In *Proceedings of the 18th National Conference on Artificial Intelligence (AAAI-02)*
AAAI/MIT Press, 2002.

Bootstrap Learning for Place Recognition*

Benjamin Kuipers and Patrick Beeson
Computer Science Department
The University of Texas at Austin
Austin, Texas 78712
{kuipers,pbeeson}@cs.utexas.edu

Abstract

We present a method whereby a robot can learn to recognize places with high accuracy, in spite of perceptual aliasing (different places appear the same) and image variability (the same place appears differently). The first step in learning place recognition restricts attention to distinctive states identified by the map-learning algorithm, and eliminates image variability by unsupervised learning of clusters of similar sensory images. The clusters define *views* associated with distinctive states, often increasing perceptual aliasing. The second step eliminates perceptual aliasing by building a causal/topological map and using history information gathered during exploration to disambiguate distinctive states. The third step uses the labeled images for supervised learning of direct associations from sensory images to distinctive states. We evaluate the method using a physical mobile robot in two environments, showing high recognition rates in spite of large amounts of perceptual aliasing.

Introduction

It is valuable for a robot to know its position and orientation with respect to a map of its environment. This allows it to plan actions and predict their results, using its map.

We define *place recognition* as identifying the current position and orientation, a task sometimes called "global localization" (Thrun *et al.* 2001). However, not every location in the environment is a "place", deserving of independent recognition. Humans tend to remember places which are distinctive, for example by serving as decision points, better than intermediate points during travel (Lynch 1960). In fact, Polynesian navigators use distinctive places as representational devices even when they cannot be physically detected, such as on the open ocean (Hutchins 1995).

Real sensors are imperfect, so important but subtle image features may be buried in sensor noise. Two complementary problems stand in the way of reliable place recognition.

- *Perceptual aliasing*: different places may have similar or identical sensory images.
- *Image variability*: the same position and orientation may have different sensory images on different occasions, for example at different times of day.

These two problems trade off against each other. With relatively impoverished sensors (e.g., a sonar ring) many places have similar images, so the dominant problem is perceptual aliasing. With richer sensors such as vision or laser rangefinders, discriminating features are more likely to be present in the image, but so are noise and dynamic changes, so the dominant problem for recognition becomes image variability. We want to use real sensors in real environments, avoiding assumptions that restrict us to certain types of sensors or make it difficult to scale up to large, non-simply-connected environments.

Unique place recognition is not always possible using the current sensory image alone. If necessary, we will use active exploration to obtain history information to determine the correct place. However, when subtle features, adequate for discriminating between different places, are buried in the noise due to image variability, we want to recover those features.

We build on the *Spatial Semantic Hierarchy* (SSH) which provides an abstraction of the continuous environment to a discrete set of *distinctive states* (dstates), linked by reliable actions (Kuipers & Byun 1991; Kuipers 2000). We assume that the agent has previously learned a set of features and control laws adequate to provide reliable transitions among a set of distinctive states in the environment (Pierce & Kuipers 1997).

A Hybrid Solution

The steps in our solution to the place recognition problem apply several different learning and deductive methods (Figure 1).

1. Restrict attention to recognizing *distinctive states* (dstates). Distinctive states are well-separated in the robot's state space.
2. Apply an unsupervised clustering algorithm to the sensory images obtained at the dstates in the environment. This reduces image variability by mapping different images of

*This work has taken place in the Intelligent Robotics Lab at the Artificial Intelligence Laboratory, The University of Texas at Austin. Research of the Intelligent Robotics lab is supported in part by the Texas Higher Education Coordinating Board, Advanced Technology Program (grant 003658-0656-2001), and by an IBM Faculty Research Award.

