Neural Networks 126 (2020) 21–35

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

NeuroBayesSLAM: Neurobiologically inspired Bayesian integration of multisensory information for robot navigation[☆]

Taiping Zeng[a,b,c,d], Fengzhen Tang[c,d], Daxiong Ji[e], Bailu Si[f,*]

[a] *Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China*
[b] *Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, China*
[c] *State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China*
[d] *Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China*
[e] *Ocean College, Zhejiang University, Zhoushan, 316021, Zhejiang, China*
[f] *School of Systems Science, Beijing Normal University, 100875, China*

ARTICLE INFO

Article history:
Received 23 October 2019
Received in revised form 10 February 2020
Accepted 28 February 2020
Available online 4 March 2020

Keywords:
Bayesian
Multisensory integration
Attractor dynamics
Head direction cells
Grid cells
Monocular SLAM

ABSTRACT

Spatial navigation depends on the combination of multiple sensory cues from idiothetic and allothetic sources. The computational mechanisms of mammalian brains in integrating different sensory modalities under uncertainty for navigation is enlightening for robot navigation. We propose a Bayesian attractor network model to integrate visual and vestibular inputs inspired by the spatial memory systems of mammalian brains. In the model, the pose of the robot is encoded separately by two subnetworks, namely head direction network for angle representation and grid cell network for position representation, using similar neural codes of head direction cells and grid cells observed in mammalian brains. The neural codes in each of the sub-networks are updated in a Bayesian manner by a population of integrator cells for vestibular cue integration, as well as a population of calibration cells for visual cue calibration. The conflict between vestibular cue and visual cue is resolved by the competitive dynamics between the two populations. The model, implemented on a monocular visual simultaneous localization and mapping (SLAM) system, termed NeuroBayesSLAM, successfully builds semi-metric topological maps and self-localizes in outdoor and indoor environments of difference characteristics, achieving comparable performance as previous neurobiologically inspired navigation systems but with much less computation complexity. The proposed multisensory integration method constitutes a concise yet robust and biologically plausible method for robot navigation in large environments. The model provides a viable Bayesian mechanism for multisensory integration that may pertain to other neural subsystems beyond spatial cognition.

1. Introduction

Simultaneous localization and mapping (SLAM), which addresses the problem of constructing a spatial map of an unknown environment while simultaneously determining the mobile robot's position relative to this map, is regarded as one of the key technologies in mobile robot navigation (Stachniss, Leonard, & Thrun, 2016). The robot, therefore, needs sensors to perceive the environment for the purpose of navigation without external positional cues such as GPS, and integrate the sensory information into a coherent metric or topological map representation of the environment (Nüchter, Lingemann, Hertzberg, & Surmann, 2007; Tully, Kantor, & Choset, 2012). However, available sensors are usually noisy and unreliable. Thus, multiple different types of sensors are required to gain better integrated information for navigation (De Almeida, Araújo, Dias, & Nunes, 1995). How to integrate multiple sensory cues to achieve accurate navigation is one of the key challenges in robotics research.

In order to solve the SLAM problem effectively and efficiently, researchers investigate the mechanisms adopted by animals in navigation (Llofriu et al., 2015; Rolls & Stringer, 2005; Strösslin, Sheynikhovich, Chavarriaga, & Gerstner, 2005), and seek inspirations to achieve better algorithms for robot navigation (Barrera & Weitzenfeld, 2008; Cuperlier, Quoy, & Gaussier, 2007; Milford & Wyeth, 2008; Milford, Wyeth, & Prasser, 2004; Mulas, Waniek, & Conradt, 2016; Sünderhauf & Protzel, 2010a; Tang, Yan, & Tan, 2017; Zeng & Si, 2017). For a review, see Madl, Chen, Montaldi,

[☆] The authors would like to thank the support from the National Key Research and Development Program of China (NO. 2016YFC0801808); Natural Science Foundation of China (NO. 51679213); CAS Pioneer Hundred Talents Program, China (NO. Y8F1160101); and the State Key Laboratory of Robotics, China (NO. Y7C120E101).

[*] Corresponding author.
E-mail addresses: zengtaiping@fudan.edu.cn (T. Zeng), tangfengzhen@sia.cn (F. Tang), jidaxiong@zju.edu.cn (D. Ji), bailusi@bnu.edu.cn (B. Si).

https://doi.org/10.1016/j.neunet.2020.02.023
0893-6080/© 2020 Elsevier Ltd. All rights reserved.

and Trappl (2015). As a matter of fact, many animals show overwhelming navigation capabilities as compared to robots. They are capable of traveling long distances and navigating in complex environments searching for food, and then returning to their nests often in shortcuts (Mandal, 2018). Studies have shown that several types of spatially responsive neurons in the neural system are involved in animal navigation, namely place cells in the hippocampus (O'keefe & Conway, 1978; O'Keefe & Dostrovsky, 1971), head direction (HD) cells (Taube, Muller, & Ranck, 1990) and grid cells in the medial entorhinal cortex (MEC), presubiculum and parasubiculum (Boccara et al., 2010; Hafting, Fyhn, Molden, Moser, & Moser, 2005). HD cells fire strongly only when the animal's head points to specific directions. The activity of HD cells constitutes an internal allocentric representation of the head directions of the animal. HD cells are considered to provide compass information for animals. In two dimensional open field, place cells become active whenever the animal enters specific locations in the environment. The sparse neural codes of place cells readily enable them to function as cognitive map of the environment. Grid cells fire at regularly spaced locations which collectively define hexagonal lattices. The grid codes of different grid cells turn out to be scaled, rotated and translated relative to each other, and could provide a metric system for navigation. Moreover, two primary mechanisms have identified in animal navigation, i.e. path integration and landmark calibration (Etienne, Maurer, & Séguinot, 1996; Milford, Wiles, & Wyeth, 2010). By path integration, animals update their internal spatial representations using self-motion cues such as vestibular inputs. Through the mechanism of landmark calibration, the internal representations of self location are corrected by detecting familiar landmarks in the environment. In brief, place cells, HD cells and grid cells are the key neural basis for cognitive map during path-integration based and landmark based navigation (Grieves & Jeffery, 2017). Place cells of the mammalian brain provide flexible graph-based map representations, while HD cells and grid cells deliver metric encoding of orientation and position respectively (Moser, Kropff, & Moser, 2008; Moser & Moser, 2008). These cells are neural substrates for the integration of multisensory cues during navigation.

However, in large natural environments, both path integration and landmark calibration are faced with uncertainty: the path integration process is subject to accumulation error, while landmark calibration is undermined by perceptual ambiguity. How these spatial selective cells deal with uncertainty of sensory information when navigating in large scale environments is availing for robot navigation.

Both vestibular and visual inputs carry information about the animal's heading directions and spatial locations in the environment. Thus, neural navigation system integrates both vestibular and visual inputs to determine the animal's head directions or positions (Chen, DeAngelis, & Angelaki, 2013; Gu, Angelaki, & DeAngelis, 2008). This kind of multisensory integration is also found in many other sensory modalities. Visual and proprioceptive cues are combined for the perception of hand position. Motion and texture cues are integrated for sensing depth information. Visual and auditory cues are combined to determine object locations (Chandrasekaran, 2017).

Neurophysiological experiments and theoretical modeling have revealed that neural systems primarily combine sensory information from self-motion cues and visual cues to estimate the animal's heading in a near Bayesian optimal manner (Zhang & Wu, 2013). Extensive studies have been conducted to understand this computation mechanism of sensory integration in animal spatial navigation.

Bayesian multisensory integration could be achieved by reciprocally connected attractor networks, each maintaining an estimate of head directions according to an independent cue, either the visual or the vestibular cue (Zhang & Wu, 2013). The disparity between cues could be represented by attractor networks connected by opposite isomapping, so that multisensory segregation of sensory cues is realized concurrently in a Bayes-optimal manner (Zhang, Wang, Wong, & Wu, 2016), enabling animals to integrate multiple cues and simultaneously sense the difference between cues. Both the saliency of visual cues and the synaptic plasticity between networks affect multisensory integration. Plastic remapping of visual cues on the HD cells layer shifts the preferred directions of HD cells due to multiple reliable experiences (Knight et al., 2012). However weaker visual cues fail to remap the preferred directions of HD cells (Knight et al., 2012). Attractor network model has been proposed to account for the interaction between landmark and vestibular cues (Page et al., 2013). Here, we hypothesis that the same Bayesian inference mechanism of HD cells maybe also exist in grid cells to represent spatial locations in a torus or a twisted torus attractor network.

Although biological plausible attractor network models have been developed to demonstrate probabilistic computational mechanisms in the brain, as far as we know, there exist only a few neurobiologically plausible Bayesian models to implement multisensory information integration in SLAM system.

In this paper, we present a novel neurobiologically inspired Bayesian attractor network model to demonstrate the potential that could be brought to robot navigation systems by emulating the computational mechanisms of animal navigation. We employ probabilistic methods to model neural population coding. The activity of HD cells in a ring attractor network is represented by a one-dimensional Gaussian distribution with periodic boundary conditions (Ben-Yishai, Bar-Or, & Sompolinsky, 1995; Zhang, 1996). The response of grid cells in a torus attractor network with a single activity peak is modeled by a two-dimensional Gaussian activity packet with periodic boundary conditions (Guanella, Kiper, & Verschure, 2007). Integrator cells are introduced to integrate either angular velocity or translational velocity. Calibration cells are incorporated to integrate visual inputs. Conflict between cues is resolved by competitive dynamics of the two populations. We implemented our model in a SLAM system and demonstrated its performance on an open-source dataset of 66 km car journey in a 3 km x 1.6 km urban area (St Lucia 2007 dataset) and on an iRat miniature robot platform in a small-sized maze (iRat 2011 Australia dataset).

The contribution of this paper is threefold. First, a novel neurobiological Bayesian attractor network model is proposed for multisensory integration inspired by neural computation mechanisms of head direction cells. The model reproduces similar cue conflict resolution behavior as that of head direction cells in multi-cue integration tasks. Second, the model adopts parametric representations of network population activities, in the form of Gaussian distributions, and only needs to update a few variables of the network activities, resulting in constant computational complexity. The constant computational complexity of the model renders its advantage in the application in energy-critical situations such as unmanned aerial vehicles. Third, the multisensory integration mechanism is incorporated into a monocular SLAM system for autonomous robot navigation in large scale environments. By demonstrating successfully on real world datasets, the proposed Bayesian multisensory integration method is robust for robot navigation in environments of various sizes, different sensory noise and uncertainty, as ubiquitously faced during the exploration of large natural environments. In summary, the proposed neurobiological Bayesian attractor network model takes advantage of the computational mechanism of spatial memory neural circuits, and sheds light on further developing novel trustable and interpretable neural network models for robot navigation tasks.

The rest of this paper is organized as follows. In Section 2, we briefly review the related work of visual SLAM systems and neurobiologically based robot navigation systems. We derive a Bayesian theoretical frame for HD cells and grid cells models in Section 3. Sections 4 and 5 present the Bayesian attractor network and the SLAM system. The detailed results are reported in Section 6. Section 7 discusses the results and future works, with a brief conclusion in Section 8.

2. Related work

Our work has roots in visual SLAM and neurorobotic navigation systems. In this section, we briefly review the state-of-the-art techniques in these two research areas.

