Learning Localisation Based on Landmarks using Self-Organisation

Kaustubh Chokshi, Stefan Wermter, and Cornelius Weber

The Informatics Centre, School of Computing and Technology, University of Sunderland
St. Peter's Campus, Sunderland SR6 0DD, United Kingdom
http://www.his.sunderland.ac.uk/

Abstract. In order to have an autonomous robot, the robot must be able to navigate independently within an environment. Place cells are cells that respond to the environment the animal is in. In this paper we present a model of place cells based on Self Organising Maps. The aim of this paper is to show that localisation can be performed even without having a built in map. The model presented shows that the landmarks are selected without any human interference. After training, a robot can localise itself within a learnt environment.

1 Introduction

Ideally an autonomous robot should have a self-contained system which allows it to adapt and modify its behaviour to all possible situations it might face. The classical method is to pre-define the internal model of the robot relating it to the external world. The problem which arises is that the external world is very complex thus making it often unpredictable [2]. With a pre-defined program the robot can navigate only within a highly controlled environment [7].

There are various ways to address the problem of localisation. The most common approach is to ignore the errors of localisation [6]. This has the advantage of being simple but a major disadvantage is that it cannot be used as a global planning method. To overcome this problem, another technique updates the robot location by bar codes and laser sensors in the environment, thus giving it an explicit location. This method was purely motivated by the "*go until you get there*" philosophy [6]. Another approach was to use topological maps where symbolic information for localisation at certain points for instance gateways was being used [3, 12, 4]. With gateways as a navigational strategy the robot can change its direction for instance at intersections of hallways. Here the robot is not concerned with how long the hallway is but it knows that at the end of the hallway there is an intersection. Therefore the gateways are critical for localisation, path planning and map making. The main disadvantage here is that it is very hard to have unique gateways. Another technique of localisation is to match raw sensor data to an a priori map [6]. Usually the sensors used are distance sensors such as sonar, infrared, laser etc. The problem here is that the sensor data rarely comes in without noise but it can be used in highly restricted environments. It is used in certain industrial robots who operate in highly controlled environments but they fail to work in dynamic environments. This problem is overcome by generating small local maps and then integrating them with larger maps. The use of small local maps motivates the need of good map building [6].

More sophisticated systems either identify added artificial landmarks or natural landmarks. Landmarks can be defined as *a distinctive object or a place in the environment which help localisation*. In other words landmarks are distinct objects in the habituated environment. In a natural environment the landmarks would be chosen as those objects that have noticeable geometric properties.

In this paper we present a model that learns to form landmarks for helping the robot to navigate. In order to generate an internal representation, we use vision to identify landmarks. This paper focuses on a type of landmark localisation which depends on extracting novel features from the environment in order to localise itself. Landmarks can be defined as distinguishable input patterns. In other words each category of input activates different output neurons. In our approach we have developed Self Organising Maps which classify the input images for localisation.

2 Self Organising Maps Based on Place Codes for Localisation

SOM networks learn to categorise input patterns and to associate them to different output neurons or a set of output neurons. Each neuron, j, is connected to the input through a synaptic weight vector $\mathbf{w}_j = [w_{j1} .... w_{jm}]^T$. At each iteration, the SOM finds a winning neuron **v** by minimising the following:

$$v(x) = arg\ min_j ||x(t) - w_j||, \qquad j = 1, 2, ...n \tag{1}$$

x belongs to an *m*-dimensional input space, $||.||$ is the Euclidean distance, while the update of the synaptic weight vector follows:

$$w_j(t+1) = w_j(t) + \alpha(t) h_{j,v(x)}(t)[x(t) - w_j(t)], \qquad j = 1, 2, ...n, \tag{2}$$

This activation and classification are based on features extracted from the environment by the network. Feature detectors are neurons that respond to correlated combinations of their inputs. These are the neurons that give us symbolic representations of the world outside. In our experiments once we get symbolic representations of the features in the environment we use these to localise the robot in the environment.

The removal of redundancy by competition is thought to be a key aspect of how the visual system operates [11]. Competitive networks also reduce the dimensions of the input vector as a set of input patterns, in our case pixels of the input image vector. The representation of a location is done by activation of a neuron.

The sparsification performed by the competitive networks is very useful for preparing signals for presentation to pattern associators and auto associators, since this representation increases the number of patterns that can be associated or stored in such networks [11]. Although the algorithm is simple, its convergence and accuracy depend on the selection of the neighbourhood function, the topology of the output space, a scheme for decreasing the learning rate parameter, and the total number of neuronal units [5].

