2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

RGB-D based Cognitive Map Building and Navigation

Bo Tian, Vui Ann Shim, Miaolong Yuan, Chithra Srinivasan[1], Huajin Tang[2], Haizhou Li[3]

Abstract— **This paper describes a cognitive map building and navigation system using an RGB-D sensor for mobile robots. A brain-inspired simultaneously localization and mapping (SLAM) system, that requires raw odometry data and RGB-D information, is used to construct a spatial cognitive map of an office environment. The cognitive map contains a set of spatial coordinates that the robot has traveled. A global path is extracted from the built cognitive map and subsequently used by a local planner to instruct the robot to navigate. The global path is a subset of the path that builds up the cognitive map. This is different from other path planning mechanisms that construct a path based on a ground-truth map. Experiment results show that the employment of the RGB-D sensor significantly improves the mapping results.**

I. INTRODUCTION

Spatial cognition is the basic ability of mammals to perform cognitive tasks including exploration, map building, localization, and navigation in an environment. The study of spatial cognition has gained a fast-paced development since the past decades, which can be attributed to a shared idea from various disciplines such as psychology and neuroscience [1], [2], [3], [4], [5]. A concept called cognitive map, which is used to acquire spatial knowledge and represent the various topological relationships, has been brought forward. This concept has enhanced the study of various topics in robotics such as obstacle avoidance, dead-reckoning, self-localization, mapping, and path planning [6].

To navigate freely and safely in an environment, a robot should be endowed with an ability to interpret a physical world. This can be achieved by providing a representative map of the physical world to the robot. It has been observed that humans do not need to recall all the details of the environment and can navigate rather effortlessly. Though the underlying principles of how humans navigate is still unclear, it can be safely claimed that the navigation of humans is not completely based on a detailed world model, but also involves path planning based on their own previous experiences. Thus, this logically questions the underlying need to build the ground-truth map for robotic navigation. As such, RatSLAM, a brain-inspired SLAM algorithm [7], which constructs a cognitive map using the information of the path traveled in its previous experiences, is of our interest.

[1]Bo Tian, Vui Ann Shim, Miaolong Yuan, and Chithra Srinivasan[1] are with the Institute for Infocomm Research, Singapore.

[2]Huajin Tang is with the Institute for Infocomm Research, Singapore, and also with School of Computer Science, Sichuan University, China. (WhoCorrespondence should be addressed to: htang@i2r.a-star.edu.sg).

[2]Haizhou Li is with the Institute for Infocomm Research, Singapore, and also with School of Electrical Engineering and Telecommunications, University of New South Wales, Australia.

Today, the release of a new generation of sensing technology, RGB-D sensor in particular, has enhanced videos quality in both color and depth with comparatively inexpensive computational power. This provides an opportunity to dramatically improve the capabilities of robot systems. Many researchers are now focusing on how to construct a perfect physical world utilizing the depth information through map stitching and loop closure. Impressive results have been achieved [8]. Even though it is able to construct a good 3D map, it is computationally expensive and not suitable for real-time robot navigation.

In this paper, we employ an RGB-D camera as a visual sensor which is able to build a more compact and accurate spatial cognitive map in indoor environments. The RGB-D information and raw odometry data are used for map building based on RatSLAM. The map does not record the details of the environment, instead is presented by a set of spatial coordinates that the robot has experienced in its past travels. We also demonstrate that the robot is able to navigate the office environment in real-time by following the path generated by a global planning module. This path is a subset of the route mapped by the robot during its mapping stage. Thus, extracting a global path for navigation is computationally inexpensive. The experimental results show that the mapping using an RGB-D sensor and navigation utilizing the global path extracted from the constructed map are promising.

This paper is organized as follows. Section II introduces related work. Section III describes our robot platform and related tasks. Technical details are given in Section IV. Section V describes the implementation details with experimental results. Finally, conclusion is given in last Section.

