A Logic-based Algorithm for Image Sequence Interpretation and Anchoring

Paulo Santos* and **Murray Shanahan**
Dep. of Electrical and Electronics Engineering
Imperial College, London
{p.santos, m.shanahan}@imperial.ac.uk

Abstract

This paper describes a logic-based framework for interpretation of sequences of scenes captured by a stereo vision system of a mobile robot. An algorithm for anchoring and interpretation of such sequences is also proposed.

1 Introduction

In this work we extend the logic-based spatial reasoning system for scene interpretation proposed in [Santos and Shanahan, 2002] and develop an algorithm that encodes the interpretation process. This algorithm also accounts for the process whereby mappings between logical symbols and sensor data are built up and maintained over time. This is an aspect of the so-called *symbol anchoring* problem [Coradeschi and Saffiotti, 2000].

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 *horizontal slice* of each snapshot. A horizontal slice is, in effect, a 2D *depth profile* of the scene before the robot, taken at a particular height. Within these depth profiles, peaks occur that are caused by nearby objects or collections of objects, these peaks are called *depth peaks*. The size and disparity values of single depth peaks, the distance between pairs of peaks and the transitions that occur in these attributes through consecutive pairs of profiles are the building blocks for our spatial reasoning theory.

Within this framework scene understanding is understood as a process of hypothesising the existence and the dynamic relationships between physical objects (and between physical objects and the observer) assuming temporally ordered sequences of depth profiles. This process recalls sensor data assimilation as abduction first proposed in [Shanahan, 1996]. In fact the initial motivation for the present research was to propose a new qualitative background theory about space-time within this framework by using notions from qualitative spatial reasoning theories such as [Randell *et al.*, 1992].

*Supported by CAPES.

2 Depth Profiles

This work assumes sequences of depth profiles as temporally ordered snapshots of the world. A sketch of a depth profile is shown in Figure 1b.

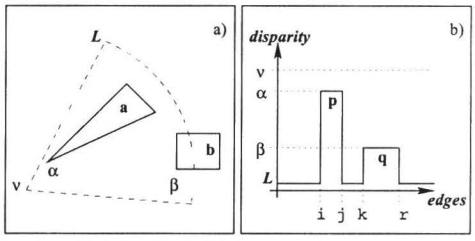

Figure 1: a) Two objects a, b noted from a robot's viewpoint ν; b) Depth profile (relative to the viewpoint ν in a) representing the objects a and b respectively by the peaks p and q.

The axis *disparity* in depth profiles is constrained by the furthest point that can be noted by the robot's sensors, this limiting value is represented by $\boldsymbol{L}$ in these charts (Figure 1b). In fact, $\boldsymbol{L}$ is determined by the specification of the robot's sensors.

Differences in the size (and/or disparity) of peaks and transitions of the size (disparity) 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. This is the initial insight upon which the scene interpretation system proposed in this paper is based. The next section presents some relations that represent transitions in depth profiles.

3 Relations on Depth Profile Transitions

This work assumes a relation $peak_of(p, b, v)$, $p_o(p, b, v)$ for short, representing that there is a peak p assigned to the physical body b with respect to the viewpoint v. In this work we are dealing with the viewpoint of a single robot, therefore we abbreviate $p_o(p, b, v)$ to $p_o(p, b)$. The relation $p_o/2$ plays a similar role of the *predicate grounding relation* defined in [Coradeschi and Saffiotti, 2000].

Assuming that the symbols a and b represent physical bodies, p and q depth peaks, and t a time point,

the following relations are investigated in this paper: $extending(p_o(p,b),t)$, states that the disparity value of a peak p is increasing at time t; $shrinking(p_o(p,b),t)$, states that the disparity value of a peak p is decreasing at time t; $approaching(p_o(p,a),p_o(q,b),t)$, represents that two peaks p and q are approaching each other at time t; $receding(p_o(p,a),p_o(q,b),t)$, states that two peaks p and q are receding each other at time t; $coalescing(p_o(p,a),p_o(q,b),t)$, states that two peaks p and q are coalescing at time t; $splitting(p_o(p,a),p_o(q,b),t)$, states the case of one peak splitting into two distinct peaks p and q at time t. These relations are hypotheses assumed to be possible explanations for transitions in the attributes of a peak (or set of peaks).

Similarly to [Santos and Shanahan, 2002], the relations described above can be connected to descriptions of transitions on sensor data and, further, to relations about changes in the robot environment by means of sets of axioms. Due to space restrictions, however, we do not present axioms for these relations.

Informally, $coalescing(p_o(p,a),p_o(q,b),t)$ and $splitting(p_o(p,a),p_o(q,b),t)$ can be related, respectively, to the event of an object a **occluding** an object b and of a **appearing from behind** b. Therefore, once a hypothesis on peak transition has been obtained, a relative hypothesis on objects in the world can be inferred from the appropriate axiom stating this connection. This is a central idea underlying our solution for anchoring. In fact, this solution assumes processes of explanation and expectation as described in the following sections.

