From Stereoscopic Vision to Symbolic Representation

Paulo E. Santos* and Murray P. Shanahan
{p.santos, m.shanahan}@ic.ac.uk
Intelligent and Interactive Systems Group
Imperial College of Science, Technology and Medicine
University of London
Exhibition Road SW7 2AZ London

Abstract

This paper describes a symbolic representation of the data provided by the stereo-vision system of a mobile robot. The representation proposed is constructed from qualitative descriptions of transitions in the depth information given by the stereograms. The goal of this paper is to investigate the consequences in the depth transitions of qualitative spatial events such as the relative motion between objects, spatial occlusion and object's deformation (expansion and contraction). This causal relationship between spatial events and depth transitions forms a background theory for an abductive process of sensor data assimilation.

Introduction

One of the most important issues in mobile robot navigation is the connection between the physical world and the internal states of the robot. The term "anchoring" (Coradeschi & Saffiotti 2000), refers to this process of assigning abstract symbols to real sensor data, and is related to the symbold grounding problem (Harnard 1990). There are three main possible approaches to this problem, either this connection is explicitly made via a knowledge representation theory, such as (Shanahan 1997)(Hahnel, Burgard, & Lakemeyer 1998), or it is hidden in the layers of neural-networks (Racz & Dubrawski 1995) or, finally, there can be no representation of the robot's internal states whatsoever, as is the case with purely reactive systems (Brooks 1991a)(Brooks 1991b). Although the last two frameworks give important solutions to certain sets of problems, they are insufficient for the construction of task planning systems using sensor information. For this reason the present work assumes a knowledge representation view of the connection between sensor information and abstract symbols.

Briefly, the goal of the present work is to develop a logic-based system for scene interpretation using dynamic qualitative spatial notions. Therefore, not only is anchoring a central topic in this work, but the scene interpretation issues discussed in this paper are also going to contribute to the understanding of how to track and predict future states of symbolic representations of objects in a robot's knowledge base.

This paper assumes a stereo vision system embedded in a mobile robot as the source of data about the world. A symbolic representation of the sensor data from the vision system is constructed assuming a simplification of the stereograms, a one-dimensional view of the scene (*horizontal cut*). The depths and edges of the bright contrasting regions in the stereogram noted in this view are, then, turned into two-dimensional charts, named *depth profiles*.

Peaks in depth profiles are connected to *sensed objects* (e.g. physical bodies, occluding physical bodies, light reflexion, sensor noise). Each depth profile is, thus, a snapshot of the robot's environment as noted by the its sensors at a time instant. An example of the output of the vision system, a horizontal cut and the corresponding depth profile is shown in figure 1.

In this paper we are interested in explaining incoming sensor data by hypothesising the existence and dynamic relationships between physical objects (and between physical objects and the observer). The incoming sensor data is assumed to take the form of temporally ordered sequences of depth profiles. This process of considering hypotheses to interpret a sequence of sensor information recalls the sensor data assimilation as abduction proposed in (Shanahan 1996). In fact the inspiration for the present work was to propose a new qualitative background theory about space for this framework. In this paper, the background theory about space takes into account dynamical relations, thus the notion of time plays an important role in this theory.

There are many approaches to qualitative theories for time and for space. Van Benthen (van Benthen 1991) presents a discussion about the main problems on defining an ontology for time. A comprehensive overview of the field of spatial reasoning is presented in (Stock 1997). Although, theories about qualitative space and time address important problems, most of them are not concerned with central issues in robotics such as how space is related to real sensor data, nor are there any concern with respect to predicting sensor data as events happen in the world. The framework presented in this paper introduces some ideas to overcome these limitations.

In terms of qualitative theories of space, particularly relevant to this work is the Region Occlusion Calculus (*ROC*)

*Supported by CAPES.

(Randell, Witkowski, & Shanahan 2001). The *ROC* is a first-order axiomatisation that can be used to model spatial occlusion of arbitrary shaped objects. The formal theory is defined on the degree of connectivity between regions representing two-dimensional images of bodies as seen from some viewpoint. This paper presents a dynamic characterization of occlusion *as noted by a mobile robot's sensors*, which could extend *ROC* by introducing time into its ontology.