2.1. Visual SLAM

As mentioned earlier, SLAM is a method to simultaneously estimate the pose of a robot and construct the map of an unknown environment while the robot is perceiving the environment with on-board sensors. Over the past 30 years, the SLAM research community has witnessed spectacular progress, enabling the applications of this technology in large scale natural environments and also a transition of it to industry. At an early age of SLAM (1986–2004), probabilistic models were introduced to solve the SLAM problem, including extended Kalman filters (e.g. MonoSLAM), Rao–Blackwellized particle filters (e.g. FastSLAM), and maximum likelihood estimation (Thrun, Burgard, Fox, & Arkin, 2005). In the following stage (2004–2015), major efforts were devoted to understanding the fundamental properties of SLAM (i.e. observability, convergence, and consistency), proposing efficient SLAM solvers, and developing open-source SLAM libraries. The SLAM research now enters into a robust perception age, with visual SLAM being a representative. During this age, attempts are focused on robust performance, high-level understanding, resource awareness, and task-driven perception (Cadena et al., 2016).

Visual SLAM, as suggested by its name, takes advantage of image as a primary source of information for localization and mapping. This type of SLAM methods can be divided into two classes: direct and indirect methods, according to the way how the image information is utilized. Indirect visual SLAM systems first extract features from the images captured by a camera, and then use the features to infer the pose of the camera and subsequently build a map. Vision was brought into SLAM community by A. J. Davison's seminal work called MonoSLAM (Davison, Reid, Molton, & Stasse, 2007), which extracts image features to represent landmarks within the Extend Kalman Filter (EKF) framework. This work becomes a standard framework for the implementation of visual SLAM systems. However, the computation load of EKF based visual SLAM increases substantially as the size of the map grows, limiting its applicability in large scale environments. Another feature-based SLAM algorithm called Parallel Tracking and Mapping (PTAM) is proposed to parallelize the motion estimation and mapping tasks (Klein & Murray, 2007). This algorithm relies on performing keyframe bundle adjustment instead of filtering, leading to higher computational efficiency. Although PTAM achieves satisfactory real time performance in small environments, its scalability does not allow direct application in large scale mapping. ORB-SLAM and ORB-SLAM2 employ improved image feature detectors and descriptors over those used in PTAM, leading to better performance than that of PTAM (Mur-Artal, Montiel, & Tardos, 2015; Mur-Artal & Tardos, 2016). These two methods parallelize tracking, mapping and loop closing to achieve impressively consistent localization and mapping.

Instead of extracting features, direct SLAM systems, on the contrary, directly perform matching in the raw image space. Dense tracking and mapping (DTAM) is the first direct visual SLAM being able to run in real time with a GPU (Newcombe, Lovegrove, & Davison, 2011). Large-Scale Direct Monocular SLAM (LSD-SLAM) employs direct tracking by image-to-image alignment to build a semi-dense map by depth estimation at pixels solely near image boundaries. This algorithm is also able to run in real time on a CPU (Engel, Schöps, & Cremers, 2014). Recently, deep recurrent convolutional neural networks are trained to infer poses and uncertainties from a sequence of raw images by automatically learning effective feature representations (Wang, Clark, Wen, & Trigoni, 2018).

2.2. Neurobiologically inspired robot navigation

Animals, especially mammals, such as bats and whales, show amazing navigation ability (Horton et al., 2011; Tsoar et al., 2011). They are able to orient in large scale, complex, and dynamic environments for a very long time. A cognitive map of the environment, i.e. an internal map-like representation in the brain, has long been proposed to support their superior navigation abilities (Tolman, 1948). Place cells and HD cells are sufficient to represent the pose of the animal including head directions and locations. Grid cells form multi-resolution representations of the position of the animal, and could provide better spatial information than place cells do (Mathis, Herz, & Stemmler, 2012).

The mechanisms of animal navigation have long motivated robot navigation. It is especially rewarding to bridge the gap between neuroscience and robotics research. Many works have been carried out to identify the critical functional components of robot navigation systems and their corresponding structures in neural systems. Research along this line has enriched substantially our understandings of the computational principles of animal navigation systems and has led to design better robot navigation algorithms for natural environments. Dated back to 1997, a model of hippocampal place has been proposed for a robot to navigate in open-field environments of various shapes by Neil Burgess and John O'Keefe (Burgess, Donnett, Jeffery, & John, 1997). A navigation model inspired by HD cells and place cells was demonstrated on Khepera robot in a small arena by Gerstner's group (Arleo & Gerstner, 2000). This work is followed by Strösslin et al. (2005), incorporating unsupervised Hebbian learning to form a cognitive map of the environment. A complex and modular computational model including navigation related neurobiological entities is proposed to learn or unlearn goal locations through changing rewards by Barrera and Weitzenfeld (2008). In Cuperlier et al. (2007), place cells are coupled to create abstract transition cells, which explicitly encode spatiotemporal transitions experienced by the robot. Its follow-up work developed a computational model composed of multisensory place cells by merging the activities of visual place cells and grid cells (Jauffret, Cuperlier, & Gaussier, 2015). This model has been demonstrated to deliver effective navigation ability in a real robot platform. RatSLAM has been developed for long term robot navigation tasks in large scale environments by employing abstract pose cells to represent the conjunction of HD and positions (Milford & Wyeth, 2008; Milford et al., 2004). This work is upgraded to OpenRatSLAM, a package of open-source SLAM libraries and publicly available datasets, making great contributions in pushing forward the development of SLAM (Ball et al., 2013). Related to our work, Sünderhauf and colleagues linked the neural network model of pose cells to Bayesian inference, and derived a novel filter scheme for the abstract pose cells in RatSLAM (Sünderhauf & Protzel, 2010a, 2010b).

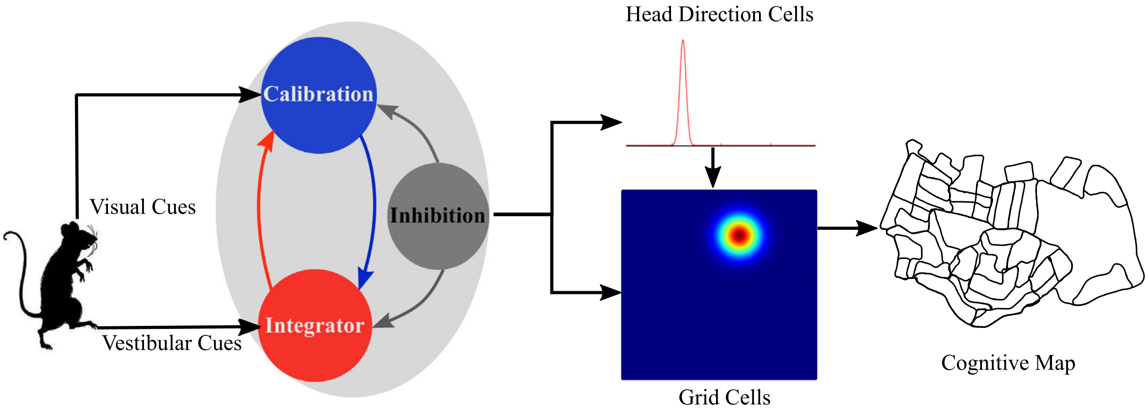

Fig. 1. Bayesian attractor network architecture of the model and information flow char. HD cells and grid cells are organized by the same computational mechanisms. Information from vestibular cues and visual cues are integrated by the populations of integrator cells and calibration cells respectively. Cue conflict is resolved by the two populations by mutual and global inhibition, resulting in single-peaked activities in the networks. The directions encoded in the HD cell network are used by the grid cell network to encode the locations of the agent. A cognitive map is built based on the location information provided by the grid cell network.

3. The Bayesian cue integration framework

Accumulating neurobiological evidence has shown that brain performs Bayesian inference in many cognitive tasks (Seilheimer, Rosenberg, & Angelaki, 2014). In this paper, we propose a cue integration framework based on Bayesian inference for robot navigation task. The probability distributions on head directions or locations are inferred by combining multiple cues together, namely vestibular cue c_{ve} and visual cue c_{vi}, so that the uncertainty in motion and perception is taken into account for robust representation of the pose.

Given the heading direction or location denoted by θ, the likelihoods of perceiving visual or vestibular cues are described by probability distributions $p(c_{\mathrm{vi}}|\theta)$ and $p(c_{\mathrm{ve}}|\theta)$ respectively. Since the noise of visual and vestibular cues are mutually independent, according to Bayes' theorem, the posterior distribution of θ with the presence of two cues, denoted by $p(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}})$, can be expressed as

$$p(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}}) \propto p(c_{\mathrm{vi}}|\theta)p(c_{\mathrm{ve}}|\theta)p(\theta), \tag{1}$$

where $p(\theta)$ is the prior probability. If there is no prior knowledge, $p(\theta)$ is uniform (Zhang & Wu, 2013), then (1) can be rewritten as

$$p(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}}) \propto p(c_{\mathrm{vi}}|\theta)p(c_{\mathrm{ve}}|\theta) \tag{2}$$

Taking the time into consideration, we can merge the past experience and the current evidence in an iterative manner:

$$p^t(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}}) \propto p^t(c_{\mathrm{vi}}|\theta)p^t(c_{\mathrm{ve}}|\theta)p^{t-1}(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}}), \tag{3}$$

where $p^{t-1}(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}})$ is the posterior distribution of θ up to time $(t-1)$ (i.e. past experience), $p^t(c_{\mathrm{ve}}|\theta)$ and $p^t(c_{\mathrm{vi}}|\theta)$ are the likelihoods of perceiving c_{ve} and c_{vi} (i.e. current evidence).

Since self-localization is decoupled into two separated processes, namely path integration and landmark calibration, Eq. (3) is spitted into two updates to integrate currently available cues accordingly:

$$p^t(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}}) \propto p^t(c_{\mathrm{ve}}|\theta)p^{t-1}(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}}), \tag{4}$$

$$p^t(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}}) \propto p^t(c_{\mathrm{vi}}|\theta)p^{t-1}(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}}). \tag{5}$$

Eq. (4) is to mimic path integration that updates the estimate of θ using vestibular cues, while Eq. (5) is to perform landmark calibration that updates the estimate of θ using visual cues. The two updates on θ do not have to occur simultaneously, nor do they have to follow a fixed order. In other words, $p^{t-1}(\theta|c_{\mathrm{vi}}, c_{\mathrm{ve}})$ in Eq. (4) may be the updated result of Eq. (5), and vice versa. The superscript $t-1$ here represents pre-update while t indicates post-update. This notation is also applicable to the rest of this paper.

4. The neurobiological Bayesian attractor model

Attractor dynamics is found to be one of the key computational mechanisms of spatial memory systems in the brain, such as HD cells, grid cells and place cells (Knierim & Zhang, 2012). Due to the recurrent connections, attractor neural networks require heavy computation when the network size is large. In order to reduce computational cost of attractor neural network, we propose an abstract and concise model, discarding the details in the connectivity between cells and in the currents into cells, yet including both attractor dynamics and Bayesian cue integration as essential computational mechanisms. The state of attractor network is simplistically represented by Gaussian distribution, while keeping the encoding properties of HD cells and grid cells. The dynamics of attractor network is modeled by the update of the mean and variance of Gaussian distribution.