4 Sensor Data Assimilation

Following the ideas proposed in [Shanahan, 1996] and [Santos and Shanahan, 2002], the task of the abductive process for sensor data assimilation is to infer the relations described in Section 3 as hypotheses given a description (observation) of the sensor data in terms of depth peaks transitions. More formally, assuming that Ψ is a description (in terms of depth profiles) of a sequence of stereo images, and Σ is a background theory comprising axioms connecting the relations in Section 3 to sensor data transitions and to changes in the robot's environment, the task of assimilation as abduction is to find a set of formulae Δ, such that

$$\Sigma, \Delta \models \Psi.$$

5 Expectation

Expected peak transitions are suggested by the conceptual neighbourhood diagram (CND) of the relations described in Section 3, shown in Figure 2[1].

Given an abduced explanation for a transition on a pair of peaks, the expected future relations involving these peaks are the neighbours of this transition in the diagram in Figure 2. As this process generates multiple competing expectations, only those that are verified by further sensor data lead to the generation of new predictions. In practice, prediction in this work is, thus, reduced to a table look-up procedure.

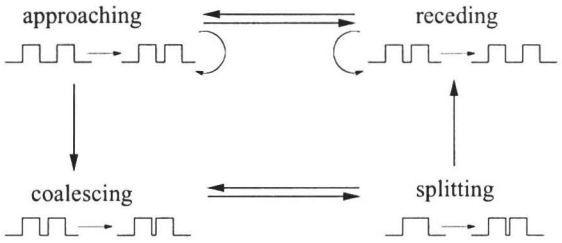

Figure 2: Conceptual neighbourhood diagram.

6 Image segmentation

From a practical point of view, horizontal slices defining depth profiles are comprised of a sequence of measurements made by the considered off-the-shelf vision system. Each horizontal slice is then segmented into *peaks* and *backgrounds* via a simple threshold mechanism on disparity values. This segmentation procedure outputs a sequence of first-order terms of the form: `peak(Lb, Disp, Size)` and `background((Lb1, Lb2), Disp1,Size1)`, where `Lb`, `Lb1` and `Lb2` are variables for $p_o/2$ terms labelling peaks. In particular `Lb1` and `Lb2` are the labels of the peaks bounding the background segment; `Disp` and `Size` are the values of disparity and size of the peak `Lb`; and, `Disp1` and `Size1` are the disparity and size values of the background segment labelled `(Lb1, Lb2)`. If it is the case that a background is bounded by one of the end points of a depth profile (borders of the field of view), a corresponding symbol is assigned to compose its label. In other words, in such cases, either the symbol `borderLeft` or `borderRight` are part of the pair defining the label of a background segment.

Subsequent pairs of horizontal-slice descriptions are input to an algorithm that provides the appropriate interpretation in terms of the dynamic relations discussed in Section 3. This algorithm is summarised in the next section.

7 The anchoring and interpretation algorithm

This section describes an algorithm for interpretation and anchoring, so called *A&I algorithm*, that works by matching `peak` segments in consecutive pairs of profile descriptions that are depicting the same object in the world. Each profile description is comprised of a finite number of segments `peak` and `background`, which are provided by the segmentation routine described in the previous section.

In order to describe the *A&I algorithm* let P_i and P_{i+1} be one consecutive pair of profiles, we assume that P_i has n segments and P_{i+1} has m. In this section we denote a segment r in a profile k by $s_{k,r}$, thus, P_i and P_{i+1} can be denoted as:

$$P_i = \{s_{i,1}, s_{i,2}, \ldots, s_{i,n}\}$$

and

$$P_{i+1} = \{s_{i+1,1}, \ldots, s_{i+1,m}.\}$$

Let $PRED$ and INT be initially empty lists containing, respectively, a set of predictions and a set of interpretations

[1]For brevity, we are omitting from this diagram the relations $extending/2$ and $shrinking/2$.

for the profile pair (P_i, P_{i+1}). We assume also a symbol o denoting a physical body.

For every subsequent pair of profile descriptions (P_i, P_{i+1}), the algorithm's task is to match pairs of segments $(s_{i,j}, s_{i+1,r})$ in (P_i, P_{i+1}) that depict the same object in the world. The segments are considered from left to right in the profile descriptions, starting with the first pair of segments $(s_{i,1}, s_{i+1,1})$. The algorithm is summarised as follows.