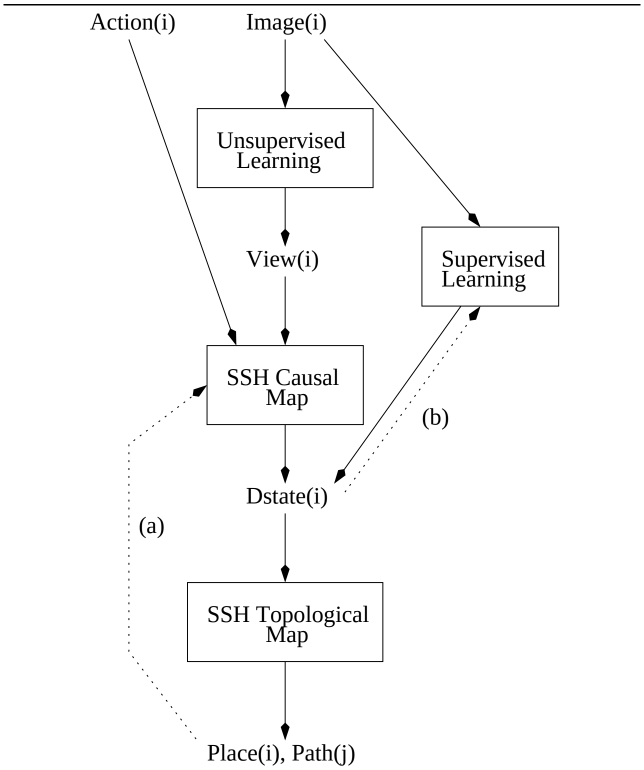

Figure 1: Bootstrap learning of place recognition. Solid arrows represent the major inference paths, while dotted arrows represent feedback.

the same place into the same cluster, even at the cost of increasing perceptual aliasing by mapping images of different states into the same cluster. We define each cluster to be a *view*, in the sense of the SSH (Kuipers 2000).

3. Build the SSH causal and topological maps — symbolic descriptions made up of dstates, views, places, and paths — by exploration and abduction from the observed sequence of views and actions (Kuipers 2000; Remolina & Kuipers 2001). This provides an unambiguous assignment of the correct dstate to each experienced image, which is feedback path (a) in Figure 1.

4. The correct causal/topological map labels each image with the correct dstate. Apply a supervised learning algorithm to learn a direct association from sensory image to dstate. The added information in supervised learning makes it possible to identify subtle discriminating features that were not distinguishable from noise by the unsupervised clustering algorithm. This is feedback path (b) in Figure 1.

We call this *bootstrap learning* because of the way a weak learning method (clustering) provides the prerequisites for a deductive method (map-building), which in turn provides the labels required by a stronger supervised learning method (nearest neighbor), which can finally achieve high performance.

Markov Localization

Markov localization has been used effectively by Thrun and his colleagues (Thrun, Fox, & Burgard 1998; Thrun *et al.* 2001) to build occupancy grid maps and to localize the robot in the grid, given observations from range sensors. The central equation for Markov localization is

$$p(x'|a,o,m) = \alpha\ p(o|x',m) \int p(x'|x,a,m)\ p(x|m)\ dx \tag{1}$$

which updates the prior probability distribution $p(x|m)$ over states x in the map m, to the posterior probability distribution $p(x'|a,o,m)$ after performing action a and observing sensory image o. $p(o|x',m)$ is the sensor model for the agent, $p(x'|x,a,m)$ is the action model, and α is a normalizing constant.

The Markov equation (1) applies whether m is an occupancy grid or a topological graph (Basye, Dean, & Kaelbling 1995), and its structure will help us compare the two representations.

Occupancy Grids

The occupancy grid representation has been popular and successful (Moravec 1988; Thrun, Fox, & Burgard 1998; Yamauchi, Schultz, & Adams 1998). Although the size of the occupancy grid grows quadratically with the size of the environment and the desired spatial resolution of the grid, this memory cost is feasible for moderate-sized environments, and modern Monte Carlo algorithms (Thrun *et al.* 2001) make the update computation tractable. Nonetheless, fundamental drawbacks remain.

- The occupancy grid assumes a single global frame of reference for representing locations in the environment. When exploring an extended environment, metrical errors accumulate. Reconciling position estimates after traveling around a circuit requires reasoning with a topological skeleton of special locations (Thrun *et al.* 1998).
- The occupancy grid representation is designed for range-sensors.[1] For a laser range-finder, an observation o consists of 180 range measurements r_i at 1° intervals around a semicircle: $o = \wedge r_i$. The scalar value stored in a cell of the grid represents the probability that a range-sensor will perceive that cell as occupied, making it relatively simple to define $p(r_i|x,m)$. Deriving a usable value of $p(o|x,m)$ is problematic, however.

