

Towards Bio-inspired Place Recognition over Multiple Spatial Scales

Zetao Chen¹, Adam Jacobson¹, Uğur M. Erdem², Michael E. Hasselmo² and Michael Milford¹

¹School of Electrical Engineering and Computer Science

Queensland University of Technology

²Center for Memory and Brain and Graduate Program for Neuroscience

Boston University

zetao.chen@student.qut.edu.au

Abstract

This paper presents a new multi-scale place recognition system inspired by the recent discovery of overlapping, multi-scale spatial maps stored in the rodent brain. By training a set of Support Vector Machines to recognize places at varying levels of spatial specificity, we are able to validate spatially specific place recognition hypotheses against broader place recognition hypotheses without sacrificing localization accuracy. We evaluate the system in a range of experiments using cameras mounted on a motorbike and a human in two different environments. At 100% precision, the multi-scale approach results in a 56% average improvement in recall rate across both datasets. We analyse the results and then discuss future work that may lead to improvements in both robotic mapping and our understanding of sensory processing and encoding in the mammalian brain.

1 Introduction

The vast majority of robotic mapping and navigation systems perform mapping at either one fixed spatial scale or over two, a local and a global scale [Bosse, et al., 2003, Kuipers, et al., 2004, Kuipers and Byun, 1991]. A range of recent high profile discoveries in neuroscience have demonstrated that animals such as rodents, and likely many other mammals including humans, encode the world using multiple parallel mapping systems each of which encode the world at a different scale [Stensola, et al., 2012, Hafting, et al., 2005]. In rodents, the mapping system scales from neurons that encode an area of a few square centimetres to neurons that encode an area of several square metres, with many intermediate scales

represented in-between. There is also currently no known upper limit on scale, as practical limitations mean experiments cannot be performed in arbitrarily large environments.

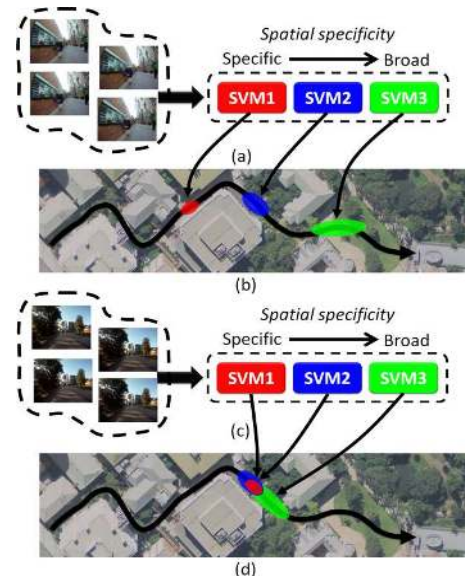


Figure 1: Our multi-scale place matcher combines the output of multiple arrays of SVMs trained to perform place recognition at different scales in order to filter out hypotheses not supported at all spatial scales. (a) Camera frames are input into each place recognition module leading to inconsistent place recognition hypotheses at different spatial scales, and hence (b) no match is reported. (c) A later set of camera frames leads to (d) consistent place recognition hypotheses at all scales, and hence the most spatially specific hypothesis is accepted.

While a number of theoretical-only investigations have hypothesized possible benefits of such a multi-scale mapping system [Burak and Fiete, 2009, Welinder, et al., 2008], no one has investigated potential mapping and place recognition performance benefits in real world environments. In this paper, we propose that these multiple scales provide a mechanism for effectively combining place recognition hypotheses with varying spatial specificities – for example, combining a spatially specific hypothesis that a rat or robot is located at a specific corner of a room with a spatially broad hypothesis that it is located “somewhere” in that room. We further propose that the

ZC, AJ and MM are with the School of Electrical Engineering and Computer Science at the Queensland University of Technology, zetao.chen@student.qut.edu.au. UE and MH are with Center for Memory and Brain and Graduate Program for Neuroscience at Boston University. This work was supported by an Australian Research Council Discovery Project DP1212775 awarded to MM, and an Office of Naval Research ONR MURI N00014-10-1-0936 and Silvio O. Conte Center Grant P50 NIMH MH094263 to UME and MEH.

reliability of spatially specific place recognition hypotheses can be improved by validating them against spatially broad hypotheses, without sacrificing spatial accuracy (Figure 1). This approach differs from traditional probabilistic robotic methods, where place recognition spatial specificity is passively driven by models of sensor uncertainty. Instead we intentionally create parallel learning systems that learn associations between sensory input and the environment at different spatial scales.

We present a biologically inspired, multi-scale framework that uses arrays of Support Vector Machines, with each array trained to perform place recognition at a different scale, and methods for combining different resolution place recognition hypotheses. We present results from two real world experiments comparing single- and multi-scale place recognition performance that demonstrate that multi-scale recognition leads to significantly improved recognition performance.

The paper proceeds as follows. Section 2 provides an overview of place recognition and mapping techniques. In Section 3 we describe the components of the multi-scale place learning system. The experiments are described in Section 4, with results presented in Section 5. Finally we conclude the paper in Section 6 and discuss ongoing and future work.

2 Background

In this section we briefly review recent evidence regarding multi-scale maps in the mammalian brain and single and two-scale robotic mapping approaches including bio-inspired methods.

