

Tina Memo No. 1999-002

Short Version published in: proc. BMVC, 800- 809, Southampton, 1998.

A Feature Representation for Map Building and Path Planning.

N.A.Thacker and A. J. Harris.

Last updated
19 / 10 / 1998

This document forms part of the **Recognition and Intelligence Series** available from www.tina-vision.net.

- 2007-001 Retinal Sampling, Feature Detection and Saccades: A Statistical Perspective.
- 2006-008 Statistical Principles for Selection of Computer Vision Algorithms as Modules for Visual Perception - Show Me the Errors.
- 1991-001 Designing a Layered Network for Context Sensitive Pattern Classification.
- 1997-002 Supervised Learning Extensions to the CLAM Network.
- 1996-003 Tutorial: Algorithms For 2-Dimensional Object Recognition.
- 1997-005 Speechreading Using Probabilistic Models.
- 2000-002 Solving Shape Based Object Recognition from a Computational Standpoint - Practical and Physiological Constraints.
- 1995-004 Assessing the Completeness Properties of Pairwise Geometric Histograms.
- 1996-004 Robust Recognition of Scaled Shapes Using Pairwise Geometric Histograms.
- 1996-005 Multiple Shape Recognition Using Pairwise Geometric Histogram Based Algorithms.
- 2007-007 Automatic Identification of Morphometric Landmarks in Digital Images.
- 1999-002 A Feature Representation for Map Building and Path Planning.
- 2001-015 Colour Image Segmentation by Non-Parametric Density Estimation in Colour Space.
- 2001-006 What is Intelligence?: Generalised Serial Problem Solving.
- 1994-002 A Correlation Chip for Stereo Vision.
- 1995-001 Specification and Design of a General Purpose Image Processing Chip.



Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Feature Representation for Map Building and Path Planning.

Abstract

We explain how most conventional approaches to the problem of visually guided robot navigation are fundamentally restricted. Starting with a specification of the requirements of a general purpose visually guided robot which includes, robustness, autonomy, flexibility and scalability, we then develop an approach to the solution of map building and path planning which operates within a learning framework. We suggest a new representation of shape using features based on geometric co-occurrence which has the capabilities of providing such metric information and is expected to have similar recognition properties to other techniques based on relational geometric measures.

1 Introduction

Many research groups have attempted to build autonomous visually guided robots. These systems can generally be divided into two categories, a top down approach based on world models [1, 2, 3] and a bottom up approach based on layered responses (such as the subsumption architecture [4]). The first of these is limited by its ability to deal with the complexities of the real world and is generally restricted to indoor or man-made environments. It has been remarked that such approaches lead to brittle inferences and that the heuristic assumptions used lead to strong domain-specific dependencies [5]. Further, such approaches are not supported by biological evidence. The second approach is more responsive, but rarely develops truly advanced capabilities. The goal of the research presented here is to design a system capable of automatically learning a visual domain so that it can subsequently execute a visually controlled navigation from any location to another when instructed to do so. The key features of this work are: the acceptance of the need for representations of features in the world which are not based on preconceptions of regular structure and the key role of learning and flexibility. The idea is to construct a system capable of organising the information from a visual domain in such a way that it can later make use of it to solve real problems in planning and control. Our approach has a similar motivation to the work of Waxman [6, 7] and is under-pinned by biological models of cognition [8]. As we will demonstrate, the problems involved with this approach involve not only having to be able to reliably recognise visual locations, but also be able to use visual information to estimate physical distance. Our list of requirements has far broader scope than many other attempts at visually guided control designed to operate in man-made or semi-structured environments. The scale of the problem is quite daunting and we will only have the space here to outline the approach and present some preliminary (though encouraging) results.

We start by defining the required capabilities of the system, which can be summarised as; flexibility, autonomy, scalability and robustness. It must be flexible in order to cope with a broad range of problems, for example indoor and outdoor path planning and a large class of visual features i.e. the ability to function in as broad a range of circumstances as possible. Autonomy is the ability to function with data collected only during the normal operation of the system. This means that we do not wish to have to use any a priori knowledge of the problem domain (including a world model or fixed topology). Scalability is simply the requirement that the algorithmic solution time and data structures should scale no worse than linearly with the physical size of the problem. In this way we may be able to implement the resulting system for real time execution on a parallel processing platform. Finally, robustness simply means that the algorithms should cope with any expected changes in visual data which may be expected from all possible application environments.

