RatSLAM: Using Models of Rodent Hippocampus for Robot Navigation and Beyond

Michael Milford, Adam Jacobson, Zetao Chen and Gordon Wyeth

Abstract

We describe recent biologically-inspired mapping research incorporating brain-based multi-sensor fusion and calibration processes and a new multi-scale, homogeneous mapping framework. We also review the interdisciplinary approach to the development of the RatSLAM robot mapping and navigation system over the past decade and discuss the insights gained from combining pragmatic modelling of biological processes with attempts to close the loop back to biology. Our aim is to encourage the pursuit of truly interdisciplinary approaches to robotics research by providing successful case studies.

1 Introduction

The brain circuitry involved in encoding space in rodents has been extensively tested over the past thirty years, with an ever increasing body of knowledge about the components and wiring involved in navigation tasks. The learning and recall of spatial features is known to take place in and around the hippocampus of the rodent, where there is clear evidence of cells that encode the rodent's position and heading. RatSLAM [1-3] is a robotic navigation system based on current models of the rodent hippocampus, which has achieved several significant outcomes in vision-based Simultaneous Localization And Mapping (SLAM), including mapping of an entire suburb using only a low cost webcam [4, 5], and navigation continuously over a period of two weeks in a delivery robot experiment [6]. These results showed for the first time that a biologically inspired mapping system could compete with or surpass the performance of conventional probabilistic robot mapping systems. The RatSLAM system has recently been open-sourced and published [7].

We have also "closed the loop" back to the neuroscience underpinning the RatSLAM system. In our research, we took a pragmatic approach to modelling the neural mechanisms, and would engineer "better" solutions whenever the underlying biology did not appear to meet the robot's needs. However, some of the modifications necessary to make the models of hippocampus work effectively over long periods in large and ambiguous environments raised new questions for further biological study, including a potential neural mechanism for filtering uncertainty in navigation [8]. The research has also led to recent experiments demonstrating that vision-based naviga-

The authors are with the School of Electrical Engineering and Computer Science at the Queensland University of Technology. E-mail: michael.milford@qut.edu.au.

tion can be achieved at any time of day or night, during any weather, and in any season using sequences of visual images as small as 2 pixels in size [9-12]. Most recently we have led collaborative research with human- and animal-neuroscience labs leading to novel human-inspired vision-based place recognition algorithms that are starting to rival human capabilities at specific tasks [13, 14].

In this paper we describe two recent biologically-inspired areas of investigation building on the existing RatSLAM system. We first provide a brief but necessary overview of the core RatSLAM system. We then describe research mimicking the hypothesized sensory calibration processes in the rodent brain and present experiments demonstrating autonomous calibration of a place recognition system, a key requirement for mapping and navigation systems. Finally we describe new research modelling the multi-scale, homogeneous mapping frameworks recently discovered in the rat brain and present results showing the place recognition performance benefits of such an approach. We conclude with a discussion of the key lessons learnt in more than a decade of pursing an interdisciplinary robotics-neuroscience research agenda.

2 RatSLAM

In this section we briefly describe the core RatSLAM algorithms upon which the new research presented here is based. RatSLAM is a SLAM system based on computational models of the navigational processes in the part of the mammalian brain called the *hippocampus*. The system consists of three major modules – the *pose cells, local view cells*, and *experience map*. Further technical details on RatSLAM can be found in [4, 6].

2.1 Pose Cells

The pose cells are a Continuous Attractor Network (CAN) of units connected by both excitatory and inhibitory connections, similar to a recently discovered class of navigation neurons found in many mammals called *grid cells* [15]. The network is configured in a three-dimensional prism (Fig. 1), with cells connected to nearby cells by excitatory connections, which wrap across all boundaries of the network. The dimensions of the cell array nominally correspond to the three-dimensional pose of a ground-based robot – x, y, and θ. The pose cell network dynamics are such that the stable state is a single cluster of activated units, referred to as an *activity packet* or *energy packet*. The centroid of this packet encodes the robot's best internal estimate of its current pose. Network dynamics are regulated by the internal connectivity as well as by input from the local view cells.

2.2 Local View Cells

The local view cells are an expandable array of cells or units. Novel scenes drive the creation of a new local view cell which is then associated with the raw sensory data (or an abstraction of that data) from that scene. In addition, an excitatory link is learnt

(one shot learning) between that local view cell and the centroid of the dominant activity packet in the pose cells at that time. When that view is seen again by the robot, the local view cell is activated and injects activity into the pose cells via that excitatory link. Re-localization in the pose cell network occurs when a sufficiently long sequence of familiar visual scenes is experienced in the correct sequence, causing constant injection of activity into the pose cells resulting in the re-activation of the pose cells that were associated with that scene the first time.