2.1 Robotic Mapping Methods

Mapping research has been a core area in robotics research for the last two decades. A large number of methods have been developed to solve the problem of mapping and localizing, often simultaneously [Dissanayake, et al., 2001], within an environment, using a range of sensor types and mapping algorithms. Here we mention some of the key systems, although the list is by no means meant to be exhaustive.

Extensive research in vision-based mapping and place recognition algorithms has resulted in methods including FAB-MAP [Cummins and Newman, 2009, Cummins and Newman, 2008], MonoSLAM [Davison, et al., 2007], RatSLAM [Ball, et al., 2013, Milford and Wyeth, 2010, Milford, 2008, Milford and Wyeth, 2008], FrameSLAM [Konolige and Agrawal, 2008] and SeqSLAM [Milford, 2013, Milford and Wyeth, 2012]. Hybrid approaches that combine local and global mapping algorithms have also been proposed. Atlas [Bosse, et al., 2003] is a hybrid SLAM algorithm which implements a hybrid metric-topological mapping framework. Similarly, a hybrid extension to the Spatial Semantic Hierarchy [Kuipers, et al., 2004, Kuipers and Byun, 1991] builds local maps using metric SLAM methods but represents the structure of large-scale space using a topological map. The mapping frameworks in these and other approaches have generally been heterogeneous, in that different types of

maps are used at different scales, and limited to two distinct scales.

2.2 A Multi-Scale Neuronal Map

In contrast to mobile robots, which are typically only capable of operating within one specific type of environment, rats are the second most widespread mammal after humans. Wild rats found in Alaska are genetically similar to “urban” rats that spend their time living in the walls of a house or sewer, as are burrow rats that never come above ground and rats living in an open air Asian market or Australian garbage dump. Rats across all these environments map and perform place recognition using the same neuronal machinery. Underpinning these impressive rodent capabilities is a sophisticated sensory processing and mapping system.

The rodent entorhinal-hippocampal (EC) formation, and more specifically the *medial entorhinal cortex* (MEC) stores spatial information (a map) in the form of a highly regular, grid-like representation of space [Hafting, et al., 2005]. The primary spatially-responsive cell in MEC is called a “grid” cell [Hafting, et al., 2005] – grid cells fire whenever the rodent (and likely other animals such as bats [Heys, et al., 2013]), is located at a vertex of a regular grid of locations over the environment. Recent evidence has shown that grid cells encode *multiple, discrete* scales of grid locations, in steps of approximately $\sqrt{2}$ [Stensola, et al., 2012]. The area encoded by a cell at each grid vertex can vary from a few square centimetres to tens of square metres. The upper limit, if there is one, is unknown. This integrated, multi-scale representation has been shown to have a number of theoretical advantages including efficient mapping of arbitrarily large environments [Burak and Fiete, 2009, Welinder, et al., 2008].

2.3 RatSLAM

RatSLAM is a computational model of the part of the rodent brain thought to be responsible for mapping called the hippocampus. The system creates a simplified single-scale neural spatial map similar to that stored in the MEC [Hafting, et al., 2005]. RatSLAM has been demonstrated in a number of experiments including mapping of the longest path by a visual SLAM algorithm at the time [Milford and Wyeth, 2008], and a long term delivery robot experiment [Milford and Wyeth, 2010]. These experimental results were achieved using only low resolution, whole image matching techniques, typically around 1000 pixels [Milford, et al., 2011] and a single scale neural map. Much of the representational power of single scale implementations like RatSLAM is drawn from the temporal dynamics of the network over time. Using multiple networks with different mapping scales offers the potential to add a powerful additional combinatorial mechanism for improving place recognition performance.

3 Approach and Methodology

In this section we describe the key place learning algorithms, the method for combining place recognition hypotheses with different spatial scales and

the two feature-based image pre-processing methods used in these experiments.

3.1 Learning Algorithms

In this paper we use arrays of Support Vector Machines (SVMs) to create a system that is capable of recognizing a specific spatial region in an environment. SVMs [Vapnik, 1998] are an effective method of finding the optimal hyperplane separating training data comprising two or more classes, and hence are suited to our task. In this context, the SVMs provide a practical means of implementing a place recognition system without the extensive tuning required of more biologically plausible grid cell models including continuous attractor networks [Milford, 2008]. This advantage is especially relevant given we are training a number of SVMs in parallel, each with its own distinct spatial scale.

We denote the data and its label as (x_i, y_i) where $x_i \in R^N$ is an N-dimensional feature and $y_i \in \{+1, -1\}$ is its label. With the assumption that the data can be separated by a hyperplane in some Hilbert space H, we search for the optimal separating hyperplane that maximizes its distance to the closest points in the training data, resulting in a discriminant function:

$$f(x) = \sum_{i=1}^m a_i y_i K(x_i, x) + b \quad (1)$$

The sign of $f(x)$ indicates the classification result and a_i and b can be found by solving a constrained minimization problem, which can be efficiently calculated using the SMO algorithm [Cristianini and John, 2000]. Because visual place recognition data is typically non-linearly separable, a slack variable is introduced to enable violations of the classification margin while minimizing the classification error. A parameter C controls this trade-off.

The kernel function $K(x_i, x)$ maps the input data to a high dimensional feature space H where the non-linearly separable data may be separated linearly. We use the popular RBF kernel here.

