

# COMMONSENSE KNOWLEDGE OF SPACE: LEARNING FROM EXPERIENCE

Benjamin Kuipers  
Department of Mathematics  
Tufts University  
Medford, Massachusetts 02155

The TOUR model is a computational model of human commonsense knowledge of large-scale space. It shows how observations are assimilated into a description, from multiple perspectives, of the spatial environment. In this paper we propose a representation for sensory events at a level suited to this investigation, and an inference strategy by which these sensory events are assimilated into descriptions of spatial structure. We also discuss the states of partial knowledge that occur during the learning process, and show that the representation exhibits graceful degradation of performance under resource limitations.

## 1. INTRODUCTION

Commonsense knowledge of space is a particularly interesting domain of human knowledge, because commonsense knowledge is a powerful and mysterious aspect of human thought, and because spatial knowledge is widely applied and distinctly different from linguistic knowledge. Knowledge of large-scale space (e.g. the "cognitive map" of a city) is particularly accessible to study because the process of assimilating new information is constrained by the speed of physical travel, so states of partial knowledge are easily observable.

The TOUR model [1,2] is a computational model of the cognitive map, exhibiting a solution to the problem:

*How can local observations acquired during travel be assimilated into a structure that permits answering route finding and relative position questions?*

The model consists of a number of different representations for spatial relationships, and inference rules for creating descriptions in one representation from information in others. The different representations can be categorized:

1. Procedural. A description specifies a procedure for travel between two given places.
2. Network. Descriptions specify the topological connectivity between places and paths, and the local geometry of intersections.
3. Geometrical. Descriptions specify relative position vectors, and the coordinate frames in which they are defined.
4. Structural. Descriptions specify containment, the role of region boundaries, parallel relations, and rectangular grids.
5. Abstraction. Descriptions link representations of spatial features at different levels of detail.

The observational input to the TOUR model is simulated as a sequence of partially specified travel instructions containing only information that could be acquired from a local observation. All other information is supplied by the assimilation process [1,2]. While this permits subsequent acquisition of order and local geometry information (and thence the rest of the cognitive map), there remain easily observable phenomena for which no explanation can be stated in terms of this representation.

1. "I can't tell you how I get there. I just know what to do." In some circumstances, a person is able to use a stored route description for physical travel, but not to anticipate up-coming landmarks or to describe the route verbally.

2. Occasionally a person will have several different internal descriptions for the same place and not realize it. Upon

*noticing the identity, his cognitive map undergoes a significant reorganization.*

The problem is to find a representation for observational input that (a) refers to sensory events rather than previously acquired descriptions, (b) serves as an adequate base for further processing, and (c) can express mechanisms to explain a wider range of phenomena. The new representation is defined (Section 2), shown to interface correctly with the existing TOUR model (Section 3), and shown to exhibit graceful degradation of performance under severe resource limitations (Section 4).

## 2. TRIPLES

We assume two primitive descriptions of sensory events for interaction with the physical world: the view, or perceptual snapshot, and the action, causing or recording physical motion and simultaneously changing the current view. More formally:

**Definition:** A view (V) is the description produced by the perceptual system of the scene observed from one vantage point. While the internal structure of a view must certainly be very complex, for our purposes it is subject to only two operations:

1. Matching. Two views can be compared to see if they describe the same place. Matching is subject to errors, primarily in the direction of failing to detect an identity (e.g. due to differences in illumination, etc.).

2. Indexing. A view can be used as the key for storage and retrieval of other internal descriptions.

**Definition:** An action (A) is the description produced by the perceptual system of a physical motion within the environment. Such a motion always produces a change in the current view. There are two types of action description, each of which may provide the observed value of an associated magnitude:

1. (TRAVEL <distance-measure> )
2. (ROTATE <angle-measure> )

A TRAVEL action corresponds to motion along some unambiguously identifiable path within the initial view (i.e. no path-selection decisions are required during the motion). A ROTATE action corresponds to a change in the direction faced without translational motion. In addition to being observed, stored, and manipulated like any other description, an action can be executed to produce physical motion.

The observations corresponding to travel in an environment will be a sequence of alternating views and actions:

$$V_1 A_1 V_2 A_2 \dots V_i A_i \dots$$

The information provided by the views and actions and their temporal sequence is used by the assimilation processes to construct the cognitive map. We will summarize the spatial properties of the physical world by defining a function relating observed views and actions:

$$\text{RESULT}(V, A) = V'$$

if the action  $A$ , taken when  $V$  is the current perceived view, results in the view  $V'$ .

**Definition:** A sequence of observations is stored as an internal description whose basic element is the Triple  $(V A V')$ . As part of an internal description, a triple corresponds to:

- a) the assertion that  $\text{RESULT}(V, A) = V'$ .
- b) the imperative to do  $A$ , when  $V$  is the current view.

