Gateways: An approach to parsing spatial domains

Eric Chown ECHOWN@BOWDOIN.EDU
Department of Computer Science, Bowdoin College, 8650 College Station, Brunswick, ME 04011 USA

Abstract

Cognitive maps are made up of two types of spatial units – landmarks and gateways. While most research has focused upon landmarks, gateways offer an attractive alternative that can open up new possibilities for building spatial representations. One example of this is the crucial role that gateways play in helping to parse environments. Parsing is a crucial step in reasoning about complex environments since it works to break large regions into smaller spaces. Further, since gateways are less directly tied to object recognition they afford a number of advantages for computational models. This paper examines the possibilities of using gateways, structures first proposed as part of the PLAN architecture, as the basis for spatial modeling. Two systems are reviewed which have implemented this methodology and extensions to the research are proposed, including spatial representations of abstract environments.

1. Introduction

It is far from obvious that researchers interested in machine learning should concern themselves with human spatial representation. At its core, the human spatial system relies on a highly developed object recognition system that is far advanced from its most sophisticated machine counterpart. This object recognition system makes constructing spatial representations (usually called "cognitive maps") relatively simple – a reasonable map of an environment consists of a topologically organized collection of landmarks. Because machines do not have the same capabilities, it can be very reasonably argued that spatial representations should be developed which are more tailored to their abilities. Robots, for example, can use sonar, can efficiently dead reckon, etc. There is a difference, however, between an architectural level description of a structures and the implementation of that same structure. It is our belief that the architecture of the human spatial system is adaptive enough that it can function effectively even with systems such as those typically found in today's robots. Human spatial structures have been tested by hundreds of thousands of years of evolution and have adapted to the point where they can function even in the face of a breakdown in the object recognition system; this happens at night, during fog, and under a variety of other circumstances. Visual object recognition is merely an implementation detail (albeit an extraordinarily complex and useful one) in the human spatial system. For a robot, or even for a human in an abstract environment, this detail can be replaced by alternative systems. The human spatial system can be usefully studied as an system even with many of its details abstracted away. The reasonableness of this approach can be found in the wide range of tasks to which humans apply spatial thinking.

Abstractly it is the task of the spatial system to be able to represent and provide reasoning capacity about environments physical or otherwise. In many environments this task is challenging because of the tremendous amount of information available. In order to effectively manage complex environments the spatial system must be capable of discarding useless information, and breaking large environments into collections of smaller environments. This process involves a kind of parsing, roughly analogous to how languages are analyzed. Whereas the spoken word might be broken down into individual sentences, then words, syllables, etc. a city can be broken into districts, neighborhoods, blocks, houses, and so on.

The most obvious, and most well studied, spatial unit is the landmark. Landmarks are the unique objects which populate an environment. A system simply armed with a good object recognition system could navigate fairly effectively just by relying upon landmarks (Chown et al., 1995). Of course humans have a superb object recognition system far beyond any computer-based system. Fortunately landmarks are only one way that humans parse environments. Further, with large environments, landmarks are not even necessarily a reasonable basis on which to build the kind of hierarchical structure necessary for effective reasoning. An alternative first proposed as part of the PLAN architecture (Chown et al., 1995) is called a *gateway.*

As the name suggests, gateways mark the transitions between different regions of space. In indoor environments doors are the most obvious gateways, in outdoor environments gateways might include mountain passes, bridges, or openings in a forest. Gateways are characteristically places where people stop to pause and look around. This may be because they occur at choice points, such as intersections, or because new information is afforded, such as when arriving at the top of a hill. Those reasons alone strongly argue for the importance of gateways as a place to anchor a cognitive structure; choice points should clearly be central to a structure used for planning. The fact that gateways are also locations which are frequently visited means that the learning process will benefit from repetition.

In the PLAN architecture cognitive maps consist of two integrated halves, one based on landmarks, the other on gateways. As befits an adaptive system navigation is possible with either alone, but works optimally if both are functioning efficiently. In both cases the spatial units are connected in a network structure. This is a good example of a basic architecture, in this case a topological network, working efficiently with two different implementations. Each implementation results in a subsystem with different strengths and weaknesses, ideally with one complementing the other in that regard. For example indoor environments are often rich with gateways, but lacking in landmarks, while the reverse might be true in many outdoor environments. Given these differences one might predict that people use different navigational strategies in each case. A number of unresolved research issues revolve around how the two systems interact, function separately, etc. as well as how different environments resonate differently to each of the structures.