3.2 Place Recognition using an Array of SVMs

To perform place recognition, image frames are grouped temporally into segments and undergo dimensional reduction. Details of image clustering are described in Section 4.2. In general, video frames are grouped into different arrays of consecutive clusters. Cluster sizes are the same within one array but different between different arrays. Figure 3, for example, contains three different arrays of clusters with cluster sizes decreasing from top to bottom. Each of these frame clusters is input into the array of trained SVMs to produce a corresponding array of firing scores, where each score represents how strongly the current segment matches the segment encoded by the SVM.

Figure 2 shows a summary of SVM firing scores over an entire test dataset that has been divided into 20 segments. Element $M(i, j)$ indicates the response of the i^{th} SVM to the j^{th} segment in a test dataset. Similarly, the j^{th} column provides a place recognition distribution

over all training segments. Firing scores are normalized to sum to one for each place recognition distribution:

$$M(i, j) = \frac{M(i, j)}{\sum_i M(i, j)} \quad (2)$$

3.3 Multi-Scale Place Matching Verification

Each array of SVMs is capable of producing place recognition hypotheses that are only as precise as the average size of a segment for that array. All other things being equal, an array of 20 SVMs will report place recognition hypotheses that are twice as spatially specific as an array of 10 SVMs encoding the same route. Here we present a two-step method for combining place recognition hypotheses at multiple spatial scales.

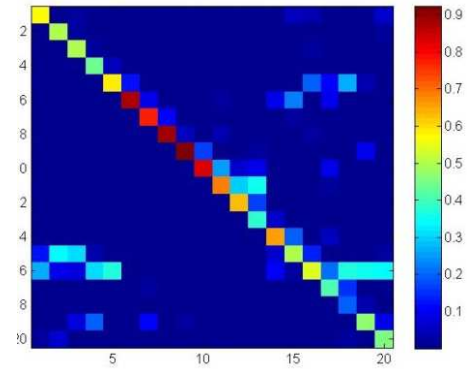


Figure 2: An example confusion matrix showing the distribution of SVM scores over an input dataset comprising 20 segments.

Finding Overlapping Hypotheses

The first step involves finding places that have overlapping place recognition hypotheses – that is, places for which every array of SVMs reports a possible place recognition match (Figure 3). To identify overlapping place recognition hypotheses with different spatial scales, we first normalize the reported place recognition matches to the scale space of the smallest segment size.

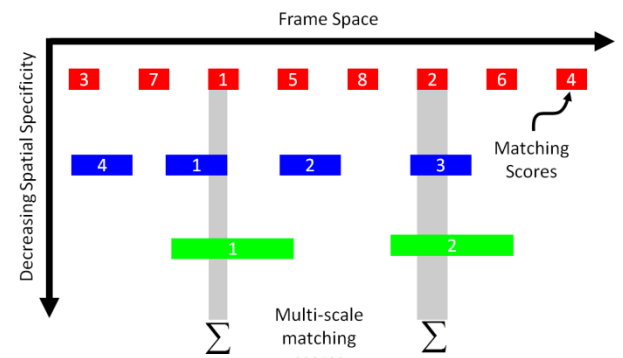


Figure 3: Overlapping SVM matching scores are combined at the smallest spatial scale in order to accept or reject place match hypotheses.

For each set of SVMs, we only search for overlapping place recognition areas from the top K place recognition candidates. For all experiments described in this paper, K was set to 1/10 of the entire dataset

length, which approximately corresponds to the spatial coverage of biological grid cells. If there are no overlapping place hypotheses, then the system reports “no coherent” match.

Finding the Best Overlapping Match Candidate

At the smallest spatial scale, there can be several competing place recognition hypotheses that are supported by all other spatial scales. To determine the most likely hypothesis, we sum the firing scores of the overlapping SVMs at each spatial scale.

3.4 Image Dimensionality Reduction

To provide a practical size input to the SVMs, we implemented two commonly used image processing methods – Principal Component Analysis (PCA) and GIST – which we briefly describe here.

3.5 Principal Component Analysis

PCA [Jolliffe, 2002] is applied to down sampled 64×48 pixel images. Analysis indicated that the top 38 principal eigenvectors encoded 90% of the variance of the training video sequences, so the first 38 principal eigenvectors were chosen as the input to the SVMs.

3.6 GIST

Human vision research has shown that global image information – the “gist” of a scene – is used to recognize real world scenes, without necessarily requiring the use of local object information [Biederman, 1988, Potter, 1975]. In this paper, we choose the model proposed by Oliva [Oliva and Torralba, 2001] to model the “gist” of an image, which provides a holistic description of the scene called Spatial Envelope. Spatial Envelope uses spectral and coarsely localized information to estimate the dominant spatial structure of a scene. It generates a low dimensional representation of the scene in which scenes that share similar semantic categories can be projected closely together. Since only global information is used, this model does not require any specific information of object shape or identity. GIST is also an attractive feature to use in a biologically-inspired approach because of the increased possibility of generating relevant insights into how the biological visual mapping system may function. Extracted GIST feature from the down sampled 64×48 images contained 512 dimensions. We then picked the top 32 principal eigenvectors, which captured approximately 90% of the total variance.

4 Experimental Setup

In this section, we describe the data sets used and the SVM training procedure.