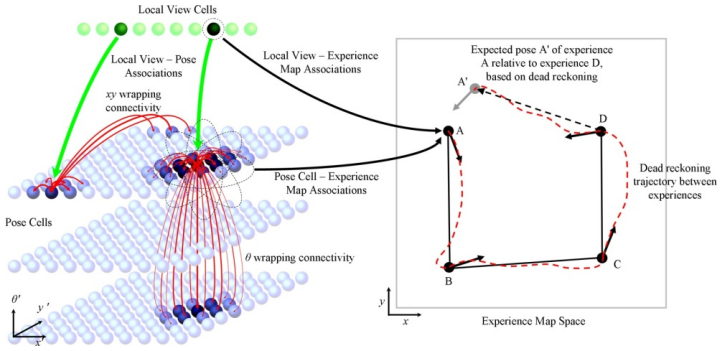

Fig. 1 The RatSLAM system, consisting of local view cells, pose cells and the experience map.

2.3 Experience Map

Initially the representation of space provided by the pose cells corresponds well to the metric layout of the environment a robot is moving through. However, as odometric error accumulates and loop closure events occur, the space represented by the pose cells becomes discontinuous – adjacent cells in the network can represent physical places separated by great distances. Furthermore, the pose cells represent a finite area but the wrapping of the network edges means that in theory an infinite area can be mapped, which implies that some pose cells represent multiple physical places. The experience map is a graphical map that provides a unique estimate of the robot's pose by combining information from the pose cells and the local view cells. A new experience is created when the current activity state in the pose cells and local view cells is not closely matched by the state associated with any existing experiences. As the robot transitions between experiences, a link is formed from the previously active experience to the new experience. A graph relaxation algorithm runs continuously to evenly distribute odometric error throughout the graph, providing a map of the robot's environment which can readily be interpreted by a human.

3 Brain-based Sensor Fusion and Calibration

Current state of the art robot mapping and navigation systems produce impressive performance under a narrow range of robot platform, sensor and environmental conditions. In contrast, animals such as rats produce "good enough" maps that enable them

to function in an incredible range of situations and environments around the world. From only four days after birth, rat pups start to *learn* how to best sense, map and navigate in their environment [16, 17]. Rat pups have been seen to demonstrate particular movement behaviours such as pivoting that are theorized to help them calibrate their sensory stream. Furthermore, adult rats rapidly adapt to changes in their own sensing equipment or in their environment during their adult life [18]. It has even been shown that it is possible to integrate novel sensory devices into a rat brain and have the rats subsequently learn to utilise this novel input [19]. We investigated the feasibility of adopting a "sensor agnostic" approach to mapping and localization inspired by the adaptation capabilities of rats.

We describe a rat-inspired multi-sensor fusion and calibration system that assesses the usefulness of multiple sensor modalities based on their utility and coherence for place recognition both when a robot is first placed in an environment through calibration behaviors [20] and autonomously while moving [21], without knowledge as to the type of sensor. We demonstrate the system on a Pioneer robot in indoor and outdoor environments with large illumination changes.

3.1 Approach

Here we present our sensor-agnostic approach to multi-sensory calibration and online sensory evaluation. The system is algorithmic in nature; however it is loosely inspired by rodent behavioural and neural processes.

3.1.1 Sensor Pre-Processing

Sensor data is pre-processed to enable agnostic evaluation of sensory information through a standardized format. All sensor data is normalized by dividing by the maximum possible sensor reading producing a value between (0, 1). Sensor data in the form of multi-dimensional arrays, such as images, are down-sampled and separated into a single line vector, for example, RGB images are converted to grayscale, down-sampled to 12×9 and separated into a single vector 108 elements long. Sensor pre-processing is applied to all sensor modalities producing a single vector for each sensor called a template.

3.1.2 Multi-Sensor Fusion

Sensor data similarity is evaluated utilizing a Sum of Absolute Differences (SAD) comparison, in order to determine the similarity between the current template and all previously stored templates. The best template match to the current sensor template is the previously learnt template with the smallest difference score. We define a template as *familiar* if a previously learnt template has a difference score less than a predetermined recognition threshold, S_{thresh}. The current sensory template is defined as *novel* if the best template match difference score is greater than the recognition threshold. Furthermore, we define a technique for dynamically evaluating the utility and reliability of sensors as the robot moves through the environment. Sensor reliability is determined using two biologically inspired metrics, *spatial coherence* and *template*

expectation similarity. These metrics are binary operators and evaluate the agreement between two sensory modalities. Each sensor is compared to each other sensor using these two metrics and combined to produce a single coherence score which is used to determine the utility of each sensor. Spatial coherence builds on the idea of using geometric information to validate place recognition and utilizes the experience map to determine the Euclidean distance between template matches. Two sensors are deemed to be spatially coherent if the Euclidean distance between the location matches is below a geometric threshold, g_{thresh}. Template expectation similarity determines the similarity between the current sensor data and a predicted sensor reading generated from another sensor. Sensors are deemed to be reliable if coherent with at least one other sensor or if no template match has been reported, otherwise the sensor is tagged as unreliable.

