

Topological Robot Localization by Training a Vision-based Transition Detector

Dandapani Radhakrishnan

The University of the South
Sewanee, TN

Illah R. Nourbakhsh

The Robotics Institute,
Carnegie Mellon University
Pittsburgh, PA

Abstract

This paper presents a mobile robot localizer that detects topological transitions using a color vision-based classification system. After a two-step training process, the localization system is shown to achieve sufficient reliability to track the position of a mobile robot across two floors of an office building using purely discrete (non-probabilistic) update. Experimental results are based on an off-the-shelf electric wheelchair fitted with a parabolic color camera.

1 Topological Localization

Approaches to mobile robot localization can be divided into two major classes. *Geometric localization* depends upon a metric internal representation, or map, of the robot and its environment. *Topological localization* uses a graph representation that captures the connectivity of a set of features in the environment. The occupancy grid approach is a special case of geometric localization which has seen increasing popularity of late [2;15;18]. This approach was also used in one of the earliest mobile robot navigators [11].

An advantage of the occupancy grid approach is that motion control for obstacle avoidance can use the same short-term occupancy grid that is being used for localization, leading to representational confluence [15;18]. Another advantage has been the straightforward application of probabilistic methods to the geometric representation, whereby robot position can be modeled as a distribution of metric coordinates in the map.

However, geometric localizers generally suffer from extremely large memory requirements, since the fidelity of the map must correspond well to the accuracy of sensor range values. Often, a two-dimensional map is used for navigation to avoid the computation and space explosion that a three-dimensional representation may entail [7;8;14;15;16;18]. In recent research, statistical

sampling is also used to alleviate this computational burden at the cost of completeness [4].

A further disadvantage of geometric localization is that it has limited the applicability of various mobile robot sensors to the localization problem. Metric maps are generally freespace maps, showing the free and occupied space of a projection of the robot's environment. As a result, sensors are often used only in their 2D freespace ranging modalities. This is a serious loss of information for a vision CCD, which can provide not only stereo ranging information but other cues such as texture, color, brightness and occlusion.

The present work turns to topological localization, which was used by the earliest successful robot navigators [5]. Topological representations are equally amenable to a probabilistic approach, as shown by [7;14;16]. Furthermore, the feature-based approach inherent in the topological map means that, in principle, any sensor cue may be harnessed for localization.

A major limitation of the topological approach stems from its coupling of localization and action behavior. Topological navigators must be able to traverse the arcs connecting a pair of features, and must terminate when reaching the arc endpoint. This has generally been accomplished through the careful design of control loops, or behaviors, that exhibit these properties robustly [1;5;14]. For instance, [5] creates an executable **Hall-Follow** program that exhibits the behavior of traveling reactively down a hall and terminating at the next intersection.

The behavior-programming process can be time-consuming, especially when the environment is diverse (e.g. hallways of varied widths), and this has been a significant disadvantage of the topological approach. Yet, if the actions are designed well, the topological map's abstraction of metric detail can mean more robustness in the face of a dynamic environment. Research in the principled design of behaviors has been conducted, employing analytical techniques to attempt a robust set of motion behaviors. [13] uses a control loop that follows the lines of a Voronoi diagram. [18] defines distinct features in terms of primitive sensor

readings, then defines control loops that move the robot to positions where these sensor readings are attained.

The present work takes the novel approach of decoupling action from localization for a topological representation. The result is a localizer that can robustly track the robot's position with respect to an abstract, topological map without the need for careful tuning of the robot's behaviors.

We attempted to create such a localizer using a color camera as the sole input sensor. Not surprisingly, a simple Bayesian classifier was inadequate at achieving reliable, passive localization. Yet, when this classifier was tuned automatically by a meta-level training algorithm, the localization accuracy quickly reached 100%. The resulting system, described here, is sufficiently accurate that it can track robot position in indoor spaces as a discrete, non-probabilistic value.