II. RELATED WORK

Place cells which are located in the hippocampus residing in the temporal lobe of the brain give evidence that a cognitive map exists in human brain [2]. After the discovery of grid cells in the dorsocaudal medial entorhinal cortex[4], extensive research analysing their properties has been carried out and computational models have been proposed to describe the brain-based spatial navigation mechanism. One of the most famous models, based on continuous attractor network (CAN), was proposed by McNaughton et al. [5]. Milford and Wyeth have successfully applied this computational model to build cognitive maps in a large area (RatSLAM) [7]. It is able to build the map of a suburb with a monocular camera. Thereafter, it has been served as a core component for any persistent mapping and navigation system on a mobile robot equipped with a panoramic camera [9]. Furthermore, it is an appearance-based SLAM system, which

978-1-4673-6357-0/13/$31.00

can filter out random false positives to guarantee a successful loop closure.

Another notable work on appearance-based SLAM is FAB-MAP [10], [11]. FAB-MAP does not rely on a filter, but depends on the accurate comparison derived from speeded up robust features(SURF) descriptors. A combination of the RatSLAM and FAB-MAP significantly improves the robustness of the mapping system [12].

In recent years, RGB-D sensors have been introduced to robotics in various domains. In virtue of the rich depth information, the RGB-D sensor has been applied to many indoor SLAM algorithms [13], [14], [8]. These RGB-D based algorithms are based on visual features to match the images and calculating the transformation using RANSAC. Finally the RGB-D images are stitched together to form a whole 3D map of the physical environment.

Although many individual topics have been well studied, building an integrated system that involves a mapping to navigation modules is still a difficult task due to the complicated system design, the uncertainties in a real world, and many other practical issues. The navigation stack [15] aims for a general navigation purpose and has been shown that it is smooth and robust in a real office environment.

III. PLATFORM AND TASKS

A. Robot Platform

NECO-II is a mobile robot, which is equipped with a Pioneer 3-DX mobile base, an RGB-D Sensor, a TP-LINK wireless router for client-master communication, and a PC/104 embedded computer. The physical views of NECO-II are shown in Figure 1. The Pioneer 3-DX mobile base consists of two front wheels and a supporting back wheel for stabilization. The front wheels are equipped with encoders that records the distance traveled. The maximum speed of this mobile base is set to 0.4 m/s.

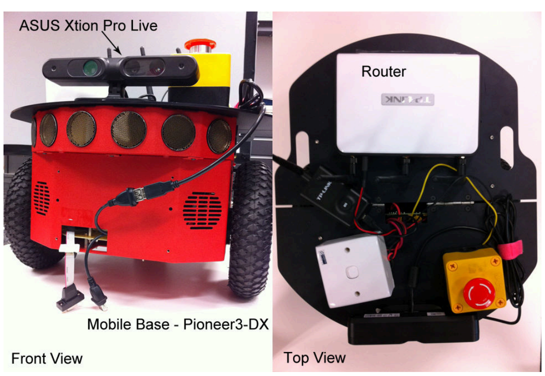


Fig. 1. Physical views of NECO-II.

An RGB-D sensor is mounted on the front-top of the Pioneer 3-DX mobile base. This sensor features as a camera as well as a fake-laser pointer in a point cloud stream. NECO-II is developed on the well-known robotic platform, i.e., Robot Operating System (ROS). Three computers construct the whole mapping and navigation platform for NECO-II. A PC/104 mounted inside the Pioneer 3-DX mobile base, running Ubuntu 12.04, is the ROS master that connects directly to input sensors. A laptop, running Ubuntu 12.04, is a client that subscribes to the topics published by the ROS master. It is used for process visualization and target location assignation. Another laptop, running Window XP, is another client that communicates with the ROS master through TCP/IP. The brain-inspired SLAM algorithm runs in this computer. Both client computers communicate through TCP/IP for global and local coordinate frame matching.

B. Tasks Description

Three main tasks are performed by a robot system, henceforth referred to Neural Cognitive Robot II (NECO-II), a successor of NECO-I [16], [17], [18].

