Action Vectors for Interpreting Route Descriptions

Junko Araki† **Takashi Ninomiya**‡* **Takaki Makino**¶ **Jun'ichi Tsujii**‡*

† Interfaculty Initiative in Information Studies, University of Tokyo
‡ CREST, JST (Japan Science and Technology Corporation)
* Department of Computer Science, University of Tokyo
¶ Department of Complexity Science and Engineering, University of Tokyo
address: Hongo 7-3-1, Bunkyo-ku, 113-0033, Tokyo phone: +81-3-5803-1697 FAX: +81-3-5802-8872
e-mail: {jun3, ninomi, mak, tsujii}@is.s.u-tokyo.ac.jp
keywords: navigation system, route description, perspective system, human spatial cognition

Abstract

This paper describes a mechanism of interpreting route descriptions in navigation. We introduce a notion of *action vectors*, which are points where an agent took some action according to the given description. In a new perspective system, *a position-centered perspective system*, an agent determines his configuration by selecting one of action vectors as a reference. Ambiguity in interpreting a description is explained by indeterminacy in the selection of an action vector. In our theory, the process of interpretation is formulated as a dynamic system of incrementally updating a set of action vectors.

Introduction

The principal aim of our research is to give an explanation of ambiguity in navigation. We introduce an *action vector* as a possible reference of instructions and show that action vectors can be defined as former positions of the agent where he turned or stopped, etc. according to instructions. By introducing an action vector, the structure of interpreting instructions can be explicitly shown.

Consider the case where an agent travels around the world in Figure 2 according to Description 1 in Figure 1. The agent interprets the instructions such as *'On the right'* or *'Go straight'* in turn, and heads for a destination. We find, however, indeterminism in the choice of *the reference (e.g., the post office)* causes the multiple interpretations of the instructions. For instance, with regard to *'On the right, you can see a flower shop'*, we can have three interpretations depending on the reference of *'On the right'*.

1. On the right of *the exit of the post office*. (Flower shop 1)
2. On the right of *the entrance of the post office*. (Flower shop 2)
3. On the right at *the crossing where he turned left a little while ago*. (Flower shop 3)

In addition, there are three interpretations for 'go straight' for each of the interpretations above. As a result, Description 1 has nine interpretations. We must know the mechanism of ambiguity in route description in order to give an unambiguous, comprehensive description to an agent.

You can see a crossing straight ahead and turn left there. On the left side there is a post office. Get me a stamp there. Then on the right, you can see a flower shop. Don't forget to buy a bouquet of roses for mother's birthday present. And go straight to get to the cafe. I'm waiting there.

Figure 1: Description 1: Instructions to get to the destination

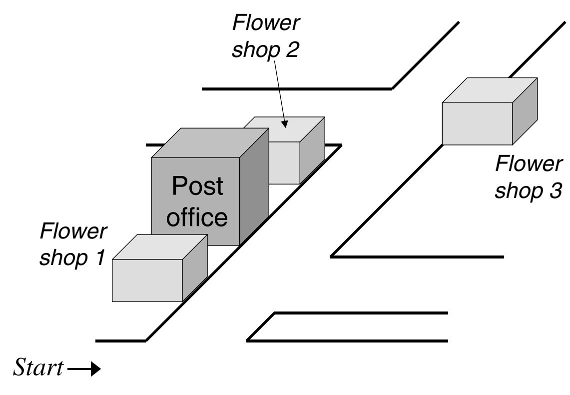

Figure 2: The world where an agent travels according to Description 1.

In former research, especially regarding robotic navigation systems, the planner assumed that the agent always interpreted each instruction in its current position where he stood (Simmons & Koenig 1995). Consequently, the problem of ambiguity of the instructions has not been dealt with. They have focused on developing their techniques, which can be roughly divided into two classes: *keeping track of the agent's position* and *globally estimating the agent's position* (Burgard *et al.* 1998). Thus they have made the navigation plan to let the agent follow each instruction in its current position. But as taking human spatial cognition into consideration, it is desirable to make a system which lets the agent

d_1	*you can see a crossing ... turn left there*
d_2	*on the left side, you can see a post office ... ... get me a stamp there*
d_3	*on the right, you can see a flower shop ... ... buy a bouquet*
d_4	*go straight to get to ...*