Many of the difficulties we expect to encounter are very similar to those associated with view based object recognition. Recognition of location can basically be seen as view based object recognition, but from the inside. The representation technique we intend to develop is based on the statistical method of geometric histograms which have recently been shown to have great potential as general purpose object recognition techniques [9]. In particular these methods can be designed to cope with variability of: scene illumination, visibility of features, cluttered scenes, scene complexity, restricted fields of view. All of which are crucial in obtaining a working recognition system and will also be necessary for map building. The dependency of the representation on viewpoint, which may be seen by some as weakness for object recognition, is needed if a mobile system is to be controlled using visual feedback. Seen this way, view based recognition itself has the analogy of visually guided assembly, implying that view-based object recognition may be of more value than the long sought after ideal of full transformation invariance in practical circumstances. Also, we may hope that our work on path planning may ultimately transfer back to solutions for intelligent visually guided robot manipulation.

The initial motivation for this work comes from our previous efforts to design a statistically based representation architecture [10, 11] or neural network. This has the ability to self-generate in a well defined manner and can be

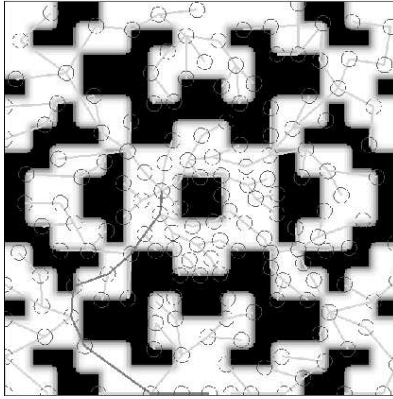


Figure 1: Results from the software prototype.

thought of as a feature based recognition system. Standard techniques for planning routes between fixed nodes in a topological framework also exist. The question then was to see if it is possible to make use of both classes of algorithm to design a highly autonomous visually guided system for map building and path planning. This way we hope to avoid the need for a predefined model of the environment or a rigorous spatial map.

2 Overview

We start by defining individual visual beacons in an image for recognition purposes. The beacons constitute well located and reproducibly observable visual features, in a manner similar to interesting features as defined by Moravec [12]. We define a location in the visual feature map as a conjunction of visual beacons. These are generated automatically during an exploration of the environment. By associating locations on the visual map only with previously visited sites and then restricting the path planning process to operate on these points we can generate candidate paths which are executed solely on known positions.

Many existing methods work with a simplified model of the environment but in order for any method to be automatic we cannot know the topology of the environment in advance. We may not, for example, know that the domain is simply two or even three dimensional. We must also have a way of comparing our location representations to determine which are spatially adjacent. Given a method of establishing costs between two visually similar locations in the world, it is possible to find a simple path from any known point to any other using a so called ‘fire-front’ algorithm which is described briefly below.

A software prototype has been developed which consists of a planar environment with simulated visual features distributed randomly across it. For this work we have assumed that the response of these features is continuous with changes in viewing angle and then generate a response vector for every possible location in the environment. Referring to figure 1, the circles indicate locations in the world which are visually dissimilar (using our location distance metric) and the dark regions are flagged as impassable to the path planner. In the real system, we will require an independent mechanism for vetoing impassable routes even though they show close visible similarity. We envisage that this may be done using a stereo vision system or trial and error learning (using collision detection) and will not be discussed further here.

The path planning algorithm executes on a mesh of known visually similar map node locations starting from the target node. For each connected node all possible connections are examined and the minimum cost path is stored at each location. This process is iterated to construct a minimum cost to all other points in the space. It is then possible to reconstruct the minimum cost path from the current location to the target by following the connections to minimum cost locations. We initially attempted to use visual similarity as the cost function for the distance between locations, but whilst this results in visually optimal paths, they are not necessarily physically optimal. Replacement of a visual measure with a physically meaningful distance yields an immediate solution to this problem (the bold path in figure 1 illustrates one such solution).