First, NECO-II is required to build a map of an office environment. Specific features in the office space are memorized through the map building process. These features serve as cues for loop closure as well as keys for spatial recognition. One of the Research Institute's office in Singapore, the ninth floor Connexis north tower of the Fusionopolis building, is used for the purpose of experiment. This is a typical office which mainly consists of working cubicles, chairs, tables, passageways, and dynamically moving people. The environment is kept in its original state without any rearrangement. The main challenge of this environment is that the visual frames of the features captured by NECO-II may be different from time to time as people are captured during the experiments. This may compromise the robustness of the system built as no advanced image processing methods are used because of requirement of real-time performance for a mobile robot system. NECO-II may fail to recognize a particular spatial coordinate in case it is not able to match the current view templates with its memorized cues.

Second, given a target destination, NECO-II is required to plan a global path connecting its current location and the final goal destination. The final goal destination can be any obstacle-free locations in the map. In order to plan the path, NECO-II has to localize itself and determine its current location. The path is also required to be the shortest path identified in the cognitive map.

Third, by following the guide of the global path, NECO-II is required to safely navigate from its current coordinate to the final goal location. That is to say that NECO-II needs a fast response mechanism to avoid obstacles, a capability to fit itself to a narrow path, and a reliable path following paradigm.

IV. METHODS

In this section, we describe the related methods used in our robot system in detail.

A. Overall Architecture

Figure 2 shows the architecture of the mapping and navigation system. Image information from a Xtion camera, odometry from the mobile base, and a goal destination decided by a user are the input data to the system. The visual templates from the Xtion camera as well as the raw

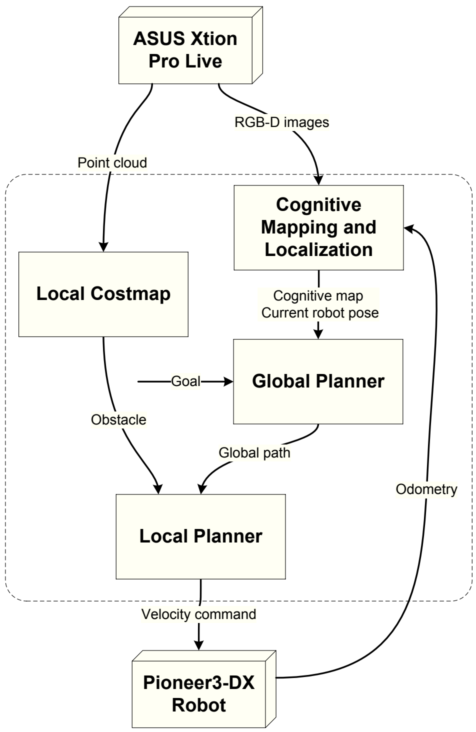


Fig. 2. System architecture.

odometry data from the mobile base are used to built the map which is then used by a global planner for path planning. The localization module matches the visual templates from the Xtion to the features observed in the map in order to determine the robot's current position. The local planner generates a velocity command if it is obstacle free. The task execution stops when the goal destination is reached or the timeout is activated.

B. Visual Processing

The RGB-D sensor provides the color image and depth image for a brain-inspired appearance-based SLAM system. It is difficult to perform an appearance-based SLAM in an environment with uniform furniture and similar decorations using a single RGB camera. Taking advantages of the depth measurements, which can avoid ambiguity caused in 2D images and it is invariant to lighting conditions, many similar indoor scenes become distinguishable. Therefore, it is possible to perform an appearance-based mapping for a challenging office environment. Both of the RGB image and depth information are converted to 8-bit mono images, shrunk to smaller size (from 640 $\times$ 480 pixels to 320 $\times$ 240 pixels) for real-time purpose, and then changed into scanline intensity profiles. A comparison between the profiles is performed for each pair of incoming RGB and depth frames. If the current profiles match previously seen profiles, it is considered as an old scene which has been seen again. Otherwise, a new visual template is created. The image comparison plays an important role in loop closure.

Figure 3 shows several mono images of processed RGB (upper) and depth (bottom) images of the office environment. This environment is challenging for a SLAM system as it is highly dynamic and consists of many near-similar views in different locations. The depth images which provide important features for image comparison may reduce the false positive comparison.