Figure 3: Instructions in Description 1

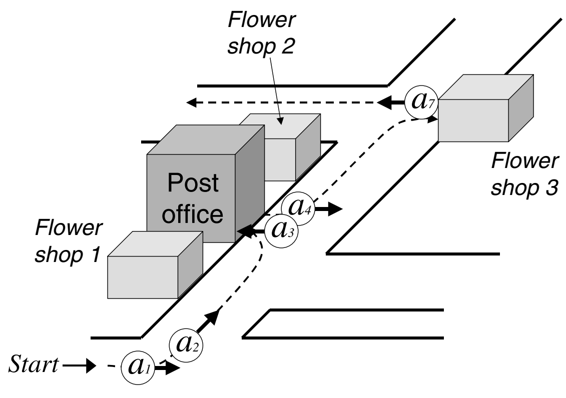

Figure 4: A route and action vectors

interpret the instructions in its former positions such as at the crossing.

In our theory, we suppose the following situations.

1. A planner may not specify the reference object of instructions (e.g., *'On the left'*).
2. An agent interprets a set of instructions in turn, which is given as one route description in advance, i.e., we suppose non real-time navigation.

The action vectors and instructions have inherent relations in navigation. Once the agent chooses an action vector for interpreting the instruction, the interpretation of the instruction is uniquely determined. The case of interest here, the result of interpreting the instruction generates new action vectors, which will be candidates of action vectors for interpreting the next instruction. Namely, the different choices of the action vector make the different contexts of the route description. In our theory, the process of interpreting a route description can be regarded as such a dynamic system.

Action Vector

This section explains an *action vector*, which is a specific point on the route where the agent turned, stopped or perceived the surroundings to fulfill its missions (to buy a stamp or a bouquet). The ambiguities in interpreting the route descriptions are well explained by supposing the action vectors.

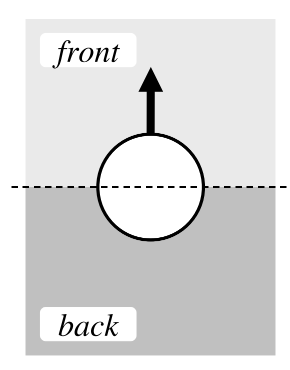

Figure 5: The front-back/right-left regions of the action vector in the position-centered perspective system

To begin with, let $\mathcal{L}$ ($\subset \mathbf{R}^3$) be the set of all possible vectors of positions on the map, $\mathcal{O}$ ($\subset \mathbf{R}^3$) be the set of all possible vectors of orientation, and $\mathcal{C}$ ($= \mathcal{L} \times \mathcal{O}$) be the set of all the possible pairs of a position vector and an orientation vector, which defines the set of action vectors. Let D be the description, which is the sequence of the instructions $d_1 d_2 d_3 \cdots d_n$. Suppose the following situation:

- A planner gives a description D to an agent to let him do some goal-oriented task. For example, Description 1 in Figure 1 is an example of the description.
- At the time t, an agent has executed $d_1 d_2 \cdots d_{t-1}$, and is about to execute d_t.

Let c_t be the context that includes the route and any actions which the agent had taken at the time t, and $actvec$ be the function that takes a context and returns a set of *action vectors*:

$$
\begin{aligned}
actvec(c_t) = \{ \quad & a_{1,1}, a_{1,2}, \cdots, a_{1,m_1} \\
& a_{2,1}, a_{2,2}, \cdots, a_{2,m_2}, \\
& \cdots, \\
& a_{t,1}, a_{t,2}, \cdots a_{t,m_t} \qquad \}
\end{aligned}
$$

where action vectors $a_{i,1}, \cdots, a_{i,m_i}$ correspond to an instruction d_i, and each action vector $a_{i,j}$ ($\in \mathcal{C}$) is a pair of the position where the agent takes an action according to d_i, and the orientation for which the agent is heading according to d_i.

Figure 3 shows the instructions in Description 1. Figure 4 shows an example of the route that the agent follows by reading the instructions (dashed line in the figure) and an example of $actvec(c_3)$ for the route. In the figure, five action vectors are defined: the crossing before and after he turned (a_1 and a_2) corresponding to d_1, the entrance and the exit of the post office (a_3 and a_4) for d_2 and the exit of the flower shop 3 (a_7) for d_3.