4.1. Overview of the proposed model

As shown in Fig. 1, the proposed Bayesian attractor model is composed of a network of HD cells and a network of grid cells. Each network consists of a population of integrator cells and a population of calibration cells, which receive vestibular cues from MSTd area and visual cues from VIP area, respectively. In the HD network, each population maintains a one-dimensional Gaussian distribution with the mean being a representation of the agent's head direction, and the variance describing the reliability of the corresponding cues. The two populations are interconnected through mutual inhibition and global inhibition. The visual cues and vestibular cues are then integrated through Bayesian inference. Conflict between cues is solved by the competitive dynamics between integrator cells and calibration cells. The population of integrator cells integrates the current evidence from vestibular cues and the past experience from both vestibular cues and visual cues. Under the periodic boundary conditions of the HD neural space, the population maintains a bump or packet of activity, similar to the activity state of a ring attractor. The agent's real angle of HD in the physical environment is represented by the center of the activity bump. The population of integrator cells holds and updates memories of head directions, mimicking the ring attractor network of HD cells.

A similar framework of Bayesian attractor network is proposed to simulate the coding scheme of grids cells. However, instead of one-dimensional Gaussian distributions, the two populations, namely integrator cells and calibration cells, manipulate two-dimensional Gaussian distributions on a neural space with periodic boundary conditions. The two-dimensional Gaussian distributions maintained by the grid cells in the network represent single-peaked activity bumps in torus attractor networks. Different from previous works in which the neural space is twisted to produce triangular firing patterns, our model does not focus on reproducing neural responses exactly as those observed in neurobiological experiments, but explores the potential to solve real world spatial cognition problem based on neural encoding mechanisms of spatial memory systems. Due to the periodic boundary conditions, the bump center of the torus attractor represents the real location of the robot in the physical environment up to a modulo operation.

In summary, HD cells and grid cells with single-peaked activities jointly represent the pose of the robot in the environment and further a cognitive map of the physical environment could be constructed.

4.2. Model of head direction cells

In this section, we present the Bayesian attractor model of head direction cells in detail. For computational efficiency, the activities of integrator cells and calibration cells are modeled by Gaussian distributions, which represent the beliefs in head directions. Due to the difference in the incoming cues to the integrator cells and the calibration cells, their beliefs may not be the same. Integrator cells and calibration cells of the HD network separately maintain their beliefs in head directions. The belief of each population has the following form

$$p(\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-|\theta-\mu|^2/2\sigma^2}, \tag{6}$$

where $\theta \in [0, 2\pi)$ is the label of a cell. θ lies in a one-dimensional neural space. $|\cdot|$ takes the difference of two angles on a circle. The mean μ is the maximum likelihood estimation of the head direction given by the population, and the variance σ^2 is the uncertainty of the belief. With this parametric representation, it is sufficient to update the four variables for the two populations in the HD cell network, i.e. the mean and the inverse uncertainty for each of the population in the HD network (ref Table 2).

4.2.1. Attractor dynamics

To resolve cue conflicts, the two populations interact by global inhibition and mutual inhibition, endowing the whole network with the behavior of attractor dynamics.

The global inhibition is formulated as follows:

$$\frac{1}{(\sigma^t_{\text{inte}})^2} = \frac{E}{W}\frac{1}{(\sigma^{t-1}_{\text{inte}})^2}, \tag{7}$$

$$\frac{1}{(\sigma^t_{\text{cali}})^2} = \frac{E}{W}\frac{1}{(\sigma^{t-1}_{\text{cali}})^2}, \tag{8}$$

where $(\sigma^{t-1}_{\text{inte}})^2$ and $(\sigma^{t-1}_{\text{cali}})^2$ are the uncertainty of the belief of the integrator cells and the calibration cells before update, respectively. $1/(\sigma^{t-1}_{\text{inte}})^2$ and $1/(\sigma^{t-1}_{\text{cali}})^2$ are the fisher information with respect to head direction. W is the total fisher information and works as a normalization factor:

$$W = \frac{1}{(\sigma^{t-1}_{\text{inte}})^2} + \frac{1}{(\sigma^{t-1}_{\text{cali}})^2}. \tag{9}$$

E is a predefined constant, which is the total fisher information with respect to head direction contained in the two population. Since the fisher information with respect to head direction measures the extent to which the belief is peaked, we term it as head direction *reliability*.

The mutual inhibition is to guarantee that there is only one stable bump with single peak existing in the network over time. The mutual inhibition between the two population is defined by

$$\frac{1}{(\sigma^t_{\text{inte}})^2} = \frac{1}{(\sigma^{t-1}_{\text{inte}})^2} - \varDelta_{\text{inte}}\frac{1}{(\sigma^{t-1}_{\text{cali}})^2}, \tag{10}$$

$$\frac{1}{(\sigma^t_{\text{cali}})^2} = \frac{1}{(\sigma^{t-1}_{\text{cali}})^2} - \varDelta_{\text{cali}}\frac{1}{(\sigma^{t-1}_{\text{inte}})^2}, \tag{11}$$

where $\varDelta_{\text{inte}}$ and $\varDelta_{\text{cali}}$ are the strength of inhibition to the integrator cells and the calibration cells respectively.

Eqs. (7)–(11) implement competition between the reliability of two populations. Under mutual and global inhibition, the network converges to a prominent single-peaked activity bump and a weak single-peaked activity bump. This competition behavior resembles the dynamics of an attractor network. In order to let the weak population recover at the time when reliable cues are available, we set a lower bound value U for the reliabilities of the two populations.

Depending on the relative magnitudes of reliability, the time duration of competition varies. The competition allows the network to temporally accommodate different beliefs, and reach coherent representations by integrating more information. On top of the attractor dynamics, the two populations undergo vestibular cue integration and visual cue calibration, which are explained in the following sections.

4.2.2. Vestibular cues integration

In this paper, path integration is not performed by network mechanisms such as proposed in Si, Romani, and Tsodyks (2014) and Zeng and Si (2017). Instead, activity bumps are updated by directly shifting the means of the Gaussian distributions given the information from vestibular inputs. Path integration denoted by Eq. (4) is implemented as follows:

$$\begin{aligned} \mu^t_{\text{inte}} &= \mod(\mu^{t-1}_{\text{inte}} + v^t\varDelta t, 2\pi) \\ \mu^t_{\text{cali}} &= \mod(\mu^{t-1}_{\text{cali}} + v^t\varDelta t, 2\pi), \end{aligned} \tag{12}$$

where μ^t_{inte} and μ^t_{cali} are the mean encoded in the integrator cells and the calibration cells. v^t is the velocity from vestibular or movement system. $\varDelta t$ is the time interval between time steps t and $t-1$. mod is the modulus operation in the domain of real number, it returns the remainder and maps the new phase into the range $[0, 2\pi)$.

Note that the uncertainty during path integration in this work does not increase, which is different from classical probabilistic SLAM.

4.2.3. Visual cues calibration

The activity of head direction cells is calibrated by familiar visual cues. If the current view is observed previously, it excites calibration cells according to Eq. (5):

$$p^t_{\text{cali}}(\theta) \propto p^t_{\text{inject}}(\theta)\, p^{t-1}_{\text{cali}}(\theta), \tag{13}$$

where $p^t_{\text{inject}}(\theta)$ is the likelihood function, and describes the location and the spread of the current injected from visual cues to the network. Since the two distributions on the right hand side of Eq. (13) are Gaussian by assumption, it can be implemented by

$$\frac{1}{(\sigma^t_{\text{cali}})^2} = \frac{1}{(\sigma^{t-1}_{\text{cali}})^2} + \frac{1}{(\sigma^t_{\text{inject}})^2} \tag{14}$$

$$\mu^t_{\text{cali}} = \mod\left(\frac{(\sigma^t_{\text{cali}})^2}{(\sigma^{t-1}_{\text{cali}})^2}\mu^{t-1}_{\text{cali}} + \frac{(\sigma^t_{\text{cali}})^2}{(\sigma^t_{\text{inject}})^2}\mu^t_{\text{inject}}, 2\pi\right), \tag{15}$$

where $1/(\sigma^t_{\text{inject}})^2$ is the reliability of the visual cue, μ^t_{inject} is the location where the current is injected to the one dimensional neural manifold of HD cells.

The combined belief of the head direction is then given by merging the beliefs of both integrator and calibration cells

$$p^t_{\text{f}}(\theta) \propto p^t_{\text{cali}}(\theta)\; p^t_{\text{inte}}(\theta). \tag{16}$$

Using the Gaussian assumptions, Eq. (16) is given by

$$\frac{1}{(\sigma^t_{\text{f}})^2} = \frac{1}{(\sigma^t_{\text{inte}})^2} + \frac{1}{(\sigma^t_{\text{cali}})^2} \tag{17}$$

$$\mu^t_{\text{f}} = \quad \text{mod}\left(\frac{(\sigma^t_{\text{f}})^2}{(\sigma^t_{\text{inte}})^2}\mu^t_{\text{inte}} + \frac{(\sigma^t_{\text{f}})^2}{(\sigma^t_{\text{cali}})^2}\mu^t_{\text{cali}},\, 2\pi\right), \tag{18}$$

where $1/(\sigma^t_{\text{f}})^2$ is the reliability and μ^t_{f} the center of the belief.

If the maximum likelihood estimation of the head direction μ^t_{f} is close to the location of the current injection μ^t_{cali}, i.e. $|\mu^t_{\text{f}} - \mu^t_{\text{cali}}| < \delta$, where δ is a threshold, the visual cue has strong calibration effect on the belief, and the decision that the agent revisits a familiar location is made. This means a loop closure is detected. In this case, the belief $p^t_{\text{inte}}(\theta)$ of integrator cells is reset by assigning the combined belief $p^t_{\text{f}}(\theta)$ to $p^t_{\text{inte}}(\theta)$. If, however, the maximum likelihood estimation of the head direction μ^t_{f} is very different from the location of the current injection, the belief of the integrator cells is not reset by the combined belief. The integrator cells and calibration cells continue integrating cues and inhibiting each other. The reset of the belief $p^t_{\text{inte}}(\theta)$ by $p^t_{\text{f}}(\theta)$ imitates the remapping of visual information onto the HD cells (Knight et al., 2012; Page et al., 2013).