The topological map representation (i) uses a set of dstates vastly smaller than an occupancy grid, (ii) does not assume a single global frame of reference, (iii) does not embed assumptions about the nature of the sensors in the representation, and (iv) clusters images o into views v giving a natural meaning to $p(v|x,m)$. We are particularly interested

[1] Minerva (Thrun *et al.* 2001, sect. 2.7) used Markov localization with particle filters using visual images from a vertically-mounted camera to localize in a "ceiling map." The ceiling map can be represented in an occupancy-grid-like structure because of the way nearby images share content. This trick does not appear to generalize to forward-facing images.

in a uniform framework for place recognition that will generalize from range-sensors to visual images (cf. (Ulrich & Nourbakhsh 2000)).

Abstraction to Distinctive States

The Spatial Semantic Hierarchy (Kuipers 2000) builds a topological map by abstracting the behavior of continuous control laws in local segments of the environment to a directed graph of *distinctive states* and actions linking them.

A distinctive state is the isolated fixed-point of a hill-climbing control law. A sequence of control laws taking the robot from one dstate to the next is abstracted to an *action*.

Starting at a given distinctive state, there may be a choice of applicable *trajectory-following* control laws that can take the agent to the neighborhood of another distinctive state. While following the selected trajectory-following control law, the agent detects a qualitative change indicating the neighborhood of another distinctive state. It then selects a *hill-climbing* control law that brings the agent to an isolated local maximum, which is the destination distinctive state. The error-correcting properties of the control laws, especially the hill-climbing step, mean that travel from one distinctive state to another is reliable, i.e., can be described as deterministic.

The directed link $\langle x, a, x' \rangle$ represents the assertion that action a is the sequence of trajectory-following and hill-climbing control laws that leads deterministically from x to x', both distinctive states. The directed graph made up of these links is called the *causal map*. The *topological map* extends the causal map with places, paths, and regions.

Since actions are deterministic, if the link $\langle x, a, x' \rangle$ is in the causal map, then $p(x'|x, a, m) = 1$, while $p(x''|x, a, m) = 0$ for $x'' \neq x'$. This lets us simplify equation (1) to get

$$p(x'|a, o, m) = \alpha \; p(o|x', m) \sum \{p(x|m) : \langle x, a, x' \rangle\} \quad (2)$$

A topological map represents vastly fewer values of x than an occupancy grid, so evaluating the sum in equation (2) will be very efficient.

Distinctive states are well-separated in the environment. Intuition suggests, and our empirical results below demonstrate, that sensory images collected at distinctive states are well-separated in image space, with the possibility of multiple states sharing the same cluster.

Unfortunately, one can construct counterexamples to show that this is not guaranteed in general. In particular, if sensory images are collected at states evenly distributed through the environment (Yamauchi & Langley 1997; Duckett & Nehmzow 2000), then image variability will dominate the differences due to separation between states, and well-separated clusters will not be found in image space. Restricting attention to a one-dimensional manifold or "roadmap" within the environment (Romero, Morales, & Sucar 2001) reduces image variability significantly, but not as much as our focus on distinctive states.

Cluster Images Into Views

A realistic robot will have a rich sensory interface, so the sensory image o is an element of a high-dimensional space, and $p(o|x, m)$ is so small as to be meaningless. Therefore, we cluster sensory images o into a small set of clusters, called *views* v. The views impose a finite structure on the sensory space, so $p(v|x, m)$ is meaningful, and in fact can be estimated with increasing accuracy with increasing experience observing images o at position x. This lets us transform equation (2) into the more useful:

$$p(x'|a, v, m) = \alpha \; p(v|x', m) \sum \{p(x|m) : \langle x, a, x' \rangle\} \quad (3)$$

In addition, our place recognition method clusters images aggressively, to eliminate image variability entirely even at the cost of increasing perceptual aliasing. That is, for a given distinctive state x, there is a single view v such that, for every sensory image o observed at x, $o \in v$. We describe this situation by the relation $view(x, v)$. This means that $p(v|x, m) = 1$ and $p(v'|x, m) = 0$ for $v' \neq v$, allowing us to simplify equation (3) further:

$$p(x'|a, v, m) = \alpha \sum \{p(x|m) : \langle x, a, x' \rangle \wedge view(x', v)\} \quad (4)$$

Intuitively, this means that prior uncertainty in $p(x|m)$ is carried forward to $p(x'|a, o, m)$, except that alternatives are eliminated if the expected view v is not observed. The probability mass associated with that alternative is distributed across the other cases when the normalization constant α is recomputed.