Position-centered Perspective System

In cognitive studies, many researchers (Tversky & Hermenway 1984; Herskovits 1989; Levelt 1986; Retz-Schmidt

1988; Olivier 1996; Gapp 1994) have proposed several perspective systems such as deictic (agent-centered) , intrinsic (object-centered), extrinsic and absolute perspective systems. They, however, all deal with only the static directional constraint on the located objects relative to the reference objects, and the question of the dynamic perspective like navigation is still open.

This section explains a new perspective system, *a position-centered perspective system*. The position-centered perspective system is centered at an action vector and adopts the orientation of the action vector. That is, a world is partitioned by the action vector (Figure 5). Thus, by using the orientation of the action vector in the position-centered perspective system, we have new explanations of the spatial relations between objects and the agent.

In the position-centered perspective system, an action vector is regarded as a reference of an instruction. In the former studies, reference positions are regarded as only the current position of the agent. For example, the planner believed that the agent, traveling around the world, followed the given instructions like '*Turn left*' or '*to the right* !¢in turn in his current position where he had finished the preceding instruction. But it is insufficient to suppose only the current position to interpret the instructions. For example, given an instruction '*to the left*', the meaning of it would be regarded as '*to the left of the action vector*'. Consider the situation where the agent facing at the post office heads for the flower shop 3, according to the description 'To the right, there is a flower shop' (Figure 4). The flower shop is located to the right of the post office neither intrinsically nor deictically at the agent's current position. In this way, with this perspective system, several configurations of objects which cannot be explained by other perspective systems are well explained.

Interpreting Route Description

The Planner is often unaware of ambiguous instructions in the route description. There are three factors to cause ambiguity.

1. the choice of the perspective systems
2. the choice of the action vector
3. the choice of the reference objects

In this section we examine interpretations of Description 1 by using the action vector. Consider the world in Figure 2 and Description 1 again.

The agent often confuses where to practice the instruction 'Go straight' or '*to the left*' and '*to the right*'. For example, the agent, standing at the exit of the post office, may interpret '*to the right*' as '*to the right from the crossing he turned a little while ago*', and consequently choose flower shop 3 to buy a bouquet. After he got out from the flower shop 3, he may suddenly turn right without going straight ahead. As we examine possible interpretations of the directives and dimensional prepositions in Description 1, there have been actually nine interpretations.

The agent is located on one of the action vectors actually or imaginably when he/she interprets the instructions. Let

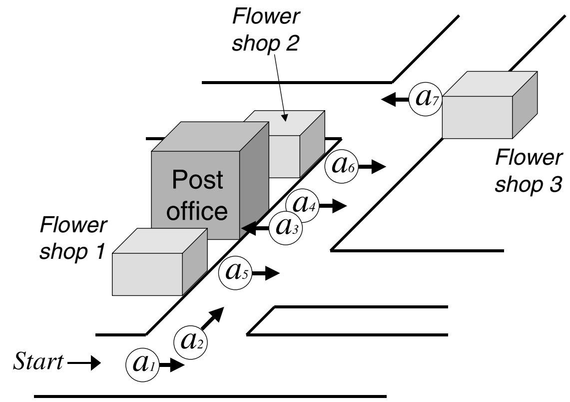

Figure 6: All action vectors

P be a set of perspective systems, i.e., $P = \{int, dei, pos\}$, where int is the intrinsic perspective system, dei is the deictic perspective system, and pos is the position-centered perspective system. Given a set of action vectors $actvec(c_t)$ and an instruction d_{t+1}, the agent interprets d_{t+1}. As a result of interpretation, the agent has a *view-frame* v_{t+1} that is defined as $v_{t+1} = \langle p, a, r\rangle$ where p ($\in P$) is the perspective system that he takes, and a ($\in actvec(c_t)$) and r are the action vector and the reference object for the perspective system respectively. The meaning of a view-frame for each perspective system is given as follows: (a) $\langle int, \_, r\rangle$ means an intrinsic system defined for a reference object r, (b) $\langle dei, a, r\rangle$ means a deictic system where r is a reference object and a is a view point, and (c) $\langle pos, a, \_\rangle$ means an position-centered system defined for an action vector a.