4.1 Datasets

Datasets were obtained from two different environments, summarized in Table I. In each environment, the dataset consisted of two traverses along the same route. For both environments, full resolution images were first converted to gray-scale and then histogram normalized to reduce the effect of illumination variations.

The *Rowrah* dataset (Figure 4) was sourced from YouTube and was collected using a forward facing camera mounted on the motorbike chassis. The *Campus* dataset was collected using a GoPro Hero 1 camera mounted on a bicycle pushed by an experimenter walking around the Queensland University of Technology campus. While the Rowrah environment was relatively uniform, the Campus dataset included a wide variety of environments including building interiors, grassy parks and concrete walkways. The Campus dataset was collected during a weekday and included small dynamic changes due to pedestrians and minor lighting variations.

TABLE I DATASET DESCRIPTIONS			
Dataset Name	Single Traverse Distance	Number of Frames per Traverse	Resolution
Rowrah	1000 m	1570	320×240
Campus	800 m	1000	1280×960

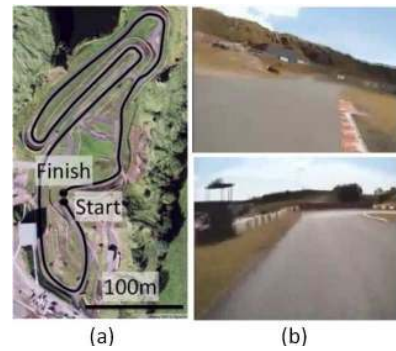


Figure 4: (a) The Rowrah dataset consisted of video from a camera mounted on a motorbike chassis ridden around a racing track. (b) Some sample frames from the captured video.

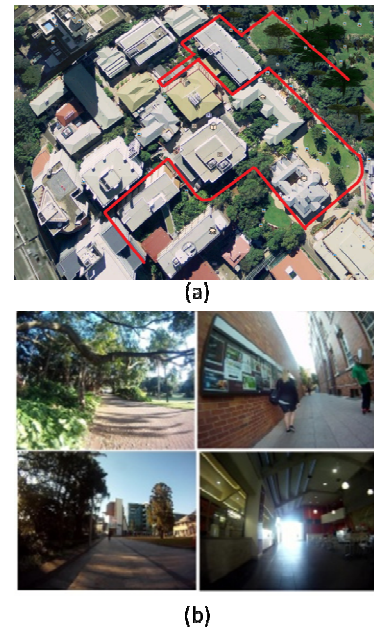


Figure 5: The Campus data set showing the (a) path that was traversed twice and sample frames from the environment that included parkland, outside walkways and building interiors.

4.2 Training Procedure

Images from the first traverse of the environment were used for training while images from the second traverse

were used to evaluate performance. The overall training procedure consisted of the following steps:

1. Dataset Segmentation
2. Feature Extraction
3. SVM Training

In the following sections we describe these steps in more detail.

Dataset Segmentation

The images from the first and second traverses of each environment were grouped into segments of varying sizes to form a total of S segments per dataset ($S / 2$ segments per traverse). Larger values of S corresponded to more segments per dataset and hence each segment representing a smaller, more spatially specific area. For example, splitting the Campus dataset into 170 segments (85 segments per traverse) resulted in an average segment size of approximately 9.4 metres. Although no odometry information was available, each dataset was captured at an approximately constant camera speed. Using the known route length, the average segment size in metres can be estimated and is listed in Table II. In the discussion we describe how odometry could be incorporated to group frames into segments.

The range of segmentation values was somewhat arbitrary and chosen to provide reasonable coverage over a range of segment sizes. In the Discussion we discuss methods for a more comprehensive search through scale space.

TABLE II
DATASET SEGMENTATION AND SEGMENT SIZES

Number of Segments	Rowrah (m)	Campus (m)
30	66.67	53.33
40	-	40
43	46.51	-
56	-	28.57
60	33.33	-
80	-	20
85	23.53	-
90	22.22	-
120	16.67	13.33
170	-	9.41
200	10	-
240	-	6.67
300	-	5.33

Feature Extraction

Feature extraction enabled us to accelerate the training and testing stages by providing a reduced dimensionality input into the SVM models. Two feature types (as discussed in Sections 3.4 and 3.5) were extracted from each dataset. The feature vectors from all frames in a segment were consolidated into a single vector and input into each of the SVMs.

SVM Training

The third step involved training a SVM model for each segment. When training each segment model, the frames corresponding to that segment were labelled as positive examples. Ideally, all other ($S-1$) groups would be used as negative examples. However, since in real world situations it may not be possible to train on the entire training dataset, we instead arbitrarily set N to be

9, indicating for each segment, we use 1 frame group as a positive example and 9 other frame groups as negative samples.

Two parameters need to be determined for an RBF kernel SVM: C and σ . A parameter search was conducted before training each SVM to find optimal parameter values. We followed a “grid-search” strategy using cross-validation. C and σ were tested in exponentially growing order from 2^{-1} to 2^{10} and from 2^{-15} to 2^{-1} , respectively. Each (C, σ) pair was evaluated, and the pair with the best cross-validation accuracy was chosen.

4.3 Ground Truth

GPS-based ground truth was not available with either dataset (due to the Rowrah video being sourced online and because GPS was unreliable or not available near and inside buildings in the Campus environment). To obtain ground truth for each dataset, each segment in the second run was manually parsed and matched back to the corresponding segment in the first run. Linear ground truth plots were generated (see Figure 10) where progression along the ground truth line corresponds to progression along the complex route travelled through the environment.

