Available online at www.sciencedirect.com

ScienceDirect

Robotics and Autonomous Systems 55 (2007) 403–410

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Learning spatial concepts from RatSLAM representations

Michael Milford*, Ruth Schulz, David Prasser, Gordon Wyeth, Janet Wiles

School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, Queensland 4072, Australia

Available online 29 December 2006

Abstract

RatSLAM is a biologically-inspired visual SLAM and navigation system that has been shown to be effective indoors and outdoors on real robots. The spatial representation at the core of RatSLAM, the experience map, forms in a distributed fashion as the robot learns the environment. The activity in RatSLAM's experience map possesses some geometric properties, but still does not represent the world in a human readable form. A new system, dubbed RatChat, has been introduced to enable meaningful communication with the robot. The intention is to use the "language games" paradigm to build spatial concepts that can be used as the basis for communication. This paper describes the first step in the language game experiments, showing the potential for meaningful categorization of the spatial representations in RatSLAM.

Keywords: Spatial conceptualization; RatSLAM; SLAM; Experience mapping

1. Introduction

Recent research in mobile robotics has been dominated by the problem of Simultaneous Localization And Mapping (SLAM). Roboticists have investigated a wide range of approaches to solving the problem and have created a number of probabilistic methods that can perform SLAM under appropriate assumptions [2,7,14]. However, by focusing on the SLAM problem other considerations such as map usability have mostly been neglected. Traditional metrics such as accuracy are starting to be supplanted by concepts such as map usability and communicability.

Geometric space representations are a natural choice for many robot mapping and localization methods, but are dissimilar to the more abstract ways in which humans view their environments. Humans can conceptualize their environment in terms of concepts such as rooms: "the bathroom", "the kitchen"; or objects: "behind the couch", or "on top of the table". If humans are to easily and naturally interact with robots, the robots must be able to understand and process such concepts. For instance, one goal for domestic robots would be the ability for a human to tell a robot to "clean the bathroom", rather than specifying a range of geometric co-ordinates.

Conceptualization and communication can be investigated using embodied agents that develop languages for labeling objects in their environment, or for coordinating signaling and motor behaviors for the completion of a task. Embodied agents have been implemented as both static and mobile robots, using communication based on synthetic and natural languages. While the majority of studies investigating communication in embodied agents involve robot–robot communication, some involve interaction with humans. By interacting with humans, robots can be taught to understand natural language for labels [13], descriptions [11], or can play games [12], where the human is another agent with which the robot can interact. For a review on communication in embodied agents see [8].

Conceptualizations are typically grounded in the image domain or in the simple sensory perceptions of the robots. The representations used as the basis for conceptualization include vision, where scenes are segmented and processed for concepts of color, position, and size [12], minimal proximity sensors, and light sensors [3]. Mobile robots can be taught spatial descriptions such as left, right, front, and back, and can follow natural language instructions relating to movement, such as "move forward" [11]. Because these concepts are derived from the physical world through the robot's sensors, they are said to be physically grounded.

* Corresponding author. Fax: +61 7 3365 4999.
E-mail addresses: milford@itee.uq.edu.au (M. Milford), ruth@itee.uq.edu.au (R. Schulz), prasserd@itee.uq.edu.au (D. Prasser), wyeth@itee.uq.edu.au (G. Wyeth), wiles@itee.uq.edu.au (J. Wiles).

0921-8890/$ - see front matter
doi:10.1016/j.robot.2006.12.006

To develop and appropriately generalize spatial concepts, such as rooms, the agent will need to access representations based in space rather than directly from a sensor. This study uses a SLAM system, called RatSLAM, to produce spatial representations. RatSLAM is a vision-based system inspired by models of mapping and navigation in the rodent hippocampus [5], which has been demonstrated to autonomously explore, SLAM, navigate to goals, and adapt to simple environment changes. The spatial representations that it produces and uses do not conform to the constraints of typical Cartesian mapping, but rather coalesce the robot's physical experiences, based on sensing and actuation. The spatial representations, as a consequence, have strong physical grounding but are not effective for robot–human communication.

RatChat is a proposed language system extension to RatSLAM to enable a population of robots (and potentially humans) to form a language by playing language games. Preliminary studies have investigated the robot representations, and how agents can generalize from the training sets to novel meanings and terms [10]. In this work a teacher–student paradigm is used to teach the robot to recognize distinct rooms and corridors in a large indoor environment. The aim is to form spatial concepts that are strongly tied to the physical world to allow appropriate learning and generalization.