Stereo-Vision System

The stereo-vision system used in this work is Small Vision System (*SVS*), which is an efficient implementation of an area correlation algorithm for computing range from stereo images. The images are delivered by a stereo head *STH-MD1*, which is a digital device with megapixel imagers that uses the 1394 bus (*FireWire*) to deliver high-resolution, real time stereo imagery to any PC equipped with a 1394 interface card [1].

Figure 1 shows a pair of input images, the resultant stereo disparity image from the vision system and a depth profile of a scene.

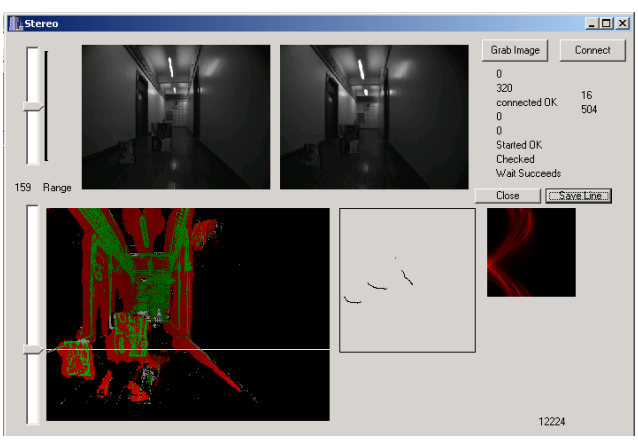

Figure 1: Images of a scene (top pictures); stereogram given by *SVS* (bottom left picture); and depth profile of a scene's horizontal cut (bottom centre image).

Depth Profiles

In order to transform the stereograms from the vision system into a high-level interpretation of the world, a simplification of the initial data is considered. Instead of assuming the whole stereogram of a scene, only a horizontal line of it (*horizontal cut*) is analysed. Each horizontal cut is represented by a two-dimensional chart (*depth profile*) relating the *depths* and the extreme *edges* of bright contrasting regions in the portion of the stereogram crossed by the cut. An example of a depth profile is shown in figure 2b.

Depth profiles can be interpreted as snapshots of the environment, each profile taken at a time instant. Thus, a sequence of profiles is a sequence of temporally ordered snapshots of the world. For brevity, assume throughout this paper that a profile P_i represents the snapshot taken at the time point t_i, for $i \in N$.

Information about the objects in the world, as noted by the robot's sensors, is then encoded as peaks (*depth peaks*) in depth profiles (figure 2b).

The variables *edge* and *depth* in the depth profiles range over the positive real numbers. The edges of a peak in a profile represent the image boundaries of the visible objects in a horizontal cut.

The depth of a peak is assumed to be a measure of the relative distance between the observer (robot) and the objects in its field of view. Further investigation should be done in order to extend this notion to take into account the depths in the objects' shape. This would lead to the treatment of the peak's format as an additional information in the theory.

The axis *depth* in a profile is constrained by the furthest point that can be noted by the robot's sensors, this limiting distance is represented by L in the charts (figure 2b). The value of L is, thus, determined by the specification of the robot's sensors[2].

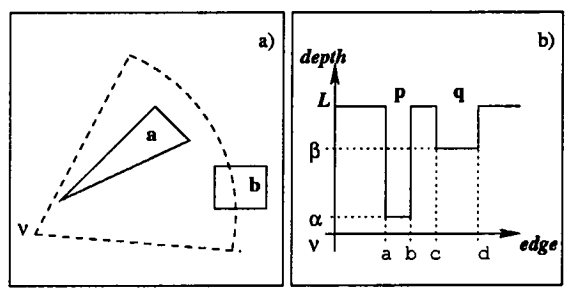

Figure 2: a) Two objects a, b noted from a robot's viewpoint v; b) Depth profile (relative to the viewpoint v in a) representing the objects a and b respectively by the peaks p and q.

An important information embedded in the depth profiles is the *size* of the peaks. The size of a peak is given by the modulo of the subtraction between the values of its *edges* (e.g. in figure 2b the size of the peak q is defined by $|c-d|$). Thus, the size of a peak is connected to the angular distance between two objects in the world with respect to a viewpoint.