It is important to note that when creating precision recall graphs using the ground truth information, the decision for whether a reported match was deemed a true or false positive was made based on whether the correct, most spatially specific *segment* match was reported. Consequently the spatial tolerance for a correct match varies, depending on what the spatial size of the most spatially specific segment size in an experiment was. Because we are primarily focusing on the relative performance change between single- and multi-scale place recognition, we believe this is a valid metric. In addition, we provide metric segment sizes in Table II that indicate the implied error tolerance for each precision-recall graph.

4.4 Illustrative Segment Samples

In this section we present representative segments input into the SVM arrays. Each figure shows the location and coverage of segments of varying sizes in the aerial image and the range of camera frames belonging to each segment.

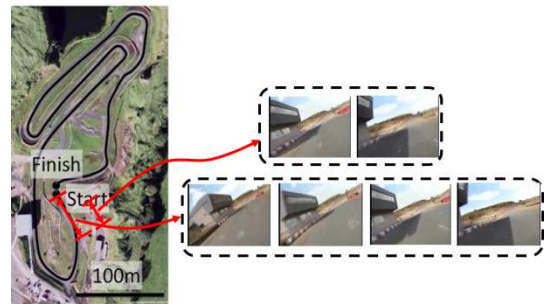


Figure 6: Examples of camera frames from two different segment sizes for the Rowrah dataset.

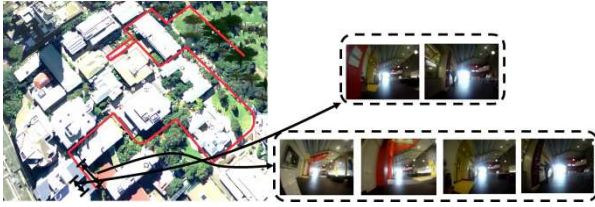


Figure 7: Examples of two different segment sizes from a building interior encountered in the campus dataset.

5 Results

This section presents results showing precision recall graphs and ground truth plots comparing single- and multi-scale place matching performance as well as quantitative performance metrics.

5.1 Single- and Multi-scale Place Recognition

This section presents precision recall curves results for the single- and multi-scale place recognition experiments. Each precision recall curve was generated by performing a sweep over the ordered segment matching rankings. For example, over a 20 segment long traverse, there would be 20 top ranked place match hypotheses. If 18 of these matches were correct and 2 were incorrect, precision was 90% and recall was also 90%. By then expanding the sweep to include the 20 second ranked place matches over those 20 segments, there might now be a total of 40 matches, of which 20 were correct and 20 others were incorrect – precision would be 50% and recall would be 100%. Because the ranked matches are mutually exclusive (the top ranked and second top ranked match can't both be to the same segment), if the top ranked match is already correct, all other ranked matches are incorrect. This results in a rapid drop in precision as the sweep is expanded. The curves also appear discretised because the parameter value can only be varied in integer steps.

The first, perhaps expected observation is that for both single- and multi-scale matching, it is easier to perform place recognition when trying to match a spatially broad segment than when trying to match a spatially specific segment. This disparity is most likely due to two reasons; firstly because performance is bound to increase when the false positive spatial error tolerance is bigger, and secondly, because the larger segments are trained on a larger number of frames.

In 10 of 12 experiments (83%), precision-recall performance at all except very low precision levels improved significantly. At 100% precision, the recall rate improved by an average of 56.6% across all experiments.

We observed one experiment where the multi-scale approach performed worse than the single-scale approach. For the Campus dataset, combining the 300 segment system with the 30 and 80 segment system resulted in a significant degradation in performance. Examining the results, it became apparent that the place recognition hypotheses from the 300 and 80 segment systems had minimal overlap, causing the reduction in combined performance. The 300 segment system had the smallest segment size (5 metres), and it is possible the poor and inconsistent place recognition

output was due to difficulties with training the SVM on a small number of frames.

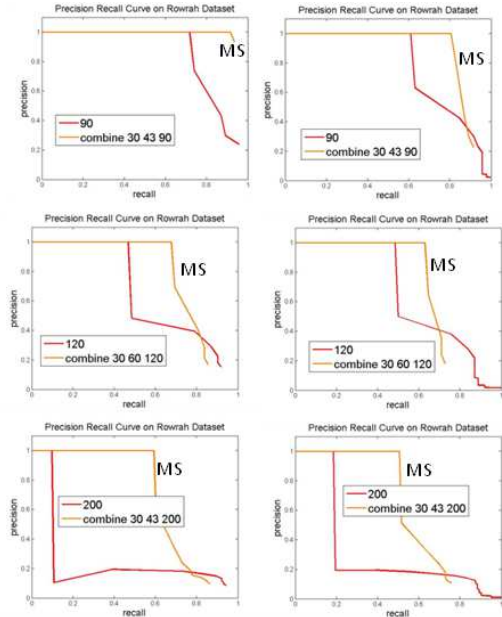


Figure 8: Precision recall curves demonstrating the single- and multi-scale place recognition performance for the Rowrah dataset ("MS" marks the multi-scale curve). The left figures are results using GIST features. The right figures are results using PCA features.