The remainder of the paper is structured as follows. Section 2 describes the core RatSLAM system and its representation of robot pose and vision. Section 3 presents an algorithm known as experience mapping, which creates robot usable maps from the core RatSLAM representations. The spatial conceptualization process is described in Section 4, showing how a neural network can be used to label and recognize distinct regions of the environment. Section 5 describes the experimental setup and procedure, and is followed by a presentation of the results in Section 6. Section 7 discusses the results, before the paper concludes in Section 8.

2. RatSLAM

The RatSLAM system is centered around two core components; the *pose cell* structure, which encodes the robot's position and orientation, and the *local view* structure, which encodes visual information. These two structures and their interaction are briefly described here (for a more detailed description see [5,4]).

2.1. Pose cells

The robot's pose is represented by a three-dimensional continuous attractor neural network called the pose cells (shown in the right of Fig. 1). Each pose cell is associated with a specific robot pose within the environment, and to a lesser degree with proximal robot poses. Cells are fully interconnected by excitatory links, with stronger link values between nearby cells. Active cells therefore tend to activate neighbouring cells, creating a cluster of active cells known as an *activity packet*. Global inhibition and normalization ensure that activity does not spread in an unbounded manner. Multiple activity packets can coexist and compete against one another for extended periods of time, with each packet encoding a robot pose hypothesis. Path integration is achieved by using information from the robot's wheel encoders to shift the current activity in the pose cells. Translation causes activity to shift in the (x', y') plane of the pose cells, while rotation causes activity to shift in the θ' direction. The network equations are given in [5].

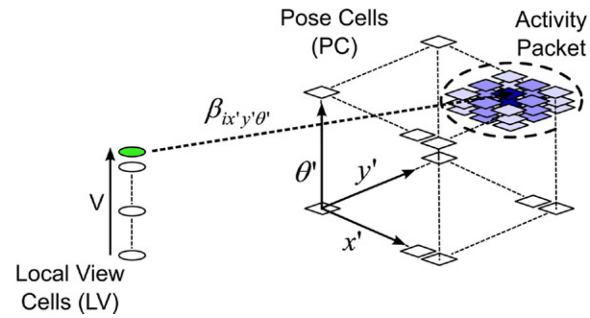

Fig. 1. The local view and pose cell structures. The local view encodes visual information about the environment. The pose cells represent the robot's pose. Co-activated local view and pose cells form associative links. A familiar visual scene activates a local view cell which in turn injects activity into the pose cells associated with it. Re-localization occurs when the activity packet caused by visual input becomes dominant.

In indoor environments the path integration model parameters result in each pose cell initially corresponding to an area of 250 × 250 mm and an orientation range of 10°. As the robot moves around its environment, path integration errors and re-localization driven by visual input causes the pose cells' spatial correspondence to vary — a cell may become associated with a larger or smaller physical area and orientation range. The pose structure boundary conditions wrap activity across opposing faces of the pose cell structure, allowing the robot to map much larger environments than would be possible with a one-to-one correspondence between cell and space. This varying representation can result in collisions, where one cell represents multiple places in the environment, and multiple representations, where multiple cells are associated with the same place. However the robot is still able to function by using the experience mapping algorithm to create a coherent global map (discussed in Section 3).

2.2. Local view

Visual information is stored in a one-dimensional array of cells called the local view (shown on the left of Fig. 1). Each local view cell is associated with a distinct visual scene. RatSLAM recognizes locations from external sensor information provided by an appearance-based view recognition system [6]. The role of this system is to create patterns of activity in the local view cells which depend on robot location, using the camera data. Camera information is matched against a growing database of learnt images, each of which has an associated cell in the local view. Recognition of an image activates the appropriate cell. Unrecognized camera images are added to the database so that the system is able to explore a previously unseen environment.

In indoor environments the vision system uses low resolution (24 × 18) normalized grayscale images. A sum of absolute differences matcher is used to compare the current image to those stored in the database on a pixel by pixel basis. Learnt views that are within a threshold distance of the current view have their local view cells activated in inverse proportion to their distance from the current view. A completely novel view creates a new local view cell. Computation scales with the number of images learned, which scales approximately with the area of the environment. During a typical hour long experiment in an indoor environment, the robot may learn tens of thousands of template images. More sophisticated image processing methods such as edge detection have been found to provide no increase in functionality in indoor environments using a small field of view camera. However, in outdoor environments with a panoramic camera, a histogram matching procedure in red–green space is used to provide rotational invariance and some robustness to illumination change [9].