We are making use of a simple scenario for our symbolic representations of the world model. For this experiment the overall goal for the robot is to navigate between two desired locations. When it is navigating between those two locations, it should be able to localise itself. In the environment we provide a random colour coded rectangles on the wall, along with some distinguishable features like cubes and pyramids randomly kept in the environment like in the figure 1(**b**). During experimentation, the robot was able to make its own symbolic representation of the world model based on unsupervised learning.

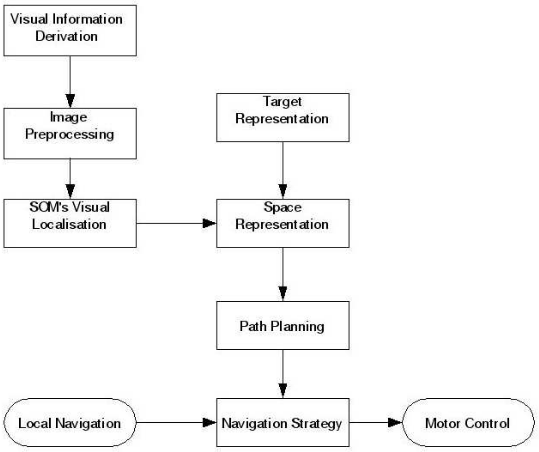
Hippocampal pyramidal cells called place cells have been discovered that fire when an animal is at a certain location in its environment [8]. In our model, place cells can provide candidate locations to the path integrator and place cells could localise the robot in a familiar environment. Self-localisation in animals or humans often refers to the internal model of the world outside. As seen in a white water maze experiment [10] even though a rodent was not given any landmarks, it could still reach its goal by forming its own internal representation of landmarks of the world outside. It is seen in humans and animals that they can create their own landmarks, depending on the firing of place cells. These cells change their firing patters in an environment when prominent landmarks are removed from the environment. With this evidence from computational neuroscience it is reasonable to assume that place cells could prove to be an efficient way of localisation using vision.

2.1 Architecture

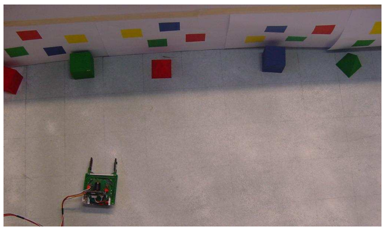
In figure 1 **(a)** visual information derivation is a module which is responsible for getting the images from the robot's camera. The image preprocessing module is responsible for normalising the images for the network. The SOM is responsible for the localisation of the robot in the environment. There are various ways in which the robot can be instructed as to where its target for navigation is. We are currently experimenting as to how to translate the place code output by the SOM's network into a spatial representation. Once we can map both the target representation and the current location of the robot into the same spacial representation, then we can have a simple path-planning module which directs the robot to its goal. The path planning module provides output to the motors. This forms the global navigation strategy. We have implemented the local navigation strategy using reactive behaviour. Both the global navigation strategy and the local navigation strategy meet each other in the navigation strategy module which is mostly responsible for choosing which navigation strategy to choose. Accordingly it chooses the output from either the global navigation strategy or local navigation strategy into motor control commands.

2.2 Visual Data Pre-processing

The images in original format were of 360 x 480 pixels. A total of 75 images were used for learning. Images were in 24 bit colour RGB format. In order to reduce the dimensionality of the images without losing valuable information, we first reduced the



(a) Overall Architecture



(b) Birds eye view of Scenario used

Fig. 1. (a) Architecture of Navigation strategy being developed. The focus of this paper is on visual landmark recognition. **(b)** This figure provides the birds eye view of the overall scenario which is being used for experiments. In this picture, we can see Khepera robot looking at the wall which have colour rectangles and pyramids and cubes kept adjacent to the wall.

dimensions of the colours within the images. We chose blue of the the RGB cube as given in the equation 4, here N represents the whole picture.

$$N = (a_{ijk}) \in \mathbb{A}^{3mn}, \quad i = 1, 2, \ldots, m \quad j = 1, 2, \ldots, n \quad k = 1, 2, 3 \tag{3}$$
$$B = (b_{ij}) \in \mathbb{A}^{mn}, \quad i = 1, 2, \ldots, m \quad j = 1, 2, \ldots, n \quad b_{ij} := a_{ij3} \tag{4}$$

$\mathbb{A}$ is is the set of possible pixel values. The values are between 0 and 255, and can be represented as $\mathbb{A} = \{0, 1, 2, \ldots, 255\}, \qquad m = 17, \qquad n = 27.$