Where does prior uncertainty come from, since this process can only decrease it? If the initial problem is global localization, then initial ignorance of position is reflected in the distribution $p(x|m)$. Alternatively, if the robot is exploring and building a map of an unknown environment, then sometimes it will be at a dstate x performing an action a such that $\langle x, a, x' \rangle$ is unknown. A view v is observed, but the resulting probability mass must be distributed across dstates x' such that $view(x', v)$.

How Many Clusters?

We cluster images using k-means (Duda, Hart, & Stork 2001), searching for the best value of k. We use two different metrics to assess the quality of clustering: one for the agent to use to select a value of k, and one for omniscient researchers to use to evaluate the agent's selection.

The *decision metric* M uses only information available to the agent, so the agent can select the value of $k > 1$ that maximizes M. After exploring several alternatives, we adopted the following formulation of this metric which rewards both tight clusters (the denominator in equation (5)) and clear separation between clusters (the numerator).

$$M = \frac{\min_{i \neq j}[min\{dist(x, y) : x \in c_i, y \in c_j\}]}{\max_i[max\{dist(x, y) : x, y \in c_i\}]} \quad (5)$$

The *evaluation metric* U uses knowledge of the true dstate x associated with each image o to allow the researchers to assess the quality of each cluster v. The agent, however, does not have access to U. The *uncertainty coefficient* $U(v|x)$ measures the extent to which knowledge of dstate x predicts the view v (Press *et al.* 1992, pp. 632–635). (Here,

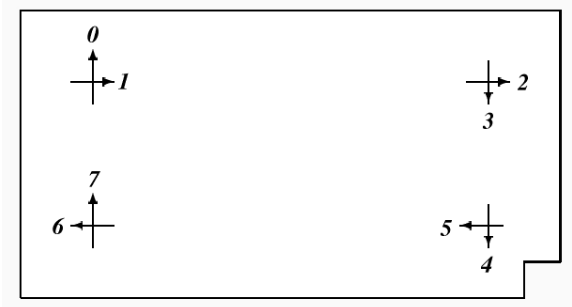

Figure 2: Simple environment for testing image variability, perceptual aliasing, and dstate disambiguation.

$p_{i,j}$ is the probability that the current view is v_i and the current dstate is x_j.)

$$
\begin{aligned}
U(v|x) &= \frac{H(v) - H(v|x)}{H(v)} \\
H(v) &= -\sum_i p_{i*} \ln p_{i*} \text{ where } p_{i*} = \sum_j p_{i,j} \\
H(v|x) &= -\sum_{i,j} p_{i,j} \ln \frac{p_{i,j}}{p_{*j}} \text{ where } p_{*j} = \sum_i p_{i,j}
\end{aligned}
$$

$U = 1$ means that image variability has been completely eliminated. As k increases, perceptual aliasing decreases, so the ideal outcome is for the value of k selected by the decision metric M to be the largest k for which $U = 1$.

A Simple Experiment

We begin testing our method in the simplest environment (Figure 2) with a distinguishing feature (the notch) small enough to be obscured by image variability.

Lassie is a RWI Magellan robot. It perceives its environment using a laser range-finder: each sensory image o is a point in R^{180}, representing the ranges to obstacles in the $180°$ arc in front of the robot. So that the Euclidean distance metric we use for clustering will emphasize short distances over long ones, we apply a “reciprocal transform”, replacing each r_i in o with $1/r_i$.