2.3. Visual association and re-localization

Two simultaneous interactions occur between the local view cells and pose cells: mapping via associative learning, and re-localization by the injection of activity into the pose cells. Visual scenes are associated with the robot's estimate of its current pose by a learning function which increases the connection strengths between co-activated local view and pose cells. The connection strength, $\beta^{t+1}_{ix'y'\theta'}$, between a local view cell and a pose cell, is given by:

$$\beta^{t+1}_{ix'y'\theta'} = \max\left(\beta^{t}_{ix'y'\theta'}, \lambda V_i P_{x'y'\theta'}\right) \tag{1}$$

where V_i is the activity level of the local view cell, λ is an association constant, and $P_{x'y'\theta'}$ is the activity level of the pose cell.

The associative links between local view and pose cells can be used to inject activity into the pose cells, which is the mechanism that RatSLAM uses to maintain or correct its estimate of its current pose. A familiar visual scene activates a local view cell, which in turn projects activity into the pose cells with which it is associated, by an amount proportional to the association strengths. The change in pose cell activity, $\Delta P_{x'y'\theta'}$, is given by:

$$\Delta P_{x'y'\theta'} = \sum_i \beta_{ix'y'\theta'} V_i. \tag{2}$$

Visual ambiguity or redundancy in the environment is accounted for by the view to pose associations, which enable one view to correspond to multiple physical locations and vice versa. The network dynamics ensure that re-localization can only occur after extended visual support of an alternative pose hypothesis, which filters the effects of incorrect recognition and visual ambiguity.

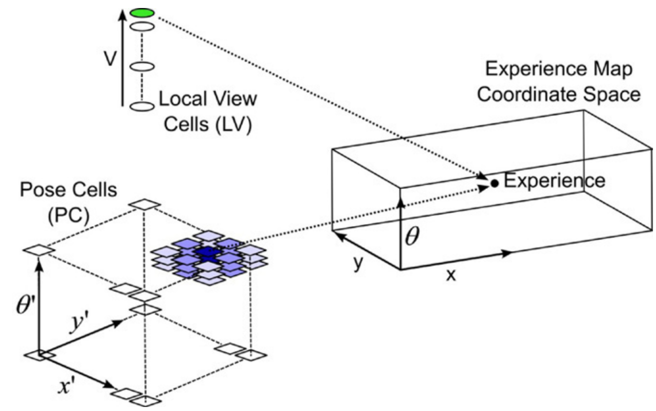

Fig. 2. Experience map co-ordinate space. An experience is associated with certain pose and local view cells, but exists within the experience map's own (x, y, θ) co-ordinate space. x', y', and θ' describe the location of the cells within the pose cell matrix associated with the experience, and V describes the local view cell associated with the experience.

3. Experience mapping

Experience mapping is a process that creates a spatially coherent and continuous map from both the pose and local view cell structures, which can be used for higher level tasks such as goal navigation. The method is centered on a process of map correction, which creates a map free of the spatial discontinuities and multiple representations present in the pose cell representations. As understanding the experience map is essential to appreciating the representations used in spatial concept learning (Section 4), this section gives a brief overview of the process (for a more detailed explanation of the algorithm see [4]).

3.1. Experiences

Activity in the pose cells and local view cells drives the creation of experiences. Experiences are represented by nodes in (x, y, θ) space connected by links representing transitions between experiences. Each experience represents a snapshot of the activity within the pose cells and local view cells at a certain time, and in effect represents a specific spatial and visual robot experience (see Fig. 2).

Experiences have an activation level that depends on how close the activity peaks in the pose and local view networks are to the cells associated with each experience. Each experience has a *zone of association* within the pose cells and local view cells. When the activity peaks in each network are within these zones, the experience is activated. Within the pose cells the zones are continuous, so an experience can be associated with many pose cells. In the local view cells the zone is discrete, allowing the experience to only be associated with one visual scene. A component of an experience's activity level is determined by the current pose cell activity, $E_{x'y'\theta'}$, and is calculated by:

$$E_{x'y'\theta'} = \begin{cases} 0 & \text{if } r'_r > 1; \\ 0 & \text{if } \theta'_r > 1; \\ 2 - r'_r - \theta'_r & \text{otherwise} \end{cases} \tag{3}$$

$$r_r' = \frac{\sqrt{\left(x_{pc}' - x_i'\right)^2 + \left(y_{pc}' - y_i'\right)^2}}{r_a} \tag{4}$$

$$\theta_r' = \frac{|\theta_{pc}' - \theta_i'|}{\theta_a} \tag{5}$$

where x_{pc}', y_{pc}', and θ_{pc}' are the co-ordinates of the maximally active pose cell, x_i', y_i', and θ_i' are the co-ordinates of the pose cells associated with the experience, r_a is the zone constant for the (x', y') plane, and θ_a is the zone constant for the θ' dimension.