4.3. Model of grid cells

In this section, we extend the above HD cell network to model grid cells. The same mechanism of HD cell network is used to represent 2D locations. Similar to the HD cell network, the grid cell network is composed of a population of integrator cells and a population of calibration cells. Neuroscience recordings have revealed that grid cells are distributed in multiple brain regions, such as MEC, pre- and parasubiculum (Boccara et al., 2010; Sargolini et al., 2006). Grid cells in these areas may function differentially as calibration cells and integrator cells. Different from HD cell network model, the grid cells in the model form 2D torus attractors with single-peaked activity profiles. More specifically, the activity of the integrator cells and the calibration cells of the grid cell network is defined by two-dimensional Gaussian distributions

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-(|x-\mu_x|^2/2\sigma_x^2 + |y-\mu_y|^2/2\sigma_y^2)}, \tag{19}$$

where $x, y \in [0, 2\pi)$ are the coordinates of the cells in the respective 2D neural manifold. (μ_x, μ_y) is the spatial phase encoded by the cells. Due to the periodic boundary conditions of the neural manifold, the spatial phase encodes the actual location of the robot up to a modulo operation. The periodic torus structure of the neural manifold provides one mechanism to map large physical space into periodic neural representations. In this study, we only use for each of the population of integrator cells and the population of calibration cells one neural manifold. In this simplest setup, the actual location of the robot can be recovered by unwrapping the spatial phase considering the periodic boundary conditions (e.g. through counting the jumps at the boundaries in the whole history of the spatial phase). $1/\sigma_x^2$ and $1/\sigma_y^2$ are the reliabilities of spatial phase estimation in each dimension. Note that we assume that the two dimensions of the spatial representation in the model are independent, and therefore the cross-correlation between the two dimensions vanishes, as shown in Eq. (19). This assumption is reasonable under the conditions where sensory cues are available for integration. Taking together, the grid cell model boils down two four variables, spatial phase and reliability of the two populations, for each spatial dimension, resulting in eight variables (ref Table 2). Both the spatial phases and the corresponding reliabilities of the two populations are updated in the same way as those in the HD cell network. The translational velocity input to the grid cell network is obtained by projecting velocity signal to the head direction estimated by the HD cell network.

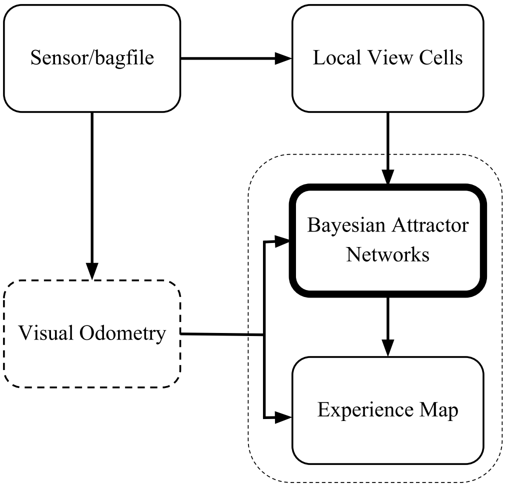

Fig. 2. The software architecture of the NeuroBayesSLAM system. The sensor/bagfile node provides images and odometry information. Visual odometry provides velocity estimation for pure visual datasets. The local view cell node determines whether the current view is familiar or not. Bayesian attractor network node performs path integration and makes decisions of loop closures. The experience map node generates the topological map.

5. Robotic implementation

To test the ability of the proposed Bayesian attractor model in solving realistic SLAM problems, we implemented the model in C++ language on Robot Operating System (ROS). Our SLAM system is termed as NeuroBayesSLAM. Adopting some fundamental modules of the publicly available OpenRatSLAM system (Ball et al., 2013), we organized the NeuroBayesSLAM system into five nodes, as shown in Fig. 2, namely Sensor/Bagfile, Local View Cells, Visual Odometry, Bayesian Attractor Networks, and Experience Map. Detailed description of functionality of each node is given as follows:

- The sensor, i.e. camera, captures the view that the robot can observe and sends the view to the local view cells and visual odometry node.
- The local view cells compare the view from the sensor with visual templates stored by the local view cells, and determine whether this view is familiar or not. If this view is new, a new local view cell is created and added to the system. The newly added local view cell is associated with the visual feature of the view as a new local view template, the head direction estimation of the head direction network and the spatial phase estimation of the grid cell network. The connections between the local view cell and the cells in the HD cell network and grid cell are strengthened via Hebb's rule (Hebb, 1949). Instead, if the view is one that the robot has encountered previously, the local view cell associated with the current view can be reactivated and injects energy into HD cells and grid cells through learned connections via Eq. (13).

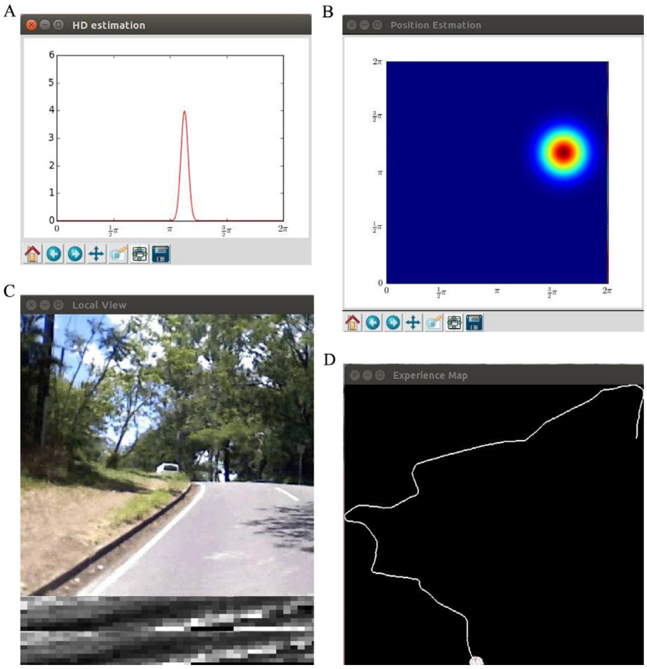

Fig. 3. Screenshots of the NeuroBayesSLAM system running on the St Lucia suburb dataset. (A) Neural activities of the integrator cells in the HD network; (B) Neural activities of the integrator cells in the grid cell network shown as a heat map. Activity from zero to maximum is color-coded from blue to red; (C) Input visual scene of a road with ascending slop (top), the local view template of the scene (middle), and the best matched template (bottom); (D) Experience map. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- The visual odometry node estimates angular speed and linear speed by comparing consecutive image arrays. If inertial measurement unit (IMU) exists to provide self-motion information for the robot, this node would not be enabled.
- The Bayesian attractor network node consists of the HD cell network and the grid cell network proposed in Sections 4.2 and 4.3, respectively. This node receives odometry and view templates as inputs in the form of ROS messages. The velocity information drives the activity bumps along ring manifolds in the HD cell network and torus manifolds in the grid cell network to encode the pose of the robot. The belief is also generated about whether creating a new experience or closing a loop in the experience map.
- The experience map node generates a coherent semi-metric topological map that includes many vertices connected by edges. Every vertex is characterized by a local view cell, a head direction phase coded by HD cells, and a location phase coded by grid cells. Since locations are represented by the phases on torus, infinite area can be mapped by grid cells with periodic codes. When the distance between the current phase on the torus and the phase of the previous vertex is long enough to meet a threshold, the Bayesian attractor network would send a ROS message to the experience map node to create a new vertex and a new edge connecting it to the previous vertex. When a loop closure occurs, a new edge would be created to connect to an existing vertex. Then, a graph relaxation algorithm (Duckett, Marsland, & Shapiro, 2002) is used to distribute odometry error throughout the topological map, and a map readily readable to a human can be provided to present the absolute location in the physical environment.

Table 1
Values of parameters.

Parameter	HD cells	Grid cells
E	100	1
$\frac{1}{\sigma^2_{\textbf{inject}}}$	40	0.4
Δ_{cali}	0.05	0.05
Δ_{inte}	0.005	0.005
U	0.001	0.001

Table 2
Initial values of the variables for integrator cells and calibration cells. The variables of the grid cell populations are initialized as the same value for each spatial dimension. For the ease of display, only the variables for one spatial dimension are shown and the subscripts for spatial dimensions are omitted.

Variable	HD cells	Grid cells
$\frac{1}{\sigma^2_{\textbf{inte}}}$	100	1
μ_{inte}	0	0
$\frac{1}{\sigma^2_{\textbf{cali}}}$	10	0.1
μ_{cali}	0	0

Python scripts are written to visualize the live state of the NeuroBayesSLAM system. The neural activity of HD cells and grid cells, the visual inputs and the local view templates, and the experience map are shown during running.

6. Results

We run the NeuroBayesSLAM system on two publicly available datasets, namely the St Lucia suburb dataset and the iRat Australia dataset (Ball et al., 2013). For each dataset, we present firing rate maps of example HD cells and grid cells and the cognitive maps build by the system. For the St Lucia dataset, we demonstrate cue conflict resolutions between visual cues and vestibular cues when loop closures occur.

6.1. St Lucia suburb dataset

The St Lucia dataset was recorded in the suburb area of St Lucia in Brisbane, Australia using a web camera mounted on a vehicle (Ball et al., 2013; Milford & Wyeth, 2008). The trajectory is 66 km long through a wide range of terrain types, and spans an area of 3 km by 1.6 km in east–west and north–south directions, respectively.

Fig. 3 shows the interface of the NeuroBayesSLAM system running on the St Lucia dataset. The complete mapping process is given by the Supplementary video S1. Key parameters and initial values of the integrator and calibration cells used in the system are summarized in Tables 1 and 2, respectively. Following the findings in Butler, Smith, Campos, and Bülthoff (2010), more weight is given to the vestibular cues than the visual cues in cue integration. This is realized by assigning larger value to mutual inhibition intensity Δ_{cali}, as compared to Δ_{inte}.

6.1.1. Neural representations

Example activities of HD cells and grid cells in the model are shown in Fig. 4.

Panels A and C of Fig. 4 give the neural activities of the integrator cells in the HD network and the grid cell network respectively, at the beginning of the experiment, when the robot was at the origin. The localized activity bumps represent the belief of the robot's pose. The head direction of the robot in the physical environment is encoded by the center of the bump in Fig. 4A. The width of the bump in Fig. 4A gives the uncertainty of the belief. In a similar way, the center of the bump in Fig. 4C

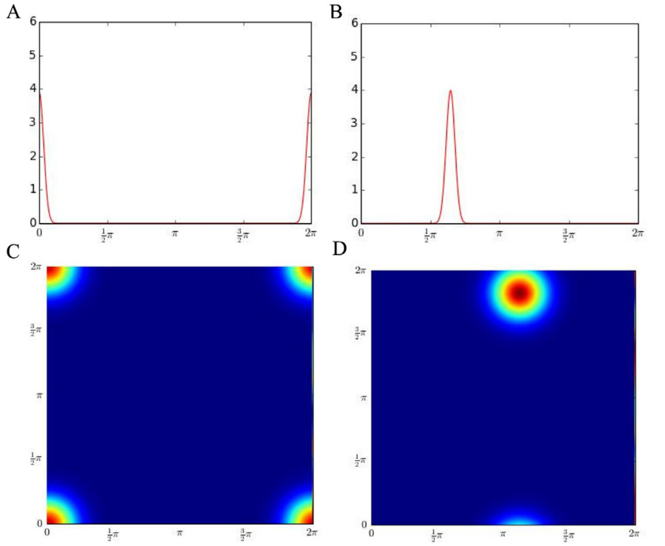

Fig. 4. The activities of the integrator cells of the HD cell network and grid cell network in the NeuroBayesSLAM system. (A,C) In the beginning of the experiment, the neural activities of the integrator cells of the HD network and grid cell network are centered at the origins. (B,D) During navigation, the activity bumps are updated to encode the pose of the robot. At this moment, the activity bump of the integrator cells of the HD network is centered at 2.02, about 115.64 degrees in (B). The integrator cells of the grid cell network have peak activity at (3.56, 5.68) in (D).

encodes the location of the robot in the physical environment, and the width of the bump reflects the reliability of the belief.