Sensor data is fused together through the implementation of "super templates", formed by concatenating each sensor template into a single vector. When comparing super templates, the component of the overall matching score corresponding to each sensor is normalized by the number of readings for the sensor to remove any effect of varying sensor vector sizes.

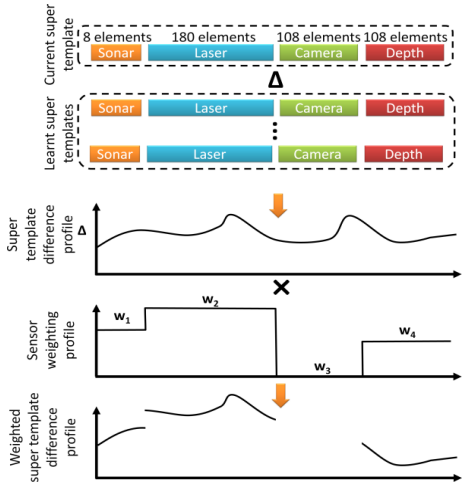

Fig. 2 Super templates are created by the concatenation of individual sensor data and compared to previously learnt super templates using a weighted SAD. Super templates allow the storage of sensory information for a particular scene, allowing all sensory data to be processed in a uniform manner.

3.1.3 Movement-driven Autonomous Calibration

Autonomous calibration of the place recognition processes for each sensor is achieved by mimicking the pivoting behavior of young rat pups when calibrating their sensors. The main requirement of a robot is that it is capable of safely performing two donuts within the operating environment and that the environment is primarily static for the calibration behaviors. The performance of two donut behaviors is required to allow

the sensory equipment to experience an environmental scene twice, allowing the distinction of novel and familiar sensory data.

Place recognition calibration is performed by monitoring the difference scores between the current and previous sensory snapshots as the robot completes two revolutions, the first a "novel' revolution and the second a "familiar" revolution, since the robot is repeating a previous movement. The place recognition threshold is set to the maximum difference score for the familiar region of the calibration behavior. This method captures the largest possible variance in difference score for a *familiar* template match. This process is a conservative one –while it is likely the system will miss place matches in more perceptually challenging environments, false negatives are generally less catastrophic than false positives. The system also calculates a threshold quality score based on analysis of the difference score distribution over the two revolutions.

3.2 Experimental Setup

All the dataset acquisition and testing was performed in ROS groovy, all datasets ROS bags are available for readers to download and process at: https://wiki.qut.edu.au/display/cyphy/Michael+Milford+Datasets+and+Downloads. Detailed system parameters are provided in [20, 21].

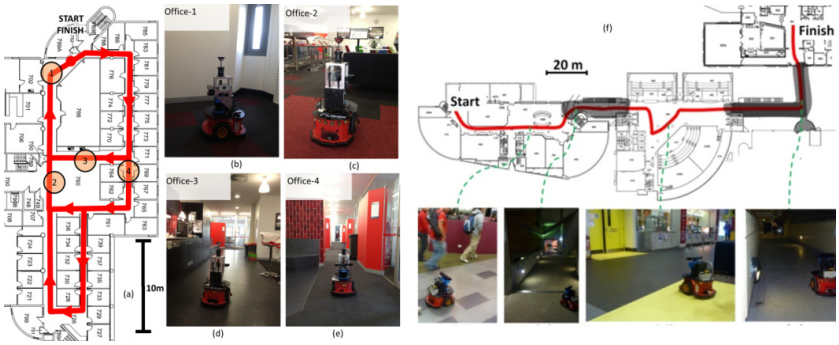

Fig. 3 (a) Map indicating the calibration locations and robot path for the office environment. (b-e) show photos of the calibration locations used within the office environment, which varied between open plan space, corridors and a kitchen. (f) Campus environment. The route was traversed during both day- and night-time conditions, with snapshots of the robot in the environment shown along the route.

3.2.1 Testing Environments

The testing environments were diverse and included a university campus and an office building floor. The campus dataset was traversed during day and night conditions to test the system's ability to handle varying environmental conditions.

3.2.2 Robot Platforms

The office robot configuration was built on an Adept MobileRobots Pioneer 3DX utilizing a FireWire PointGrey Camera with Catadioptric mirror, 16 ultrasonic range sensors, SICK laser range finder and Microsoft Kinect with RGB and Depth images. The campus robot configuration was also assembled on the Adept MobileRobots Pioneer 3DX using 16 ultrasonic range sensors, SICK laser range finder and Microsoft Kinect with RGB and Depth images.