Previous experience with self generating object recognition systems tells us that the automatic generation of the associative memory needed to encode this map is not likely to be a difficult problem, particularly if we are free to design our own exploration strategy. As pointed out in [13], the main obstacle therefore in the use of such a path planning process is the lack of a suitable visual based representation. This should provide well localised objects which are sufficiently unambiguous for reliable recognition and also a means for estimating distances between

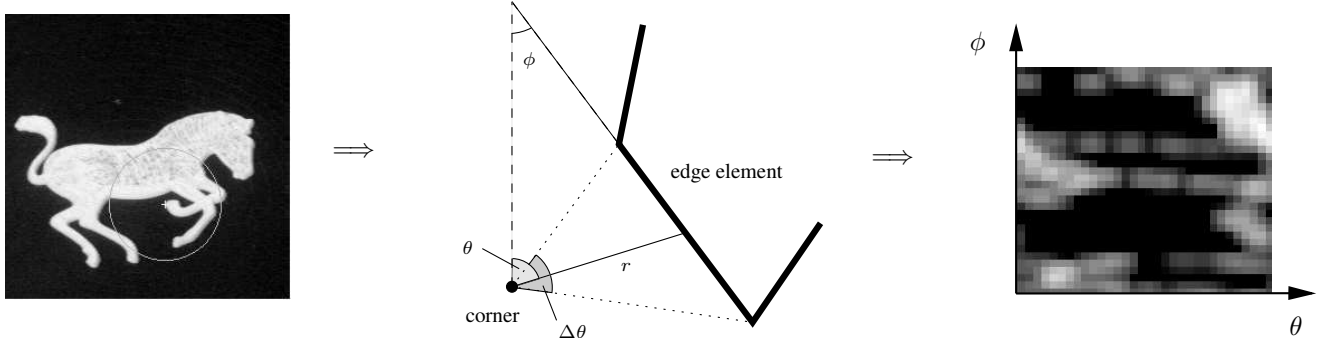


Figure 2: Constructing the cornerwise geometric histogram.

visual locations. Therefore, in this paper, we have concentrated on addressing this problem. A new technique for representing local shape, which we call Corner-centred Geometric Histograms (CGH) has been designed with a matching response behaviour which makes distance calibration, and visual feedback control, possible.

3 Corner-centred Geometric Histograms

We needed a method for encoding visual features from specific views in the world in a manner which is robust to noise, lighting effects and some occlusion. For this work, we use a novel form of geometric histogram which we call the Corner-centred Geometric Histogram (CGH). Corners and edges are extracted from the scene and a histogram generated for each chosen corner, encoding the relative orientation and distance between a corner with its surrounding edges.

We define the weighting for an entry at a given θ , ϕ co-ordinate as:

$$w_{\theta,\phi} = \sum_i \frac{\Delta\theta g_{\theta,\phi}}{r_i^n + k}$$

where i is the number of edge elements, r is the distance from the corner to the mid point of the edge element, $\Delta\theta$ is the angle presented by the edge element at the corner, $g_{\theta,\phi}$ is a linear measurement kernel and k is a small constant added to remove infinities.

A radial weighting is used to reduce computation and limit the effects of our finite image size. Less importance is attributed to those edges further from the corner which are more numerous due to the larger areas involved. This strategy has some biological precedence, as the log-polar sampling on the retina results in a similar weighting strategy. The $\Delta\theta$ weighting ensures that edge features that pan directly through the centre corner do not give infinite weighting.

As with the work in [9], the dimensions of the bins of the histogram must be carefully chosen to reflect the variance of noise in the system. If the bins are too small, the noise will produce different histograms for the same corner, too large and discriminability is compromised. Ideally entries should be made in the histogram which reflect probabilities of geometric co-occurrence in agreement with our previous research. By following this procedure the generated histograms are expected to be; robust to illumination, rotation invariant and generate histograms which respond to scale change (i.e. viewing distance) in a manner which is dominated by an overall normalisation change.