Differences in the sizes (and/or depths) of peaks in the same profile and transitions of the size (depth) of a peak in a sequence of profiles encode information about dynamical relations between objects in the world and between the objects and the observer. The main goal of this work is to develop methods to extract this information for scene interpretation.

It is worthy pointing out that, for the purposes of this work, the variables *depth* and *edge* do not necessarily have to represent precise measurements of regions in a horizontal cut of a scene, but approximate values of these quantities within a predefined range. However, the values of these

[1]Both *SVS* and *STH-MD1* were developed in the SRI laboratories, http://www.ai.sri.com/ konolige/svs/index.htm.

[2]The practical limit (L) for depth perception in the apparatus available for this work is about 5 meters, readings further than this has shown to be very unreliable.

variables should be accurate enough to represent the order of magnitude of the actual measurements of depth and size of objects in the world *from the robot's perspective.*

In this work, depth profiles are assumed to be the only information available to the robot about its environment. The purpose of this assumption is to permit an analysis of how much information can be extracted from the vision system alone, making precise to which extent this information is ambiguous and how the assumption of new data (such as robot's proprioception) can be used to disambiguate it.

Motivating Examples

In order to motivate the development of a theory about depth information two examples are presented.

Example 1: Approaching an Object Assuming the two profiles in figure 3 as constituting a sequence of snapshots of the world taken from the viewpoint of one observer (ν). Common sense makes us suppose that there is an object in the world represented by the peak p which, either:

1. *expanded* its size, or
2. was *approached by* the observer, or
3. is *approaching* the observer.

It is worthy pointing out that any combination of the previous three hypothesis would also be a plausible explanation for this profile sequence. The choice about the best set of hypotheses to consider should follow a model of priority, which is not developed in the present paper.

The transition in the profile sequence that suggests these hypotheses account for the decrease in the relative depth of the peak p from $P1$ to $P2$. The task of a reasoning system, in this domain, would be to hypothesise similar facts given sequences of depth profiles.

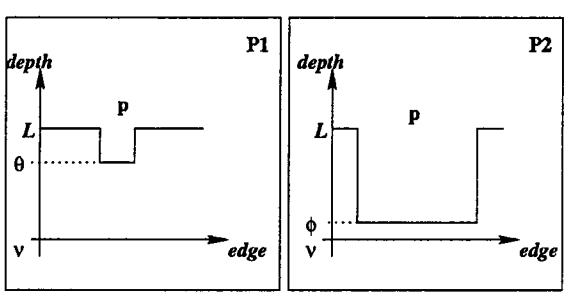

Figure 3: Depth profile sequence P1, P2.

There are other plausible hypothesis to explain this profile sequence that were not considered in this example. For instance, it could be supposed that in the interval after $P1$ and before $P2$ a different object was placed in front of the robot's sensors that totally occludes the object represented in $P1$. This is not a weakness of the present work, but an example of its basic assumption concerning the incompleteness of information. For the sake of this simple example it might have being possible to enumerate all of the possible hypothesis to explain the scenario in figure 3, however, in a more complex situation the amount of hypothesis and their logical descriptions would be too extensive to be taken into account completely. Thus, the strategy of the reasoning system is not to always providing a correct answer, but to conclude a *possibly correct* interpretation of the sensor transition. If this interpretation turns out to be false, another hypothesis should be searched for.

Example 2: Occlusion between two objects Figure 4 is another example of a depth profile sequence. In this case the profiles contain two peaks, p and q, possibly representing two different objects in the observer's field of view.

