A Vision-Based Robot Navigation System

Daniel Roth

What: We are building a vision-based robot navigation system that constructs a high level topological representation of the world. A robot using this system learns to recognize rooms or spaces and to navigate between them by building models of those spaces and their connections. The intent of this project is to create a system that allows a robot to be able to explore and to navigate in a wide variety of environments in a way that facilitates path planning and goal-oriented tasks.

Why: Many current robot navigation systems make restrictive assumptions about the environment the robot will navigate in, such as assuming a two-dimensional world or assuming the world consists of hallways and lobbies. Vision provides a rich set of features which should allow a robot to navigate and map a larger variety of environments.

Many robot navigation systems also focus on producing a detailed metric map of the world using expensive hardware, such as a laser scanner. After this map is built the robot then has to solve a complicated path-planning problem. A topological representation facilitates path-planning and does not require expensive range sensors.

Humans can readily demonstrate the ability to navigate visually without the supplement of metric information [3]. This ability has inspired the development of view-based robot navigation systems, which represent an environment as a set of snap shots and also a set of control sequences or algorithms for navigating between those snapshots. View-based approaches have been shown to be effective for both exploration and navigation tasks [1].

View-based approaches operate well in static environments but are inherently not robust to a dynamic environment. View-based approaches also scale poorly since the topological representation tends to be fine grained and therefore requires storing many views. Experiments on humans indicate that people do not represent space as a collection of views but instead reason about landmarks in an environment at a higher level of abstraction [2]. A more abstract model-based approach has the potential to deal with dynamic environments and to compress the information stored in collections of views.

How: We are building on work by Torralba, Murphy, Freeman and Rubin which demonstrated that localization at the room level is robust and practical. [4]. The robot is trained on a new environment by capturing sequences of images from a camera mounted on the robot as the robot traverses an environment. The images are labeled to delineate separate spaces and processed to extract texture information. The robot builds and uses a hidden Markov model of the environment trained on the image to perform localization. Experiments with the localization system on the robot have shown the system to be responsive and robust to changes in illumination and view point.

To navigate between rooms, images captured by the robot will be further labeled to indicate the room that the robot is approaching. These images can then be used to construct a visual gradient which the robot can then use to home in on neighboring rooms. The advantage of using a visual gradient is that it defines a very general policy for navigating between rooms.

Progress: Our group is using an ER1 robot from Evolution Robotics and the ERSP robot programming interfaces provided from the same company. For an obstacle avoidance we are experimenting with simple reactive obstacle avoidance using infrared sensors, and also doing local path planning using a stereo vision camera to detect obstacles.

We have already programmed the robot to wander, avoid obstacles and recognize its location. We have performed some initial experiments in having the robot follow a visual gradient by building a mixture of gaussians model of the visual gradient and then giving the robot a left and right eye so that it can follow the gradient using a simple differential controller.

Future: Currently, images are hand labeled and the models of the environment are built as a preprocessing step. We are working on devising ways to allow the robot to build these models on its own and explore and learn new environments.

The localization system is accurate and robust in static environments, but still very brittle when the world changes (for example, when people walk in front of the robot's cameras). More work needs to be done to determine how to build models of places and their connections in a manner which is robust in dynamic environments.

We also hope to improve the quality of interactions between humans and the robot by experimenting with expression of emotion and speech recognition. We envision developing the robot to the point where it can give guided tours in new unstructured environments.

Research Support: This research is supported in part by NASA under contract NCC2-1237 and in part by DARPA under contract DABT63-99-1-0012.

References:

[1] Matthias O. Franz, Bernhard Scholkopf, Hanspeter A. Mallot, and Heinrich H. Bulthoff. Learning view graphs for robot navigation. *Autonomous Robots*, pages 111–125, 1998.

[2] Hanspeter A. Mallot and Sabine Gillner. View-based vs. place-based navigation: What is recognized in recognition-triggered responses?, 1998.

[3] M. J. O'Neill. Evaluation of a conceptual model of architectural legibility. *Environment and Behavior*, 23:259–284, 1991.

[4] Antonio Torralba, Kevin P. Murphy, William T. Freeman, and Mark A. Rubin. Context-based vision system for place and object recognition. In *Internation Conference on Computer Vision*, 2003.