Fig. 4B and D show the neural activities of the integrator cells in the HD cell network and grid cell network when the robot moved to a different location. As can be seen in Fig. 4B, the activity bump of HD cells shifted its center to 2.02, indicating the current HD of the robot in the physical environment is at 115.64 degrees with respect to the initial head direction. Similarly, in Fig. 4D, the activity bump of grid cells changed its center to (3.56, 5.68). Due to the periodic boundary condition of the neural space, this grid cell representation can then be unwrapped iteratively by considering possible jumps at the boundaries of the neural space to obtain an estimation of the robot's location in the physical environment.

6.1.2. Cognitive map

The cognitive map constructed by the NeuroBayesSLAM system is depicted as a semi-metric topological map (Fig. 5B), generated by graph relaxation algorithm, taking into account the physical distance between related vertices through the activity bump of grid cells. The vertices of the topological graph are depicted by green circles (as the circles are densely located, they overlap and appear to be a thick green line). The fine blue line comprises the edges connecting related vertices. Due to the local metric information contained in the edges, the topological graph captures the structure of the road network of the environment, as can be seen by comparing with the ground truth map (Fig. 5A). All loops, intersections, and corners are correctly preserved in the graph. Due to the error in speed estimation, the path of the cognitive map is slightly geometrically distorted as compared with the ground truth map.

6.1.3. Cue conflict resolution

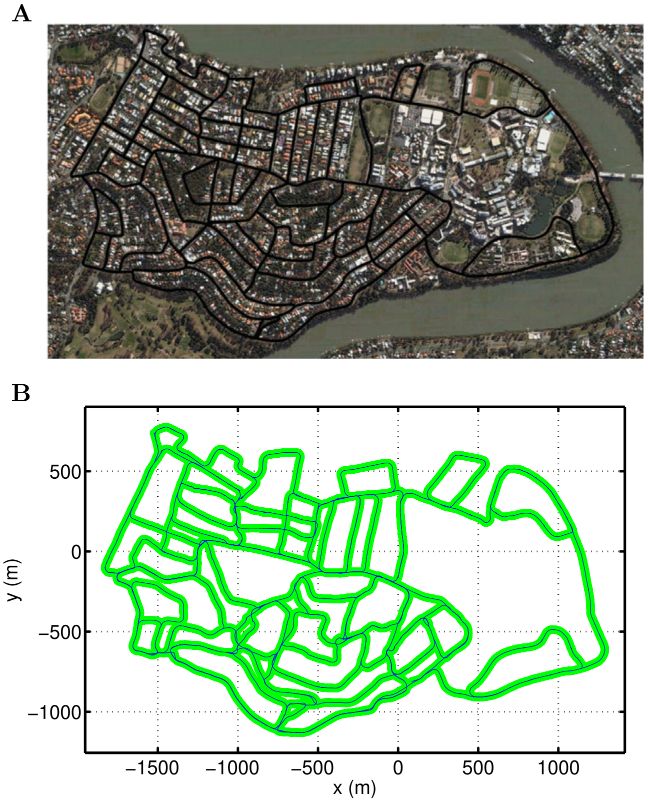

Fig. 5. The semi-metric topological map (B) constructed by the NeuroBayesSLAM system for St Lucia suburb (A). The green thick line comprises topological graph vertices, and the blue thin line consists of edges between related vertices in the graph. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Strong visual cues parameters.

Parameter	HD cells	Grid cells
E	100	1
$\frac{1}{\sigma^2_{\mathbf{inject}}}$	20	0.2
Δ_{cali}	0.01	0.01
Δ_{inte}	0.001	0.001

Table 4
Weak visual cues parameters.

Parameter	HD cells	Grid cells
E	100	1
$\frac{1}{\sigma^2_{\mathbf{inject}}}$	1.1	0.011
Δ_{cali}	0.01	0.01
Δ_{inte}	0.001	0.001

During navigation, the motion errors accumulate and lead to conflict between the vestibular cues and the visual cues when loop closure happens. Cue conflict resolution is key to the formation of robust spatial representations. Cue conflict here is solved by plastic remapping of visual information onto HD cells and grid cells. The activities of HD cell network and grid cell network are calibrated by a consecutive sequence of familiar landmarks from visual cues, over a period of several minutes. Thus, the visual landmarks gain cue control. The plastic remapping process causes a shift in preferred head directions and locations undershooting those corresponding to visual landmarks, which are then inherited by HD and grid cells during path integration (Knight et al., 2012; Page et al., 2013).

In order to demonstrate clearly the importance of visual cues in plastic remapping process, we varied the reliability parameter of the visual cues ($1/\sigma^2_{inject}$) in the simulation (Table 3 vs. Table 4).

At about 181 s, the robot was about to close a loop, located in the south-east part of Fig. 5. During loop closure, conflict

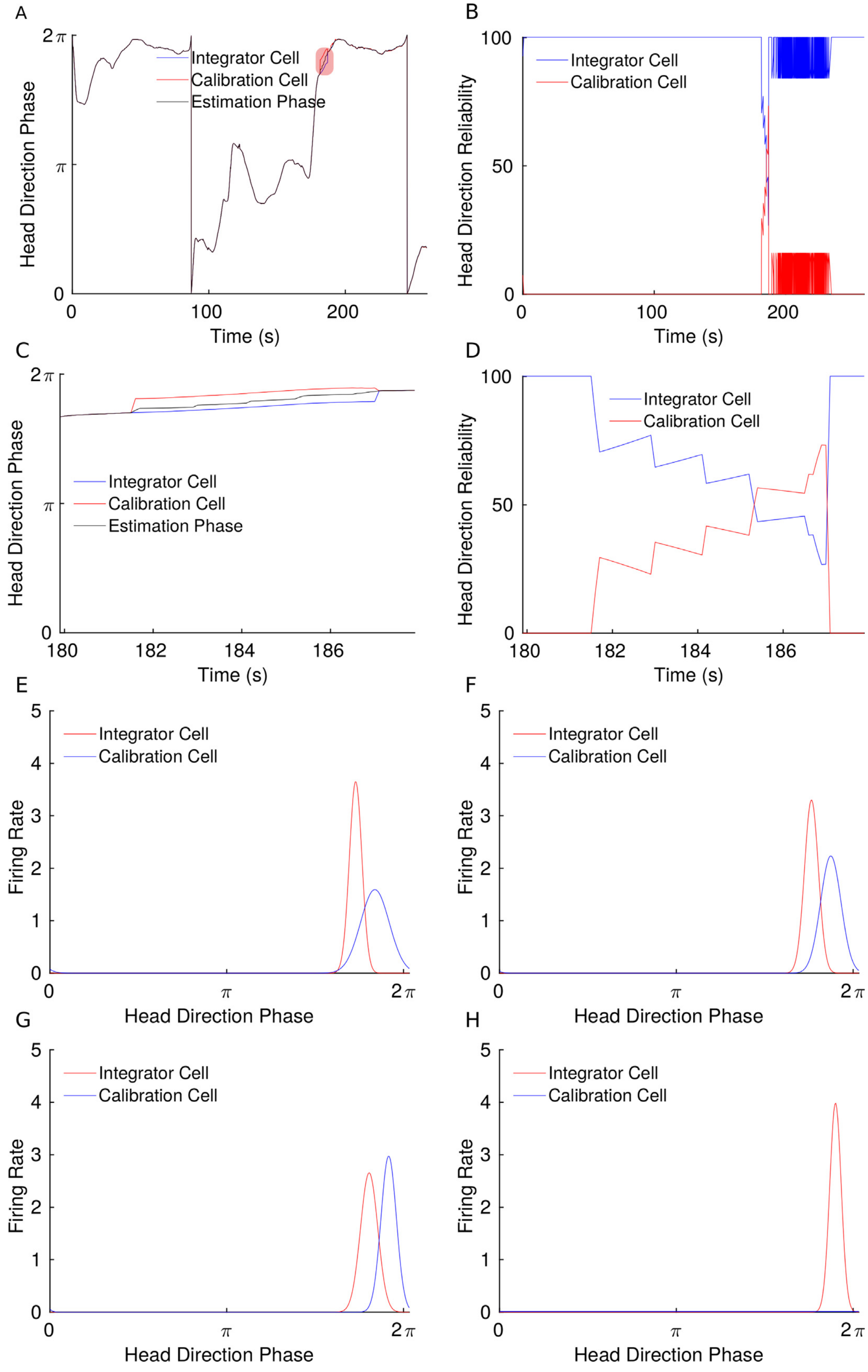

Fig. 6. An example of cue conflict resolution for HD cells. (A) shows the HD phase changing overtime; (B) shows reliability changing overtime. The read shadow time window in (A) is shown in (C) and (D). The HD cell neural activities of four different time stamps of (C) and (D) are shown in (E), (F), (G), and (H). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

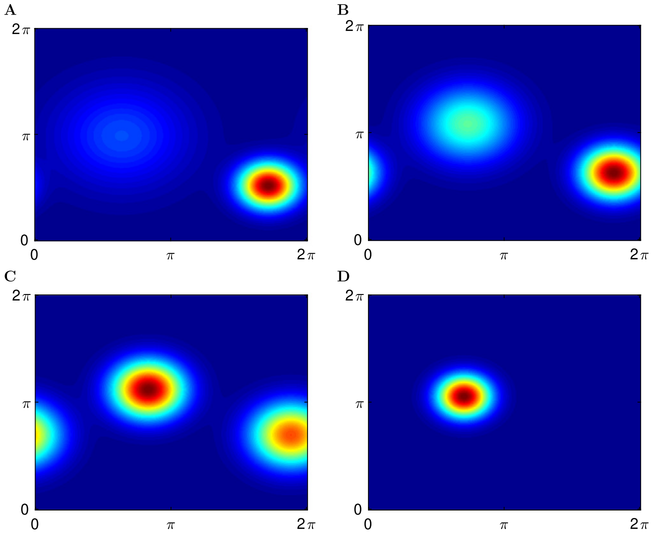

Fig. 7. An example of cue conflict resolution for grid cells. The grid cell neural activities of four different time stamps of Fig. 6 C and D are shown in (A), (B), (C), and (D).

between the integrator cells and calibration cells occurred, as shown by the difference between the phases encoded by the two populations (highlighted by the red shadow area in Fig. 6A). With strong visual cue (supplementary video S2), the integration of the visual landmark information into the calibration cells increased the reliability of calibration cells (Eq. (14)), as shown in Fig. 6B. A close-up of the red shadow area shows the evolution of the phases of the integrator cells and calibration cells. The final estimation of the phase was closer to that of the integrator cells in the beginning and was taken over by the phase of the calibration cells in the end (Fig. 6C), due to the increasing reliability of the calibration cells (Fig. 6D). During loop closure, the neural activities of the integrator and calibration cells competed and the activity of the calibration cells finally dominated (Fig. 6E–H). The visual landmark therefore gained cue control.

The same remapping process can be observed in grid cells, as shown in Fig. 7. In order to ease the description, the neural activities of the integrator and calibration cells at the same time step are summed to be presented in a single panel, given by Fig. 7A–D, the time steps of which exactly correspond to those of E–H in Fig. 6, respectively. In the beginning, the integrator cells, whose activity bump was located in the right-bottom of Fig. 7A, were dominant in the determination of activity bump center of grid cells. As the calibration cell received consecutively the energy injection from local view cells, the neural activity of the calibration cell rapidly increased due to the attractor dynamics (Fig. 7B and C). At last, the calibration cells completely overrode the integrator cells (Fig. 7D), meaning a successful remapping of visual cues onto grid cells happened.