Lassie explores a rectangular room (Figure 2) whose only distinguishing feature is a small notch out of one corner. Image variability arises from position and orientation variation when Lassie reaches a distinctive state, and from the intrinsic noise in the laser range-finder. Perceptual aliasing arises from the symmetry of the environment, and the lack of a compass. The notch is designed to be a distinguishing feature that is small enough to be obscured by image variability.

As Lassie performs clockwise circuits of its environment, it encounters eight distinctive states, one immediately before and one immediately after the turn at each corner. In 50 circuits of the notched rectangle environment (Figure 2), Lassie experiences 400 images. Applying the decision metric (5) of cluster quality, Lassie determines that $k = 4$ is the clear winner (Figure 3(top)). Figure 3(bottom) shows us that $k = 4$ is also optimal to the evaluation metric.

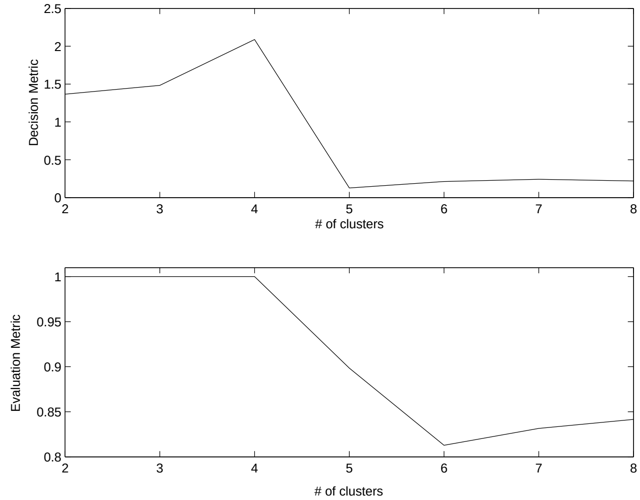

Figure 3: After Lassie explores the notched rectangle, $k = 4$ is selected as the best number of clusters by the decision metric M (top), and is confirmed as optimal by the evaluation metric U (bottom).

The notch in the rectangle is clearly being treated as noise by the clustering algorithm, so diagonally opposite dstates have the same views. In this environment, the four views correspond to the following eight dstates.

view	v_0	v_1	v_2	v_3
dstate	x_0, x_4	x_1, x_5	x_2, x_6	x_3, x_7

Build the Causal and Topological Maps

As the robot travels among distinctive states, its continuous experience is abstracted, first to an alternating sequence of images o_k and actions a_k, then images are clustered into views v_k, and finally views are associated with dstates x_k.

t_0		t_1		$\cdots$		t_n
o_0	a_0	o_1	a_1	$\cdots$	a_{n-1}	o_n
v_0		v_1		$\cdots$		v_n
x_0		x_1		$\cdots$		x_n

Clustering images into views eliminates image variability, but retains or increases perceptual aliasing:

$$
\begin{aligned}
view(x, v_1) \wedge view(x, v_2) &\rightarrow v_1 = v_2 \\
view(x_1, v) \wedge view(x_2, v) &\not\rightarrow x_1 = x_2
\end{aligned}
$$

The problem is to determine the minimal set of distinctive states x_i consistent with the observed sequence of views and actions. (Remolina & Kuipers 2001; Remolina 2001) provide a non-monotonic formalization of this problem and the axioms for the SSH causal and topological maps.

The approach is to assert that a pair of dstates is equal unless the causal or topological map implies that they are unequal. Of course, dstates with different views are unequal. But how do we conclude that $x_0 \neq x_4$ even though they share the same view v_0? When the topological map is constructed, dstate x_0 is at a place that lies on a path defined by dstates x_1 and x_2. x_4 is at a place that lies to the right of that same path, so $x_4 \neq x_0$. Similarly for the other pairs of diagonally opposite states in Figure 2. Lassie thereby determines that the four views are part of a topological map with eight dstates, four places, and four paths.