3.3 Results

For reasons of brevity, here we present only the maps produced in each experiment – which reveal whether the system was able to produce topologically correct maps without any false connectivity between map locations. Further results can be found in [20, 21].

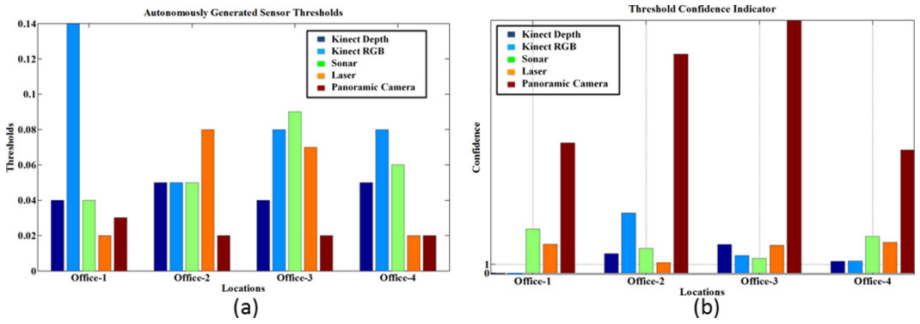

Fig. 4 (a) Autonomously calibrated thresholds from office calibration locations 1-4. Each group of five bars corresponds to the five sensor calibrations at one calibration location. (b) Corresponding calibration confidence scores.

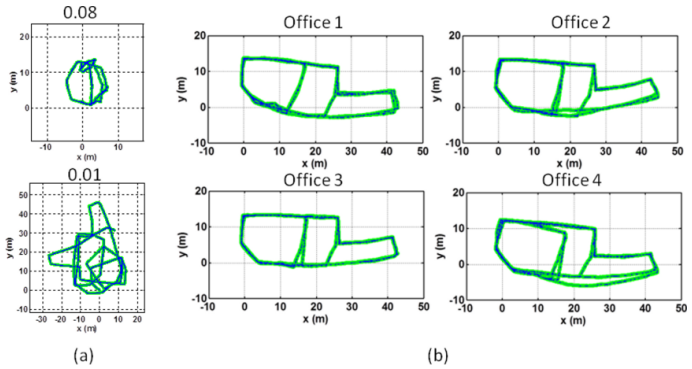

Fig. 5 OpenRatSLAM experience maps for the office environment generated with wheel encoder self-motion information using the (a) manually selected super template thresholds of 0.08 and 0.01 and reliable autonomously calibrated thresholds from office calibration locations 1-4.

3.3.1 Office Environment

The calibration behavior was performed in the office environment resulting in the generation of the four sets of sensor thresholds and confidence scores shown in Figure 4. Evaluation of Figure 4b illustrates that all the autonomously generated place recognition thresholds are reliable (have a confidence score above 1), except the thresholds for sensors 1 and 2 in office calibration location 1. These low confidence scores were most likely due to the approximately equally distant and bland white walls of office calibration location 1. Figure 5 shows the experience maps are all topologically correct and have no incorrect loop closures, including the map created using only 3 reliable sensors from office calibration location 1.

3.3.2 Campus Environment

Here we present results for the campus environment experimentation produced from traversing the campus environment twice, first during the day and the second at night. Place recognition thresholds calibrated in the office calibration locations that resulted in a full set of trusted sensors (locations 2-4) were used for testing in the campus environment. Figure 6 shows the resultant OpenRatSLAM maps for the campus environment. All sensors were down weighted at various times during the experiment, removing large amounts of false positive matches from individual sensors. The dynamic sensor fusion system also removed some true positive matches, which resulted in some regions of the map not being connected together. All the maps are topologically correct although the recall rate for Figure 6c is less than ideal. A reference map without sensor weighting is shown in Figure 6d.

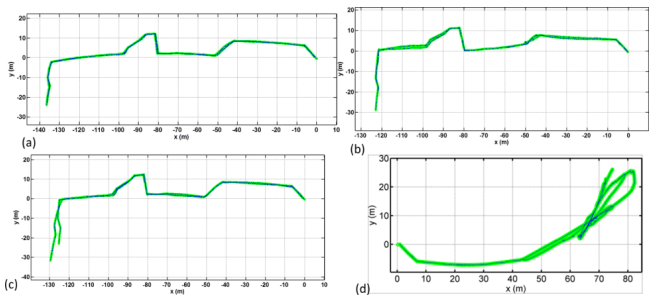

Fig. 6 (a-c) OpenRatSLAM experience maps of the campus environment generated with wheel encoder self-motion information using reliable autonomously calibrated thresholds from office calibration locations 2-4. (d) Map without online sensor weighting.