The rest of this paper consists of a brief review of several salient aspects of the PLAN architecture. That is followed by a description of systems which implement significant portions of PLAN. Finally there is a discussion of extending the architecture to domains other than navigation, including abstract domains.

2. Cognitive Structures

Within the spatial representation community there is one widely agreed upon feature of cognitive maps – a topologically organized collection of landmarks usually called a "route map" (Siegel & White, 1975; Chown et al., 1995). For a system with a highly developed object recognition system route maps are easy to construct, represent space relatively efficiently, and are simple to use in navigation. It is not surprising that roboticists have extensively researched route maps as well. As a biologically-based model, route maps are implemented as connectionist spreading activation networks in PLAN. Planning with this type of system consists of activating the network node corresponding to the current location as well as the one corresponding to the goal. Activity spreads from each node and when the waves of activity meet a path has been found. Tests comparing systems employing this style of planning against human data have found the results to be extremely comparable (O'Neill, 1990).

Topological structures, such as networks, are attractive because they capture sequence in a natural way. Since numerous tasks, including navigation, are inherently sequential this serves as an important way of encoding spatial information. Because this also happens to be perhaps the most fundamental way that humans encode spatial information, and also the fastest to learn, it has been the focus of the most research on cognitive mapping.

Gateway-based mapping develops in parallel with the landmark network, but probably at a slower pace (though this is undoubtedly heavily influenced by environmental factors). Each of these system arises out of one of the two main visual pathways. The landmark network from the "what" pathway, and the gateway network (in PLAN gateways are called *local maps* and networks of them are called *regional maps*) from the "where" pathway. Again these are different implementations of the same architectural idea. Landmarks stand out to the object recognition system because they are perceptually distinct from their surroundings. Gateways are defined by perceptual distinctiveness as well. A simple way to detect a gateway is to look for places where the perceived background distance suddenly changes. For example, when a vista that had been occluded suddenly opens up. In their own way gateways are landmarks defined by structural properties.

Topological networks do not naturally scale. Searching for paths in a structure consisting of a small number of landmarks is fairly simple, but the search grows more difficult as the number of landmarks increases. This is a problem currently being addressed in reinforcement learning systems which also process networks of states (Dietterich, 2000). Unfortunately there is still only a rudimentary understanding of hierarchy in reinforcement learning, and it is clearly central to human spatial representation. The PLAN proposal is that spatial hierarchies arise most naturally out of the structural properties of a given environment. Physical boundaries, for example, can serve as an intrinsic mechanism for separating neighborhoods. It is worth noting that since reinforcement learning systems completely ignore such properties it is difficult to see how they will ever build the same sorts of hierarchical structures as humans. In the PLAN architecture, on the other hand, it is postulated that gateways serve to differentiate regions from one another. Passing through a gateway, as the name implies, results in moving from one region to another. In PLAN gateways are defined according to a set of highly visual criteria; this is not a necessary condition however, and indeed it has been proposed that gateways can serve as the basis of building spatial structure in abstract and nonvisual environments as well (Chown, 1999b).

Gateways allow for natural hierarchies and scaling in several important ways. First, although the structure of a regional map is topological its content is spatially rich. Corresponding to each gateway is a stored collection of scenes. Scenes provide information on relative distance and direction that is otherwise lacking in a purely topological structure. This renders concepts like one object being to the left of another meaningful. Second, gateways provide a useful and simple heuristic for parsing space: the things that can be seen from the current gateway are part of the same region of space, things which cannot be seen are part of a different region. This heuristic can even be applied in a recursive fashion to imagined views to create higher levels in the hierarchy (Chown et al., 1995).

Implicit in this representation is the fact that a spatial system built on this scheme is egocentrically organized. There is substantial evidence that humans organize space this way (Shelton & McNamara, 1997) and there is a great deal of payoff for such a scheme. Since gateways occur at critical, often visited, environmental locations they are exactly at the places where decisions need to be made, turns taken, etc. The information is stored in a semantically transparent fashion (Smolensky, 1988) and is therefore simply extracted. It may be possible to construct an allocentric representation from a collection of gateways but there is no obvious payoff and potentially a great deal of costs in terms of extra processing and the introduction of errors.

Because gateways serve as a natural way to parse environments and because computers and robots can recognize them much more simply than they can identify landmarks, it is reasonable to focus on gateways as the basis for navigation systems.

3. R-PLAN and A-PLAN

We have built two systems which are based more heavily upon gateways than on landmarks. One system is a simulator, the other was implemented on a mobile robot, and neither require sophisticated object recognition. The first, called A-PLAN (for Agent-PLAN) was implemented on top of the Quake III software engine and has been tested on simulated environments. The second, called R-PLAN (for Robot-PLAN) has been tested on indoor environments (Kortenkamp, 1993).