We were fortunate in this case that the prescribed exploration route provided the necessary observations to resolve the potential ambiguity. In general, it may be necessary to search actively for the relevant experience, using "homing sequences" from deterministic finite automaton learning (Rivest & Schapire 1989) or the "rehearsal procedure" (Kuipers & Byun 1991).[2]

Supervised Learning to Recognize Dstates

With unique identifiers for distinctive states (dstates), the supervised learning step learns to identify the correct dstate directly from the sensory image with high accuracy. The supervised learning method is the nearest neighbor algorithm (Duda, Hart, & Stork 2001). During training, images are represented as points in the sensory space, labeled with their true dstates. When a test image is queried, the dstate label on the nearest stored image in the sensory space is proposed, and the accuracy of this guess is recorded. Figure 6 shows the rate of correct answers as a function of number of images experienced. In two test environments, accuracy rises rapidly with experience to 100%.

The purpose of the supervised learning step is to resolve cases of perceptual aliasing,

$$view(x_1, v) \wedge view(x_2, v) \wedge x_1 \neq x_2,$$

by identifying a subtle distinction $v = v_1 \cup v_2$ such that $view(x_1, v_1) \wedge view(x_2, v_2)$. The effect of this in the Markov localization framework is that the probability distributions in equation (3) will be sharper and the sets in equations (4) will be smaller.

In general, of course, it is impossible to eliminate every case of perceptual aliasing, since there can be different dstates whose distinguishing features, if present at all, cannot be discerned by the robot's sensors. In this case, the robot must use historical context, via equation (4), to keep track of its location.

A Natural Office Environment

A natural environment, even an office environment, contains much more detail than the simplified notched-rectangle environment. To a robot with rich sensors, images at distinctive states are much more distinguishable. Image variability is the problem, not perceptual aliasing.

Lassie explored the main hallway on the second floor of Taylor Hall (Figure 4). It collected 240 images from 20 distinctive states. The topological map linking them contained seven places and four paths. When clustering the images, the

[2]We take comfort from the following qualified endorsement: *"Given a procedure that is guaranteed to uniquely identify a location if it succeeds, and succeeds with high probability, ... a Kuipers-style map can be reliably probably almost always usefully learned ... "* (Basye, Dean, & Vitter 1997, p. 86).

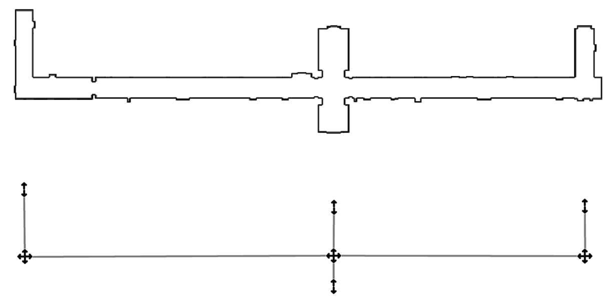

Figure 4: Taylor Hall, second floor hallway (top). The actual environment is 80 meters long and includes trash cans, lockers, benches, desks and a portable blackboard. The causal/topological map (bottom) has 20 dstates, 7 places, and 4 paths.

decision metric M had its maximum at $k = 10$. The evaluation metric U shows that higher values of k could still have eliminated all image variability (Figure 5). By building the causal and topological map the robot is able to disambiguate all twenty distinctive states, even though there are only ten different views. Given the correct labeling of images with dstates, the supervised learner reaches high accuracy (Figure 6(b)). In these environments, with rich sensors, perfect accuracy is achievable because sensory features are present with the information necessary to determine the dstate correctly, but supervised learning is required to extract that information from natural variation.

Conclusion and Future Work

We have established that bootstrap learning for place recognition can achieve high accuracy with real sensory images from a physical robot exploring among distinctive states in real environments. The method starts by eliminating image variability by focusing on distinctive states and doing unsupervised clustering of images. Then, by building the causal and topological maps, distinctive states are disambiguated and perceptual aliasing is eliminated. Finally, supervised learning of labeled images achieves high accuracy direct recognition of distinctive states from sensory images.

In future work, we plan to explore methods for robust error-recovery during exploration, by falling back from logical inference in the topological map to Markov localization when low-probability events violate the abstraction underlying the cognitive map. Once further exploration moves $p(v|x, m)$ and $p(x'|x, a, m)$ back to extreme values, the abstraction to a logical representation can be restored.

We are also exploring the use of local metrical maps, restricted to the neighborhoods of distinctive states, to eliminate the need for physical motion of the robot to the actual location of the locally distinctive state.