However, weaker visual landmarks failed to gain the control (Fig. 8, supplementary video S3). To distinctively show the difference, mush smaller reliability $1/\sigma^2_{\text{inject}}$ is used (Table 4). When the phases of the calibration and integration cells were different from each other during loop closure, the final estimated phase was consistently coincided with that of integration cell (Fig. 8A). Weak reliability of the visual cues was not able to inhibit the vestibular cues. As a result, insufficient weight was assigned for visual information in the integration, leading to unsuccessful resolve of cue conflicts.

6.1.4. Firing rate maps

In order to see the neural response of a single cell, we divided the ring attractor manifold of head direction equally into 36 units and the torus attractor manifold of grid cells equally into 36 × 36 units.

The neural activities of single units (equivalent to a single cell) are shown on top of the cognitive map in Fig. 9.

HD unit 1 fires strongly when the robot moves in directions close to southwest (Fig. 9A), and shows degraded activity over similar directions due to the width of the bump in the HD cell model (Fig. 4). The unit does not fire on many roads that are parallel to the southwest direction, since the robot moves on those roads in the direction of northeast, i.e. opposite to the preferred firing direction of the unit. This is confirmed by the firing rate map of HD unit 19, whose preferred firing direction is opposite to that of HD unit 1 (9B). This unit fires when the robot travels towards northeast and its firing rate map does not overlap with that of HD unit 1 shown in Fig. 9A, suggesting that it keeps silent when the unit in Fig. 9A is active. On bending paths, as the robot turns toward and away from the preferred head directions of the units, their firing rates increase first and then decrease (Fig. 9A–B), with changing speed being coincided with the angular velocity of the robot. Globally, HD units only maintain consistent preferred directions in local regions, due to the fact that the error in path integration is correctly only in local loops.

Due to the torus structure of the grid cell model, a single grid unit fires at multiple distinct locations in the environment. Fig. 9C shows the firing rate map grid unit (1, 1). In general, the activity of a grid unit gradually increases when the robot enters the center of its firing fields, and decreases slowly as the robot leaves its fields. The multiple firing fields of the grid unit do not distribute in a triangular grid structure, as the large-scale physical environment explored here is only traversed for a few times. As a contrast, the maze used in the recording of grid cells is explored by the animal extensively, a situation may facilitate the emergence of globally consistent grid maps.

6.2. iRat Australia dataset

The iRat Australia dataset is obtained by iRat, a miniature mobile robot with similar size and shape as a rat, via a web camera in a maze mimicking Australian geography, including famous Australian landmarks, such as the Sydney Opera House and Uluru. The camera images, odometry messages, and overhead images are all provided by the iRat ROSbag dataset. The parameters of the NeuroBayesSLAM system are set the same as before, according to Tables 1 and 2.

The mapping process is presented in supplementary video S4. The interface of the NeuroBayesSLAM system running on this dataset is shown in Fig. 10, with the display of overhead image in panel B.

Since the behavior of the HD and grid cell models has been already discussed in detail and the mechanism of cue conflict resolution between visual cues and vestibular cues has been also concretely elaborated, here, we only show the results of the cognitive map and firing rate maps for iRat Australia dataset.

6.2.1. Cognitive map

Fig. 11 shows the cognitive map constructed by NeuroBayesSLAM system for the iRat Australia dataset. The path integration process is subject to the accumulation error in integrating velocity information. Visual inputs provide correction information during loop closures, and as a result, the shape of the cognitive map before and after loop closures is optimized (see video S4 in supplementary materials). As the iRat explores the whole environment, the map gets more stable with minor tweaks. Compared with the ground truth map, shown in Fig. 10B, the final cognitive map topologically captures the structure of the explored environment.

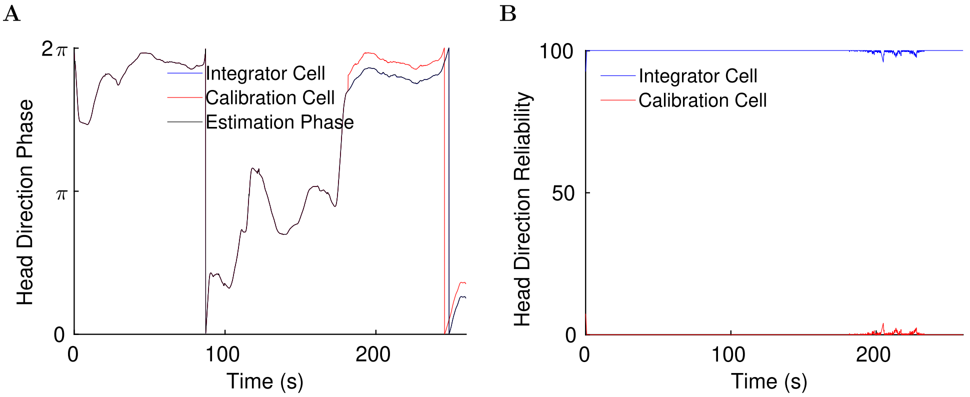

Fig. 8. An example of cue conflict resolution for weaker visual landmarks. (A) shows the HD phase changing overtime; (B) shows reliability changing overtime.

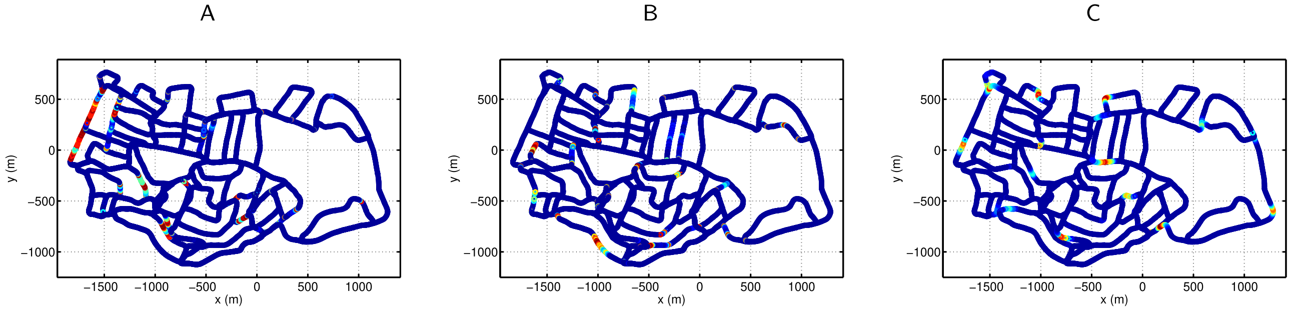

Fig. 9. Firing rate maps of example units for St Lucia suburb dataset. Firing rate is color-coded from blue to red by the same jet colormap. The firing rate of a single unit at different locations is plotted on top of the experience map. (A) The firing rate map of HD unit 1 corresponding to head direction 0 degree. (B) The firing rate map of HD unit 19 corresponding to head direction 180 degree, whose preference is opposite to HD unit 1 in (A). (C) The firing rate map of grid unit (1, 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

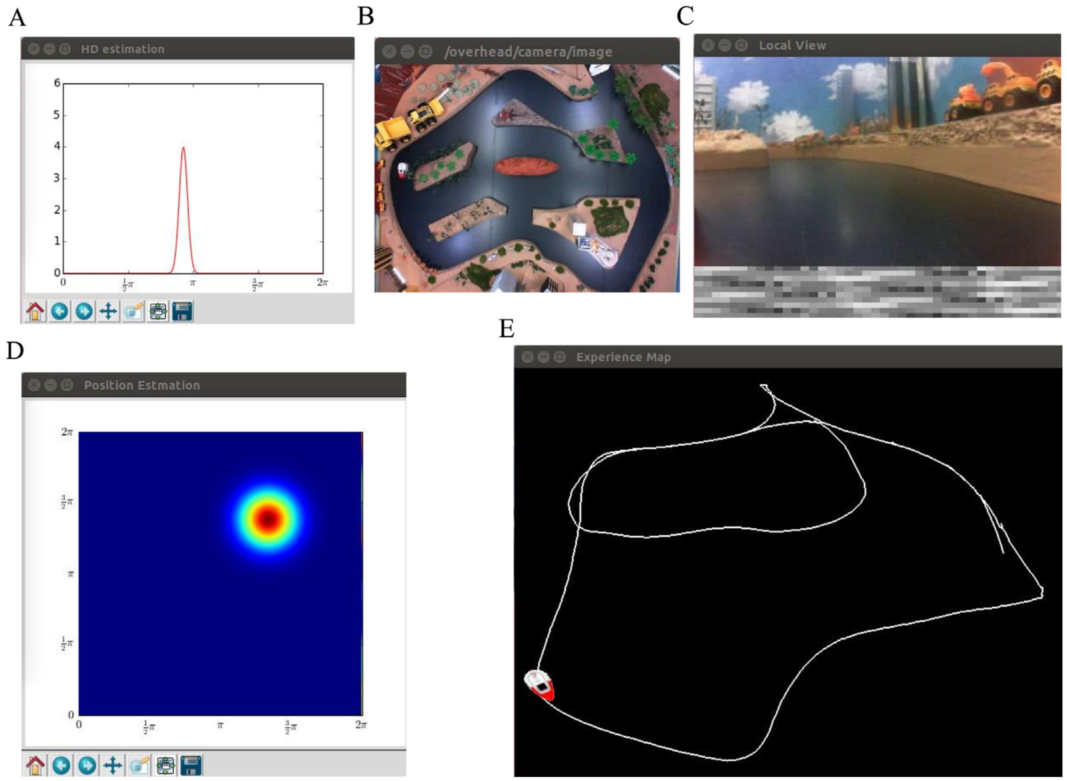

Fig. 10. Screenshots of the NeuroBayesSLAM system for iRat Australia dataset. (A) The neural activity of HD cells; (B) Overhead image is shown by ROS image_viewer; (C) Input visual scene (top); the local view template and the best matched template (bottom); (D) The neural activity of grid cells is rendered as heat map; (E) Experience map.

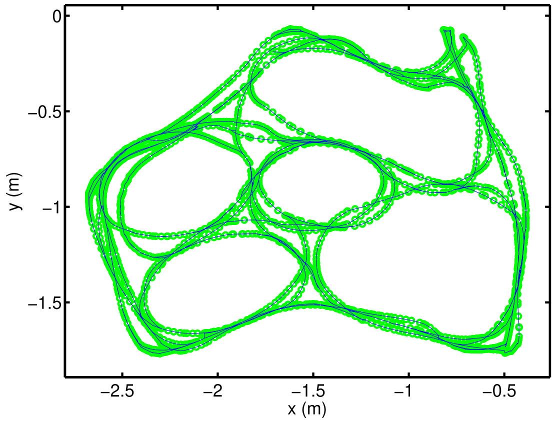

Fig. 11. Cognitive map for iRat Australia dataset. The green circles are the vertices of the topological graph, and the blue thin line consists of edges connecting related topological vertices. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6.2.2. Firing rate maps

The same way that used to obtain the firing rate map of a single unit for the St Lucia suburb dataset has been used here for the iRat Australia dataset.