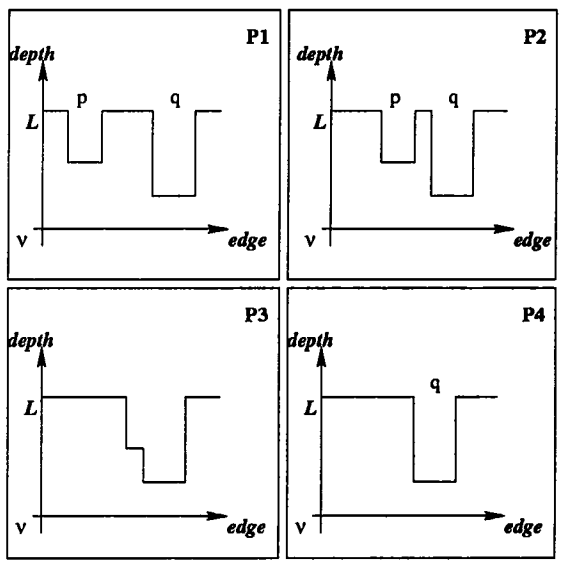

Figure 4: A depth profile sequence $P1$, $P2$, $P3$ and $P4$.

Intuitively, analysing the profile sequence in figure 4, the transition from $P1$ to $P2$ suggests that either the two objects represented by p and q were getting closer to each other or the observer was circling around p and q in the direction of decreasing the observer-relative distance between these two objects.

Moreover, from $P2$ to $P3$ it is possible to suppose that the object represented by q and p are in a relation of occlusion, since the peaks p and q in $P2$ became one peak in $P3$. Likewise, from $P2, P3, P4$, *totally occlusion* between two objects can be entailed, since the peaks p and q changed from two distinct peaks in $P2$ to one peak in $P3$ (which is a composition of the peaks in $P2$), and eventually to one single peak (in $P4$).

In the next section these notions are made more rigorous.

Peaks, Attributes and Predicate Grounding

In the present section a formal language is defined to describe depth profiles transitions in terms of relations between the objects of the world and the observer's viewpoint.

The theory's universe of discourse includes sorts for time points, depth, size, peaks, physical bodies and viewpoints. *Time points, depth* and *size* are variables that range over positive real numbers (R^+), *peaks* are variables for depth peaks,

physical bodies are variables for objects of the world, *viewpoints* are points in R^3. A many-sorted first-order logic is assumed in this work.

The relation $peak\_of(p, b, v)$ represents that there is a peak p assigned to the physical body b with respect to the viewpoint v. This relation plays a similar role of the *predicate grounding relation* defined in (Coradeschi & Saffiotti 2000).

A peak assigned to an object is associated to the attributes of depth and size of that object's image. In order to explicitly refer to particular attributes of a peak at a determined time instant t, two functions are defined:

- **depth**: *peak* x *time point* → *depth* that gives the peak's depth at a time instant; and
- **size**: *peak* x *time point* → *size*, which maps a peak and a time point to the peak's angular size.

Besides the information about individual peaks, information about the relative distance between two peaks is also important when analysing sequences of depth profiles. This data is obtained by introducing the function **dist** described below.

- **dist**: *peak* x *peak* x *timepoint* → *size*, maps two peaks and a time point to the angular distance separating the peaks in that instant. This angular distance is the modulo of the subtraction of the nearest extreme edges between two peak (e.g. in figure 2b the distance between peak p and peak q is given by $|b - c|$).

It is worthy pointing out that the depth profiles of a stereogram may contain peaks representing not only real physical objects of the world but also *illusions of objects* (e.g. shadows and reflexes) and sensor noise. This kind of data cannot, in most of the cases, easily be distinguished from the data about real objects in the robot's environment. In order to cope with this problem, the framework presented in this paper assumes, as a rule of conjecture, that *every peak in a profile represents a physical body unless there is reason to believe in the contrary.* In other words, this is the assumption that the sensors (by default) note only physical bodies in the world, illusions (and noise) of objects are exceptions. One way to state this default rule formally is by the use of circumscription in a similar way as presented in (Shanahan 1996) and (Shanahan 1997). A complete solution to this problem, however, is a subject for future research.