In indoor environments recognizing gateways is a relatively simple task. A robot moving along a hallway can simply track sonar readings looking for a large change in distance. When one occurs the most likely reason is that an intersection has been reached (or a doorway) and therefore a gateway. In tests on a mobile robot it was found that a robot could repeatably return to the same gateway location within 70 millimeters on successive trips (Kortenkamp et al., 1992). The simulator, which had the advantage of noise free sonar, could achieve essentially arbitrary accuracy.

We have experimented with several variations of what to store corresponding to each gateway. In PLAN it was proposed that people store a collection of scenes at each gateway. The resulting structure can be organized in terms of a combination of body, head and eye positions. Each scene in this scheme consists of a few salient landmarks and the contents of the scenes are useful in accessing and distinguishing individual gateways. Since landmark recognition is difficult for robots we have looked towards other representations for these tasks. In both R-PLAN and A-PLAN we have tried fairly simplistic strategies and have found that they are sufficient to achieve excellent results. In A-PLAN we took the snapshot idea literally and stored a series of bitmaps corresponding to what can be seen from each gateway. In a simulated environment with controlled lighting, and minimal environmental changes this can be an extremely effective method for doing place recognition. Such a strategy might even be more generally plausible, but would require more extensive computations involving normalizing what is viewed. For A-PLAN to determine which gateway it is at, it merely needs to center itself within the gateway, then take a series of snapshots at predetermined angles (based upon lining up along hallways) and compare those snapshots to stored gateways. The process can be done exceptionally quickly as it involves a simple vector subtraction. Most of the time A-PLAN will not need to do this as it will be able to predict where it is according to its internal map; this is mainly useful when exploring or trying out new routes. One of the intriguing things about this scheme is that it was achieved without any vision algorithms whatsoever. Gateways were derived purely using sonar, and the identification task was a simple arithmetic operation.

In R-PLAN, facing a real environment, a different approach was used (Kortenkamp, 1993). In R-PLAN, the snapshot of "landmarks" stored with a gateway consisted of a visual scene stripped of everything but vertical lines. Landmark recognition consisted of using Bayes-nets to compare the relative probabilities that the current pattern of vertical lines corresponded to stored patterns. This computationally simple trick proved to be extremely successful in allowing the robot to distinguish where it was in an environment (Kortenkamp, 1993).

R-PLAN and A-PLAN work in similar fashions. As they explore an environment they naturally will come upon gateways. When this happens they "look around" and compare what they see to any gateways they have stored previously. If the gateway is new it too is stored and connected into the cognitive map. As an agent continues to explore an environment it will build up network of gateways which it can later use to make navigational plans. A-PLAN is actually a more recent system than R-PLAN despite the fact that it was simpler to construct. The reason for building it is that it can serve as a more general purpose simulator. For example, to date both A-PLAN and R-PLAN have only been tested on indoor environments, environments naturally rich with gateways. It is anticipated that making these systems work in an outdoor setting will be much more difficult. It is simple to use sonar to pinpoint a hallway intersection, but it is another matter altogether to determine where on a hilltop a gateway should be anchored.

Both A-PLAN and R-PLAN effectively parse indoor environments without the benefit of a strong object recognition system. The representations they construct do not take direct advantage of the precision possible in computer systems, but it is altogether possible that this is one of the reasons why they are so successful (Chown, 1999a). The gateway construct gives rise to efficient representations which store knowledge in a natural way. This is undoubtedly one of the reasons that humans have been able to use their spatial systems for such a wide variety of tasks other than navigation.

4. Other domains

A-PLAN and R-PLAN represent a proof of concept of the general PLAN architecture. While both systems are true to the general spirit of the architecture, neither is a true model in the sense that they function in exactly the same way that humans do. That these systems are effective should not be a complete surprise, after all the human spatial system is used for a wide variety of tasks including many which are not visual at all in character. R-PLAN and A-PLAN abstract away from implementation details and allow an examination of the basic structure of the architecture. This provides credible evidence that this structure could be used on the same sorts of tasks that people process spatially in everyday life. The next step for this research is to apply these same structures to some of these tasks.

We have proposed (Chown, 1999b) that the spatial structures of PLAN are suited for use with tasks that share three characteristic properties:

1. There should be uniquely identifiable objects.
2. Information flow should be sequential.
3. There should be easily identifiable transitions marking context shifts.