3.4 Future Work

We are currently investigating the use of a much wider range of sensing modalities such as WiFi. One of the most interesting insights from these multi-sensor fusion experiments is that different sensor types have varying spatial specificities when used in an associative mapping framework such as RatSLAM. Cameras offer the potential

for spatially precise place recognition performance, while sensors such as WiFi offer broader spatial localization. Attempting to integrate the place recognition information provided by each of these very different sensor types using a single scale mapping framework is likely suboptimal. In the next section, we present a pilot study investigating a multi-scale, homogeneous mapping framework inspired by the multi-scale maps recently found in the rodent brain.

4 Multi-scale Mapping

Most robot navigation systems perform mapping at one fixed spatial scale, or over two scales, often locally metric and globally topological [22-24]. Recent discoveries in neuroscience suggest that animals such as rodents, and likely many other mammals including humans, encode the world using multiple but homogeneous parallel mapping systems, each of which encode the world at a different scale [15, 25]. Although investigated in a theoretical context [26, 27], the potential performance benefits of such a mapping framework have not yet been investigated in a real-world robotics context. In this study, we investigated the utility of combining multiple homogeneous maps at different spatial scales to perform place recognition [14]. The performance of the multi-scale implementation was compared to a single scale implementation using two different vision-based datasets.

4.1 Approach

Our overall approach involves a feature extraction stage, a learning stage using arrays of Support Vector Machines, and a place recognition stage that combines place recognition hypotheses at different spatial scales.

4.1.1 Feature Extraction

Dimensional reduction was performed before camera images were input to the SVMs. We implemented two commonly used feature extraction methods – Principal Component Analysis (PCA) and GIST. PCA [28] is an efficient dimension reduction method which projects the original data into the directions with largest variances. Camera images were down-sampled to 64×48 before applying PCA. The first 38 principal eigenvectors were picked which were shown to already capture 90% of the data variance. For GIST features, we chose the model proposed by Oliva [29] which provides a holistic description of the scene called Spatial Envelope. GIST was also attractive because of the possibility of generating relevant insights into how the biological visual mapping system may function. We extracted the GIST feature from down sampled 64×48 images which resulted in a 512-dimensional feature. We then extracted the top 32 principal eigenvectors, which captured approximately 90% of the total variance.

4.1.2 Learning Algorithm

Support Vector Machines (SVM) [30] were chosen as the learning algorithm for two reasons. Firstly, they are an effective method for finding an optimal hyperplane to

separate training data whilst simultaneously maximizing the classification margin, making it suited to the task of training recognition of a specific spatial segment and maximizing the difference between the training segment and other similar segments. Secondly, the use of SVMs removes the need for the extensive parameter tuning required of more biologically plausible grid cell models, such as continuous attractor networks [2], although we do intend to eventually adopt these models to maximize biological relevance.

4.1.3 Combining Multi-Scale Place Match Hypotheses

Each array of SVMs produces a firing matrix M representing the matching scores of the testing segments on the trained SVMs where element $M(i,j)$ indicates the response of the i^{th} SVM from a training dataset to the j^{th} segment in a test dataset. Firing scores in each column j are then normalized to sum to one for each segment recognition distribution:

$$M(i,j) = \frac{M(i,j)}{\sum_i M(i,j)} \tag{1}$$

Place recognition hypotheses produced by each array of SVMs are only as accurate as the average size of a segment in that array. To create a common scale in which hypotheses from different spatial scales can be compared and combined, reported place recognition matches are transformed to the scale space of the smallest segment size. For K arrays of SVMs, the matching scores after normalization of each array are:

$$M_p, \quad \mathrm{p} = 1, \ldots, K \tag{2}$$

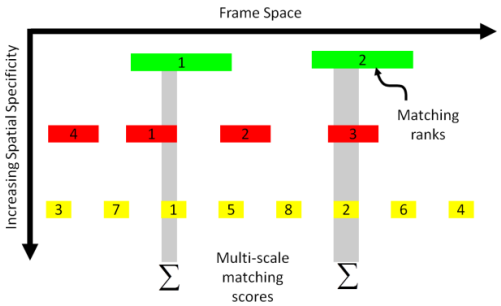

Fig. 7 Overlapping SVM matching scores are combined at the smallest spatial scale in order to accept or reject place match hypotheses. In this case, K =3.

Suppose there are L_p training segments for the matching score M_p. For a segment j in a test data set, its coherence measurement on each training segment $c(i,j), i = 1,\ldots,L_p$ is determined by whether spatially overlapping hypotheses exist over all SVMs scales. If not, the system reports "no coherent" match (c = 0):

$$c(i,j) = \begin{cases} 1, & M_p(i,j) > 0, \ \forall p \\ 0, & else \end{cases} \tag{3}$$

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 and classify segment j to the class $C(j)$ with the largest accumulated firing score:

$$C(j) = arg\max_i \sum_p M_p(i,j),\ \forall c(i,j) = 1 \quad (4)$$

4.2 *Experimental Setup*

We used two datasets (Fig. 8) to test the multi-scale algorithms, with details listed in Table I. Each dataset consists of two traverses along the same route with the first traverse used for training and the second traverse for testing. The Rowrah dataset was collected from a forward-facing camera mounted on a motorbike and can be downloaded at the following link: http://www.youtube.com/watch?v=_UfLrcVvJ5o. The Campus dataset was sourced from a GoPro Hero 1 camera mounted on a bicycle pushed by an experimenter. The bike was pushed through and in-between buildings along a mixed indoor-outdoor path approximately 800 meters long. Due to GPS not being viable, datasets were parsed frame by frame to build ground truth correspondence between testing and training data sets for each spatial scale.