To illustrate the concept of these formalizations, consider the world in Figure 6. The finite set of reference objects R is as follows:

$$R = \{ \quad \text{crossing, postoffice, flowershop1, flowershop2, flowershop3} \quad \}$$

Example of view-frame (1): The agent has traveled according to the instructions $d_1d_2d_3$ in Figure 3, and the set of action vectors $actvec(c_3)$ are defined as $\{a_1, a_2, a_3, a_4, a_5\}$. The agent goes straight from a_5 according to the instruction d_4 (= '*Go straight*').

$$v_4 = \langle pos, a_5, \_\rangle$$

Example of view-frame (2): The agent has traveled according to the instruction d_1d_2, and the set of action vectors $actvec(c_2) = \{a_1, a_2, a_3, a_4\}$ are defined. The agent enters the flower shop 2, according to the instruction d_3 (='*to the right*'). The view-frame becomes as follows.

$$v_3 = \langle int, \_, \text{postoffice}\rangle$$

Description (d_t)	v_t	$actvec(c_t)$
d_0 (= $\emptyset$)	–	$\{a_1\}$
d_1	$\langle dei, a_1, crossing\rangle$	$\{a_1, a_2\}$
d_2	$\langle pos, a_2, \_\rangle$	$\{a_1, a_2, a_3, a_4\}$
d_3	$\langle pos, a_2, \_\rangle$	$\{a_1, a_2, a_3, a_4, a_7\}$
d_4	$\langle pos, a_7, \_\rangle$	$\{a_1, a_2, a_3, a_4, a_7, a_8\}$

Table 1: An example of incremental definition of action vectors

The agent may enter the flower shop 1 from the viewpoint of a_4, then the view-frame becomes as follows.

$$v_3 = \langle pos, a_4, \_\rangle$$

The agent may also enter the flower shop 3 from the viewpoint of a_2, then the view-frame becomes as follows.

$$v_3 = \langle pos, a_2, \_\rangle$$

In this way, the interpretation by the agent can be represented by encoding the agent's behaviors B as a sequence of view-frames (namely, $B = v_1 v_2 \cdots v_n$).

Incremental Definition of Action Vectors

As the agent proceeds to read the description, the action vectors are accumulated one by one. For instance, after the agent got the stamp at the post office, two action vectors (a_3, a_4) are added to the set of action vectors. With $actvec(c_t)$, the next instruction d_{t+1} and the view-frame v_{t+1}, the next action vectors $a_{t+1,1}, \cdots, a_{t+1,m_{t+1}}$ are determined. Formally, given the instruction d_{t+1} and the view-frame v_{t+1}, we have:

$$actvec(c_{t+1}) = actvec(c_t) \cup f_a(d_{t+1}, v_{t+1})$$

where f_a is a function that corresponds to an action a. f_a takes an instruction and a view-frame and returns the set of newly defined action vectors.

Table 1 shows an example of the incremental definitions of action vectors for the route depicted in Figure 4. As seen in the table, action vectors are added by interpreting the instructions one by one. Also note that, in the view-frame v_t, an action vector is selected from the previous set of action vectors $actvec(c_{t-1})$.

Conclusion

We presented a new cognitive approach for interpreting a route description in navigation systems. The process of interpretation was formulated as a dynamic system of incrementally updating a set of action vectors. Our formulation will be helpful for developing robotic systems for navigation in natural language.

By implementing the above concept, we are developing the computational system which leads an agent to a destination by giving a route description. Contrary to traditional approaches, our system allows more natural interaction with agents and human beings because our formulation of route description interpretation is more close to the interpretation by human beings than existing approaches.

References

Burgard, W.; Derr, A.; Fox, D.; and Cremers, A. 1998. Integrating global position estimation and position tracking for mobile robots: the dynamic markov localization approach. In *proceedings of International Conference on Intelligent Robots and Systems (IROS)*.

Gapp, K. P. 1994. Basic meanings of spatial relations: Computation and evaluation in 3D space. In *AAAI-94*, 95–105.

Herskovits, A. 1989. *Language and Spatial Cognition. An Interdisciplinary Study of the Prepositions in English*. Cambridge University Press.

Levelt, W. 1986. *Speaking: From intention to articulation*. MIT Press.

Olivier, P. 1996. *Mediating the Interchange of Information between the Visual and Verbal Domains*. Ph.D. Dissertation, The University of Manchester Institute of Science and Technology.

Retz-Schmidt, G. 1988. Various views on spatial prepositions. *AI Magazine* 9(2).

Simmons, R., and Koenig, S. 1995. Probabilistic robot navigation in partially observable environments. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*, 1080–1087.

Tversky, B., and Hermenway, K. 1984. Objects, parts and categories. *Journal of Experimental Psychology: General* 113(2):169–191.