3.1 Histogram Response

Given the CGH definition, we need a way of generating meaningful recognition responses for an arbitrary corner feature from a set of stored histograms. The similarity metric we use is based on the Bhattacharyya distance [14]. We have proven this to have the correct statistical properties for comparing frequency coded distributions [15] and it forms a sensible starting point for a similarity function within an artificial neural network.

$$B = \frac{\sum_{ij} h_{ij} s_{ij}}{|\sum_{ij} h_{ij}^2|^{\frac{1}{2}} |\sum_{ij} s_{ij}^2|^{\frac{1}{2}}}$$

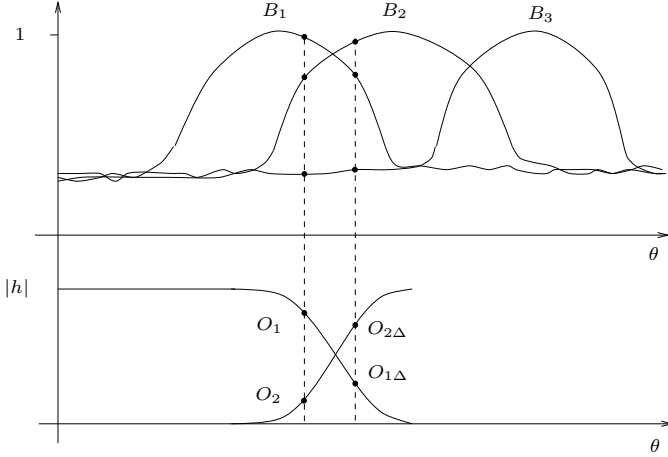


Figure 3: Competitive response function.

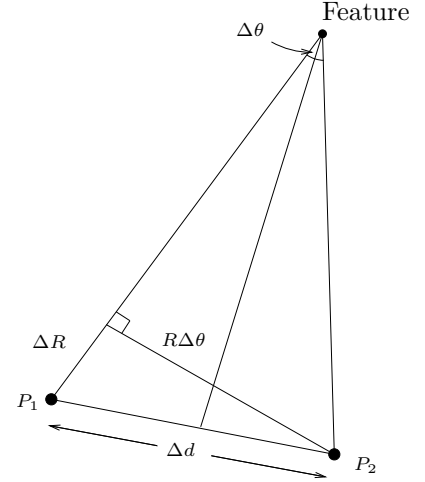


Figure 4: Polar geometry between two locations.

In order to perform distance calculations we need a mechanism for converting a set of responses to physical distance. This is not a trivial task, as the method used must be robust to large first order changes in the histograms due to effects such as occlusion. The approach taken uses what is effectively a competitive mechanism in order to generate a well behaved response function which can be approximated locally by calibration of linear factors. The equations are of the form:

$$O_1 = \frac{(B_1 - B_3)}{(B_1 - B_3) + (B_2 - B_3)} |h| \quad \text{and} \quad O_2 = \frac{(B_2 - B_3)}{(B_1 - B_3) + (B_2 - B_3)} |h|$$

where B_1 B_2 B_3 are the three largest responses in the set and $|h|$ is the modulus of the input histogram. Such competitive calculations have been shown to be good models of physiological neuronal functions.

Provided that a change in physical distance to a corner from the view location produces a strong first order change in the CGH which is effectively only a normalisation factor (scaling as some power of the distance to the corner feature n), the above outputs will behave in a way that angular and radial distance changes to the feature can be decomposed and computed from a change in the pair of outputs.

These physical dimensions can be estimated using:

$$\frac{\Delta R}{R} = 2 \left(\frac{|h_{P1}|^n - |h_{P2}|^n}{|h_{P1}|^n + |h_{P2}|^n} \right) = 2 \left(\frac{(O_2 + O_1)^n - (O_{2\Delta} + O_{1\Delta})^n}{(O_2 + O_1)^n + (O_{2\Delta} + O_{1\Delta})^n} \right) \quad (1)$$

$$\text{and} \quad \Delta\theta = k_1 \left[\left(\frac{O_2 - O_1}{O_2 + O_1} - \frac{O_{2\Delta} - O_{1\Delta}}{O_{2\Delta} + O_{1\Delta}} \right) \right] \quad (2)$$

Formulating the changes as ratios in this way produces dimensionless quantities which are less sensitive to missing data in the input histogram and easier to calibrate.