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

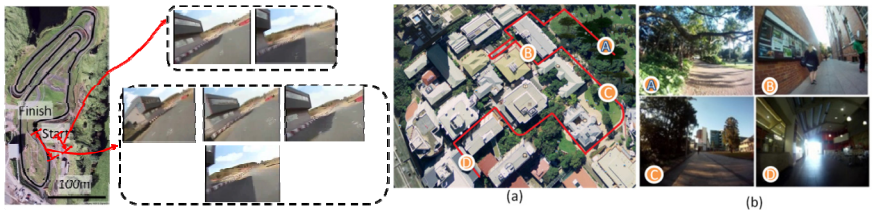

Fig. 8 The Rowrah dataset (left) and Campus data (right) with example frames.

4.2.1 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 three steps: dataset segmentation, feature extraction and SVM training.

Dataset Segmentation

The images in each dataset were grouped into a total of S subsequent segments (S/2 segments per traverse). Larger values of S result in smaller size of each segment. For the sake of intuition, we refer to different SVMs by the size of each segment, not the number of segments. For example, each traverse in the Campus dataset is approx-

imately 800 meters and therefore splitting the Campus dataset into 170 segments (85 segments per traverse) resulted in an average segment size of approximately 9.4 meters. We then use "9.4 meter system" to refer to the SVMs with 170 segments.

Feature Extraction

Two feature types (as discussed in Section IIIA) were extracted from each dataset. The feature vectors from all frames in a segment were combined into a single vector and input into each of the SVMs.

SVM Training

To train a SVM model for each segment, we manually labeled the images in that segment as positive examples and those from the other N segments as negative 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 adjacent frame groups as negative samples

4.3 Results

We show three key sets of results – comparison between single and multi-scale place recognition, ground truth plots and illustrative multi-scale place recognition combination plots.

4.3.1 Single- and Multi-scale Place Recognition

This section presents precision recall (PR) curves for the single- and multi-scale place recognition experiments. Each PR curve was generated by sweeping the accepted range in each hypothesis rank. For both single- and multi-scale matching, it is, not surprisingly, 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.

Precision-recall performance at all except very low precision levels improved significantly across all experiments. At 100% precision, the recall rate was improved by an average of 74.79% across all experiments. The biologically-inspired feature GIST slightly outperformed PCA – at 100% precision, the recall rate for GIST was improved by an average of 81.7% over all experiments, versus 67.9% for PCA.

4.3.2 Ground Truth Plots

Figures 11a and b present 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 datasets without (a) and with (b) multi-scale combination. Straight lines connect the matching segments.

4.3.3 Multi-hypothesis Combination Plots

Fig. 11c-f show examples of how place match hypotheses at varying scales are combined together. In general, a large number of false positives at the smallest spatial scale (yellow color) are eliminated due to lack of support from larger spatial scales. The examples in (c-d) show how secondary ranked spatially specific matches are correctly chosen as the overall place match due to support from other spatial scales. In (e) the best ranked spatially specific match is correctly supported by the other spatial scales, while (f) shows a failure case where the incorrect 4th ranked spatially specific match is more strongly supported by the other spatial scales that the 1st ranked and correct spatially specific match. Interestingly, the most common failure mode of the system is to report a "minor" false positive match – a place match to a different location at the smallest spatial scale but within the correct place at a larger spatial scale.

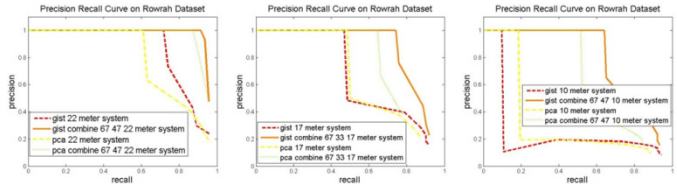

Fig. 9 Precision recall curves demonstrating the single- and multi-scale place recognition performance for the Rowrah dataset.