Begin: *(A&I algorithm)*

While (P_i and P_{i+1} are non-empty)

1. if $s_{i,j}$ and $s_{i+1,r}$ are `peak` segments; then
   (a) if $s_{i,j}$ is bound to a term $p_o(j, o)$, then assign the term $p_o(r, o)$ to the label of $s_{i+1,r}$;
   (b) else, create two new terms $p_o(j, o)$ and $p_o(r, o)$ and assign them respectively to the labels of $s_{i,j}$ and $s_{i+1,r}$;
   (c) compare the size and disparity measurements in both segments and put into INT the appropriate interpretation according to the relations *extending* and *shrinking*, in Section 3;
   (d) from these interpretations and the conceptual neighbourhood diagram in Figure 2 obtain the predicted predicates and insert them into $PRED$;
   (e) return the next pair of segments: $(s_{i,j+1}, s_{i+1,r+1})$;
2. if $s_{i,j}$ and $s_{i+1,r}$ are `background` segments; then
   (a) unify their labels;
   (b) compare the sizes of $s_{i,j}$ and $s_{i+1,r}$ and consider the difference between these values, which gives the difference in the distance between the two peaks bounding $s_{i,j}$ and $s_{i+1,r}$;
   (c) interpret this difference in terms of the relations *approaching* and *receding*, and insert the appropriate interpretation into INT;
   (d) from this interpretation and the CND in Figure 2 insert the relative predictions into $PRED$;
   (e) return the next pair of segments: $(s_{i,j+1}, s_{i+1,r+1})$;
3. if $s_{i,j}$ is a `background` segment and $s_{i+1,r}$ is a `peak` segment then check in the prediction set $PRED$ whether the two peaks bounding $s_{i,j}$ have been expected to be *coalescing* or *splitting*,
   (a) if so, assume this expectation as explanation for $s_{i,j}$ and $s_{i+1,r}$, inserting it into INT. Execute the appropriate unifications on peak variables and terms $p_o/2$ as explained in step 1a and 1b. Then, obtain the predicted predicates, inserting them into $PRED$. Return the pair $(s_{i,j}, s_{i+1,r+1})$ to be considered for interpretation;
   (b) else explain away the peak $s_{i,j}$ as noise and assume the pair $(s_{i,j+1}, s_{i+1,r})$ for interpretation;
4. if $s_{i,j}$ is a `peak` segment and $s_{i+1,r}$ a `background` segment then proceed analogously to the previous case.

End while.

- If P_i is empty, get the next pair of profile descriptions (P_{i+1}, P_{i+2}).
- else, if P_{i+i} is empty but P_i is not, then explain away the remaining `peak` segments in P_i and consider the next pair of profile descriptions (P_{i+1}, P_{i+2}).

The while loop above is repeated until there are no more depth profile descriptions.

End.

Informally, the algorithm above considers consecutive pairs of depth profile descriptions as lists. Pairs of elements from these lists are compared and a match for `peak` segments is obtained according to the expected transitions in depth profiles. The result of this process is the interpretation of peak transitions, which are related to changes in the objects depicted by peaks.

The algorithm main processing is the while loop that considers, in pairs, every segment in the depth profile descriptions. The running time of this algorithm is, thus, linear on the size of the list containing these descriptions. However, as the number of segments in these lists is related to the number of objects depicted by the considered depth profiles, and there are only a finite number of objects in each scene, the asymptotic upper bound of this algorithm is $\mathcal{O}(1)$.

8 Conclusion

This paper presents a new step towards an abductive framework for sensor data interpretation of sequences of stereovision images obtained by a mobile robot's vision system. We proposed a symbolic representation of the stereovision data based on depth profiles obtained from horizontal slices of snapshots of the world. These profiles encode information about objects in the world as peaks. Each profile was, then, segmented into first-order statements that were input of an algorithm whose task was to unify pairs of segments, providing the interpretation for the occurred transition in depth peaks.

Further research has to consider two main open issues in this framework. First, we have to develop of a qualitative theory about object's shape from multiple horizontal slices of a scene. This theory would enhance the representation of the environment and, therefore, improve the capabilities of the reasoning system. A second issue is the development reasoning modules capable of explaining away noise patches in the image sequences according to common-sense knowledge about the world and about the sensor's limitations.

References

[Coradeschi and Saffiotti, 2000] S. Coradeschi and A. Saffiotti. Anchoring symbols to sensor data: Preliminary report. In *Proc. of AAAI*, pages 129–135, Austin, U.S., 2000.

[Randell *et al.*, 1992] D. Randell, Z. Cui, and A. Cohn. A spatial logic based on regions and connection. In *Proc. of the KR*, pages 165–176, Cambridge, U.S., 1992.

[Santos and Shanahan, 2002] P. Santos and M. Shanahan. Hypothesising object relations from image transitions. In Frank van Harmelen, editor, *Proc. of ECAI*, pages 292–296, Lyon, France, 2002.

[Shanahan, 1996] M. Shanahan. Robotics and the common sense informatic situation. In *Proc. of ECAI*, pages 684–688, Budapest, Hungary, 1996.