Now referring to figure 4, assuming $\Delta d/R$ is small then:

$$\Delta d_2 = k_2 (\Delta R^2/R^2 + \Delta\theta^2) \quad \text{where} \quad k_2 = R^2$$

The constants k_1 and k_2 are linearisation factors which can be stored locally for each pair of responding features. The angular constant can be calibrated by observing the movement of corner features on the image plane during the exploration, as the rotation about a feature is equivalent to an angular motion on the image plane for a fixed viewing axis. The remaining factor can be calibrated provided we have a means of measuring small physical movements, e.g. via odometry or a stereo vision system. Once these constants are determined we can now successfully compute the local distance to any visual location on the basis of each feature visible from any given location (see below). This redundancy supports reliable distance calculation via a simple voting strategy, allowing the method to be robust to incorrectly recognised features. Local calibration permits representations of similar visual objects to be re-used at different sites in the topological map without any adverse affects from feature ambiguity.

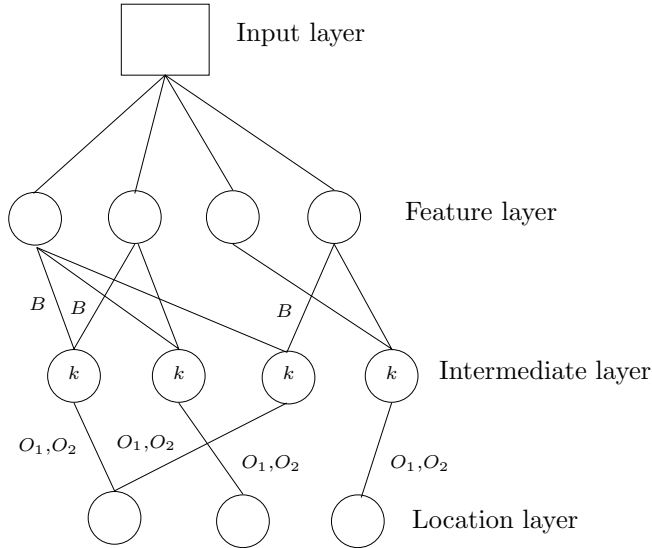


Figure 5: The CLAM architecture for this application.

4 Map Building

The map building process involves exploring the environment, recording new visual features and locations from where these features are visible. The map can be totally self-generating and updating with new features and locations added as required and unreliable ones eventually discarded. The map is to be stored using algorithms based on a novel neural network, the Contextual Layered Associative Memory network (CLAM), originally designed as the basis of a learning view-based recognition system. The CLAM network is a self-generating structure which implements a probabilistic encoding scheme and employs training algorithms geared towards the unsupervised generation of layerable associative memory [10]. The CLAM network is based upon sound statistical principles and can be guaranteed to learn any problem provided the statistical requirements on the behaviour of the input data are satisfied. These conditions effectively require a frequency based data encoding, such as a histogram or other probability estimators.

The structure proposed is illustrated in figure 5. Feature representation nodes will be added when a CGH is generated which is sufficiently different from any other in the network. When adding a new location node, we add an intermediate node for each CGH generated from that location. This node links the highest responding feature nodes to the new location, storing the calibration constants (k) and competitive response function results (O).

5 Navigation

In order to navigate successfully in the world, we require the abilities of: navigation from any (arbitrary) location to a visually similar location in our map (local navigation); navigation between any two locations stored in our map (global navigation).

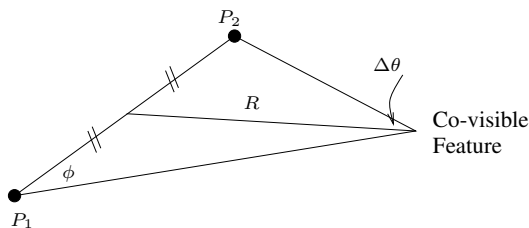


Figure 6: Navigating between two points.

