consider the range data given by the stereo-vision system in the reasoning process.

6 Discussion

This paper described three sets of axioms for logic-based scene interpretation. These axioms form a hierarchy. 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 at 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 use of abstract predicates recalls the idea of abstract reasoning [2][6]. 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 [5]. 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 more specific equally-plausible hypotheses to explain particular transitions. Therefore, not only can abstraction interleave *planning and execution* (as proposed in [10]) but also it can interleave sensor data interpretation and planning. Further examination of this issue is a subject for future research.

Another potential topic for further investigation is the possibility of incorporating feedback and expectation into the sensor data interpretation process described in this paper, along the lines proposed in [18].

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

[1] A. G. Cohn, B. Bennet, J. Gooday, and N.M. Gotts, 'Representing and reasoning with qualitative spatial relations about regions', in *Spatial and Temporal Reasoning*, ed., Oliviero Stock, 97 – 134, Kluwer Academic Publishers, (1997).

[2] L. Console and D. T. Dupre, 'Abductive reasoning with abstraction axioms', in *LNCS*, volume 810, Springer-Verlag, (1994).

[3] I.J. Cox and G.T. Wilfong, *Autonomous Robot Vehicles*, Springer-Verlag, 1990.

[4] P. Gärdenfors, *Conceptual Spaces: the geometry of thought*, The MIT Press, 2000.

[5] F. Giunchiglia, 'The use of abstraction in automatic inference', DAI research paper 454, University of Edimburgh, (1990).

[6] F. Giunchiglia and T. Walsh, 'Abstract theorem proving', in *Proc. of IJCAI*, pp. 372–377, (1989).

[7] D. Hernández, E. Clementini, and P. di Felice, 'Qualitative distances', in *Spatial Information Theory*, volume 988 of *LNCS*, 45 – 57, Springer-Verlag, (1995).

[8] J. C. Latombe, *Robot Motion Planning*, Kluwer Academic Publishers, 1991.

[9] J. C. Lowe and S. Moryadas, *The Geography of Movement*, Houghton Mifflin, 1975.

[10] I. R. Nourbakhsh, 'Using abstraction to interleave planning and execution', in *Proceeding of the third biannual world automation congress*, (1998).

[11] D. Randell, Z. Cui, and A. Cohn, 'A spatial logic based on regions and connection', in *Proc. of KR*, pp. 165–176, (1992).

[12] D. Randell, M. Witkowski, and M. Shanahan, 'From images to bodies: modeling and exploiting spatial occlusion and motion parallax', in *Proc. of IJCAI*, (2001).

[13] R. Reiter, *Knowledge in Action*, The MIT Press, 2002.

[14] P.E. Santos and M.P. Shanahan, '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*, (2001).

[15] M. Shanahan, 'Prediction is deduction but explanation is abduction', in *Proc. of IJCAI*, (1989).

[16] M. Shanahan, 'Noise and the common sense informatic situation', in *Proc. of AAAI 96*, pp. 1098–1103, (1996).

[17] M. Shanahan, *Solving the frame problem*, The MIT press, 1997.

[18] M. Shanahan, 'A logical account of perception incorporating feedback and expectation', in *Proc. of KR*, (2002).

[19] *Spatial and Temporal Reasoning*, ed., Oliviero Stock, Kluwer Academic Publishers, 1997.

[20] Johan van Benthem, *The Logic of Time: a model-theoretic investigation into the varieties of temporal ontology and temporal discourse*, Kluwer Academic, 1991.