The visual scene V acts like a switch for the experience, turning it on or off. The total activity level of the ith experience, E_i, is given by:

$$E_i = \begin{cases} 0 & \text{if } V_{\text{curr}} \neq V_i; \\ E_{x'y'\theta'} & \text{if } V_{\text{curr}} = V_i \end{cases} \tag{6}$$

where V_{curr} is the current visual scene, and V_i is the visual scene associated with experience i. New experiences are created when the set of existing experiences is insufficient for describing the pose and local view cells' activity state.

Each experience also has its own (x, y, θ) state which describes its location within the co-ordinate space of the experience map. This co-ordinate space is completely separated from the pose and local view cell co-ordinate spaces. The first experience learned is initialized with an arbitrary (0, 0, 0) position within the experience map. Subsequent experiences are assigned a position based on the position of the last experience and the robot movement that has occurred since.

3.2. Experience transitions

As the robot moves around the environment, the experience mapping algorithm also learns experience *transitions*, which are links between experience nodes. Transitions store information about the physical movement of the robot between one experience and another, the movement behavior used during the transition, the time duration of the transition, and the traversability of the transition (see Fig. 3). This information is used to perform experience map correction, route planning and execution, and adaptation to environment change.

3.3. Map correction

The experience map correction process creates a spatially coherent map by minimizing the discrepancies between the relative spatial information stored in the experience transitions and the locations of the experience nodes in (x, y, θ) experience map space. For instance, when closing the loop, the robot's camera first detects familiar visual scenes. The vision system activates the local view cells associated with these scenes, which then inject activity into the pose cell matrix, causing the robot to re-localize its perceived pose. This causes the robot's associated location within the experience map (given by the maximally active experience) to jump from the new experience it has most recently learned to a previously learnt experience. The experience mapping algorithm learns this

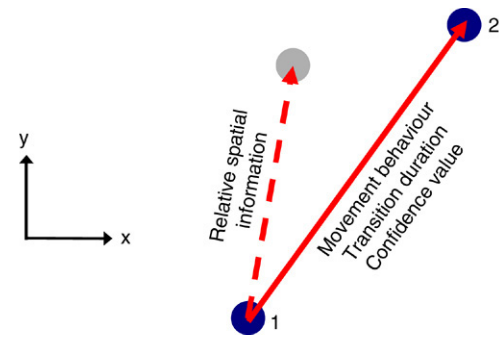

Fig. 3. A transition link between two experiences 1 and 2, shown in the (x, y) plane of the experience map. Each link represents the movement behavior required to move between the experiences, the average time taken to perform the transition, and a confidence value representing the robot's degree of certainty the transition can be traversed. Each link also stores relative spatial information obtained directly from robot odometry about the relative location of experiences to each other (shown by the dashed arrow).

new transition, which contains a large discrepancy between the transition's relative spatial information and the difference between the two experiences (x, y, θ) co-ordinates in the experience map. For example, two experiences may be positioned several meters apart in the experience map but be linked by transitions encoding much smaller distances. By minimizing the discrepancies between the relative locations of experiences in the experience map and the inter-experience spatial transition information, the experience map can become locally representative of the environment's spatial arrangement. The map correction equations are provided in [4].

4. Teacher–student spatial conceptualization using RatChat

The experience mapping process produces a map with strong geometric properties that the robot can use to navigate effectively around its environment. To facilitate human–robot interaction, a method of learning and recognizing spatial concepts is required. A teacher–student system was designed and implemented, in which a robot attempts to associate concepts provided by a human teacher with its internal representations using a single layer neural network.

4.1. Learning concepts from a teacher

Teacher–student conceptualization involves interaction with a teacher, where the different concepts that the agent is to learn are provided by that teacher. Agents use supervised learning to associate input patterns with different concepts. For the experiments presented in this paper, the RatChat agents used experiences in the experience map as the input pattern. The simplest concepts that can be formed, considering the information contained in the experience representations, are labels for locations in the world, such as specific rooms.

