Action Based Object Separation with Situated Agents

Christoph Flores

Millstätter Str. 17
D-28359 Bremen, Germany
flores@informatik.uni-bremen.de

Abstract Visual scene analysis can be augmented by proper manipulative actions like changing perspective or interacting with elements of the environment. They reveal some of the non-obvious (e.g. non-visual or visually ambiguous) intrinsic qualities of the scene's setting. A framing for actions is formalizable using situation semantics notification. This supports a non-classic view of perception: the perception of the situation *is* the action taken to inspect it, or is at least inseparably connected to it.

Keywords Action enhanced perception, action based object separation, situatedness

1 Introduction to the Idea of Action Enhanced Perception

The constraints of an environment are inevitably felt during motor actions. Therefore, we can conceive of adding a motoric component to spatial inspection (so that visual input and action feedback are seen as two sources of information that may both be related to the same exploration situation).

This provides the general idea of action enhanced perception. Perception means application of sensors and evaluation of their signals. Action means moving and mechanical manipulation of objects. They can go hand in hand, especially in the sometimes non-trivial task of "detecting objects" (as opposed to "detecting non-objects, or a background").

As long as we provide a rich enough set of actions — like "trying to pass through," "trying to get hold of a part and pull," "trying to move the object to a different position" — even in the case of more complex object constellations they would allow a robot to prove or disprove a hypothesis about objects' boundaries taken from visual analysis.

Visual clues might sometimes be "not rich enough," sometimes "too rich" for a decision. Yet there is good evidence that the fact of the objects being there is the hardest source of a force that cannot be mistaken. In a certain perspective,

Dagstuhl Seminar Proceedings 05491
Spatial Cognition: Specialization and Integration
http://drops.dagstuhl.de/opus/volltexte/2007/1030

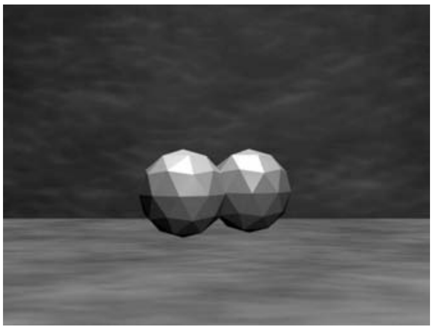


Figure 1. One Object, or Two Objects Overlapping?

looking at things might fail, but interacting cannot fail. If there is a glass wall in my pathway, I might oversee it, but I might not walk right through it.

This is why going for the perception of motoric constraints might be more promising than going for (long-distance) visual input alone — as an example will demonstrate.

2 An Example Setting in a VR Context

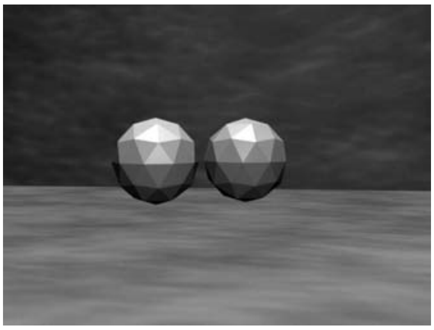
In Fig. 1 we see an artificial robot's view while navigating through an environment.

It is not obvious whether what is present is one big object, like maybe a barbell, or two smaller objects, like two balls close to each other. If it was two objects, from the current perspective they are illusionously linked but they should become discernible if the perspective changes or if the robot gets close enough.

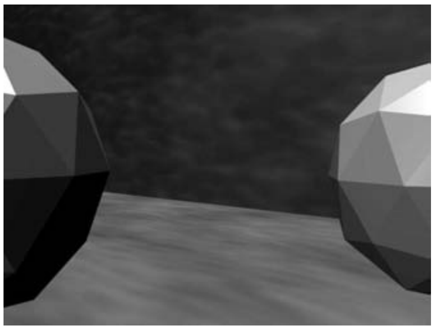
The robot could calculate expectations of possible differences for the two cases (barbell or two balls) and decide to judge the situation better by navigating to an appropriate new point of observation.

Fig. 2 and Fig. 3 show two different examples of how the image of Fig. 1 might change from such a new stand-point. The two different cases comprise empty space between the two ball-like structures (Fig. 2) or non-empty space between them (Fig. 3) like a connecting pipe or a handle.