The current unsupervised and supervised learning algorithms we use are k-means and nearest neighbor. k-means will not scale up to the demands of clustering visual images. We plan to experiment with other algorithms to fill

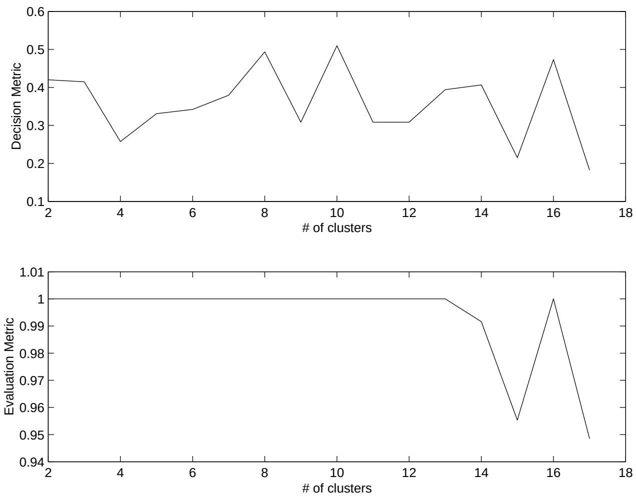

Figure 5: After Lassie's exploration of the Taylor hallway, $k = 10$ is selected as the best number of clusters by the decision metric M (top). The evaluation metric U (bottom) shows that larger numbers of views could have been selected, but $k = 10$ was still enough for supervised learning to converge to correct identification.

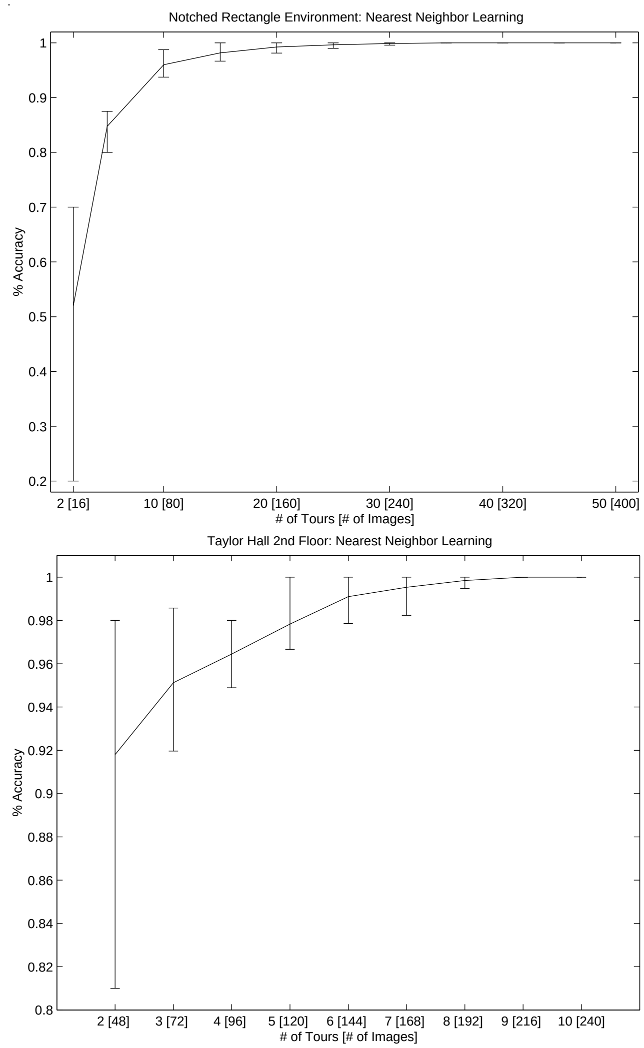

Figure 6: Learning curves (means and ranges using 10-fold cross validation) for nearest neighbor classification of dstates given sensory images for (a) the notched-rectangle environment (8 images/tour), and (b) second floor of Taylor Hall (24 images of 20 dstates/tour). Recognition rate rises rapidly to 100%, because even images of perceptually aliased dstates contain discriminating features that were averaged away by clustering, but can be identified by supervised learning.

these roles in the learning method. Other representation and clustering techniques may be more sensitive to the kinds of similarities and distinctions present in sensor images. Supervised learning methods like backprop may make it possible to analyze hidden units to determine which features are critical to the discrimination and which are noise. Using methods like these, it may be possible to discover explanations for certain aspects of image variability, for example the effect of time of day on visual image illumination.