Relations and Axioms

Assuming that the symbols a and b represent physical bodies, ν represents the observer's viewpoint, and t represents a time point, the following relations are investigated in this paper:

1. $being\_approached\_by(\nu, a, t)$, read as "the viewpoint ν is being approached by the object a at time t";
2. $is\_distancing\_from(a, \nu, t)$, read as "the object a is distancing from the viewpoint ν at time t";
3. $approaching(\nu, a, t)$, read as "the viewpoint ν is approaching the object a at time t";
4. $receding(\nu, a, t)$, read as "the viewpoint ν is receding from the object a at time t";
5. $expanding(a, \nu, t)$, read as "the object a is expanding from the viewpoint ν a at t";
6. $contracting(a, \nu, t)$, read as "the object a is contracting from the viewpoint ν a at t";
7. $getting\_closer(a, b, \nu, t)$, read as "the objects a and b are getting closer from each other with respect to ν at t";
8. $getting\_further(a, b, \nu, t)$, read as "the objects a and b are getting further from each other with respect to ν at t";
9. $circling_{<}(\nu, a, b, t)$, read as "$\nu$ is circling around objects a and b at t in the direction that the distance between their images is decreasing";
10. $circling_{>}(\nu, a, b, t)$, read as "$\nu$ is circling around objects a and b at t in the direction that the distance between their images is increasing";
11. $occluding(a, b, \nu, t)$, read as "a is occluding b from ν at time t"; and finally
12. $totally\_occluding(a, b, \nu, t)$, read as "$a$ is totally occluding b from ν at time t";

These relations are hypotheses assumed to be possible explanations for transitions in the attributes of a peak (or set of peaks). The formulae [A1] to [A6] below express the connections between the relations 1. to 12. stated above and the transitions in the peaks' attributes.

$$
\begin{aligned}
[A1] \quad & \forall a \nu t \; being\_approached\_by(\nu, a, t) \vee \\
& approaching(\nu, a, t) \vee expanding(a, \nu, t) \leftarrow \\
& \exists p t_1 t_2 (t_1 < t) \wedge (t < t_2) \wedge \\
& peak\_of(p, a, \nu) \wedge (depth(p, t_1) < depth(p, t_2))
\end{aligned}
$$

The formula [A1] states that if there exists a peak p representing the object a from the viewpoint ν such that its *depth* decreases through a sequence of profiles, then it is the case that either the observer ν is *being approached by* the object, or the observer is *approaching* the object, or the object is *expanding* itself.

The relations $is\_distancing\_from/4$, $receding/4$ and $contracting/3$ can be similarly stated. As shown in the formula [A2].

$$
\begin{aligned}
[A2] \quad & \forall a \nu t \; is\_distancing\_from(a, \nu, t) \vee \\
& receding(\nu, a, t) \vee contracting(a, \nu, t) \leftarrow \\
& \exists p t_1 t_2 (t_1 < t) \wedge (t < t_2) \wedge \\
& peak\_of(p, a, \nu) \wedge (depth(p, t_1) > depth(p, t_2))
\end{aligned}
$$

Likewise, the event of two bodies *getting closer* to each other (or the observer moving around the objects) at an instant t follows as hypothesis to the fact that the angular distance between the respective depth peaks to the two objects decreases through time. The formula [A3] below states these facts. The opposite relation, *getting further*, is stated by the formula [A4].

$$
\begin{aligned}
[A3] \quad & \forall a b \nu t \; [getting\_closer(a, b, \nu, t) \vee \\
& circling_{<}(\nu, a, b, t)] \wedge (a \neq b) \leftarrow \\
& \exists p q t_1 t_2 (t_1 < t) \wedge (t < t_2) \wedge peak\_of(p, a, \nu) \wedge \\
& peak\_of(q, b, \nu) \wedge (dist(p, q, t_2) < dist(p, q, t_1))
\end{aligned}
$$

$$[A4] \quad \forall abvt \; [getting\_further(a,b,\nu,t) \vee circling_{>}(\nu,a,b,t)] \wedge (a \neq b) \leftarrow \exists pqt_1t_2(t_1 < t) \wedge (t < t_2) \wedge peak\_of(p,a,\nu) \wedge peak\_of(q,b,\nu) \wedge (dist(p,q,t_2) > dist(p,q,t_1))$$

Assume the relation $connected(p,q,t)$, read as "peak p is connected to peak q at time t"; $connected(p,q,t)$ holds if the distance between the peaks p and q at time t is zero.

$$\forall pqt \; connected(p,q,t) \equiv (dist(p,q,t) = 0)$$

Note that the previous relation is only defined for peaks, it has no relation to the degree of connectivity between two physical bodies. In this article the relation $connected/3$ is only used to characterise the $occluding/4$ relation, as stated in formula $[A5]$ below.