The output representation of the language agents should group input patterns into categories or concepts. The simplest output representation is to have each concept associated with a single output unit, also called one-hot encoding (not to

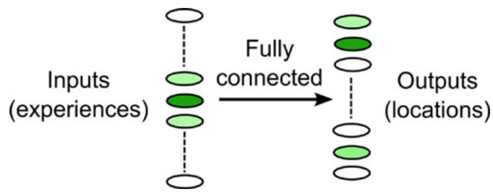

Fig. 4. The fully connected single layer neural network of the language agent takes experiences as inputs and has outputs associated with labels for locations in the world. The most active output is the label associated with the active experiences.

be confused with one-shot which is a learning technique). For locations in an indoor environment, the concepts are specifically labeled rooms and corridors, with each location associated with a single output unit. The conceptualization process for RatChat agents is the association of the input patterns of experiences with the output representations of locations.

4.2. Association mechanism

The agents learn this association using a fully connected single layer neural network shown in Fig. 4, with experiences as inputs and a set of output units referring to the different locations in the world. The transfer function of the network is linear. The network is initialized with small random weights and biases (uniformly between −0.1 and 0.1). The network is trained using gradient descent with momentum and an adaptive learning rate. The change in weights and biases, $\mathrm{d}X$, is given by:

$$\mathrm{d}X = m.\mathrm{d}X_{k-1} + \eta(1 - m)g \tag{7}$$

where m is the momentum constant, $\mathrm{d}X_{k-1}$ is the previous change in weights and biases, η is the learning rate, and g is the gradient of the network's performance with respect to its weight and bias values. After each epoch, if the performance has decreased, the learning rate is increased, and if the performance has increased significantly, the learning rate is reduced. Specific parameter values are given in the experimental procedure described in Section 5.3.

During recall, the concept associated with a pattern of experiences is the concept related to the most active output unit. The agent is considered to be 'uncertain' if the activation of the second most active unit is more than $2/3$ the activation of the most active unit. Preliminary experiments determined that a value of $2/3$ provided an appropriate balance between concept uncertainty and incorrect guessing of concepts at room boundaries.

5. Experimental setup and procedure

The experiments used a Pioneer 2 DXE mobile robot with a forward facing camera to explore a test environment. The resulting dataset was processed by the RatSLAM model and experience mapping algorithm in order to provide the input for the spatial conceptualization method. The spatial conceptualization process was then applied in an offline manner following the construction of the RatSLAM representations.

5.1. Environment and robot

The environment was one floor of a university building consisting mostly of open-plan offices and corridors (shown in Fig. 5). The robot was manually driven along a repeated path through the environment. The robot visited every place on its path at least twice, providing an opportunity for both learning and recognition.

A dataset was acquired with camera images logged at 7 Hz and on-board odometry data logged at 12 Hz. The data set contained 20,350 monochrome images with associated odometer readings covering a period of almost 40 min. The data set was then 'played back' to the RatSLAM system at real-time speed.

5.2. RatSLAM parameters

Experiments used a pose cell structure that was sufficiently large to avoid wrapping of the activity across the structure boundaries. The pose cell structure measured 200 × 100 × 36 cells (720,000 cells in total) in (x', y', θ'). Under the path integration model, the pose cell numbers correspond to nominal environment dimensions of 50 × 25 m.

5.3. Spatial conceptualization training and testing

The conceptualization process was implemented offline following the construction of the pose cell and experience maps. The environment was manually categorized by a human teacher into four rooms and two corridors, as shown in Fig. 5.

The route of the robot was divided into two sections of about 20 min duration each. Each section corresponded to the robot exploring and then revisiting one half of the building floor. These two sections were further divided into learning and recognition phases. The learning phase, in which the robot first visited an area, was used for the training set, while the recognition phase, where the robot revisited an area, was used to test if the concepts had been learnt. This was equivalent to the areas being labeled on the first circuit of the environment, and testing whether the robot had learnt these labels on later circuits.

The language agent's fully connected single layer neural network used experiences as inputs and had six output units corresponding to the concepts of four rooms and two corridors. The activations of the experience inputs was determined by including the current experience and those experiences within 1 m, with an activation relative to how close the experience was to the current experience.

Targets were created with a single active output unit corresponding to the current location of the robot. Transitions between rooms and corridors occurred at doorways and turns. The first learning phase comprised 403 time steps, with 233 in the second learning phase, 398 in the first recognition phase, and 187 in the second recognition phase.