The RGB-D sensor also works as a range finder by transforming the depth information to point cloud stream. The point cloud stream is used for obstacle detection when navigating the office environment. Though the angle of view of the RGB-D sensor is narrow compared to a laser scanner, it can sufficiently enable the robot to perform a local path planning.

C. Map Building

The map of the office environment is built based on RatSLAM algorithm [7]. The core mecnahism of RatSLAM is a three dimensional CAN [3], [5]. It is known as a pose cell model in which the robot pose information (x, y, θ) is encoded to cells representing positions (x,y) and direction (θ).

The RatSLAM system takes in images and raw odometry data, and generates a cognitive map which consists of discrete individual experiences. When similar scenes are encountered again by the robot and distance between pose cells are with in a predefined threshold, the two experiences are considered as the same and a loop closure will be performed. The pose cells guarantee that a false positive image comparison result can be filtered out and a leap in pose cell activities happens if several similar scenes are continuously seen even with the accumulated odometric errors. Once the loop is closed, the map will be corrected to adjust the positions of the experiences.

D. Localization

In robotics, global localization is used to localize a robot from a random starting position within a familiar environment. It is an essential part of a robot mapping and navigation system.

In this work, an additional global localization phase is employed. In this phase, the map remains unchanged. The activities of the pose cell are updated based on visual inputs only. After re-initialization, the robot restores to its previous status and enters the global localization phase. If a previous experience is matched, the global localization phase ends.

E. Global Planner

The generation of the global path is based on the cognitive map built by RatSLAM. It is a subset of the experienced path that NECO-II has traveled during the map building process. Each path is denoted as an array of points $P = (p_1, ..., p_n)$, n is the number of points which represents the coordinate frame $p_i = (p_i^x, p_i^y)$ in the experience map.

In most cases, the cognitive map may have several paths linking the current and goal locations. Some paths are relatively close to each others as more than one path may be generated in the same passageway. A temporal map is constructed to mate the paths that are within d_m. This is done

Fig. 3. Examples of RGB and depth images of the office environment.

by taking average of the coordinate points that are within a distance d_m. Each point in the map is then assigned an index ($I = (I_1, ..., I_n)$) for path planning. It should be noted that the points that are within the distance d_m are assigned a same index.

The global path linking the current location (L_{start}) and goal destination (L_{goal}) is a sub-route in P. The index for L_{start} (I_{start}), is assigned as the index of a point in P (I_a) that has the closest distance to L_{start}. Similarly, the index for L_{goal} (I_{goal}) is the index of point in P (I_b) that has the closest distance to L_{goal}. The indices for the global path is denoted as:

$$I_{global} = (I_{start=a}, I_{start+1}, ..., I_{goal-1}, I_{goal=b})$$

We perform the shortest path search that returns a path with the shortest distance connecting both coordinates. The path is then published over the ROS and the local planner will subscribe it before navigation process starts. A transformation mechanism is then executed to convert the global coordinate frame of the experience map space to the local coordinate frame of the robot-centered real environment. The navigation stops when the goal destination is reached within offset distance d_o.

F. Local Planner

The local planner creates a local path, connecting the current location to a local goal destination, that follows the global path closely. It generates the velocity commands to drive NECO-II to navigate in an indoor environment. For a safe drive, local planner takes into account the obstacles and dynamic movements of the mobile base. The detected obstacles are stored in a costmap. It is built and maintained, using the point cloud information, when NECO-II navigates the environment. For a safe movement, the velocity commands are generated using the dynamic window approach (DWA) [19]. A cost function consists of variables of distance to obstacles, distance to the global path, and motion speed are used to generate the velocity commands. A more detailed implementation of the local planner are referred in [15].