2 Vision-Based Transition Detection

The computer vision community has spent considerable time on the problem of image classification, as for instance would be required in image retrieval from large databases. Image similarity metrics have been applied with success to this problem; examples can be found in [6]. In the landmark recognition community, [17] construct a visual landmark model using a series of labeled example images to train a neural network. Once trained, the neural net can classify test images to determine if the same landmark is being viewed.

In both of the above applications, the image classification problem is extremely challenging because the space of test images can be large; an algorithm to retrieve red sports cars from a database must filter out not only blue and green sports cars, but also red flowers and red sunsets.

In the present work, the image comparison is used for transition detection, and so the number of actual classes is limited at any point in time to the maximum branching factor of the topological map. This constraint means that image classification need not be globally reliable, but only locally reliable with respect to potential topological transitions.

At the same time, the performance requirements are far more stringent in this application. Whereas, in image retrieval, a 96% hit rate would be excellent, the topological transition detector must have a false positive rate of zero and the ability to *eventually* detect every transition before the ensuing transition occurs.

Our strategy for achieving this conservative, high-reliability performance is a two-tiered one in which the classifier is trained, then the confidence thresholds for each transition are trained independently. We employ a

Naive Bayes classifier in which each axis of pixel color information is tessellated to 64 possible values, yielding 64^3 , or 262,144 terms in the dictionary.

During training, the mobile robot is driven at each topological node to collect *labeled* training data. This data comprises the prior $P(C_i)$. In computing $P(class)$, the probabilities are smoothed using an m-estimate to mitigate the impact that a single pixel value may have [10].

Once the classifier has been trained for all topological nodes, it can be used to classify test images. The test image is given to the classifier as a set of 64,000 pixels. Since Naive Bayes is being used; the order and locality of the pixels do not concern the system.

The three classification steps are as follows, where Narray records the amount of training data and NcArray records a count of the number of observations of each pixel value at each topological node:

$$1 \text{ Calculate } P(C_i) = \frac{(\text{NArray}[i]+1)}{(\text{totalInstances}+n)}$$

$$2 \text{ Calculate } P(E_j|C_i) = \frac{(\text{NcArray}[j,i]+1)}{(\text{NArray}[i]+m)}$$

$$3 \text{ Calculate classification} = \text{argmax } P(C_i) \prod_{j=1}^{64,000} P(E_j|C_i)$$

As stated earlier, the topological localizer must detect transitions from a current, known location. Therefore, the only places that need to be taken into consideration are adjacent nodes. Since most nodes typically have 1, 2 or 3 adjacencies, the number of locations that need to be considered by the classifier at any point in time is relatively small.

The topological map of a section of Robotics building D at Carnegie Mellon University is shown in Figure 1. This map and building were used during testing.

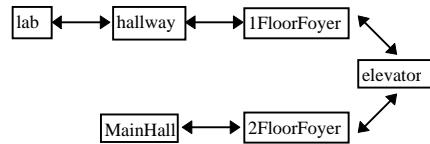


Figure 1: Adjacency map of Building D, first and second floors

The robot begins at a known location on its map. Using this information, it decides which locations need to be considered for the classifier. For instance, in the above map, if the robot knows that it is in the hallway, then

the only locations considered during classification are lab, hallway and 1FloorFoyer.

As long as the NBC classifies the current image as the current location (hallway), the robot will remain at the same map location and consider the same three locations. When the classification differs from the current location, the robot makes a transition to the new location. For instance, if the classification is 1FloorFoyer, then the robot transitions from the hallway to 1FloorFoyer and then considers hallway, 1FloorFoyer and elevator as possible locations.

3 Automatic Transition Threshold Selection

The reader may predict that simple classification as presented in section 2 is not sufficiently robust for arbitrary topological navigation.

We propose an improved approach based on an analysis of the confidence values generated by the NBC, defined as the ratio of the probability of the best classification over the probability of the second best classification.