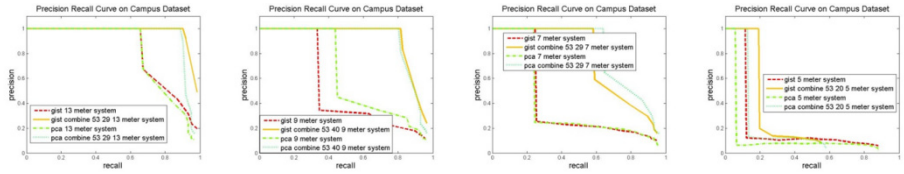

Fig. 10 Precision recall curve demonstrating the results with and without combination for Campus dataset using gist features and PCA features.

4.4 Discussion and Future Work

Place recognition performance was improved by combining the output from parallel systems, each trained to recognize places at a specific spatial scale. Although here we presented a specific implementation of both the vision processing and place recognition framework, we believe that the novel multi-scale combination concept should generalize to other systems. In future work, we will incorporate an odometry source to enable the system to allocate segments directly based on distance travelled rather than (in effect) time. Odometry information may enable us to expand our current system to two-dimensional unconstrained movement in large open environments. Testing the system in open field environments will be more analogous to many current rodent experiments and may increase the likelihood of generating neuroscience insights. An obvious extension to the sensor fusion work presented here and elsewhere [21, 31] would be to use a multi-scale mapping framework approach to exploit the variable spatial specificity of different sensor modalities, such as cameras, range finders and WiFi.

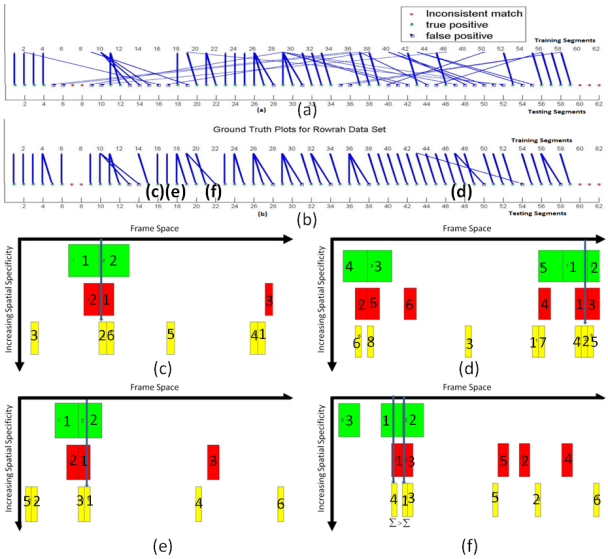

Fig. 11 Ground truth plots for the (a) single and (b) multi-scale Campus dataset. (c-d) show examples of secondary-ranked spatially specific place matches (yellow) that became the primary overall place match hypothesis due to support from other spatial scales. In (e) the first ranked spatially specific match is supported, while (f) shows a failure case where a secondary ranked spatially specific match is incorrectly chosen as the overall match due to more significant support from the other spatial scales than the correct, first ranked spatially specific match.

5 Achieving Balance in Interdisciplinary Research

If we were asked to identify the single key issue involved in conducting interdisciplinary (especially biologically-inspired) robotics research it would be this:

How can research achieve the appropriate balance between maintaining a faithful representation of the modelled biological systems and producing state of the art performance in the robotics domain at a relevant task?

To discuss this issue concisely in a paper such as this, one must necessarily make some generalizations. Research focusing on maintaining fidelity to the underlying source of biological inspiration often produces performance that is inferior to conventional mathematical approaches, but can lead to novel insightful predictions about biological systems. Conversely, research that readily abandons any relevance to the biology may lead to better robotics performance but is rarely the cause of new discoveries in biological research. In addition, it becomes an increasingly painful process to generate relevant testable predictions or insights in the biological field as the model becomes more and more abstracted.

In the initial stages of the RatSLAM project, we started with what was then a state of the art neural network model of the mapping processes observed in the rodent brain. As we tested the algorithms in larger and more challenging environments and over longer periods of time, we were forced to make some pragmatic modifications to the

algorithms to produce good mapping performance. These modifications seemingly moved the model further away from biology. One example would be the pragmatic decision to engineer the *pose cells*, artificial neurons that encode the complete three-dimensional (x, y, ϑ) pose of a ground-based robot and are re-used at regular intervals to efficiently encode large environments. The decision to move to *pose cells* was made because the neuron types known at that time – *place cells* which represent (x, y) location – and *head-direction cells* which represent orientation – were unable to represent and correctly update multiple robot location hypotheses. Subsequently neuroscientists discovered a new type of spatial neuron called a *grid cell* in the rodent brain sharing similar although not identical characteristics [15, 32]. This discovery demonstrated that a functionally driven investigation (engineering a new cell to produce better mapping performance) could lead to relevant insights or predictions in another discipline, in this case neuroscience. It is interesting to speculate that, had we abandoned the biological neural network completely and moved to a conventional technique such as a particle or Kalman filter, it may have been harder to make this specific prediction. Conversely, if we had maintained a more biological faithful model, we may never have been able to test it in environments that were sufficiently challenging to require the ability to encode and propagate multiple location hypotheses. At least in this particular example, it was only by following the "middle ground" that we were able to make some contribution to both fields.