Fig. 12A and B give the firing rate maps of head direction unit 1 and unit 19. These two units have completely opposite preferred directions. In the same local region, the two units express opposite firing fields (Fig. 12A vs. B). Due to the fact that the iRat can turn its head in any locations in the physical environment, not like St Lucia suburb dataset, the firing fields of unit 1 and 19 do not always maintain opposite to each other globally.

Fig. 12C shows that grid unit (1, 1) also fires at multiple distinct locations in the physical environment, due to the periodic boundary conditions of the grid cell model. While the iRat passes the fields through different topological trajectories in the cognitive map, the grid unit gets activated, meaning that consistent firing fields of grid units are maintained locally. The global grid structure is missing however, due to the fact that local view cells anchor grids to local visual cues when the environment is only covered by several passages.

7. Discussion

Inspired by single neuron recording presented in Knight et al. (2012), we developed in this paper a novel Bayesian attractor network model, called NeuroBayesSLAM, to replicate the multisensory integration mechanism found in animal navigation. Integrator cells and calibration cells are introduced to model HD cells or grid cells in parahippocampal region. Movement information and visual information, encoded by the integrator cells and the calibration cells respectively, were integrated by Bayesian inference. The performance of the proposed NeuroBayesSLAM model was verified in both indoor and outdoor environments of different scales. Coherent semi-metric topological maps were successfully constructed for both environments, under the same parameter setting. The NeuroBayesSLAM model reproduced similar responses of HD units and grid units in the two environments as those of HD cells and grid cells observed in rodents.

The integrator cells and the calibration cells in the HD cell model could correspond to the neurons in the VIP area and the MSTd area of the head direction neural system, respectively. We hypothesis that the corresponding brain areas for integrator cells and calibration cells of the grid cell system may reside in MEC, pre- and parasubiculum (Boccara et al., 2010). This hypothesis is supported by connectomics results showing that there is a segregation of afferent inputs to these regions. MEC receives strong egocentric movement information from regions like postrhinal cortex, parietal cortex. Pre- and parasubiculum is innervated mainly by retrosplenial cortex, which relays strong inputs from visual cortex.

To concisely describe the neural population activity of HD cells and grid cells, Gaussian profiles are applied to represent the activity packet of the cells. This simple representation preserves the single bump activity of the attractor network without the need to resort to the structured recurrent connections (Pastoll, Solanka, van Rossum, & Nolan, 2013; Tsodyks & Sejnowski, 1995). Robust estimation of head directions and locations of the robot is obtained by competitive dynamics between integrator cells and calibration cells.

Our study contributes to both neuroscience and robotic fields. In the field of neuroscience, the proposed Bayesian attractor network provides a descriptive model on conflict resolution between visual cues and vestibular cues. Successful localization and mapping in outdoor and indoor environments demonstrate that Bayesian integration might be a general rule in combining information from different sensory modalities with different reliabilities. In the proposed Bayesian attractor network, cue conflicts are resolved by plastic remapping of visual cues from calibration cells, which leads to a shift in preferred head directions of HD cells (Fig. 6E to H) and preferred positions of grid cells (Fig. 7A to D). The remapping depends on the reliability of visual cues. Strong and continuous visual cues gain cues control more easily (HD cells in Fig. 6 and grid cells in 7), in which visual cues inject higher energy into HD cells and grid cells by calibration cells (Table 3). However, weak and incontinuous visual landmarks lose cue control with parameters in Table 4 (Fig. 8).

In the field of robotics, we proposed an efficient method with biological plausibility to solve the SLAM problem. The proposed Bayesian attractor network has a simple structure, without

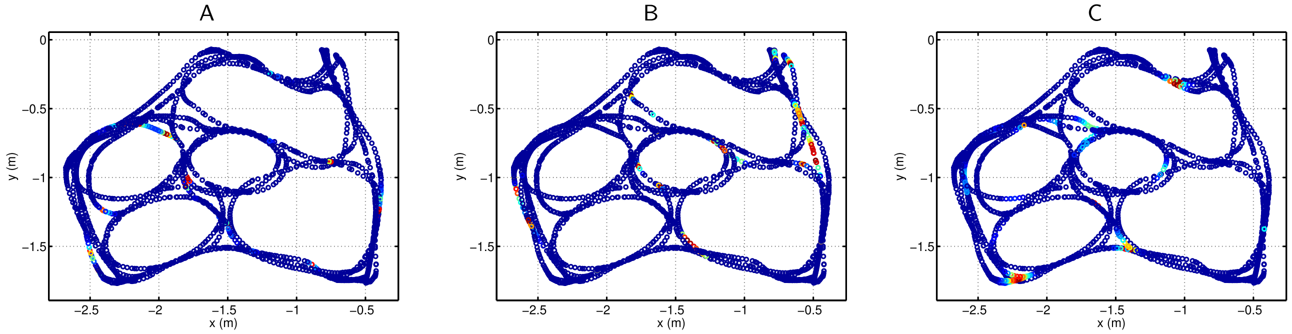

Fig. 12. Firing rate maps of example units for iRat Australia dataset. (A) The firing rate map of the HD unit 1. (B) The firing rate map of the HD unit 19, whose preference is opposite to the HD unit 1 in (A). (C) The firing rate map of grid unit (1, 1).

the need of simulating recurrent connections between neurons. This concise yet effective and biologically plausible method is computationally much more efficient than models based on recurrent networks (Milford & Wyeth, 2008; Zeng & Si, 2017), and is attractive for robot applications of long term autonomy in energy critical situations. Besides, fewer parameters are required in the model, compared with previous continuous attractor network (Milford & Wyeth, 2008; Si et al., 2014), alleviating the troublesome parameter tuning process for different scenarios. Thus, it is easier to apply the model in realistic robotic navigation tasks.

In our model, grid units and HD units form locally consistent firing fields, without coherent field structure globally (Figs. 9 and 12). The main reason is that local view cells anchor grid units and HD units to local cues (Derdikman et al., 2009), breaking the regular grid patterns for grid units, and the consistent preferred head direction firing patterns for HD units. One possible mechanism to obtain a global grid structure for grid cells is repeated exploration in the same environment (Carpenter, Manson, Jeffery, Burgess, & Barry, 2015). In the initial exploration of the environment, grid patterns only express in local regions. After rats repeatedly exploring the environment, globally coherent representations of physical environment are acquired and the grid firing maps become globally consistent. The globally consistent representation may provide a universal spatial metric in all environments (Moser et al., 2014). However, recent discoveries show that grid patterns are very often fragmented and distorted, due to the diversity of local features in the environment (Krupic, Bauza, Burton, Barry, & O'Keefe, 2015; Stensola, Stensola, Moser, & Moser, 2015). We would further extend our model to investigate the formation of global patterns of both HD units and grid units.

The core component of our NeuroBayesSLAM system, i.e., Bayesian attractor network, is essentially different from the pose cell network previously proposed in Ball et al. (2013), although some components related to vision and experience map are reused in our system. First, the Bayesian attractor network in this paper maintains a continuous representation of space by parametric probabilistic distributions. The same network can be used for exploring different environments of various scales under the same parameter setting. The pose cell network in Ball et al. (2013) would require the number of cells in the network to scale linearly with the size of the environment, or set the number of cells to accommodate the largest possible environment. Second, the Bayesian attractor network in this paper is endowed with competitive dynamics between integrator cells and calibration cells, and integrates multisensory information to resolve cue conflicts. The dynamics of the pose cell network in Ball et al. (2013) is implemented by the balance of local excitation and global inhibition. Third, the computational complexity of the proposed Bayesian attractor network is low, due to its concise probabilistic representations, requiring less computation time and resources as compared to attractor networks based on recurrent connections. More specifically, the head directions and positions of the agent are represented by twelve variables, i.e. four (angular phase and its reliability of integrator cells and calibration cells) for the head direction network and eight for the representation of two dimensional positions in grid cell network. At each time step, the time and space complexity of the proposed method are both constant $\mathcal{O}(1)$. Compared with the pose cell network in RatSLAM, it encodes the conjunction of head direction and position of the agent using a three-dimensional matrix of size $N_X \times N_Y \times N_{THETA}$, and requires to update every pose cell for each time step (Milford et al., 2004). The time and space complexity of the pose cell network are both $\mathcal{O}(N_X N_Y N_{THETA})$. In the meantime, fewer number of parameters is needed, leading to increased applicability and usability in applying to new environments. Finally, the proposed model encodes head directions and positions separately, consistent with the fact that HD cells and grid cells form complementary neural circuits. The proposed model reproduces similar effects of cue conflict resolve as observed in neurobiological experiments (Knight et al., 2012; Page et al., 2013).

There are also several limitations in this study. First, place cells, which are found to play an important role in spatial cognition, are absent in our network. Second, our network is an abstract representation of the HD cells and grid cells system, neglecting detailed connection structures within neural populations. Third, for simplicity, the same energy intensity is injected to Bayesian attractor network when visual cues calibrate spatial representations, without considering the degree to which visual cues and local view templates are matched.

Future work may extend the current model along, but not limited to, four lines. First, a population of place cells could be included into our system. Place cells combine information from multiple modules of grid cell networks with diverse spacings and orientations (Solstad, Moser, & Einevoll, 2006). The resulting model could be used to investigate the computational mechanisms of sparse spatial representations in the entorhinal–hippocampal circuit. Second, stereo visual information could be considered for integration, in order to obtain more accurate representations of the environment. Third, mechanisms for self-adaptation of the parameters could be studied, so as to break the curse of manual parameter tuning to a large extent in the SLAM system. Fourth, systematic comparison of the proposed model with other models in large environments would give an evaluation of the performance for further development of the model (Li et al., 2019).

8. Conclusion

In summary, we developed a concise yet biologically plausible model, NeuroBayesSLAM, based on spatial cognitive mechanisms of mammalian brains to solve the SLAM problem. The proposed model successfully built coherent cognitive maps both in large scale outdoor and in small indoor environments. By modeling the dynamics of multisensory information integration, the network is able to resolve cue conflicts. The units in the model express firing fields in the environment similar to those recorded in neurobiological experiments. Different from the black-box modeling principle of classical deep convolutional neural networks, the proposed model is interpretable, since the network activity is selective to spatial quantities in physical environments. In addition, the multisensory integration method proposed in the model may constitute a general computational principle of other neural subsystems, other than the head direction system. As an attempt to reach better understanding of how mammalian brains work for navigation, the proposed model sheds light on developing more practical and reliable systems for robot navigation. Our work also indicates that brain-inspired algorithm is an intriguing direction towards the development of autonomous robot systems with high efficiency.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neunet.2020.02.023.

References

Arleo, A., & Gerstner, W. (2000). Modeling rodent head-direction cells and place cells for spatial learning in bio-mimetic robotics. *From Animals to Animats*, *6*, 236–245.

Ball, D., Heath, S., Wiles, J., Wyeth, G., Corke, P., & Milford, M. (2013). OpenRatSLAM: an open source brain-based SLAM system. *Autonomous Robots*, *34*, 149–176.

Barrera, A., & Weitzenfeld, A. (2008). Biologically-inspired robot spatial cognition based on rat neurophysiological studies. *Autonomous Robots*, *25*, 147–169.

Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, *92*, 3844–3848.