Suppose that an aberrant region covers part of the image. If that aberrant portion may be classified as belonging to the wrong location, the overall confidence should be generally lower than the classification of an image truly belonging to that alternate location, for in the latter case most of the pixels would be a good “fit.” Therefore, the aberrant confidence during misclassification should generally be lower than correct classifications for the same location.

The solution was to build a second-tier trainer to be used on top of NBC to classify NBC’s classifications as being good or bad. This second tier partitions the confidence space for *every* location.

Once again, the mobile robot is driven throughout the environment and the transitions are manually labeled. During this run, the localizer chooses confidence transition thresholds that are well above those that cause potentially incorrect topological transitions during the training session.

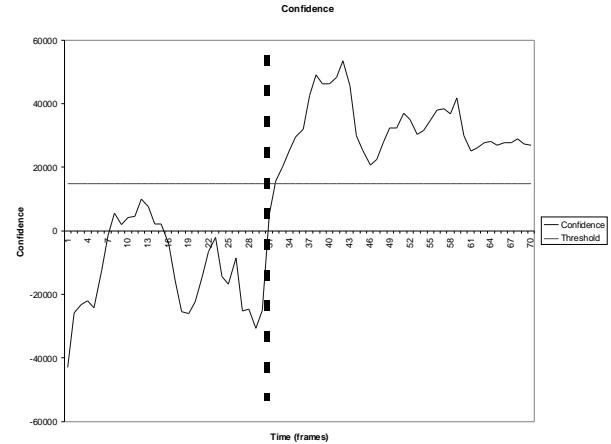


Figure 2: Wheelchair transition from hallway (left half) to 1FloorFoyer, with the maximum measured confidence values shown.

Note that the transition thresholds will be directional, as symmetry does not hold for topological transitions. Figure 2 shows the confidence levels as the robot transitions between two nodes. The sign on the Y axis denotes classification as hallway (negative) and 1FloorFoyer (positive). The overall confidence levels in hallway (left half) are somewhat lower than the confidence levels in 1FloorFoyer (right half). The potential transition from hallway to 1FloorFoyer should use a different transition threshold than the possible transition from 1FloorFoyer to hallway.

Between frames 7 and 15 there is a local peak in confidence levels that is, in fact, a misclassification of hallway as 1FloorFoyer.

However, note that the misclassifications have significantly lower confidence than the correct classifications of 1FloorFoyer to the right of the physical transition. Thus, a *relative* threshold, specific to 1FloorFoyer, is possible. The second tier trainer uses this difference to distinguish between good and bad classifications on this transition, setting a threshold line at 15007 (log scale).

The second-tier training process was undertaken in Building D’s first and second floors as well, resulting in the following transition thresholds. Note the variety and asymmetry of threshold values. Values of negative infinity indicate that a threshold was unnecessary (i.e., every classification in that direction is correct because the environmental cues are visually unique).

Table 1: Transition Thresholds (log scale)

| <i>from</i> | <i>to</i> | <i>threshold</i> |
|-------------|-----------|------------------|
| lab | hallway | $-\infty$ |
| hallway | lab | $-\infty$ |

| | | |
|-------------|-------------|-----------|
| hallway | 1FloorFoyer | 15007 |
| 1FloorFoyer | hallway | $-\infty$ |
| 1FloorFoyer | elevator | 16224 |
| elevator | 1FloorFoyer | 39961 |
| elevator | 2FloorFoyer | 38505 |
| 2FloorFoyer | elevator | 59463 |
| 2FloorFoyer | MainHall | 38505 |
| MainHall | 2FloorFoyer | $-\infty$ |

4 Results

The topological localizer was implemented on a computer-controlled, electric wheelchair (Figure 3) at the Mobot Programming Lab in The Robotics Institute. The wheelchair hardware is a product of KIPR (Maryland). An on-board 68332 processor performs motor velocity control operations on the wheelchair. Visual navigation is performed on a Toshiba Satellite Pro 440CDT laptop computer that communicates with the 68332 via serial link. The Hitachi KPD50 color CCD camera, in turn, connects to the laptop by way of a PCMCIA framegrabber (MRT VideoPort Pro).