The images were reduced to a size of 17 x 27. An example of the reduced image with only one colour pane of the cube can be seen in figures 2 and 3. This image in turn was converted into a single vector to be presented to the network. It was done as explained from equations 5 to 7.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \qquad A_i = (a_{i1}, \ldots, a_{in}) \qquad i = 1, 2, \ldots, m \tag{5}$$

$$V = (v_l) := (A_1, \ldots, A_m) \in \mathbb{A}^{mn} \qquad l = 1, 2, \ldots, mn \tag{6}$$

Equation 6 is a concatenation of A_i of A. In other words,

$$v_{(i-1)n+j} := a_{ij} \qquad i = 1, 2, \ldots, m \qquad j = 1, 2, \ldots, n \tag{7}$$

2.3 Self Organisation of Landmarks

The SOM learns to classify patterns. The network receives an input image of size 17 x 27 and the output is the activation of a cell to indicate the place where the robot is. The aim of the experiment was to study whether the robot can develop an internal representation of the world model at a symbolic level. It was studied by providing the robot with an environment where it can create its own internal representations, rather than giving explicit landmarks to the robot.

Each activation in the output layer of the SOM represents place codes. The results, which are given in figure 2 and 3 show how the SOM has classified different landmarks in different sections of the map, with neighbouring images having the same landmark; there are clusters of sections which are formed. Each cluster on the maps represents a place code, related to its landmark. Each neuron activation gives us a representation in x, y and z co-ordinates of the environment, where x represents the vertical coordinate of the environment, y represents the horizontal coordinate and z represents the head direction of the robot. These three inputs do not participate in the input of the neural network of the SOM's. They are purely deduced by the output of the SOM.

Cluster Formation The basic property of a SOM network is to form clusters of information relating to each other, in our case landmarks. In the next step we take advantage of these clusters to know which local area the robot is in and to give directions to the robot.

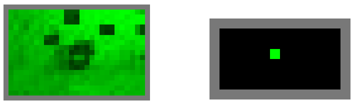
A cluster is a collection of neurons which are next to each other representing the same landmark. Figure 2 shows that when the robot was approaching the desired landmark, there were activations in the neighbouring neurons. This is due to clustering of similar images around the landmark. There are multiple similar images that are being represented by a single neuron, making the cluster a bit smaller and richer in information. On the other hand figure 3 shows the landmarks which were at a distance to the location represented in figure 2. Two landmarks that were given to the robot at a distance would be mapped not only in different clusters but also distant from each other.

By their very definition, landmarks are features in the environment. This was the reason behind a formation of these clusters by SOM's. The landmarks that were chosen by the SOM were quite significant in the image and distinguished features from the rest of the environment and other landmarks.

Noise handling by the Network A natural environment is full of noise. For a robust model of navigation the robot must still be able to navigate. At the present stage there are unknown clusters formed when the landmarks are modified or removed from the environment. This is similar to how place cells behave [8, 9, 1]. These clusters are also formed if there is noise added to the environment i.e. by adding or replacing objects in the environment.

(a) Input Image **(b)** Input Image **(c)** Neuron Activation **(d)** Input Image **(e)** Neuron Activation

Fig. 2. Here image **(a)** and image **(b)** have the same neuron activation as seen in **(c)**. These two images were taken at a close distance overlooking the same landmark. Image **(d)** has a different activation in the neighbourhood as seen in **(e)**, because the images and location were not very different from (a) and (b).



(a) Input Image **(b)** Activation

Fig. 3. The input image represented in figure **(a)** is different from those images in figure 2 and as we can see the activations are also at a distance to each other i.e. the cluster of this image is at a significant distance compared to the cluster of the images represented in figure 2.

In figure 4 the first half of the input image was cleared and black colour was presented in the image as seen in the figure. Even with 50% covered the landmark is still visible to the network. This behaviour was also observed with other landmarks and clusters. The network is able to do pattern completion and is trained to have its focus on a particular significant landmark in the environment.We have seen that the network starts to become uncertain once the noise level reaches around 50%.

(a) Image with 50% noise **(b)** Activation **(c)** Image with 80% noise **(d)** Activation

Fig. 4. In image **(a)** we see that even if there is high amount of noise the activation as seen in **(b)** is in the neighbourhood of the original activation as seen in figure 2 whereas there is a random, but specific activation with images with 80% of noise added to them

The aim of the experiment was to explore the place code to localise the robot in the area. Therefore it is less important which neuron in the cluster is active, but the neuron in the cluster relating to place code is important for localisation. Therefore our measurement of robustness for a particular place code falls into two categories either right or wrong. In figure 5 we examine the network's response to the noise for a particular place code. We have shown the effectiveness of the networks averaged over all the place codes in the network shown in figure 5.