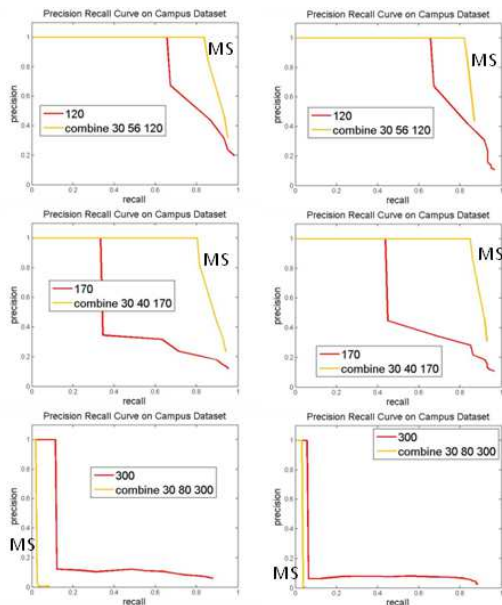


Figure 9: Precision recall curve demonstrating the results with and without combination for Campus dataset ("MS" marks the multi-scale curve). The left figures are results using gist features. The right figures are results using PCA+gray features.

5.2 Ground Truth Plots

Figure 10 presents ground truth plots showing the true positives (green circles), false positives (blue squares) and false negatives (red stars) output by the single and multi-scale systems for the Rowrah and Campus datasets for similar recall rates. Straight lines connect the matching segments. In both datasets, the multi-scale improvement contribution is clear – it eliminates a large number of false positives at the smallest spatial scale.

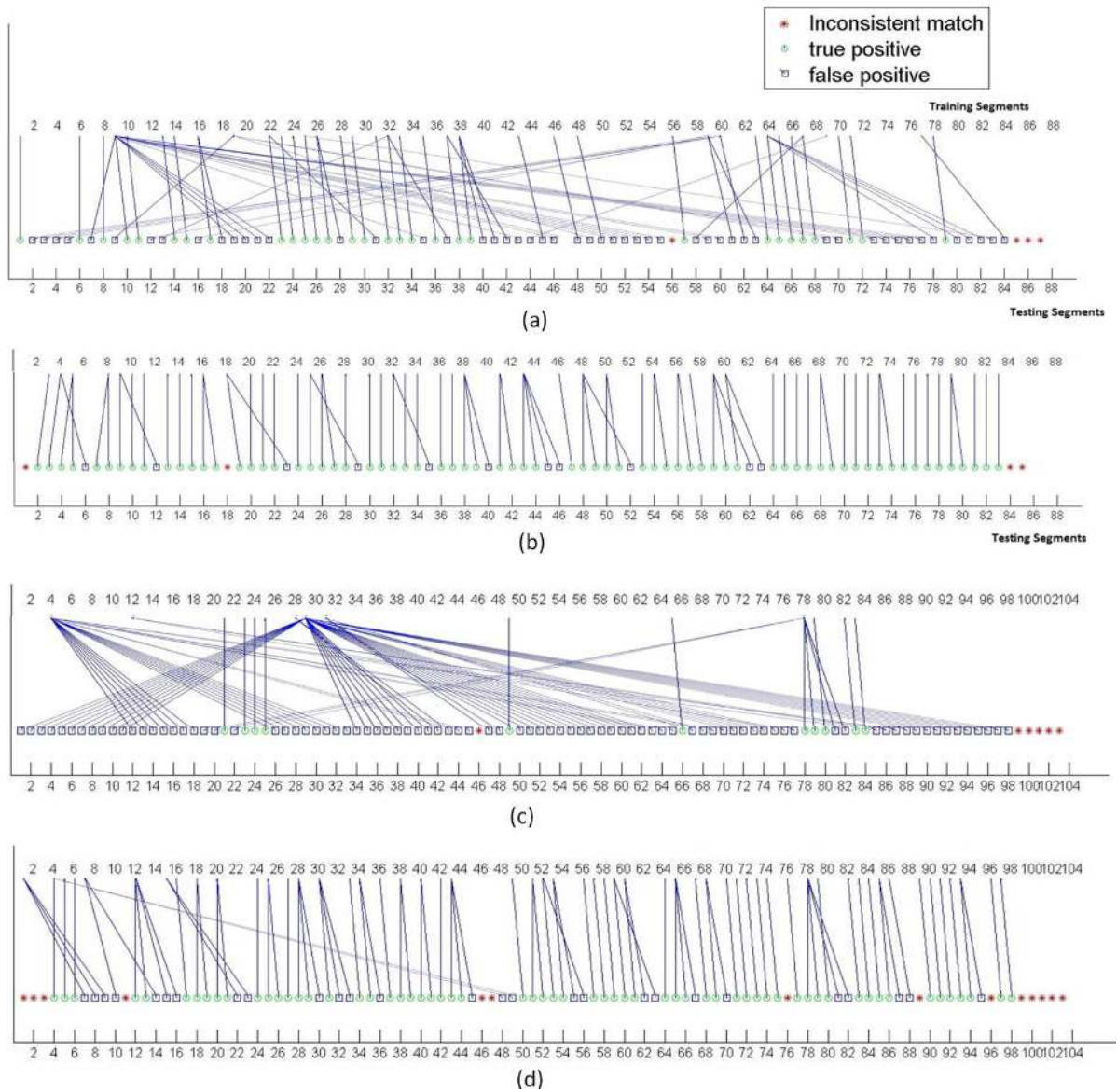


Figure 10: Ground truth plots for the (a) single and (b) multi-scale Rowrah dataset and (c) single and (d) multi-scale Campus dataset. In each case, the multi-scale approach enables the elimination of a large number of false positives matches and the addition of new correct matches.