Figure 3: The electric wheelchair base with the color camera.

In the experiments described below, the Hitachi is fitted with a ParaCamera by Cyclovision. This parabolic camera affords a complete 360 degree view to the robot, eliminating rotational variation. Figure 4 compares a standard lens to the parabolic mirror's view.

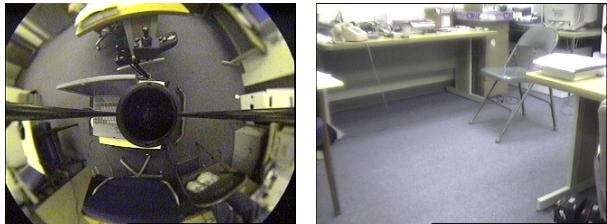


Figure 4: Omnicam (left) and standard pictures of lab

A benefit of having such a large field of view is that any single feature in a space occupies only a very small portion of the image space, enabling the system to be robust to somewhat dynamic office environments. All software was developed using Microsoft Visual C++ 5.0.

The complete system was tested on the first and second floors of Building D at The Robotics Institute. First, all six locations were trained, as described above. Then, a test route was chosen:

lab \rightarrow hallway \rightarrow 1FloorFoyer \rightarrow Elevator \rightarrow 2FloorFoyer \rightarrow MainHall \rightarrow 2FloorFoyer \rightarrow Elevator \rightarrow 1FloorFoyer \rightarrow hallway \rightarrow lab.

The transition point between 2FloorFoyer and MainHall was decided arbitrarily, without giving due attention to the physical layout of the locations. 2FloorFoyer should have been divided into two discrete nodes because it consisted of two sub-regions which appeared very different from one-another. To add to this difficulty, the sub-region adjoining the Elevator looked extremely similar to MainHall. This flaw in drawing the topological map was deliberately made in order to study the effects of poor map-making.

Once training was complete, a set of 15 trials of the entire test run were performed by allowing the localizer to identify the robot's current location while the robot was servoed, by hand, through the test course. The accuracy of the classifier in correctly tracking the robot's location was noted. The geometric accuracy of the classifier in detecting each transition at the proper location was also noted.

Each trial consists of 10 physical transitions. An important distinction between navigation and image retrieval involves the consequences of error: whereas an image misclassification is a single, insulated mistake, a transition misclassification can have disastrous consequences, as the robot's future localization attempts will be adversely affected.

Of the 150 total transitions tested, the robot achieved a perfect score: 150 transitions were detected, and 0 false positives occurred (where the classifier detected a nonexistent transition). During this test, approximately 600 images were posed to the localizer in real time. Moreover, each trial involved continuous localization from the lab to the mainHallway and back, without any pausing or resetting.

The average distance between the actual transition points and the navigation transition detection points were computed for every transition in the test route. All transitions except 2FloorFoyer-MainHall and MainHall-2FloorFoyer were made within the wheelchair's footprint from the correct transition point. The two transitions that failed are the ones that were deliberately

flawed. In the case of 2FloorFoyer-MainHall, the localizer detected the transition 8 feet beyond the physical transition point (but did so sufficiently consistently that the robot was never lost). This highlights the importance of drawing a suitable map where the locations are adequately differentiable.

Because a higher confidence implies a more successful separation of the different locations, the average confidence was computed as a suitable metric for comparing alternative camera parameters (RGB space versus HSI space; parabolic versus standard lens systems). The average confidence values for each of 6 camera scenarios are given in Figure 5.

The hyperbolic mirror in conjunction with the YUV color space was found to be the most informative camera scenario. In addition, the parabolic mirror provided higher quality information than the standard lens for every color space tested.

Most surprising, however, was the fact that the regular lens, although less informative, provided sufficiently high-quality information for navigation to proceed without faults. Indeed, every combination tested performed the navigation task with almost 100% reliability. The standard lens provides less than 20% of the field of view of the parabolic mirror, and therefore its relative success is impressive.