Boccara, C. N., Sargolini, F., Thoresen, V. H., Solstad, T., Witter, M. P., Moser, E. I., et al. (2010). Grid cells in pre- and parasubiculum. *Nature Neuroscience*, *13*(987).

Burgess, N., Donnett, J. G., Jeffery, K. J., & John, O. (1997). Robotic and neuronal simulation of the hippocampus and rat navigation. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, *352*, 1535–1543.

Butler, J. S., Smith, S. T., Campos, J. L., & Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Vision*, *10*, 23.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., et al. (2016). Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, *32*, 1309–1332.

Carpenter, F., Manson, D., Jeffery, K., Burgess, N., & Barry, C. (2015). Grid cells form a global representation of connected environments. *Current Biology*, *25*, 1176–1182.

Chandrasekaran, C. (2017). Computational principles and models of multisensory integration. *Current Opinion in Neurobiology*, *43*, 25–34.

Chen, A., DeAngelis, G. C., & Angelaki, D. E. (2013). Functional specializations of the ventral intraparietal area for multisensory heading discrimination. *Journal of Neuroscience*, *33*, 3567–3581.

Cuperlier, N., Quoy, M., & Gaussier, P. (2007). Neuro- biologically inspired mobile robot navigation and planning. *Frontiers in Neurorobotics*, *1*(3).

Davison, A. J., Reid, I. D., Molton, N. D., & Stasse, O. (2007). MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (29).

De Almeida, A. T., Araújo, H., Dias, J., & Nunes, U. (1995). Multi-sensor integration for mobile robot navigation. In *Artificial intelligence in industrial decision making, control and automation* (pp. 537–554). Springer.

Derdikman, D., Whitlock, J. R., Tsao, A., Fyhn, M., Hafting, T., Moser, M. B., et al. (2009). Fragmentation of grid cell maps in a multicompartment environment. *Nature Neuroscience*, *12*, 1325–1332.

Duckett, T., Marsland, S., & Shapiro, J. (2002). Fast, on-line learning of globally consistent maps. *Autonomous Robots*, *12*, 287–300.

Engel, J., Schöps, T., & Cremers, D. (2014). LSD-SLAM: Large-scale direct monocular SLAM. In *European conference on computer vision* (pp. 834–849). Springer.

Etienne, A. S., Maurer, R., & Séguinot, V. (1996). Path integration in mammals and its interaction with visual landmarks. *Journal of Fish Biology*, *199*, 201–209.

Grieves, R. M., & Jeffery, K. J. (2017). The representation of space in the brain. *Behavioural Processes*, *135*, 113–131.

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque mstd. *Nature Neuroscience*, *11*, 1201–1210.

Guanella, A., Kiper, D., & Verschure, P. (2007). A model of grid cells based on a twisted torus topology. *International Journal of Neural Systems*, *17*, 231–240.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*, 801–806.

Hebb, D. (1949). *The organization of behavior:a neuropsychological theory*. Wiley.

Horton, T. W., Holdaway, R. N., Zerbini, A. N., Hauser, N., Garrigue, C., Andriolo, A., et al. (2011). Straight as an arrow: humpback whales swim constant course tracks during long-distance migration. *Biology Letters*, *7*, 674–679.

Jauffret, A., Cuperlier, N., & Gaussier, P. (2015). From grid cells and visual place cells to multimodal place cell: a new robotic architecture. *Frontiers in Neurorobotics*, 9.

Klein, G., & Murray, D. (2007). Parallel tracking and mapping for small ar workspaces. In *Mixed and augmented reality, 2007. ISMAR 2007. 6th IEEE and ACM international symposium on* (pp. 225–234). IEEE.

Knierim, J. J., & Zhang, K. (2012). Attractor dynamics of spatially correlated neural activity in the limbic system. *Annual Review of Neuro- Science*, *35*, 267–285.

Knight, R., Piette, C. E., Page, H., Walters, D., Marozzi, E., Nardini, M., et al. (2012). Weighted cue integration in the rodent head direction system. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, *369*, 20120512.

Krupic, J., Bauza, M., Burton, S., Barry, C., & O'Keefe, J. (2015). Grid cell symmetry is shaped by environmental geometry. *Nature*, *518*, 232–235.

Li, L., Wang, X., Wang, K., Lin, Y., Xin, J., Chen, L., et al. (2019). Parallel testing of vehicle intelligence via virtual-real interaction. *Science Robotics*, *4*, eaaw4106.

Llofriu, M., Tejera, G., Contreras, M., Pelc, T., Fellous, J. M., & Weitzenfeld, A. (2015). Goal-oriented robot navigation learning using a multi-scale space representation. *Neural Networks*, *72*, 62–74.

Madl, T., Chen, K., Montaldi, D., & Trappl, R. (2015). Computational cognitive models of spatial memory in navigation space: A review. *Neural Networks*, *65*, 18–43.

Mandal, S. (2018). How do animals find their way back home? a brief overview of homing behavior with special reference to social hymenoptera. *Insectes Sociaux*, *65*, 521–536.

Mathis, A., Herz, A. V., & Stemmler, M. (2012). Optimal population codes for space: grid cells outperform place cells. *Neural Computation*, *24*, 2280–2317.

Milford, M. J., Wiles, J., & Wyeth, G. F. (2010). Solving navigational uncertainty using grid cells on robots. *PLoS Computational Biology*, *6*, e1000995.

Milford, M. J., & Wyeth, G. F. (2008). Mapping a suburb with a single camera using a biologically inspired SLAM system. *IEEE Transactions on Robotics*, *24*, 1038–1053.

Milford, M. J., Wyeth, G. F., & Prasser, D. (2004). Ratslam: a hippocampal model for simultaneous localization and mapping. In *IEEE international conference on robotics and automation, 2004. Proceedings. ICRA'04 2004* (pp. 403–408).

Moser, E. I., Kropff, E., & Moser, M. B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annual Reviews of Neuroscience*, *31*, 69–89.

Moser, E. I., & Moser, M. B. (2008). A metric for space. *Hippocampus*, *18*, 1142–1156.

Moser, E. I., Roudi, Y., Witter, M. P., Kentros, C., Bonhoeffer, T., & Moser, M. B. (2014). Grid cells and cortical representation. *Nature Reviews Neuroscience*, *15*, 466–481.

Mulas, M., Waniek, N., & Conradt, J. (2016). Hebbian plasticity realigns grid cell activity with external sensory cues in continuous attractor models. *Frontiers in Computational Neuroscience*, *10*(13).

Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, *31*, 1147–1163.

Mur-Artal, R., & Tardos, J. D. (2016). ORB-SLAM2: an open-source SLAM system for Monocular, Stereo and RGB-D Cameras. arXiv preprint arXiv:1610.06475.

Newcombe, R. A., Lovegrove, S. J., & Davison, A. J. (2011). DTAM: Dense tracking and mapping in real-time. In *Computer vision (ICCV), 2011 IEEE international conference on* (pp. 2320–2327). IEEE.

Nüchter, A., Lingemann, K., Hertzberg, J., & Surmann, H. (2007). 6d slam—3d mapping outdoor environments. *Journal of Field Robotics*, *24*, 699–722.

O'keefe, J., & Conway, D. (1978). Hippocampal place units in the freely moving rat: why they fire where they fire. *Experimental Brain Research*, *31*, 573–590.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*, 171–175.

Page, H. J. I., Walters, D. M., Knight, R., Piette, C. E., Jeffery, K. J., & Stringer, S. M. (2013). A theoretical account of cue averaging in the rodent head direction system. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, 369.

Pastoll, H., Solanka, L., van Rossum, M. C., & Nolan, M. F. (2013). Feedback inhibition enables theta-nested gamma oscillations and grid firing fields. *Neuron*, *77*, 141–154.

Rolls, E. T., & Stringer, S. M. (2005). Spatial view cells in the hippocampus, and their idiothetic update based on place and head direction. *Neural Networks*, *18*, 1229–1241.

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M. B., et al. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, *312*, 758–762.

Seilheimer, R. L., Rosenberg, A., & Angelaki, D. E. (2014). Models and processes of multisensory cue combination. *Current Opinion in Neurobiology*, *25*, 38–46.

Si, B., Romani, S., & Tsodyks, M. (2014). Continuous attractor network model for conjunctive position-by-velocity tuning of grid cells. *PLoS Computational Biology*, *10*, e1003558.

Solstad, T., Moser, E. I., & Einevoll, G. T. (2006). From grid cells to place cells: a mathematical model. *Hippocampus*, *16*, 1026–1031.

Stachniss, C., Leonard, J. J., & Thrun, S. (2016). Simultaneous localization and mapping. In *Springer handbook of robotics* (pp. 1153–1176). Springer.

Stensola, T., Stensola, H., Moser, M. B., & Moser, E. I. (2015). Shearing-induced asymmetry in entorhinal grid cells. *Nature*, *518*, 207–212.

Strösslin, T., Sheynikhovich, D., Chavarriaga, R., & Gerstner, W. (2005). Robust self-localisation and navigation based on hippocampal place cells. *Neural Networks*, *18*, 1125–1140.

Sünderhauf, N., & Protzel, P. (2010a). Beyond ratslam: Improvements to a biologically inspired slam system. In *2010 IEEE 15th conference on emerging technologies & factory automation (ETFA 2010)* (pp. 1–8). IEEE.

Sünderhauf, N., & Protzel, P. (2010b). From neurons to robots: towards efficient biologically inspired filtering and slam. In *KI 2010: Advances in artificial intelligence* (pp. 341–348).

Tang, H., Yan, R., & Tan, K. C. (2017). Cognitive navigation by neuro-inspired localization, mapping, and episodic memory. *IEEE Transactions on Cognitive and Developmental Systems*, *10*, 751–761.

Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *The Journal of Neuroscience*, *10*, 420–435.

Thrun, S., Burgard, W., Fox, D., & Arkin, R. (2005). Probabilistic robotics. In *Intelligent robotics and autonomous agents*. MIT Press.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(189).

Tsoar, A., Nathan, R., Bartan, Y., Vyssotski, A., DellOmo, G., & Ulanovsky, N. (2011). Large-scale navigational map in a mammal. *Proceedings of the National Academy of Sciences*, *108*, E718–E724.

Tsodyks, M., & Sejnowski, T. (1995). Associative memory and hippocampal place cells. *International Journal of Neural Systems*, *6*, 81–86.

Tully, S., Kantor, G., & Choset, H. (2012). A unified bayesian framework for global localization and slam in hybrid metric/topological maps. *International Journal of Robotics Research*, *31*, 271–288.

Wang, S., Clark, R., Wen, H., & Trigoni, N. (2018). End-to-end, sequence-to-sequence probabilistic visual odometry through deep neural networks. *International Journal of Robotics Research*, *37*, 513–542.

Zeng, T., & Si, B. (2017). Cognitive mapping based on conjunctive representations of space and movement. *Frontiers in Neurorobotics*, 11.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *The Journal of Neuroscience*, *16*, 2112–2126.

Zhang, W. H., Wang, H., Wong, K. M., & Wu, S. (2016). "Congruent" and "Opposite" neurons: Sisters for multisensory integration and segregation. In *Advances in neural information processing systems* (pp. 3180–3188).

Zhang, W. H., & Wu, S. (2013). Reciprocally coupled local estimators implement Bayesian information integration distributively. In *Advances in neural information processing systems* (pp. 19–27).