References

Basye, K.; Dean, T.; and Kaelbling, L. P. 1995. Learning dynamics: system identification for perceptually challenged agents. *Artificial Intelligence* 72:139–171.

Basye, K.; Dean, T.; and Vitter, J. S. 1997. Coping with uncertainty in map learning. *Machine Learning* 29(1):65–88.

Duckett, T., and Nehmzow, U. 2000. Performance comparison of landmark recognition systems for navigating mobile robots. In *Proc. 17th National Conf. on Artificial Intelligence (AAAI-2000)*, 826–831. AAAI Press/The MIT Press.

Duda, R. O.; Hart, P. E.; and Stork, D. G. 2001. *Pattern Classification*. New York: John Wiley & Sons, Inc., second edition.

Hutchins, E. L. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press.

Kuipers, B. J., and Byun, Y.-T. 1991. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. *Journal of Robotics and Autonomous Systems* 8:47–63.

Kuipers, B. J. 2000. The spatial semantic hierarchy. *Artificial Intelligence* 119:191–233.

Lynch, K. 1960. *The Image of the City*. Cambridge, MA: MIT Press.

Moravec, H. P. 1988. Sensor fusion in certainty grids for mobile robots. *AI Magazine* 61–74.

Pierce, D. M., and Kuipers, B. J. 1997. Map learning with uninterpreted sensors and effectors. *Artificial Intelligence* 92:169–227.

Press, W. H.; Teukolsky, S. A.; Vitterling, W. T.; and Flannery, B. P. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, second edition.

Remolina, E., and Kuipers, B. 2001. A logical account of causal and topological maps. In *Proc. 17th Int. Joint Conf. on Artificial Intelligence (IJCAI-01)*, 5–11. San Mateo, CA: Morgan Kaufmann.

Remolina, E. 2001. *Formalizing the Spatial Semantic Hierarchy*. Ph.D. Dissertation, Computer Science Department, University of Texas at Austin.

Rivest, R. L., and Schapire, R. E. 1989. Inference of finite automata using homing sequences. In *Proceedings of the 21st Annual ACM Symposium on Theoretical Computing*, 411–420. ACM.

Romero, L.; Morales, E.; and Sucar, E. 2001. An hybrid approach to solve the global localization problem for indoor mobile robots considering sensor's perceptual limitations. In *Proc. 17th Int. Joint Conf. on Artificial Intelligence (IJCAI-01)*, 1411–1416. San Mateo, CA: Morgan Kaufmann.

Thrun, S.; Gutmann, S.; Fox, D.; Burgard, W.; and Kuipers, B. J. 1998. Integrating topological and metric maps for mobile robot navigation: A statistical approach. In *Proc. 15th National Conf. on Artificial Intelligence (AAAI-98)*, 989–995. AAAI/MIT Press.

Thrun, S.; Fox, D.; Burgard, W.; and Dellaert, F. 2001. Robust Monte Carlo localization for mobile robots. *Artificial Intelligence* 128:99–141.

Thrun, S.; Fox, D.; and Burgard, W. 1998. A probabilistic approach to concurrent mapping and localization for mobile robots. *Machine Learning* 31(1–3):29–53.

Ulrich, I., and Nourbakhsh, I. 2000. Appearance-based place recognition for topological localization. In *IEEE International Conference on Robotics and Automation*, 1023–1029. IEEE Computer Society Press.

Yamauchi, B., and Langley, P. 1997. Place recognition in dynamic environments. *Journal of Robotic Systems* 14:107–120.

Yamauchi, B.; Schultz, A.; and Adams, W. 1998. Mobile robot exploration and map-building with continuous localization. In *IEEE International Conference on Robotics and Automation*, 3715–3720.

Papers from our research group are available at http://www.cs.utexas.edu/users/qr/robotics/.