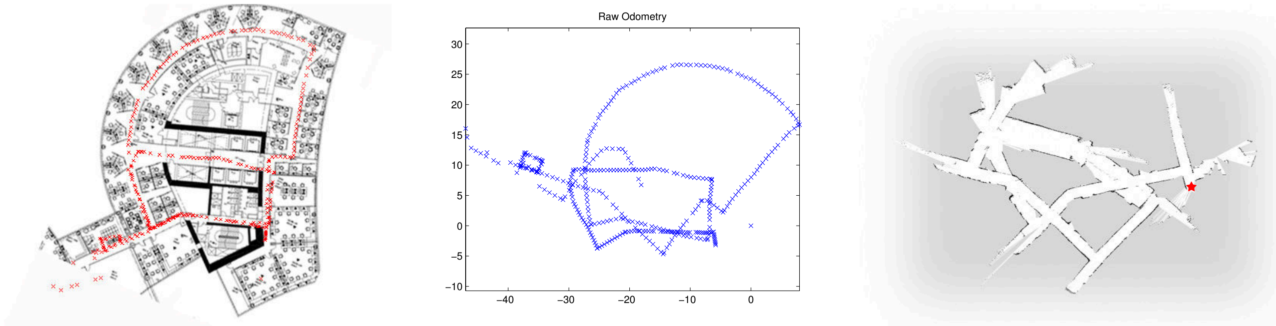
V. RESULTS

A. Implementation

During the map building process, we manually drive NECO-II using a remote controller to explore the office environment. The average motion speed is set to 0.2 m/s. The odometry data and RGB-D streams captured from the Xtion camera are recorded and fetched into RatSLAM. After the completion of the map building process, the resulting map is used for path planning and autonomous navigation with a specified target destination. The target can be any location in the map specified by a user. A temporal map is built, instead of using the cognitive map, for the global path planning. In order to construct an array of points that are at least 0.5m apart in the global path plan, we set the d_m parameter to 0.5. NECO-II is deemed to have reached its goal when it is 0.3m (d_o=0.3) from the target destination.

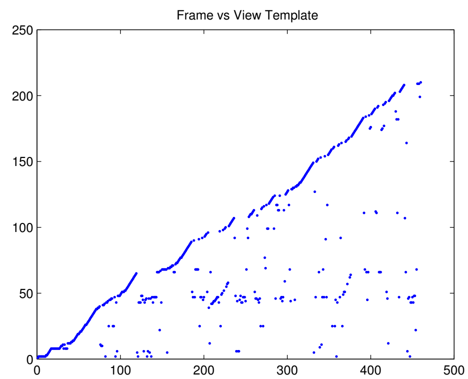
B. Mapping Results

As mentioned above, depth information provides important features for visual templates comparison. Figure 4 shows the frame versus view template and cognitive map generated from RatSLAM using RGB-only information. The y-axis is the number of visual templates and the x-axis is the number of input images captured. As can be observed from the figure, many false positive matches have been generated resulting in wrong loop closures and catastrophic failures in spatial representation. Hence, it is desirable to use RGB-D information to generate a more accurate map. In Figure 5, the frame versus view template and cognitive map using RGB-D information is presented. It can be concluded that the use of RGB-D information can significantly increase the true positive matches, resulting in correct loop closures and spatial representations as shown in Figure 6(a). Figure 6(b) shows that the map built directly using the raw odometry data fails to represent the real environment due to the accumulated errors. For comparison purpose, a map is built by gmapping [20], as shown in Figure 6(c), in a similar environment using the point cloud stream converted from depth information. It has been observed that gmapping failed to perform a loop

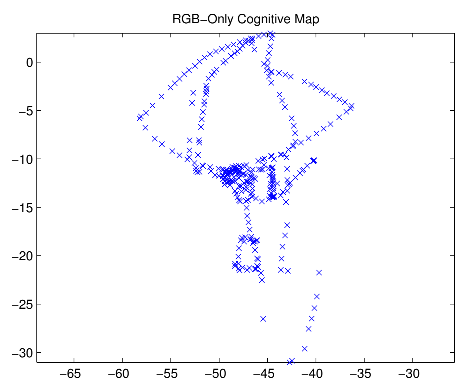


(a) (b) (c)

Fig. 6. (a) Cognitive map generated from RatSLAM using an RGB-D sensor superimposed with the real office environment (b) raw odometry data, and (c) map built from gmapping.