$$[A5] \quad \forall abvt \; occluding(a,b,\nu,t) \wedge (a \neq b) \leftarrow \exists pqt_1(t_1 < t) \wedge peak\_of(p,a,\nu) \wedge peak\_of(q,b,\nu) \wedge \neg connected(p,q,t_1) \wedge connected(p,q,t) \wedge [size(q,t) < size(q,t_1) \vee size(p,t) < size(p,t_1)]$$

The formula $[A5]$ states that one of the objects a and b is *occluding* the other at time instant t, if the peaks p and q (connected at t) are not connected at some time point before t and the size of one of the peaks is shrinking through the given profile sequence.

Similarly to $occluding/4$, if at an instant t_1 ($t_1 < t$) it was the case that $occluding(a,b,\nu,t)$, and at t the peak representing b disappears from the profile (i.e. $size(q,t) = 0$) then one of the objects is totally occluding the other. As represented in formula $[A6]$.

$$[A6] \quad \forall abvt \; totally\_occluding(a,b,\nu,t) \wedge (a \neq b) \leftarrow \exists t_1 \forall q(t_1 < t) \wedge occluding(a,b,\nu,t_1) \wedge peak\_of(q,b,\nu) \wedge (size(q,t) = 0)$$

The formulae $[A1]$ to $[A6]$ represent a set of *weak connections* between depth peak transitions and dynamic relations between objects and between the observer and then objects. By weak connections we mean that some (but not all) of the hypotheses to explain the sensor data are represented within the formulae. This set of hypotheses, however, does not exhaust all of the possible configurations of the objects as noted by each sequence of snapshots (as briefly discussed in the example 1 above). The purpose of this incomplete set of hypotheses is, initially, to give a *hint* about what possibly could be a high-level explanation for the sensor data this high-level interpretation could then be used by a planning system in order to give a possibly executable plan.

As a matter of fact, there are two ways of reading the formulae above. Firstly, if the right-hand side of one of these formulae unifies with a logic description of the sensor data, then the predicates in the left-hand side of it should be inferred. Due to the fact that the set of hypotheses is not complete, this inference process may lead to false conclusion, that can be reviewed if new data is available. For this reason we assume abduction as the inference rule. A second way of reading the formulae above is assuming the predicates in the left-hand side as actions to be manipulated by a planning system. In this case, the formulae $[A1]$ to $[A6]$ can be understood as rewriting rules for these actions in terms of the expected sensor transition.

As the present paper is concerned with the first interpretation, sensor data assimilation via an abductive process is briefly discussed in the next section.

Sensor Data Assimilation

As already mentioned, this work follows a similar approach to that proposed in (Shanahan 1996), the task of sensor data assimilation of a stream of sensor data is considered to be an abductive process on a background theory about the world.

Basically, abduction is the process of adopting hypotheses to explain a set of facts according to a knowledge theory. Given a set of formulae about a certain field of knowledge and a set of observations, the task of the abductive inference is to *suppose* the causes of these observations from the knowledge available. Abduction, thus, is a non-monotonic inference (Boutilier & Becher 1995) since the explanation inferred from the available knowledge may come out to be false provided new information is acquired.

Taking into account the formal system described in the previous section, the task of the abductive process is to infer the conclusions of formulae $[A1]$ to $[A6]$ as hypotheses, given a description (observation) of the sensor data in terms of depth peaks transitions.

More formally, assuming that Σ is a background theory comprising formulae $[A1]$ to $[A6]$ and Ψ a description (in terms of depth profiles) of a sequence of stereo images, the task is to find a set of formulae Δ, such that

$$\Sigma, \Delta \mid\!\sim \Psi, \tag{1}$$

where the symbol $\mid\!\sim$ refers to the non-monotonic logic consequence.

In order to make this clearer, examples 1 and 2 are recalled using the formal language and the concept of abduction presented above.