Apart from becoming visible, this pipe or handle would block the way of the robot if it should try to cross it. Therefore, if still in doubt after visual inspection alone, the robot could carry out additional inspections like trying to move around or in-between the visible structures. (A trajectory for moving around one of the two ball-like structures would include, at some state of the process, moving in between them — which is only possible if they are not linked.)

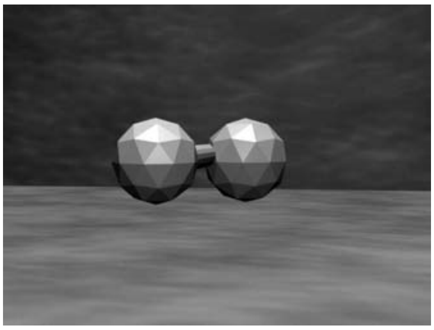


(a) View after Perspective Change: *Two* Objects?

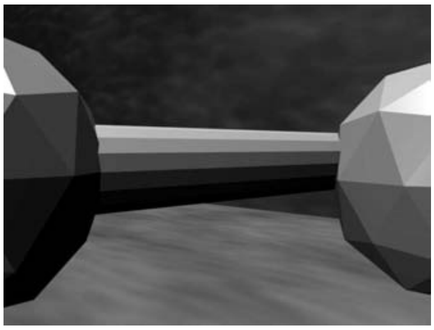


(b) Closing up and Walking through — Success

Figure 2. Two Objects Case



(a) View after Perspective Change: *One* Object?



(b) Closing up and *no* Walking through (but Bumping into One Single Object's connecting Bar)

Figure 3. One Object Case

Any of the proposed actions:

1. trying to change perspective,
2. trying to close up,
3. trying to move around (one) structure,
4. trying to pass in-between (two) structures

would eventually create an enhanced exploration situation which comprises both, action and perception, in a systematically linked way. Actions 3. and 4. do not only change (i.e., enhance) the conditions for visual perception but also provide new channels of input (potentially blocked motor activities may reveal information about the constraints of the environment) — motor activities become part of the exploration situation in which actions enhance perception.

3 The Situation in which Actions take Place has Indeterminates

Application of [Barwise and Perry, 1999] in the field of spatial cognition is, to my knowledge, unprecedented. Yet I know of no better tool to track situational settings as frames of possible actions, integrating the aspect of a self-motivated agent to be both, motivated and constrained by former events in the immediate context of the haecceities of things, here and now, to use a Peircean term. Hence my proposal for application of this powerful set of expressions for situational semantics.

Using the Barwise and Perry notification system, the *situation* of *walking around an object* becomes characterizable by a *universal constraint type* that we can specify as:

$$C := \text{at } l_u : \text{involves, } E, E' \text{ ; yes,} \tag{1}$$

where $E(\dot{\boldsymbol{a}}, \dot{\boldsymbol{b}})$ and $E'(\dot{\boldsymbol{a}})$ are given by

$$E := \text{at } \dot{\boldsymbol{l}}\text{: walk around, } \dot{\boldsymbol{b}}\text{ , } \dot{\boldsymbol{a}}\text{ ; yes}$$

$$E' := \text{at } \dot{\boldsymbol{l}}\text{: object, } \dot{\boldsymbol{a}}\text{ ; yes}$$

The encircling of some structure is the event E. It brings to existence the event E' of something being an object at a position in space-time. Take note that all entities potentially involved are given as indeterminates; the situation's structure is presentable without the necessity of it being factually instantiated at a given position in time and space and by given objects and actors. The situation type provides us with some kind of a frame for understanding, or, if you will, an abstract concept that informs us about what kind of relations have to hold between a certain number of relata if such a situation is to become real.

An agent that instantiates such a situation with its actions, does so by being an actor b walking around some object a somewhere in space. For the purpose of evaluating its own actions and their impact on possible knowledge about its environment, it should understand that it factually instantiates a walking-around-something situation. This allows for further conclusions to be drawn, because if a walking-around-something situation of the classified type eventually appears, this indicates that a thereby-defined position l is the position of an object.