We have use the same place code as shown in the images in figure 2. In figure 5 we can clearly see that the correct individual neuron is activated even with 40% noise which demonstrates effective noise handling by the network. At the cluster level we can see that the noise handling by the network is even better as it can handle noise upto 60%. Noise resistance could be improved if the responses were averaged over several images.

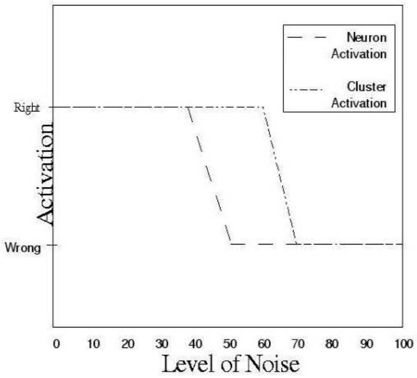


Fig. 5. Here the correctness of the activation of a location shown in figure 2. The x-axis represents the 10% of noise introduced and the y-axis represents the whether the activation was correct or not. The correctness of activations were examined at two levels the first being correctness of a neuron being active and the second being the correctness of a neuron in the right cluster being active.

Unknown clusters caused by high volumes of noise, would not matter a lot in navigation as long as most of the landmarks in the environment are not removed. The property of place cells shows us that an animal or a human can navigate even if there are significant landmarks missing from the environment, since the place cells continue to fire. It is only when a series of landmarks are not found that the animal or a human would get lost, which is similar to the behaviour of our model.

3 Conclusion

In this paper we have described a place cell model based on a SOM for navigation. The model was successful in learning the locations of the landmarks even when tested with distorted images. Visual landmarks were associated with the locations in a controlled environment. This model clusters of neighbouring landmarks next to each other. The landmarks that are away from each other are also relatively away from each in the map. Rather than preprogramming localisation algorithms as internal modules, our SOM's architecture demonstrates that localisation can be learnt in a robust model based on external hints from the environment. The noise handling of the network is also very efficient. This model was developed to learn landmarks in an environment, by having maps divided into cluster of neurons for different parts of the environment. This model is successful in localising the robot in an environment.

4 Acknowledgements

This research is partially sponsored by a FET-IST grant for MirrorBot project, by the EU to Prof. Wermter, under grant IST-2001-35282. We are grateful to Christo Panchev, Wolfgang Schmidle and Garen Arevian for their support on the paper.

References

1. Angelo Arleo. *Spatial Learning and Navigation in Neuro-Mimetic Systems, Modeling the Rat Hippocampus*. PhD thesis, Swiss Federal Institute of Technology, Lausanne, EPFL, Switzerland, 2000. http://diwww.epfl.ch/mantra/team/arleo/.
2. Angelo Arleo and Wulfram Gerstner. Spatial cognition and neuro-mimetic navigation: A model of hippocampal place cell activity. *Biological Cybernetics, Special Issue on Navigation in Biological and Artificial Systems*, 83:287–299, 2000.
3. Peter Bonasso and Robin Murphy, editors. *Artificial Intelligence and Mobile Robotics: Case Studies of Successful Robot Systems*. The MIT Press / AAAI Press, London, 1998.
4. W. Burgard, A.B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2):3–55, 1999.
5. Michel Haritoppulos, Hujun Yin, and Nigel M. Allinson. Image denoising using self-organisising map-based non-linear independent component analysis. *Neural Networks: Special Issue, New Developments in Self Organising Maps*, 15(8-9):1085–1098, October - November 2002.
6. Robin R. Murphy. *Introduction to AI Robotics*. The MIT Press, London, England, 2000.
7. Ulrich Nehmzow. *Mobile Robotics: A Practical Introduction*. Springer Verlag, London, 2000.
8. A. D. Redish and D. S. Touretzky. Navigating with landmarks: computing goal locations from place codes. 1997.
9. A. David Redish. *Beyond Cognitive Map*. PhD thesis, Carnegie Mellon University, Pittsburgh PA, August 1997.
10. A. David Redish. *Beyond Cognitive Map from Place Cells to Episodic Memory*. The MIT Press, London, 1999.
11. Edmund Rolls and Gustavo Deco. *Computational Neuroscience of Vision*. Oxford University Press, New York, 2002.
12. S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Hähnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. Probabilistic algorithms and the interactive museum tour-guide robot minerva. *International Journal of Robotics Research*, 19(11):972–999, 2000.