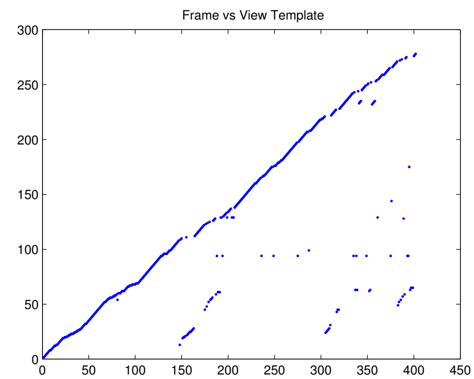


(a)

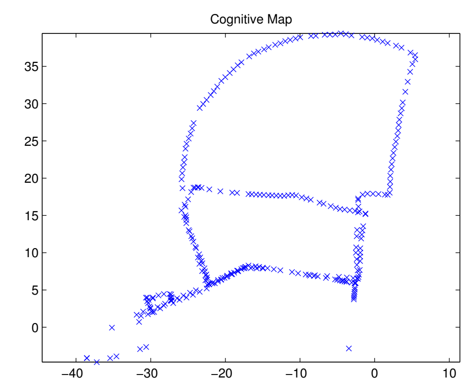


(b)

Fig. 4. (a) Frame vs view template and (b) cognitive map generated from RatSLAM using RGB-only information.



(a)



(b)

Fig. 5. (a) Frame vs view template and (b) cognitive map generated from RatSLAM using RGB-D information

closure and unable to construct a correct map of the office environment.

C. Navigation Results

The global path generated from the cognitive map is shown in Figure 7. This path is used to guide the local planner in navigation. Figure 8 shows the navigation process visualized in rviz [21]. The ground-truth-map, indicated by the light grey region, is built by the gmapping module. This map is not used by the navigation stack and it is generated for a clear visualization of the path traveled by NECO-II. NECO-II has successfully navigated from the given starting location to the goal destination following the global path obtained from our previously built cognitive map.

VI. CONCLUSIONS

We have successfully demonstrated a robotic navigation system integrated a state-of-the-art biologically inspired

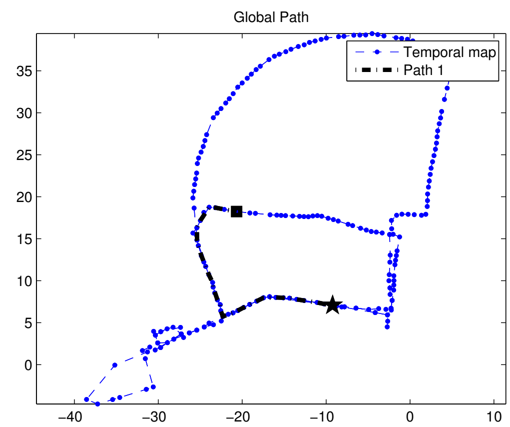


Fig. 7. Global path generated from the cognitive map. The star point is the current location and the square point is the goal destination.

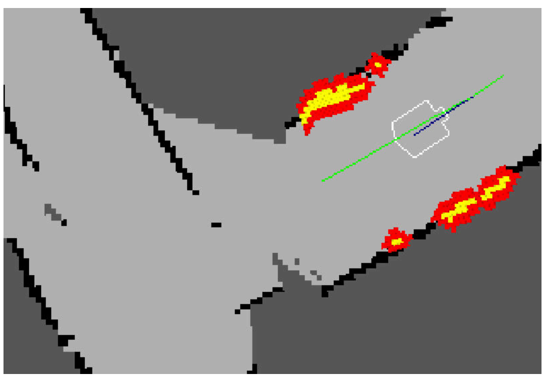


Fig. 8. Visualization of the navigation process. The green line indicates the global path extracted from the cognitive map. The blue line shows the local path generated by the local planner. The footprint of NECO-II is shown in white. The yellow and red regions are the detected obstacles.