A triple can be partially specified:  $(V - -)$  and  $(V A -)$  are both meaningful triples, though they do not carry the same assertional or imperative weight as a full triple. A view or action can be inserted in an empty slot of a partially-filled triple.

**Definition:** A route description (RD) is a set of full or partial triples  $(V A V')$ . It is subject to two operations:

- a) A new triple can be added to the set.
- b) A triple in the set can be retrieved given its first element. (In LISP,  $(\text{ASSOC } V \text{ RD}) \Rightarrow (V A V')$ .)

The assimilation process that creates internal route descriptions from sequences of observations simply creates triples from temporally associated views and actions and collects them into a set. In humans, this process is extended over several traversals of the same route, and involves creating a set of partial triples and later filling in empty slots in those triples. The assimilation process works by using the current perceived view to retrieve a triple (if any) from the route description, then filling empty slots if possible. The state of partial knowledge represented by the route description passes through three qualitatively distinct phases of behavior as assimilation proceeds. Similar phases can be observed in human behavior.

1. Landmark Recognition. When the route description consists of partial triples of the form  $(V - -)$ , each triple serves only to confirm that the given view does in fact occur along that route. It carries no further assertional or imperative value, and cannot be used for self-guided travel or for assimilation of information into the other perspectives in the cognitive map.

2. Self-Guided Travel. When the route description consists of partial triples of the form  $(V A -)$ , each triple can be interpreted as an imperative. If sufficiently many are available to form a sequence

$$V_1 A_1 V_2 A_2 \dots V_n$$

where  $(V_i A_i -)$  is in RD, and  $\text{RESULT}(V_i, A_i) = V_{i+1}$ , and  $V_1$  and  $V_n$  are the views at the endpoints of the route, then self-guided travel is possible along the route. However, the assertions  $(V_i A_i V_{i+1})$  are not available in the internal description, so it is not possible to rehearse the route without physical travel, or to assimilate information to other perspectives.

3. Mental Rehearsal. When sufficiently many full triples

$(V_i A_i V_{i+1})$  exist in RD to form the sequence

$$V_1 A_1 V_2 A_2 \dots V_n$$

the entire route description can be reviewed from the mental description alone. Both the assertional and imperative import of the triples are available, supporting self-guided travel, rehearsal of the route without physical travel, and assimilation of information to other perspectives.

Insight into the contrast between phases 2 and 3 can come from seeing a route description as consisting of decision information and sequence information. In phase 2, the decision information is represented internally, while sequence information is represented only by the physical structure of the world. Thus physical travel is possible because information contained in the structure of the environment is accessible, but mental rehearsal is not. In phase 3, both kinds of information are represented mentally. A route description in phase 2 accounts for the first phenomenon cited in the introduction: "I can't tell you how I get there. I just know what to do."

### 3. INTERFACE WITH THE TOUR MODEL

A complete set of fully specified triples can be used to create the partially specified route instructions previously assumed as input to the TOUR model [4]. Thus, the previously described assimilation processes can operate with the new form of observations. The two procedural route descriptions are nonetheless necessary, since the Triples are stated in terms of sensory events, while the instructions are stated in terms of internal descriptions of places and paths. Furthermore, the order of the steps in a route description cannot be observed trivially, but must be acquired in the process of assimilation from the associatively retrieved Triples to the sequentially ordered route instructions.

### 4. GRACEFUL DEGRADATION

One important property of a representation for commonsense knowledge is graceful degradation of performance under resource limitations: as conditions get bad, performance should become gradually worse, rather than failing catastrophically. It has its inverse in learning: as more information is acquired or more assimilation takes place, performance should gradually improve, rather than remaining constant until a large threshold is passed, then improving dramatically. This property is fundamental to McCarthy's Advice Taker [7], stated in terms of learning. Graceful degradation was explicitly presented as a design goal by Marr [6], and used as an explanatory device for several psychological phenomena by Norman and Bobrow [8]. Kuipers [3] presents a specific hierarchy of catastrophes of different levels of severity, and discusses the way different representations achieve graceful degradation of performance under resource limitations.

In this section we reformulate the Triples representation as a database accessible to three associative retrieval functions, to see the level of performance resulting from various kinds of errors. We show that the level of performance degrades gracefully as errors become more serious and frequent.