Agents were initially trained on the first learning phase and tested on the first recognition phase. Agents were then trained on both the first and second learning phase and tested on the

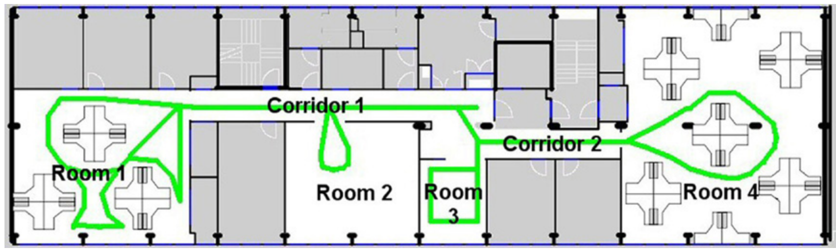

Fig. 5. Floor plan of the area used for the experiment and the approximate trajectory of the robot. Shaded areas were impassable by the robot. The environment measured approximately 43 by 13 m.

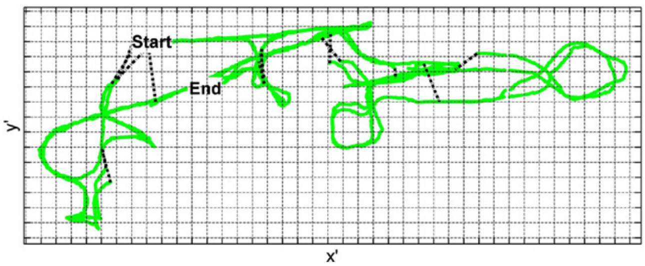

Fig. 6. Trajectory of the most highly activated pose cell during the experiment. Thick dashed lines show re-localization jumps driven by visual input. Each grid square contains 4×4 pose cells in the (x', y') plane.

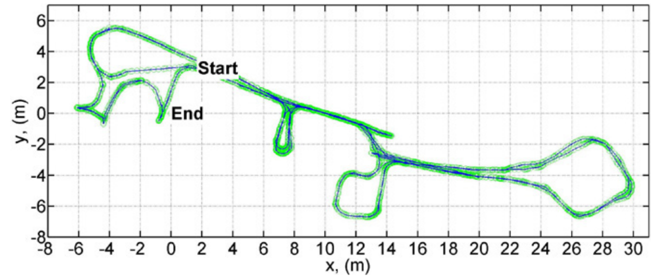

Fig. 7. The experience map. The map is continuous and has a high degree of correspondence to the spatial arrangement of the environment shown in Fig. 5. The multiple representations in the pose cell representation (Fig. 6) have been grouped into overlapping representations, and there are no spatial discontinuities.

second recognition phase. For each training segment, agents were trained for 2000 epochs using Eq. (7) with the momentum constant $m = 0.9$, and an initial learning rate $\eta = 0.01$ with increasing ratio $= 1.05$ and decreasing ratio $= 0.7$. The performance of the agents was tested on the first and second recognition phase by considering the concepts used by the agents for each location.

6. Results and analysis

The pose cell and experience map representations built by RatSLAM were collated and analyzed (Section 6.1). The results from the conceptualization experiments were grouped into the two learning and two recognition phases and then analyzed (Section 6.2).

6.1. Pose cell representations and experience maps

The pose cell representation produced by RatSLAM contains both discontinuities and multiple representations of the same place (see Fig. 6). The discontinuities were caused by visually driven re-localization jumps after long periods of exploration where the robot relied only on wheel odometry to remain localized. Odometric drift and delayed re-localization created multiple representations, where more than one group of pose cells represented the same physical location. There are also collisions in the pose cell structure, where the same cluster of cells represents more than one physical place. During the entirety of the experiment only a small fraction – 35,402 out of 720,000 – of the pose cells ever became active. This sparseness resulted from the use of a pose cell structure that was large enough to represent the environment without wrapping.

The experience mapping algorithm produced a spatially continuous map (cf. Figs. 6 and 7), with multiple representations grouped into overlapping areas of the map (see Fig. 7). During the experiment the robot learned 2384 experiences, which is significantly fewer than the number of activated pose cells, as each experience was associated with several pose cells.

6.2. Conceptualization using the experience map

In the first learning phase of the conceptualization process based on the experience map, 98.26% of the instances were labeled correctly, with 90.45% labeled correctly in the recognition phase (Fig. 8(a), (b)). The majority of the errors occurred in Room 1 where a different trajectory was taken during the second pass. In this case, the agent was uncertain about the label, rather than labeling the instances incorrectly.