To execute a local navigation, we first generate an estimate of the vector between the points from each co-visible feature, so a high visual similarity constitutes a more accurate vector. Referring to figure 6, we see that:

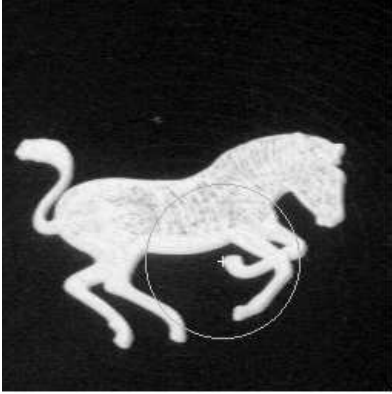


Figure 7: Horse test image

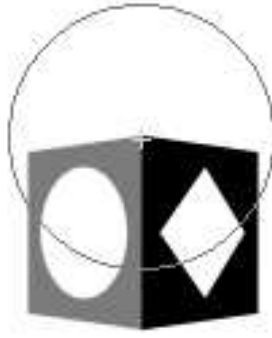


Figure 8: Cube test image

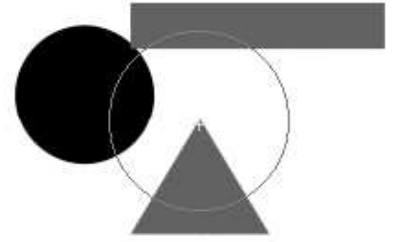


Figure 9: Shapes test image

$$\phi = \tan^{-1} \left(\frac{R\Delta\theta}{\Delta R} \right)$$

therefore, ϕ can be found from equations 1 and 2 using the responses found from the current location and those stored in the map for the target location. As with the mechanism for distance calibration, one estimate of heading is available from each co-visible feature, and the best heading direction can be estimated using a robust voting process. The vector can then be traversed using visual feedback to confirm the trajectory. Global navigation is achieved by executing a series of local navigations specified by the path planner.

6 Results

The preceding sections are based on two assumptions regarding the behaviour of a CGH match response to angular and radial distance movements relative to the viewed beacon (feature). In particular, in order that the angular and radial distances can be estimated the response of each histogram to each of these motions must be decoupled. We assume that a radial movement causes a first order re-normalisation and that the angular response is smooth and differentiable. Provided that these two statements are correct then the above algorithms for distance calibration must be correct at some scale as they are based on local linearised approximations to feature change.

To investigate the performance of our representation we used both simulated and real images. The image in figure 7 was captured using a test frame which allows us to rotate the camera on a fixed radius from the object. The distance response was measured by re-sampling the image at different scales. The images in figures 8 and 9 are synthetic which enabled us to accurately place the camera in the scene to produce both angle and distance responses. Plots 10 and 11 show similar results from both methods.

The angle results meet our requirements in that they are continuous, but due to the different gradients, reiterate the need for calibration of feature map nodes. The distance response shows a level of invariance which would allow the decoupling of angle and radial movement.

7 Conclusions

We have outlined a set of algorithms which would enable a mobile camera platform to store a visual map of a world, to plan and execute a visually controlled navigation from any one location to another. The use of an associative memory gives flexibility allowing the system to adapt and incrementally store new information, proportionally with the complexity of the environment. The new geometric histograms are based on a representation which may be constructed from radial polar imaging systems, such as the retina. They have the required statistical properties for use with a probabilistic network and we have shown that it is possible to compute robust estimates of physical distance and estimates of headings between map locations. This property makes possible, both path planning within the associative memory map and visual feedback control for navigation. The combination of algorithms thus provide a basis for a flexible, autonomous, robust system, with the potential of implementation on fast parallel hardware, meeting our initial criteria.