Given the framing of (1), information is still limited. Finding out more about what object it is would need further actions, perhaps a more thorough visual analysis — which again might be action-enhanced (and so on until a desired level of preciseness is reached). The same is true for the position l. It becomes instantiated by a framing in space and time, but it is not further characterized in a specificatory manner that would raise it to some ontological status other than playing an indexical rôle. Although characteristics of the object might help to further classify l, this kind of information is not focussed here, nor are the spatial dimensions of l. They could be revealed, though: the agent could steadily test where it gets stopped on the path, thereby finding out about objects' boundaries.

This would help to spatially define l further. This would also help to avoid mistakes of taking a non-object for an object. The laws that control the encircling actions should ensure that some checking is included like trying to cross the encircled array at various positions.

4 Understanding Situations' Impacts on Environmental Knowledge

Our describing a situation type that is linked to an action in an environment can only be the first step, because what is needed for the action to become effective is a robot's principal understanding of the situation, that is, its knowing that 'being able to walk around a certain something' means 'this something is a separate object.' Such knowledge could be characterized using an indexed constraint:

$$\begin{aligned} &C/(\dot{\boldsymbol{i}},\dot{\boldsymbol{h}}) := \text{at } l_u \text{ involves}/\dot{\boldsymbol{i}},\dot{\boldsymbol{h}},E(\dot{\boldsymbol{i}},\dot{\boldsymbol{h}},\dot{\boldsymbol{b}}),E'(\dot{\boldsymbol{i}},\dot{\boldsymbol{h}}) \text{ ; yes,} \\ &\text{where} \\ &E(\dot{\boldsymbol{i}},\dot{\boldsymbol{h}},\dot{\boldsymbol{b}}) := \text{at } \dot{\boldsymbol{h}}\text{: walk around, } \dot{\boldsymbol{b}}\text{ , } \dot{\boldsymbol{i}}\text{ ; yes} \\ &E'(\dot{\boldsymbol{i}},\dot{\boldsymbol{h}}) := \text{at } \dot{\boldsymbol{h}}\text{: object, } \dot{\boldsymbol{i}}\text{ ; yes} \end{aligned} \tag{2}$$

The advantage of using action types is the reduction in possible relations between input data that need to be foreseen. Instead of trying to compute over an infinite number of possible instantiations for variables like time, place, object-like or subject-like entity (agens or patiens), we define a limited number of general frames of situational relations. For instance, not a certain set of coordinates reveals that there is some object A at position L. Rather, a whole set of possible actions would each describe the fact of some object A (or A′ or A″ or . . .) being at a position L, namely within an area surrounding L. L becomes marked by all actions that successfully circumscribe it, providing an encircling path that ends again at the starting position.

Provided actions are chosen such that they are discriminative enough and revealing enough[1] *they can actually make an existential claim: that there is something, and that it is there, at L, and not not there.*

This entails a concept of "perception" that leads away from classical in-take / notification conceptions. It is action-perception-cycles rather than perception alone that reveal existential information of a surrounding. Potential knowledge of a setting emerges out of situated action.

Despite finding a quite obvious and unspectactular application in the field of (artificial) robotics, action enhanced perception may be a concept with generally high biological plausibility. Thus it may be an attractive field of investigation for biologically inspired research. Lately, thoughts on the relations of action and perception tend to support the non-passive, non-intake view of perception as it is plausible from our situational framing and its interpretation.

[1] We shall discuss an algorithm in the addendum.

This is also in line with current tendencies to mention the importance of motor components in exploration, and to place them aside sensor components. An example of such a view, that at the same time expresses the idea of situatedness, is given by [Garbarini and Adenzato, 2004]: "Acting in the world, interacting with objects and individuals in it, representing the world, perceiving it, categorizing it, and understanding its significance are perhaps simply different levels of the same relational link that exists between organisms and the local environments in which they operate, think and live." (p. 105) The perceiver and actor is "embedded" in the environment and interwoven with it. When we do something, the environment reacts, and that's how we perceive.[2]

5 Conclusion

We concentrated on working out the general idea that revealing actions can help in the exploration of an unknown environment. A robot can associate explorative actions, if they are successful, as well as if they are not successful, with knowledge about the situation at the inspected area. If a certain action is successful (sometimes in connection with other actions being successful or non-successful), this tells something about it and the area linked to it. If it is not successful (sometimes in connection with other actions being successful or non-successful), this tells something else about this area. In a certain perspective, we are justified to say that the perception of the situation is prominently connected to the action taken to inspect it.