For example a website is populated with uniquely identifiable objects (the individual pages) and information normally flows sequentially both within a page and between pages. Transitions from one page to the next are easily located. It is easy to imagine using a PLAN-style system to represent the information on a website. The general possibilities would seem to be virtually limitless. Music, for example, consists of sequences of notes with transitional markers like pauses, crescendos and the like. The regions between these markers consist of various themes and melodies which the composer is free to navigate back to later in a piece. Different musical formats even consist of fairly generic spaces where the paths are fixed, but the content changes. Lakoff and Johnson have done extensive research on the pervasiveness of spatial metaphors in language, providing insight into just how much of cognition is organized on spatial principles (Lakoff & Johnson, 1980).

The next step in the development of PLAN is to pursue these avenues, applying the architecture to disparate domains. There will be a number of pertinent issues to explore along the way. With navigation we have been working on a task which is fairly well understood and for which there is a great deal of empirical data. As we shift away from navigation to other domains the mapping onto the spatial system will not be so clear cut. While it has been relatively simple to think of domains, like music, for which a spatial representation seem natural we have not systematically explored this idea. Indeed one area of research we plan to pursue involves making predictions of how hard various environments (abstract and otherwise) would be for a human to learn. Such a system would have clear value in helping guide the design process for environments (for example software systems) in order to make them user friendly. Ultimately the goal should be to develop an understanding of how people use their spatial systems for such a variety of tasks. This knowledge should bring the added benefit of a better understanding of how to build spatial representations on computers.

So far our approach has focused on how people parse environments. The gateway construct is crucial to this process because it suggests how people break up large continuous environments into smaller more manageable chunks. It is also useful in a machine setting because recognizing gateways appears to be less challenging computationally than recognizing objects. Recognizing gateways is just one step towards a full-scale spatial representation however. There is still the question of how people organize information around gateways. In visual environments scenes are an obvious choice, but in abstract environments this may not be the case. What is a "snapshot" of a musical theme for example? Should it be stored roughly as it is heard, or translated into a more spatial format? There are a number of these questions to be researched.

5. Concluding Remarks

In the past computational modelers of cognitive maps have focused mainly upon route maps. This has been a useful exercise which has led to the development of a number of important systems, but it has probably also created built-in limitations as to what these systems can achieve; this is both because such systems do not capture the entire range of spatial information and because of the still developing state of computer vision research. Shifting from an object approach to a more purely locational approach opens up a number of interesting new possibilities. In navigation systems it is possible to build effective systems which do not need to rely upon computationally costly and relatively inaccurate vision systems. Further, it provides a much more natural way to approach the issue of building hierarchical structure. Finally, moving away from the focus on object recognition opens up the possibilities of translating spatial systems for use in nonvisual domains. This research is still at a relatively early stage of development, but the results achieved so far, in systems like R-PLAN and A-PLAN, are extremely promising.

Acknowledgements

The author would like to thank Doug Vail, David Kortenkamp and Steve Kaplan for their invaluable help and guidance in this research.

References

Chown, E. (1999a). Error tolerance and generalization in cognitive maps: Performance without precision. In R. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and spatial behavior*, 349–369. The Johns Hopkins University Press.

Chown, E. (1999b). Making predictions in an uncertain world: Environmental structure and cognitive maps. *Adaptive Behavior*, 1–17.

Chown, E., Kaplan, S., & Kortenkamp, D. (1995). Prototypes, location and associative networks: Towards a unified theory of cognitive mapping. *Cognitive Science*, *19*, 1–52.

Dietterich, T. (2000). State abstraction in maxq hierarchical reinforcement learning. In S. A. Solla, T. K. Leen and K.-R. Muller (Eds.), *Advances in neural information processing systems*, 994–1000. MIT Press.

Kortenkamp, D. (1993). *Cognitive maps for mobile robots: A representation for mapping and navigation.* Doctoral dissertation, The University of Michigan.

Kortenkamp, D., Baker, L., & Weymouth, T. (1992). Using gateways to build a route map. *IEEE/RSJ International Conference on Intelligence Robots and Systems.*

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* University of Chicago Press.

O'Neill, M. J. (1990). Computer simulation of the cognitive map: A validation study. In R. I. Selby, K. H. Anthony, J. Choi and B. Orland (Eds.), *Coming of age: Proceedings of the twenty-first annual conference of the environmental design research association.* EDRA.

Shelton, A. L., & McNamara, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review*, *4*, 102–106.

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. In H. W. Reese (Ed.), *Advances in child developomeent and behavior*, vol. 10. Academic Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*, 1–74.