6 Discussion and Future Work

The results in this paper demonstrate the potential for improving place recognition performance by combining the output of parallel systems, each trained to perform place recognition at a different spatial scale. By validating spatially specific place recognition hypotheses against spatially broad hypotheses, it is possible to improve performance without sacrificing spatial specificity. Although we presented a specific implementation of both the mapping framework and the vision processing system, we believe the core multi-scale concept should be generally applicable to many systems. In this section we discuss several areas of current and future work.

6.1 Incorporating Odometry Information and Moving to Unconstrained 2D Motion

Incorporating an odometry source will provide a number of potential performance and applicability benefits. To create segments, both for training and testing, we assumed that the camera was moving at a

relatively constant velocity through the environment. It should be straightforward to use odometry information to allocate segments based on spatial distance travelled rather than (in effect) time.

Furthermore, because odometry is generally relatively reliable over short time periods, it should be feasible to use it to allocate two-dimensional areas when expanding the system to unconstrained movement in large open areas. Testing the system in open field environments will be most closely analogous to the majority of the rodent experiments that show evidence of a multi-scale neuronal map, increasing the possibility of generating useful neuroscience insights. Finally, odometry will enable the system to perform dead reckoning, in order to update the robot's localization estimate in the short periods between segment matches.

6.2 Scaling Ratios, Limits and Combination

Although we did not hit a limit in our experiments, it is likely there is an upper limit to the segment length (or area size) that can be efficiently represented by a

learning and recognition system, whether an array of SVMs like the approach in this paper or a continuous attractor neural network like in RatSLAM. Since it is currently infeasible to perform rodent experiments in environments much larger than a single room, the use of robots offers the possibility for exploring possible map scaling limits. We speculate that the failure case may be caused by the inability to learn places smaller than a certain size – rodent mapping scales seem to have a lower bound as well [Stensola, et al., 2012].

Current neural evidence suggests that the ratio between consecutive map scales is approximately $\sqrt{2}$ [Stensola, et al., 2012]. While that ratio is perhaps theoretically appealing, the practical implementations are not fully known. We intend to conduct a systematic series of experiments and parameter studies that determine the performance trade-offs of using different scaling ratios and consequently using different numbers of discrete map scales. Furthermore, it seems (both intuitively and from a sensory perspective) that the transition from representing a single room or corridor intersection to an entire building floor would represent a scale boundary, and this is indeed the approach that has been used in previous robotic mapping approaches. We will run a series of studies over varying spatial scaling ratios and with different absolute scales.

Finally, the current approach requires a consensus between all map scales. In future work we will seek to introduce an absolute confidence metric in order to evaluate the quality of both single- and multi-scale place recognition hypotheses. The confidence metric may enable the system to output place matches if only some, but not all mapping scales are in agreement.

6.3 Variable Localization Accuracy

While ideally a robot's location is always reported precisely, this outcome may not be possible in all situations. It may be possible to gracefully degrade place recognition hypotheses, so that in situations where spatially specific visual cues are not available, the localization system can use broad visual cues to estimate the robot's general location.

6.4 Multi-sensor Fusion

The final, potentially most profitable area of research is in the area of multi-sensor fusion. Recent work using RatSLAM has shown that biologically inspired algorithms can perform online sensor fusion to enable place recognition in changing environmental conditions, such as over day-night cycles [Jacobson, et al., 2013, Jacobson and Milford, 2012]. An obvious extension to this research would be to use a multi-scale mapping framework to exploit the variable spatial specificity of different sensor modalities, such as cameras, range finders and WiFi. By integrating these multi-sensor fusion systems with a biologically-inspired, multi-scale mapping framework, it may be possible to combine their functional capabilities to produce a highly capable, general purpose mapping and navigation system.

References:

- [M. Bosse, P. Newman, J. Leonard, M. Soika, W. Feiten, and S. Teller, 2003] M. Bosse, P. Newman, J. Leonard, M. Soika, W. Feiten, and S. Teller, "An atlas framework for scalable mapping," in *International Conference on Robotics and Automation*, Taipei, Taiwan, 2003, pp. 1899-1906.
- [B. Kuipers, J. Modayil, P. Beeson, M. MacMahon, and F. Savelli, 2004] B. Kuipers, J. Modayil, P. Beeson, M. MacMahon, and F. Savelli, "Local Metrical and Global Topological Maps in the Hybrid Spatial Semantic Hierarchy," in *International Conference on Robotics and Automation*, New Orleans, USA, 2004.
- [B. Kuipers and Y. T. Byun, 1991] B. Kuipers and Y. T. Byun, "A Robot Exploration and Mapping Strategy Based on a Semantic Hierarchy of Spatial Representations," *Robotics and Autonomous Systems*, vol. 8, pp. 47-63, 1991.
- [H. Stensola, T. Stensola, T. Solstad, K. Froland, M. Moser, and E. Moser, 2012] H. Stensola, T. Stensola, T. Solstad, K. Froland, M. Moser, and E. Moser, "The entorhinal grid map is discretized," *Nature*, vol. 492, pp. 72-78, 2012.
- [T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, 2005] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, "Microstructure of a spatial map in the entorhinal cortex," *Nature*, vol. 11, pp. 801-806, 2005.
- [Y. Burak and I. R. Fiete, 2009] Y. Burak and I. R. Fiete, "Accurate path integration in continuous attractor network models of grid cells," *PLoS Computational Biology*, vol. 5, 2009.
- [P. E. Welinder, Y. Burak, and I. R. Fiete, 2008] P. E. Welinder, Y. Burak, and I. R. Fiete, "Grid cells: the position code, neural network models of activity, and the problem of learning," *Hippocampus*, vol. 18, pp. 1283-1300, 2008.
- [G. Dissanayake, P. M. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba, 2001] G. Dissanayake, P. M. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localisation and map building (SLAM) problem," *IEEE Transactions on Robotics and Automation*, vol. 17, pp. 229-241, June, 2001.
- [M. Cummins and P. Newman, 2009] M. Cummins and P. Newman, "Highly scalable appearance-only SLAM - FAB-MAP 2.0," in *Robotics: Science and Systems*, Seattle, United States, 2009.
- [M. Cummins and P. Newman, 2008] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," *International Journal of Robotics Research*, vol. 27, pp. 647-665, 2008.
- [A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, 2007] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-Time Single Camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1052-1067, 2007.
- [D. Ball, S. Heath, J. Wiles, G. Wyeth, P. Corke, and

- M. Milford, 2013] D. Ball, S. Heath, J. Wiles, G. Wyeth, P. Corke, and M. Milford, "OpenRatSLAM: an open source brain-based SLAM system," *Autonomous Robots*, pp. 1-28, 2013/02/21 2013.
- [M. Milford and G. Wyeth, 2010] M. Milford and G. Wyeth, "Persistent Navigation and Mapping using a Biologically Inspired SLAM System," *International Journal of Robotics Research*, vol. 29, pp. 1131-1153, 2010.
- [M. J. Milford, 2008] M. J. Milford, *Robot Navigation from Nature: Simultaneous Localisation, Mapping, and Path Planning Based on Hippocampal Models* vol. 41. Berlin-Heidelberg: Springer-Verlag, 2008.
- [M. Milford and G. Wyeth, 2008] M. Milford and G. Wyeth, "Mapping a Suburb with a Single Camera using a Biologically Inspired SLAM System," *IEEE Transactions on Robotics*, vol. 24, pp. 1038-1053, 2008.
- [K. Konolige and M. Agrawal, 2008] K. Konolige and M. Agrawal, "FrameSLAM: From Bundle Adjustment to Real-Time Visual Mapping," *IEEE Transactions on Robotics*, vol. 24, pp. 1066-1077, 2008.
- [M. Milford, 2013] M. Milford, "Vision-based place recognition: how low can you go?," *International Journal of Robotics Research*, vol. 32, pp. 766-789, 2013.
- [M. Milford and G. Wyeth, 2012] M. Milford and G. Wyeth, "SeqSLAM: Visual Route-Based Navigation for Sunny Summer Days and Stormy Winter Nights," in *IEEE International Conference on Robotics and Automation*, St Paul, United States, 2012.
- [J. G. Heys, K. M. MacLeod, C. F. Moss, and M. E. Hasselmo, 2013] J. G. Heys, K. M. MacLeod, C. F. Moss, and M. E. Hasselmo, "Bat and rat neurons differ in theta frequency resonance despite similar coding of space," *Science*, vol. 340, pp. 363-367, 2013.
- [M. Milford, F. Schill, P. Corke, R. Mahony, and G. Wyeth, 2011] M. Milford, F. Schill, P. Corke, R. Mahony, and G. Wyeth, "Aerial SLAM with a Single Camera Using Visual Expectation," in *International Conference on Robotics and Automation*, Shanghai, China, 2011.
- [V. Vapnik, 1998] V. Vapnik, "The support vector method of function estimation," *Nonlinear Modeling*, pp. 55-85, 1998.
- [N. Cristianini and S. T. John, 2000] N. Cristianini and S. T. John, *An introduction to support vector machines and other kernel-based learning methods*: Cambridge university press, 2000.
- [I. T. Jolliffe, 2002] I. T. Jolliffe, *Principal Component Analysis*, 2 ed.: Springer, 2002.
- [I. Biederman, 1988] I. Biederman, "Aspects and extension of a theory of human image understanding.," *Computational processes in human vision: An interdisciplinary perspective*, 1988.
- [M. C. Potter, 1975] M. C. Potter, "Meaning in visual search," *Science* vol. 187, pp. 965-966, 1975.
- [A. Oliva and A. Torralba, 2001] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, pp. 145-175, 2001.
- [A. Jacobson, Z. Chen, and M. Milford, 2013] A. Jacobson, Z. Chen, and M. Milford, "Brain-based Sensor Fusion for Navigating Robots," in *IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, 2013.
- [A. Jacobson and M. Milford, 2012] A. Jacobson and M. Milford, "Towards Brain-based Sensor Fusion for Navigating Robots," in *Proceedings of the 2012 Australasian Conference on Robotics & Automation*, 2012.