[2] This view has lately become famous albeit being quite obvious. It had simply been forgotten in the areas dominated by "sense data" theories, and, no contradiction, behaviorism — which, to use Kantian terms, dispensed with *spontaneous intentions* and vivid, speculative *vorstellungen*.

Addendum

Discriminative actions need to be planned. Qualitatively speaking, we have to take into account an element of time, because encircling an object includes starting at one point in space, and coming back to this same point later in time. Conceptually, we can replace a reference to time by including qualitative characterizations such as "smooth / continuous movement," but defining these themselves in quantitative terms would include introduction of an element of time (unless we define them only geometrically, but then again time would inherently be present in the indication of a *sequence* of paths). We define an area that is circumscribed by a trajectory of a point that is at location l at time t_0 and that is not at location l but at location l' at time t_1 and that is neither at l nor at l' at time t_2, and again at location l at time t_3. In between two time points t_1 and t_3 there is some following of a trajectory that is a closed line without intersections. The encircled area might contain an object.

We must be careful not to try to define (one) quantitatively precisely described encircling route, though, because it might not match with the boundaries or the size of a given object. Again, we need a qualitative framing that will allow us to identify an encircling, yet also allow us to adapt to differing object boundaries. We could either use an implementation oriented framing — which might ipso facto be more restricted concerning the shape of the encircling movement — or a more general, more abstract qualitative description. An appropriate example for an implementation oriented framing is given by [Dylla and Moratz, 2005] in a notification that characterizes "going round the Kaaba." It formalizes the encircling of a rectangle monument with four solid walls that determine the four sub-routes to travel along. In this case, at a given moment, knowledge of two walls meeting in a corner is sufficient. The four-cornered object, describable as

$$\mathrm{R0}(\mathit{errs}\text{–})\mathrm{R1} \wedge \mathrm{R1}(\mathit{errs}\text{–})\mathrm{R2} \wedge \mathrm{R2}(\mathit{errs}\text{–})\mathrm{R3} \wedge \mathrm{R3}(\mathit{errs}\text{–})\mathrm{R0},$$

will be exploited in parts, for each corner. An appropriate example for a more abstract qualitative description is given in the work of [Whege, 2005]. If we take elements of his Qualitative Trajectory Calculus, using a sub-variant for two objects moving relative to another, and set one of the objects as at least temporarily non-moving (fixed in space), we can define an encircling action such that it satisfies a QTC movement pattern. An appropriate QTC criterion helps to discriminate it from other types of movement which would not be an encircling (but something else instead). In our case, our wanted encircling action will have to follow a route that satisfies the condition $\{0\ 0\ \mathrm{C}\}_{\mathrm{B22}}$. We choose a number of consecutive movements, which we can assume here as being linear (we can adapt to the shape of any curve using many lines, leading to up to infinitesimally small segments). If the conceived movement cannot be carried out (unsuccessful encircling), it can be changed step by step, by keeping to the criterion of remaining a closed movement in the end (if possible). So if our robot, while carrying out the linear movements satisfying $\{0\ 0\ \mathrm{C}\}_{\mathrm{B22}}$, becomes stopped by an object, it could back-trace to the beginning and start again with longer distances (that is, less steps for one full encircling action), which would give its route a different shape, or it could try to back-trace always to the last step and try prolonging it in a straight line, or omit it instead and go on to the next. In order to keep track of the error thereby introduced, for the reason of being able to satisfy a criterion of closedness (so that the robot will get back to its starting point after travelling the full route), it would have to try to reverse the former deviation from a standard route. It could try to introduce such a reverse step at a position in space that would be on a line, thought to go from the robot's position at the time of the first deviation, through to the estimated center of mass of the to-be-encircled object. If that fails, other positions might be tried. We could think of trying to insert the correction movement at the end of the encircling action, and to shift it to an earlier position if there is a blocking at that end position. An algorithm for the encircling action "walk around" as a concept of our formula (1) would then be according to algorithm A (see next page).