Acknowledgements

This work was supported by an Australian Research Council Discovery Project DP120102775 and Microsoft Research Faculty Fellowship to MM, and an ARC & NHMRC Thinking Systems grant TS0669699 to GW.

References

[1] G. Wyeth and M. Milford, "Spatial Cognition for Robots: Robot Navigation from Biological Inspiration," IEEE Robotics & Automation Magazine, vol. 16, pp. 24-32, 2009.

[2] M. J. Milford, Robot Navigation from Nature: Simultaneous Localisation, Mapping, and Path Planning Based on Hippocampal Models vol. 41. Berlin-Heidelberg: Springer-Verlag, 2008.

[3] M. J. Milford, G. Wyeth, and D. Prasser, "RatSLAM: A Hippocampal Model for Simultaneous Localization and Mapping," in IEEE International Conference on Robotics and Automation, New Orleans, USA, 2004, pp. 403-408.

[4] 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.

[5] M. Milford and G. Wyeth, "Single Camera Vision-Only SLAM on a Suburban Road Network," in International Conference on Robotics and Automation, Pasadena, United States, 2008.

[6] 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.

[7] 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.

[8] M. Milford, J. Wiles, and G. Wyeth, "Solving Navigational Uncertainty Using Grid Cells on Robots," PLoS Computational Biology, vol. 6, 2010.

[9] M. Milford, I. Turner, and P. Corke, "Long exposure localization in darkness using consumer cameras," in Proceedings of the 2013 IEEE International Conference on Robotics and Automation, 2013.

[10] M. Milford, "Vision-based place recognition: how low can you go?," International Journal of Robotics Research, vol. 32, pp. 766-789, 2013.

[11] 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.
[12] M. Milford, "Visual Route Recognition with a Handful of Bits," in Robotics: Science and Systems VIII, Sydney, Australia, 2012.
[13] M. Milford, E. Vig, W. Scheirer, and D. Cox, "Towards Condition-Invariant, Top-Down Visual Place Recognition," in Australasian Conference on Robotics and Automation, Sydney, Australia, 2013.
[14] Z. Chen, A. Jacobson, U. M. Erdem, M. E. Hasselmo, and M. Milford, "Towards Bio-inspired Place Recognition over Multiple Spatial Scales," in Australasian Conference on Robotics and Automation, Sydney, Australia, 2013.
[15] 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.
[16] I. Golani, G. Bronchti, D. Moualem, and P. Teitelbaum, ""Warm-up" along dimensions of movement in the ontogeny of exploration in rats and other infant mammals," Proceedings of the National Academy of Sciences of the United States of America, vol. 78, pp. 7226-7229, 1981.
[17] P. Stratton, M. Milford, G. Wyeth, and J. Wiles, "Using Strategic Movement to Calibrate a Neural Compass: A Spiking Network for Tracking Head Direction in Rats and Robots," PLoS One, vol. 6, 2011.
[18] A. Cheung, D. Ball, M. Milford, G. Wyeth, and J. Wiles, "Maintaining a Cognitive Map in Darkness: The Need to Fuse Boundary Knowledge with Path Integration," PLoS Computational Biology, vol. 8, 2012.
[19] E. E. Thomson, R. Carra, and M. A. Nicolelis, "Perceiving invisible light through a somatosensory cortical prosthesis," Nat Commun, vol. 4, p. 1482, Feb 12 2013.
[20] A. Jacobson, Z. Chen, and M. Milford, "Autonomous Movement-Driven Place Recognition Calibration for Generic Multi-Sensor Robot Platforms," presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 2013.
[21] M. Milford and A. Jacobson, "Brain-based Sensor Fusion for Navigating Robots," in IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 2013.
[22] 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.
[23] 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.
[24] 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.
[25] 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.
[26] Y. Burak and I. R. Fiete, "Accurate path integration in continuous attractor network models of grid cells," PLoS Computational Biology, vol. 5, 2009.
[27] 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.
[28] I. T. Jolliffe, Principal Component Analysis, 2 ed.: Springer, 2002.
[29] 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.
[30] V. Vapnik, "The support vector method of function estimation," Nonlinear Modeling, pp. 55-85, 1998.
[31] A. Jacobson and M. Milford, "Towards Brain-based Sensor Fusion for Navigating Robots," in Proceedings of the 2012 Australasian Conference on Robotics & Automation, 2012.
[32] M. Fyhn, S. Molden, M. P. Witter, E. I. Moser, and M.-B. Moser, "Spatial Representation in the Entorhinal Cortex," Science, vol. 27, pp. 1258-1264, 2004.