Examples Revisited

Example 1: Approaching an Object Consider the profile sequence in figure 3. A description of the sensor data described by the transitions between the peak p in $P1$ and $P2$ is the set

$$\Psi = \begin{cases} \exists p \; a \; \nu \; t_1 \; t_2 \quad peak\_of(p,a,\nu) \wedge \\ (size(p,t_1) < size(p,t_2)) \wedge \\ (depth(p,t_1) > depth(p,t_2)) \end{cases}$$

Where $peak\_of(p,a,\nu)$ is the result of the default assumption that every peak represents a physical body.

¿From Ψ and the formula $[A1]$, abduction infers that $being\_approached\_by(\nu,a,t)$ or $approaching(\nu,a,t)$ or $expanding(a,\nu,t)$ for some time point $t \in [t_1,\ldots,t_2]$. Considering the formula (1) in the previous section, the set $\{being\_approached\_by(\nu,a,t),\ approaching(\nu,a,t),\ expanding(a,\nu,t)\}$ constitutes the set of explanations Δ.

Example 2: Occlusion between two objects Similarly to the example 1, a description of the profile sequence represented in figure 4 can be written as the set

$$\Psi' = \begin{cases} \exists a\ b\ o_1\ o_2\ t_1\ t_2\ t_3\ t_4\ \nu\ \ t_1 < t_2 \wedge t_2 < t_3 \wedge \\ t3 < t4 \wedge peak\_of(a, o_1, \nu) \wedge peak\_of(b, o_2, \nu) \wedge \\ (dist(a, b, t_1) > dist(a, b, t_2)) \wedge \neg connected(a, b, t_1) \\ \wedge \neg connected(a, b, t_2) \wedge connected(a, b, t_3) \wedge \\ \forall t \in [t_1, t_4]\ \ (depth(a, t) > depth(a, t)) \wedge \\ size(a, t_4) = 0 \end{cases}$$

¿From Ψ' and the set of formulae Σ (as defined in the previous section), an abductive explanation for Ψ' would include some of the following relations:

$$\Delta' = \begin{cases} getting\_closer(a, b, \nu, t) \wedge t_1 < t \wedge t < t_2 \\ circling(\nu, a, b, t) \wedge t_1 < t \wedge t < t_2 \\ occluding(a, b, \nu, t_3) \\ totally\_occluding(a, b, \nu, t_4) \end{cases}$$

Image segmentation: future plans

This work assumes a bottom-up approach to image segmentation. The implementation of the system is in its very early stages. Up to now there is an edge detection task followed by a region segmentation (and labeling) step on the stereograms. Therefore, depth profiles are going to be constructed from the vertical edges extracted from the image of the scene.

The central idea is to hypothesise the existence of objects from these image regions, and propose tests in order to verify the hypotheses or to search for more complete explanations *whenever it is needed.* This *necessity* in searching for more complete explanations should be decided by a planning system capable of handling priorities within the hypotheses. This issue, however, is outside the scope of this paper.

Discussion and Open Issues

This paper proposed a symbolic representation of the data from a mobile robot's stereo-vision system using depth information and dynamic notions relative to a viewpoint.

The symbolic representation proposed is based on a horizontal cut of the image given by the stereo-vision system. Transitions in the depth profiles were, then, linked to high-level descriptions of the environment by first-order axioms. The main purpose of this work is to develop a mechanism of inference to conclude hypothesis about the relative motion between objects, spatial occlusion and object's deformation from sensor data transition. This relationship between spatial events and their effects in the depth profile transitions constitutes a background theory for a logic-based abductive process of sensor data assimilation.

The reason for assuming a logic-based approach in this work is to assume a solid common language for visual data interpretation and task planning. Other options for bridging the gap between vision and planning rely on the use of special purpose algorithms or on the development of a common language, instead of first-order logic, to solve this issue. A limitation of the first approach is that it treats visual interpretation and planning as two distinct processes, while the present work assumes that they can be handled by a similar abductive mechanism, as proposed in (Shanahan 1996). On the other hand, by assuming the second option, the development of a common language distinct from logic may have to cope with most problems that were already solved through the long tradition of logic-based systems for A.I.