In the second learning phase, 98.43% were labeled correctly, with 89.84% labeled correctly in the recognition phase (Fig. 8(c), (d)). The errors in this case all occurred at the boundaries between rooms and corridors. The RatChat agents were successful in clustering the experiences appropriately, with some uncertain errors on the borders between areas (see Fig. 8(d)). At all locations, except for those on borders between areas, and those not visited during the learning phase, the agents were able to appropriately label their current location.

7. Discussion

The RatSLAM system readily and reliably produces spatially coherent representations of the robot's environment

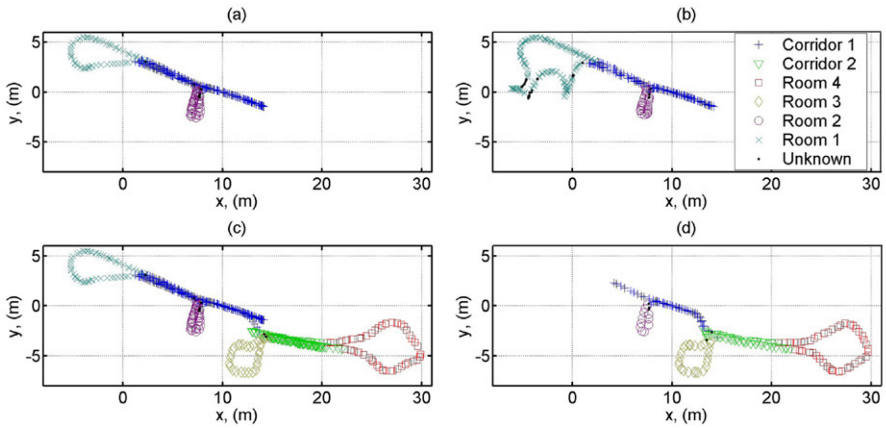

Fig. 8. Conceptualization of the agent using experiences in (a) the learning phase of section 1, (b) the recognition phase of section 1, (c) the learning phase of section 2, and (d) the recognition phase of section 2. In the learning phases there are uncertain areas in the Room 1/Corridor 1, Room 2/Corridor 1, and Room 3/Corridor 2 borders. In the recognition phases there are uncertain areas in Room 1, and in the Room 1/Corridor 1, Room 2/Corridor 1, and Room 3/Corridor 2 borders.

after a period of exploration. RatSLAM learns, builds and adapts its representations from its sensory and behavioral experiences, and integrates those experiences into the experience map. The use of the robot's own physical experiences as the basis for the spatial representations is a powerful principle: it has long been argued that agents need a strong physical grounding to interact effectively with the real world [1]. It needs to be recognized though that physically grounding the representations of the robot in its own unique experiences comes at a cost. No two robots share the same experience map, nor can the map be defined by the user as this would violate the physical grounding of the map. This creates difficulties for communication between robots, or between a human and the robot, as there is no shared symbolic base for communication.

The RatChat project is developing a reliable way of forming higher level concepts about space that are usable in robot–robot and human–robot interaction. Abstract spatial concepts, such as rooms, are identified during training only by entry and exit times. The robot does not explicitly recognize specific features that identify a room type, but rather learns the spatial correspondence between physical place and concept through the intermediary layer of the experience map. As the experience map evolves over time by integrating new robot experiences, the trained concepts generalize to new experiences in the same space. The physical grounding of the concepts is maintained.

Given the quality of the spatial representations formed by the experience map, it is perhaps unsurprising that the language agents could readily learn the spatial categories from a teacher. Nevertheless the result is important as it illustrates that it is possible to form exchangeable spatial concepts without violating the physical grounding principle. The system has immediate applications as well. It enables the robot to communicate its location in a useful way to a human supervisor. A symbol representing "I'm in the laboratory" is more meaningful to a human than "I'm experiencing experience number 1324".

The spatial conceptualization process described in this paper is an implicit one. Future work will investigate learning more complex concepts and also involve a community of robots evolving a lexicon of locations using language games. The results from this paper show that the experience map provides a sound representation as the basis for evolution of a spatial lexicon.

8. Conclusion

This paper has investigated a method for learning and recognizing abstract spatial concepts using the RatSLAM model and experience mapping algorithm. Using a simple neural network, the conceptualization process associates physically grounded experiences with spatial concepts. The experiments demonstrate that it is possible to learn abstract spatial concepts such as rooms and corridors and then generalize about these concepts when revisiting the physical areas.

Acknowledgments

The authors thank the Australian Research Council for partial funding of the RatSLAM and RatChat projects, and the reviewers for their helpful comments regarding the earlier drafts of this paper.