In the next experiment, the aim was to identify the most informative aspects of the Y, U and V bands of that most successful color space. The experiments mentioned above were repeated, again initializing the mobile robot and then training and testing its classification confidence. However, instead of providing Y, U and V as a set, subsets were provided to the classifier throughout training and classification. The resulting 6 navigators were separately tested and, once again, confidence levels were logged. Figure 6 shows the average confidence for each subset of the YUV color space.

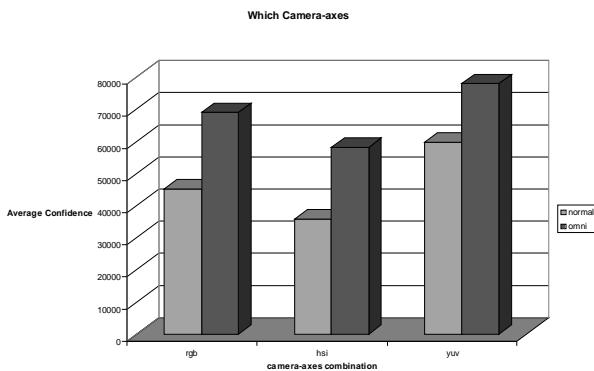


Figure 5: A comparison of the average confidence achieved by six lens-color space combinations

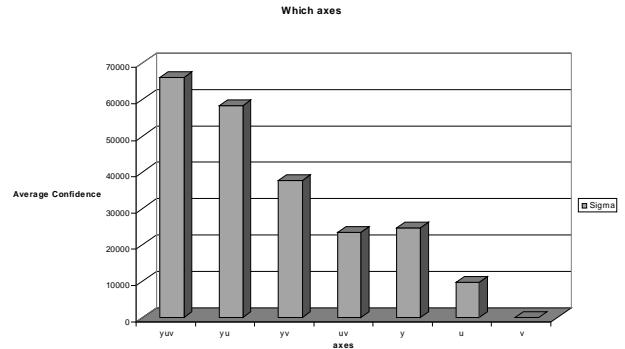


Figure 6: A comparison of the average confidence values achieved by subspaces of YUV

The results imply that most of the information is contained in the Y axis. This means that the classifier which uses the YUV transformation will be very sensitive to luminance.

5 Future Work

An important future step involves taking this localizer outdoors. This change of venue will introduce several potential challenges, including luminance constancy and non-local features. Our experiments show that the classifier is most sensitive to luminance. While this is acceptable in artificial environments where lighting is fairly controlled, it will introduce unacceptable errors as we transition to outdoor and mixed-lighting environments. A classical strategy with which we will begin involves placing an object of known luminance in the visual field of the robot, then calibrating on-line.

Indoor spaces generally consist of open areas (rooms) separated by apertures (doorways and openings). These spaces naturally afford the vision system local features with which to recognize a particular topological node. Indeed, traversal of a doorway can often result in a large number of occlusions and so the visual field can change grossly.

Outdoor spaces do not have this property. The width of an appropriate topological node may be significantly smaller than the distance from the robot to an appropriate physical feature that is useful in recognizing that node. We are considering approaches to limiting the robot's vertical field of view in order to force transitions to be detected on the basis of nearby feature changes. The appropriate method for dealing with such open spaces remains to be discovered.

Currently, the positions of the topological nodes and their connectivity is entered manually; training is used only to initialize the vision system to recognize transitions in this static map. An important next step is to choose topological nodes and establish connectivity automatically, during a zeroth level training iteration.

6 Conclusions

We have demonstrated a topological localization system that decouples action behavior from localization successfully. The internal map is represented strictly as an abstract, topological data structure. Indeed, most topological navigation systems do have approximate metric information regarding arc lengths and angles [14;16]. The present system, however, navigates indoor spaces robustly with no metric information whatsoever. It has been demonstrated empirically using a series of 600 images, classified in real time to yield a localizer with zero positioning errors.