Information in the route description RD is used by calling the retrieval functions. Information is assimilated by indexing it under the appropriate key and retrieval function.

$$\text{EN-ROUTE}(V, \text{RD}) = \begin{cases} T & \text{if } V \text{ is observed on the route;} \\ \text{nil} & \text{otherwise.} \end{cases}$$

$$\text{ACTION-AT}(V, \text{RD}) = \begin{cases} A & \text{if } A \text{ is known to be the action to} \\ & \text{perform when } V \text{ is observed;} \\ \text{nil} & \text{otherwise.} \end{cases}$$

$$\text{NEXT-VIEW}(V, \text{RD}) = \begin{cases} V' & \text{if } V' \text{ is known to be} \\ & \text{RESULT}(V, \text{action-at}(V, \text{RD})); \\ \text{nil} & \text{otherwise.} \end{cases}$$

Travel along a route corresponds to a sequence of alternating views and actions:

$V_1 A_1 V_2 A_2 \dots V_n$   
such that  $\text{RESULT}(V_i, A_i) = V_{i+1}$ , for  $i=1,2, \dots, n-1$ . (Recall that RESULT summarizes a property of the physical world, not of a stored description of it. The definitions of the retrieval functions allow us to define precisely the different levels of performance we observed above for the Triples representation.

1. Landmark Recognition: (each landmark is familiar)  
 $\text{EN-ROUTE}(V_i, \text{RD}) = T$ , for  $i=1,2, \dots, n-1$ .
2. Self-Guided Travel: (can use RD to get from A to B)  
 $\text{EN-ROUTE}(V_i, \text{RD}) = T$ ,  
 $\text{ACTION-AT}(V_i, \text{RD}) = A_i$ , for  $i=1,2, \dots, n-1$ .
3. Mental Rehearsal: (can use RD for assimilation)  
 $\text{EN-ROUTE}(V_i, \text{RD}) = T$ ,  
 $\text{ACTION-AT}(V_i, \text{RD}) = A_i$ ,  
 $\text{NEXT-VIEW}(V_i, \text{RD}) = V_{i+1}$ , for  $i=1,2, \dots, n-1$ .

The level of performance provided by a given route-description can be intermediate between these thresholds, with higher performance along some parts of the route than can be sustained along its length.

It is clear from the definitions given above that a failure of one of the associative retrieval functions results in a drop, for that point in the route, to the next lower level of performance. Thus, we can easily determine the level of performance resulting from any given failure. Note that an error on retrieval causes a failure of a particular call to a retrieval function, which may succeed on a later attempt. An error during storage causes failures of all subsequent attempts to retrieve that information until it is observed again and successfully assimilated. Thus errors during storage are intrinsically more serious than errors during retrieval.

We can give intuitive descriptions of the result of a failure as follows. Notice that none of these errors crash the system or propagate to invalidate other stored information. A failure of NEXT-VIEW provides the mildest possible catastrophe. A failure of ACTION-AT is more serious, and EN-ROUTE is the most critical of all.

EN-ROUTE: We're lost. We may or may not be on the desired route. The current travel goal must be abandoned and special processing invoked to find familiar territory. The remainder of the route-description is still intact, however, and subsequent observations can rebuild the description.

ACTION-AT: We don't know what to do. While still on the route, explicit guidance is needed to reach the next landmark, so the current travel goal need not necessarily be abandoned. The remainder of the route description still exists, and can be rebuilt.

NEXT-VIEW: We can't anticipate the next landmark. The current trip is unaffected, but no further deductions are possible. An attempt at mental rehearsal must be abandoned. Assimilation of the next-observed view in a physical trip will repair the failure.

We can see from the definitions of the levels of performance how different failures affect performance. To complete an argument for graceful degradation of performance we must show that the most likely errors cause the mildest catastrophes. This demonstration proceeds by considering first the levels of performance encountered in the normal learning process, second the effects of errors during storage, and finally the effects of errors during retrieval. This discussion can be found in [4]. Thus we put a precise meaning to the claim that the Triples representation exhibits graceful degradation of performance under resource limitations.

## REFERENCES

- [1] Kuipers, B. J. Representing Knowledge of Large-Scale Space. MIT Artificial Intelligence Laboratory TR-418, 1977. (Doctoral dissertation, Department of Mathematics, MIT).
- [2] Kuipers, B. J. "Modelling spatial knowledge." *Cognitive Science* 2:2 (1978) 129-153.
- [3] Kuipers, B. J. "On representing commonsense knowledge." In N. V. Findler (Ed.), *Associative Networks---The Representation and Use of Knowledge by Computers*. New York: Academic Press, 1979.
- [4] Kuipers, B. J. "Commonsense Knowledge of Space: Learning from Experience." Tufts University Cognitive Science Working Paper 1, 1979. (An expanded version of this paper.)
- [5] Lynch, K. *The Image of the City*. Cambridge, MA: MIT Press, 1960.
- [6] Marr, D. "Early processing of visual information." *Phil. Trans. Roy. Soc. Lon. B* 275(942): (1976) 483-524. (Also MIT AI Memo 340.)
- [7] McCarthy, J. "Programs with common sense." In M. Minsky (Ed.), *Semantic Information Processing*. Cambridge, MA: MIT Press, 1968.
- [8] Norman, D. A. and Bobrow, D. G. "On data-limited and resource-limited processes." *Cognitive Psychology* 7 (1975) 44-64.