After carrying out this algorithm, the success flag will tell about whether encircling was possible or not. If encircling failed, one strategy might be to increase n (the number of steps to be calculated), or to vary the starting angle of the beginning line segment.

The rule for understanding the just carried out action according to the demands of our formula (2) would be: "If after action: finish flag is set, then encircling was successful."

If no back-tracing was necessary in the sequence (because the robot nowhere got stopped on its way), we do not yet know whether there is an object in the encircled area, or just empty space. So for to test whether there really is an object in the encircled area, a test for being blocked by it has to be made in all cases in which after carrying out the above steps, m is equal to n, for this indicates that no blocking state was reached while trying to encircle the area, hence no object was yet detected on the path. The testing for an object could be ensured by algorithm B (see next page)

If after finishing the processing of this algorithm the success flag is set, then the encircled area contained an object, else not (as long as n was chosen big enough).

It is not important for our purposes whether we have found a very effective or a very thorough algorithm. Indeed, we could think of better ways for planning what steps to carry out in an attempted encircling action. In the algorithms presented above, (A) 6. would for example need a more precise formalization. One possibility for enhancement might be to use proposals from robotic navigation research. To give but one example, for computing an encircling trajectory for rectilinear polygons, [Isler, 2001] introduces an algorithm that, on the basis of such a path, computes the encircling.

Algorithm A

1. Choose a starting point outside of the to-be-encircled structure. Set m and retry counter to 0. Reset finish flag and success flag.
2. Calculate equally distributed straight line trajectories for n steps to go around the structure (the trajectory will be attempted to be carried out and modified on the fly if deserved).
3. While $m \leq n$ and retry counter $< n$:
   (a) carry out m^{th} step (calculated in 2.). If successful increment m, else:
      i. back-trace and prolong last step by one segment of the same length and in the same direction. If successful, increment m else back-trace and carry out $(n+1)^{\text{th}}$ step (calculated in 2.). If successful, increment m else increment retry counter.
4. If $m = n$ set success flag.
5. If $m = n$ or if finish flag set, stop.
6. Try to find path that leads back to the beginning by proposing a maximum number of diagonals with corresponding angles, and a final horizontal or vertical line, if necessary, calculating (m - retry counter) steps like in 2, replacing the old proposals. Set $m :=$ retry counter. Set finish flag. Goto 3.

Algorithm B

1. Take starting point from above. As in (A), calculate straight line trajectories with n steps to go around the structure. Set m to 1 and counter to 0. Reset success flag.
2. Omit m^{th} element of sequence.
3. While counter $< n$:
4. Try to carry out step indexed by counter.
5. If that fails, set success flag[3], else if $m \leq n$ increment m and goto 3.
6. Stop.

References

[Barwise and Perry, 1999] Barwise, J. and Perry, J. (1999). *Situations and attitudes.* CSLI Publications Stanford, CA. enh. edn.

[Dylla and Moratz, 2005] Dylla, F. and Moratz, R. (2005). Exploiting Qualitative Spatial Neighborhoods in the Situation Calculus. In *Spatial Cognition IV Reasoning, Action, Interaction*, volume 3343 of *Lecture Notes in Artificial Intelligence*, pages 304–322. Springer, Berlin.

[Garbarini and Adenzato, 2004] Garbarini, F. and Adenzato, M. (2004). At the root of embodied cognition: Cognitive science meets neurophysiology. *Brain and Cognition*, 56:100–106.

[Isler, 2001] Isler, I. V. (2001). Theoretical Robot Exploration. Written Preliminary Exam, Computer and Information Sciece, University of Pennsylvania.

[Whege, 2005] Whege, van de, N. (2005). *Representing and Reasoning about Moving Objects: A Qualitative Approach.* Ph.d. thesis, Ghent University, Faculty of Sciences, Department of Geography.

[3] In this case, failing to carry out a step means being blocked inside the formerly encircled area, which is needed to detect an object, hence failure at this stage means success in object detection.