In this work the simplification of the stereograms in terms of horizontal cuts may seem to be very limiting at first sight. However, it guaranteed the development of a qualitative spatial reasoning theory intrinsically related to sensor information. The core purpose of this ongoing research is to investigate how much of a spatial reasoning theory can be extracted from poor sensor information, concluding from that what kind of robotic tasks can be accomplished within this theory.

Extending this one-dimensional view of the scene to a more general analysis could be done by considering the joint information given by depth profiles of cuts made in other directions of the scene. This analysis may lead to a better notion of individual objects and the degree of connectivity between object. This issue, however, is left for future research.

Another issue that was overlooked by the present paper due to time constraints was the consideration of the information encoded by the format of the depth peaks. Assuming profiles taken from many different points of view around the object, the format of these peaks are intrinsically related to the shape of the bodies, and to the two dimensional region this object occupy in the environment. Thus, the format of the depth peaks could play an important role in the *perceptual signature* of the objects (Coradeschi & Saffiotti 2000) and to theories about *spatial occupancy* and persistence (Shanahan 1995).

A further subject for future research is the improvement of the representational system as more precise and richer visual information is considered.

Future work should also investigate the possible transitions between the relations 1. to 12. (described in the section *Peaks, Attributes and Predicate Grounding*). This information would be useful for constraining the possible explanations in the abductive task of sensor data assimilation.

The consideration of default rules in the theory is another important open problem worthy mentioning.

In the section *Peaks, Attributes and Predicate Grounding*, the connection between the symbolic and the physical representation of the objects in the world is defined by the predicate $peak\_of/3$. The statement $peak\_of(p, a, \nu)$ can be interpreted in terms of anchoring as "the anchor p is assigned to the object a from the viewpoint ν". It is the task of the abductive process to maintain this anchor in time. For instance, recalling the profile $P4$ in the example 2 (figure 4), where only one peak is shown, and supposing that in a sequential profile $P5$ another peak appears next to peak q, the reasoning system should conclude that this new peak is

the peak p (previously involved in the profile sequence $P1$, $P2$, $P3$ and $P4$), maintaining the previously stated connection between p and the object a. This fact can be concluded assuming a default rule about the permanence of objects, another important open issue of this work.

The connection between the depth profile representation and the Region Occlusion Calculus (*ROC*) (Randell, Witkowski, & Shanahan 2001), as mentioned in the introduction of this paper, is also an interesting topic for further investigation. The main contribution of investigating the relationship between the Region Occlusion Calculus and the framework presented in this article is the possibility to extend *ROC* including dynamic sensor information in its ontology.

Acknowledgments

Thanks to David Randell and Mark Witkowski for the invaluable discussions and to the two anonymous reviewers for the useful suggestions for improving this paper.

References

Boutilier, C., and Becher, V. 1995. Abduction as belief revision. *Artificial Intelligence.*

Brooks, R. A. 1991a. Intelligence without reason. A.I. Memo 1293, MIT.

Brooks, R. A. 1991b. Intelligence without representation. *Artificial Intelligence Journal* 47:139–159.

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.

Hahnel, D.; Burgard, W.; and Lakemeyer, G. 1998. *GOLEX bridging the gap between logic (GOLOG) and a real robot.* LNCS. Springer Verlag.

Harnard, S. 1990. The symbol grounding problem. *Physica D* 42:335–346.

Racz, J., and Dubrawski, A. 1995. Artificial neural network for mobile robot topological localization. *Journal of Robotics and Autonomous Systems* 16(1):73–80.

Randell, D.; Witkowski, M.; and Shanahan, M. 2001. From images to bodies: modeling and exploiting spatial occlusion and motion parallax. In *Proceedings of the International Joint Conference on Artificial Intelligence.* To appear.

Shanahan, M. 1995. Default reasoning about spatial occupancy. *Artificial Intelligence* 74(1):147–163.

Shanahan, M. 1996. Robotics and the common sense informatic situation. In *Proceedings of the 12th European Conference on Artificial Intelligence.* John Wiley and Sons.

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

Stock, O., ed. 1997. *Spatial and Temporal Reasoning.* Kluwer Academic Publishers.

van Benthen, J. 1991. *The Logic of Time: a model-theoretic investigation into the varieties of temporal ontology and temporal discource.* Kluwer Academic.