References

[1] R.A. Brooks, Elephants don't play chess, Robotics and Autonomous Systems 6 (1990) 3–15.
[2] G. Dissanayake, P.M. Newman, S. Clark, H. Durrant-Whyte, M. Csorba, A solution to the simultaneous localisation and map building (SLAM) problem, IEEE Transactions on Robotics and Automation 17 (2001) 229–241.
[3] D. Marocco, S. Nolfi, Emergence of communication in teams of embodied and situated agents, in: A. Cangelosi, A.D.M. Smith, K. Smith (Eds.), The Evolution of Language, World Scientific, Singapore, 2006, pp. 198–205.
[4] M.J. Milford, D. Prasser, G. Wyeth, Experience mapping: producing spatially continuous environment representations using RatSLAM, in: Australasian Conference on Robotics and Automation, Sydney, Australia, 2005.

[5] M.J. Milford, G. Wyeth, D. Prasser, RatSLAM: A hippocampal model for simultaneous localization and mapping, in: International Conference on Robotics and Automation, New Orleans, USA, 2004.
[6] M.J. Milford, G. Wyeth, D. Prasser, Simultaneous localization and mapping from natural landmarks using RatSLAM, in: Australasian Conference on Robotics and Automation, Canberra, Australia, 2004.
[7] M. Montemerlo, S. Thrun, D. Koller, B. Wegbreit, FastSLAM: A factored solution to the simultaneous localization and mapping problem, in: AAAI National Conference on Artificial Intelligence, Edmonton, Canada, 2002.
[8] S. Nolfi, Emergence of communication in embodied agents: Co-adapting communicative and non-communicative behaviours, Connection Science 17 (2005) 231–248.
[9] D. Prasser, M. Milford, G. Wyeth, Outdoor simultaneous localisation and mapping using RatSLAM, in: International Conference on Field and Service Robotics, Port Douglas, Australia, 2005.
[10] R. Schulz, P. Stockwell, M. Wakabayashi, J. Wiles, Generalization in languages evolved for mobile robots, in: ALIFE X: Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems, MIT Press, 2006, pp. 486–492.
[11] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, D. Brock, Spatial language for human–robot dialogs, IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews 34 (2004) 154–167.
[12] L. Steels, The Talking Heads Experiment, in: Words and Meanings, vol. I, Best of Publishing, Brussels, 1999.
[13] L. Steels, F. Kaplan, AIBO's first words. The social learning of language and meaning, Evolution of Communication 4 (2001) 3–32.
[14] S. Thrun, Probabilistic algorithms and the interactive museum tour-guide robot minerva, Journal of Robotics Research 19 (2000) 972–999.

Michael Milford was born in 1981 in Brisbane, Australia. He holds a Ph.D. in Electrical Engineering and a Bachelor of Engineering from the University of Queensland, awarded in 2006 and 2002, respectively. He is currently a Postdoctoral Research Fellow in the Robotics Laboratory at the University of Queensland. His research interests include Simultaneous Localisation And Mapping, computational modelling of the rodent hippocampus and entorhinal cortex, and biologically inspired robot navigation.

Ruth Schulz received a dual degree in Engineering (Electrical) and Science (Computer Science) from the University of Queensland, Australia, in 2003. She is currently a Ph.D. student at the University of Queensland in the Complex and Intelligent Systems group. Her research interests include the evolution of language, spatial language, and computational models of language for mobile robots.

David Prasser received a bachelor's degree in Electrical Engineering from the University of Queensland, Australia in 2001. He is currently preparing his Ph.D. thesis in the University of Queensland's School of Information Technology and Electrical Engineering. His research interests are in the areas of biologically inspired robotics and computer vision.

Gordon Wyeth holds a Ph.D. in Electrical Engineering from the University of Queensland, and is the founder and director of its Robotics Laboratory. He has designed and constructed more than twenty robots, including flying robots, wall-climbing robots, high performance wheeled robots, manipulators and a humanoid robot. He is the President of the Australian Robotics and Automation Association, and head of the Mechatronics Engineering Program at the University of Queensland. His research interests include biologically inspired robot navigation and humanoid robotics.

Janet Wiles holds a Ph.D. from the University of Sydney, and is Professor of Complex and Intelligent Systems at the University of Queensland. She is the project leader of the Thinking Systems Project, supervising a cross-disciplinary team studying fundamental issues in how information is transmitted, received, processed and understood in biological and artificial systems. Her research interests include complex systems biology, computational neuroscience, computational modelling methods, artificial intelligence and artificial life, language and cognition.