We hope that this work demonstrates the power of classical vision algorithms when applied to a standard robot navigation problem: the results indicate perfect navigation results, albeit under a limited set of controlled circumstances, most notably controlled lighting. As mobile roboticists make use of more sophisticated forms of vision processing, we predict that visual navigation will achieve levels of robustness not yet witnessed.

Acknowledgments

Thanks to Martial Hebert at The Robotics Institute for some of the equipment required for this project, and to Daniel Huber for comments on an early version of this paper.

References

- [1] Brooks, R. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*. 1986.
- [2] Burgard, W. Cremers, A., Fox, D., Hahnel, D., Lakemeyer, G., Schulz, D., Steiner, W. and Thrun, S. 1998. The Interactive Museum Tour-Guide Robot. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*. Madison, Wisconsin. AAAI Press.
- [3] Castellanos, J.A., Tardos, J.D., Schmidt, G. Building a global map of the environment of a mobile robot: The importance of correlations. *Proceedings of the 1997 IEEE Conference on Robotics and Automation*. 1997.
- [4] Dellaert, F., Fox, D., Burgard, W. and Thrun, S. Monte Carlo Localization for Mobile Robots. In *Proceedings of the IEEE Conference on Robotics and Automation*, 1999.
- [5] Gat, E. Integrating Planning and Reacting in a Heterogeneous, Asynchronous Control Architecture In *Proceedings of the Tenth National Conference on Artificial Intelligence*. 1992.
- [6] IEEE Workshop on Content-Based Access of Image and Video Database in conjunction with CVPR '98, June 21, 1998, Santa Barbara, California
- [7] Kaelbling, L., Littman, M. and Cassandra, A. Planning and Acting in Partially Observable Stochastic Domains, *Artificial Intelligence*, Vol. 101, 1998.
- [8] Kuipers, B., Froom, R., Lee, W-K. and Pierce, D. 1993. The Semantic Hierarchy in Robot Learning. *Robot Learning*. Jonathan Connell & Sridhar Mahadevan (eds.), Kluwer Academic Publishers.
- [9] Kunz, C.; Willeke, T. and Nourbakhsh, I. 1999. Automatic Mapping of Dynamic Office Environments. *Autonomous Robots*. In print, Spring 1999.
- [10] Mitchell, T. *Machine Learning*. McGraw Hill, 1997.
- [11] Moravec H.P. and Elfes, A. 1985. High resolution maps from wide angle sonar. In *Proceedings, 1985 IEEE International Conference on Robotics and Automation*. Silver Spring, MD.
- [12] Murray, D., Jennings, C. 1997 Stereo vision based mapping and navigation for mobile robots. *Proceedings of the 1997 IEEE Conference on Robotics and Automation*. 1997.
- [13] Nagatani, K., Choset, H. and Thrun, S. Towards Exact Localization without Explicit Localization with the Generalized Voronoi Graph. In *Proceedings of the IEEE International Conference*

on Robotics and Automation. Leuven, Belgium.
1998.

- [14] Nourbakhsh, I., Powers, R. and Birchfield, S. 1995. Dervish, An Office-Navigating Robot. *AI Magazine* 16(2).
- [15] Schultz, A., Adams, W., Yamauchi, B. and Jones, M. Unifying Exploration, Localization, Navigation and Planning Through a Common Representation. Navy Ctr for App. Research in AI. Unpublished.
- [16] Simmons, R. & Koenig, S. 1995. Probabilistic robot navigation in partially observable environments. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. Morgan-Kaufmann.
- [17] Takeuchi, Y. and Hebert, M. 1998. Evaluation of Image-Based Landmark Recognition Techniques. CMU Technical Report CMU-RI-TR-98-20. Carnegie Mellon University. Pittsburgh, PA. 1998.
- [18] Thrun, S., Fox, D., Burgard, W. 1998. Probabilistic Mapping of an Environment by a Mobile Robot. In *Proceedings of the 1998 IEEE International Conference on Robotics & Automation*. Leuven, Belgium. 1998.