A Neural Network Model of View-Based Map Making

(Bachelder, I.A. and Waxman, A.M., "A view-based neurocomputational system for relational map-making and navigation in visual environments", *Robotics and Autonomous Systems*, Vol. 16, pp 267-298, 1995)

The robotic map making system of Bachelder and Waxman is based on the notion that a map has the following features:

- it qualitatively predicts what and where landmarks become visible.
- it requires neither 3D reconstruction nor precise 2D mensuration of physical structures.
- it adapts to gradual changes in the environment.

Their system learns and adapts a map based entirely on:

- views of visual landmarks (the *what*),
- imprecise egocentric gaze directions to these landmarks (the *where*)
- short-term measurements of locomotive actions

Place Learning and Recognition

In Bachelder and Waxman's map learning system, places are defined by two different *view-based* methods:

- by the spatial directions to landmarks visible from a given place ("where" approach)
- by the visual appearances of landmarks visible from a given place ("what" approach)

The system self-organizes (i.e. "clusters") an environment into place categories, each defined by an adaptable template pattern representing a coarse conjunctive coding of 2D landmark shape and spatial direction within a restricted local view of the environment.

The visual appearance of landmarks is coarsely coded by a view-based object learning and recognition system.

View-Based Object Recognition

In view-based object recognition approaches, objects are represented by pooling sets of cells that code, invariant to a host of visual transformations (scaling, translation, lighting, etc.), complex, view-specific spatial patterns.

These systems models an object as a collection of *characteristic views*, or *aspects*, combined with possible transitions between these views.

The object recognition system used by Bachelder and Waxman uses the following sequence:

- Stable object features are extracted from the image (e.g. strong edges and high curvature points)
- The image is translated so as to centre the centroid of the feature set. This is followed by a log-polar mapping, followed by another translation to the centroid in the log space. This provides invariance to translation and scale.

- The resulting feature maps are coarsely coded (9x9 array) resulting in relative insensitivity to orientation changes.
- This invariant pattern is fed into an Adaptive Resonance Theory (ART) neural network, which self-organizes these patterns into *aspect categories*.
- Finally, an Aspect Network associatively learns the allowable transitions between aspect categories.

Once the learning is done, a viewed object is quantized into one of a set of *prototypical views*, each associated with a learned aspect template.

The learned aspect categories tesselate the *explored* portion of the viewing sphere (whose coordinates correspond to possible directions from which the object can be viewed).

Note that the viewing sphere is centred on the object, not on the observor.

The view-based object recognition approach can be used not only for landmark shape processing, but also for emulating cognitive mapping.

Places can be thought of as the viewer-centreed spatial distribution of landmarks from a particular location and heading, just as object aspects can be thought of as the viewer-centred spatial distribution and appearance of object parts from a particular viewing direction and orientation.

Furthermore, action induced transitions between places are analogous to motion induced transitions between object aspects.

In short, learning a 3-D environment defined by a distributed constellation of visual landmarks can be thought of as learning an object composed of it its constituent parts *from the inside*.

Place Learning and Recognition

To coarsely encode landmark direction, the place learning and recognition process uses internal measurements of body position and head direction, and external (visual) measurements of the fixated 2D landmark feature centroid within the field of view.

To perform place recognition the robot must continually search the visual panorama for landmarks. During this process, a short-term topographic memory slowly builds up a panoramic pattern from the current conjunction of spatial and identity (shape) information. This memory resets decays with time and whenever the robot moves.

A Heading-Invariant Version

An early version of this approach implemented a heading-invariant system, in which the entire panorama was fed into an ART network, which did the self-organizing place learning and recognition processes.

In this heading-invariant version the robot learned panoramic patterns of landmark positions with respect to a global orientation frame of reference (e.g. "North").

The robot (called "MAVIN") learned each of the prospective landmarks shape individually by traversing around them on pre-programmed paths. Then place learning took place. The landmarks that were used in the testing of the MAVIN system were boxes to which arrays of lights were attached.

Due to the separate coding of landmark shape and direction the system can recognize places invariant to both the robot's heading and landmark imaging conditions. Due to the coarse coding, the place recognition exhibits insensitivity to small changes in the robot's location or in the positions and poses of the landmarks.

As a result, each of the learned place categories corresponds to a *place region* which are areas from which the visual appearance, location, and poses of the landmarks remain sufficiently similar to its learned place template.

These mutually exclusive place regions tesselate the explored 2D area of the environment.

The *activity profile* of a particular place category falls off roughly monotonically from a peak value as the robot translates away from the center of the corresponding place region to a neighboring place region. These activity profiles overlap, and intersect at the boundaries between the place regions.

These profiles qualitatively resemble the spatial firing patterns observed in hippocampus place cells in rats.

A Heading-Variant Version

A more recent version of the MAVIN place learning system does not use the landmark shape information, but instead uses the unclassified 2D invariant view pattern.

This gives the system the ability to perform place recognition without performing landmark recognition.

It also allows landmark learning and place learning to be performed concurrently.

It can deal with the partial or complete occlusion of landmarks during both learning and recognition, and, in fact, can operate with just a single landmark.

A second modification in the new version concerns the role of heading specificity. Instead of using a complete panorama, the new system uses a local viewing centred about the current heading of the robot.

This is a more efficient approach, since the robot doesn't need to obtain a panorama at each location (which requires the robot to stop at each location and turn completely around).

The local view sampling process performs a non-uniform warping of the image as seen from the current robot heading. A logarithmic transformation is used. The effecitve view width is about 30 degrees.

Place learning is done as in the previous version - the robot traverses a pre-defined exploration path, stopping at one foot intervals. At each stopping point the robot rotated 360 degrees in 10 degree increments, obtaining restricted local views.

The system produces heading-tuned place categories, which tesselate the space of explored environment positions (combination of location and heading) into place regions.