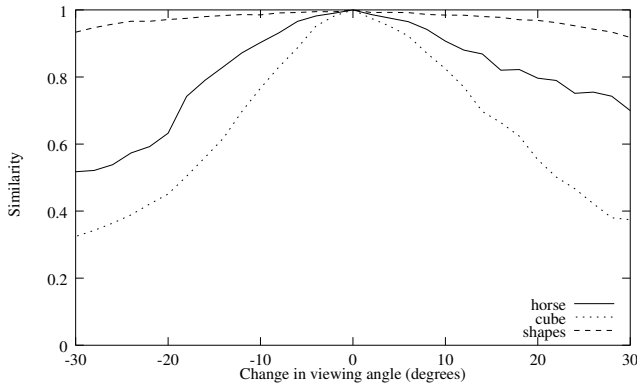


Figure 10: Angle responses

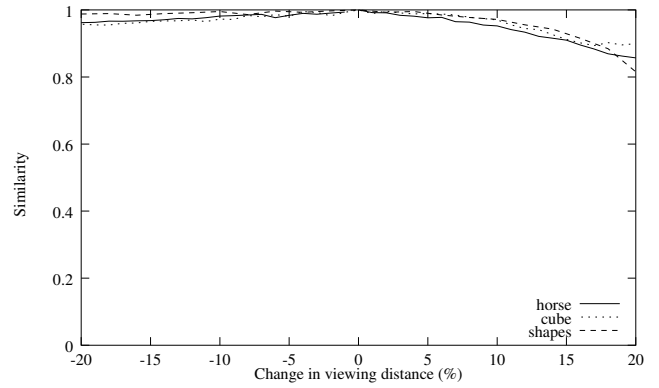


Figure 11: Normalised distance responses

We have made many assumptions in the algorithms which we realise may not always hold. A key issue with this work is however, that the vast quantity of data available implies that each result is obtained in a redundant and therefore robust manner. It is our intention that unreliable sources of information will be eliminated from consideration during the learning process.

Much of the material has been devised from previous experience and the algorithms are currently fairly immature in terms of practical demonstrators. This paper is restricted to setting out the framework for the approach and showing that the new feature representation will support the calculations needed for our task. Beyond this, the entire computational approach has a strong analogy with physiological systems, both in terms of approach and expected capabilities. What we have demonstrated so far is that feature based representational schemes can be joined with adaptive associative memories to solve the problem of map building and path planning. Our next task is to evaluate the full learning system within a simulated environment.

References

- [1] M.Asada. Map building for a mobile robot from sensory data. *IEEE Trans. Sys. Man and Cyber.*, 37(6), 1990.
- [2] C.Taylor and D.Kriegman. Vision-based motion planning and exploration algorithms for mobile robots. *Tech.rep. Yale University*, 1993.
- [3] E.Baumgartner and S.Skaar. An autonomous vision-based mobile robot. *IEEE Trans. Auto. Con.*, 39(3):493–502, 1994.
- [4] R.Brooks. A robust layered control system for a mobile robot. *IEEE Trans, Robotics and Automation*, RA-2(1), 1986.
- [5] A.Elfes. Using occupancy grids for mobile robot perception and navigation. *Computer Magasine*, June 1989.
- [6] I.Bachelder and A.Waxman. Neural networks for mobile robot visual exploration. *SPIE*, 1831(7):107–118, 1992.
- [7] A.Baloch and A.Waxman. Visual learning, adaptive expectaion, and behavioural conditioning of the mobile robot mavin. *Neural Networks*, 1(4):271–302, 1991.
- [8] J.Okeef and L.Nadel. The hippocampus as a cognitive map. *Clarendon Press, Oxford*, (ISBN0198572069), 1978.
- [9] N.A.Thacker, P.A.Riocreux, and R.B.Yates. Assessing the completeness properties of pairwise geometric histograms. *Image and Vision Computing*, 13(5):423–429, June 1995.
- [10] N.Thacker and J.E.W.Mayhew. Designing a network for context sensitive pattern classification. *Neural Networks*, 3(3), 1989.
- [11] I.A.Abraham N.Thacker and P.Courtney. Supervised learning extensions to the clam network. *Neural Networks*, 10(2):315–326, 1997.
- [12] H.P.Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. *Stanford Artif. Intel. Lab. Memo*, (AIM-340), 1991.
- [13] Krose and Eecen. A self-organising representation of sensor space for mobile robot navigation. *Proc, Int. Conf. Intel. Rob. and Sys. IROS'94*, pages 9–14, 1994.
- [14] A.Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, (35):99–110, 1943.
- [15] N.A.Thacker, F.J.Ahearne, and P.I.Rockett. The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 1997.