SLAM system, revolutionary 3D RGB-D sensor, and dominating robotic operating system in a computationally inexpensive manner. The depth information of the RGB-D sensor has significantly improved the loop closure as well as feature matching, which eventually yielded a better spatial cognitive map that represents the real environment. A global path that links the current location and target destination has been extracted from the resulting map for navigation purpose. Experiments have also been conducted to show the capability of the robot to navigate by following the path obtained from the cognitive map. The performance of the mapping and navigation modules can be further enhanced. In the map building process, a higher recording rate is desirable in order to improve the experience matching results, which can produce a more realistic map. Currently, appearance-based image matching is used in our cognitive mapping process. A better alternative is to use 2D/3D interest point matching to improve the robustness under dynamic environments. With a better map, the navigation ability can be improved.

References

[1] E. C. Tolman, "Cognitive maps in rats and men." *Psychological review*, vol. 55, no. 4, p. 189, 1948.

[2] J. O'keefe and L. Nadel, *The hippocampus as a cognitive map*. Clarendon Press Oxford, 1978, vol. 3.

[3] A. Samsonovich and B. L. McNaughton, "Path integration and cognitive mapping in a continuous attractor neural network model," *The Journal of Neuroscience*, vol. 17, no. 15, pp. 5900–5920, 1997.

[4] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, "Microstructure of a spatial map in the entorhinal cortex," *Nature*, vol. 436, no. 7052, pp. 801–806, 2005.

[5] B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser, "Path integration and the neural basis of the'cognitive map'," *Nature Reviews Neuroscience*, vol. 7, no. 8, pp. 663–678, 2006.

[6] S. Thrun, W. Burgard, D. Fox *et al.*, *Probabilistic robotics*. MIT press Cambridge, 2005, vol. 1.

[7] M. Milford and G. Wyeth, "Mapping a suburb with a single camera using a biologically inspired slam system," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1038–1053, 2008.

[8] G. Hu, S. Huang, L. Zhao, A. Alempijevic, and G. Dissanayake, "A robust rgb-d slam algorithm," in *Intelligent Robots and Systems, International Conference on*. IEEE, 2012, pp. 1714–1719.

[9] M. Milford and G. Wyeth, "Persistent navigation and mapping using a biologically inspired slam system," *The International Journal of Robotics Research*, vol. 29, no. 9, pp. 1131–1153, 2010.

[10] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.

[11] ——, "Appearance-only slam at large scale with fab-map 2.0," *The International Journal of Robotics Research*, vol. 30, no. 9, pp. 1100–1123, 2011.

[12] A. J. Glover, W. P. Maddern, M. J. Milford, and G. F. Wyeth, "Fab-map+ ratslam: Appearance-based slam for multiple times of day," in *Robotics and Automation, International Conference on*. IEEE, 2010, pp. 3507–3512.

[13] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using depth cameras for dense 3d modeling of indoor environments," in *the 12th International Symposium on Experimental Robotics*, 2010, pp. 22–25.

[14] N. Engelhard, F. Endres, J. Hess, J. Sturm, and W. Burgard, "Real-time 3d visual slam with a hand-held rgb-d camera," in *the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden*, 2011.

[15] E. Marder-Eppstein, E. Berger, T. Foote, G. Brian, and K. Kurt, "The office marathon: Robust navigation in an indoor office environment," in *Robotics and Automation, International Conference on*. IEEE, 2010, pp. 300–307.

[16] H. Tang and W. Huang, "Brain inspired cognitive system for learning and memory," in *Proc. of 18th International Conference on Neural Information Processing*. Springer-Verlag, 2011, pp. 477–484.

[17] W. Huang, H. Tang, J. Yu, and C. H. Tan, "A neuro-cognitive robot for spatial navigation," in *Proc. of 18th International Conference on Neural Information Processing*. Springer-Verlag, 2011, pp. 485–492.

[18] H. Tang, B. Tian, V. A. Shim, and K. C. Tan, "A neuro-cognitive system and its application in robotics," in *Prof. of 10th International Conference on Control & Automation*. IEEE, 2013, pp. 406–411.

[19] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics and Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.

[20] G. Grisetti, C. Stachniss, and W. Burgard, "Improved techniques for grid mapping with rao-blackwellized particle filters," *Robotics, IEEE Transactions on*, vol. 23, no. 1, pp. 1552–3098, 2007.

[21] 3D visualization tool for ROS. [Online] http://www.ros.org/wiki/rviz.