SpCoSLAM: Online Multimodal Place Categorization, Spatial Lexical Acquisition and Mapping using a Mobile Robot

Akira Taniguchi[1], Yoshinobu Hagiwara[1], Tadahiro Taniguchi[1], and Tetsunari Inamura[2]

***Abstract*— In this paper, we propose a Bayesian generative model (SpCoSLAM) that can simultaneously learn place categories and lexicons while incrementally generating an environmental map. In addition, we propose an online learning algorithm based on a Rao-Blackwellized particle filter (RBPF) for spatial concept and lexical acquisition, as well as for mapping. In the experiments, we test the online learning of spatial concepts, lexicons, and a map in a novel environment. We demonstrate that a robot without a pre-existing lexicon or map can learn spatial concepts and an environmental map incrementally.**

I. INTRODUCTION

Robots that coexist with humans in various environments are required to adaptively learn and use the spatial concepts and lexicon related to various places. However, spatial concepts are such that their target domain may be unclear compared with object concepts, and they may differ depending on the user and the environment. Therefore, it is difficult to manually design spatial concepts in advance, and it is desirable for robots to autonomously learn spatial concepts based on their own experiences.

In this study, we assume that the robot has not acquired any vocabulary in advance, and that it can recognize only phonemes or syllables. We represent the spatial area of the environment in terms of a position distribution. Furthermore, we define a spatial concept as a place category that includes place names, scene-image features, and the position distributions corresponding to those names. We develop a method that enables mobile robots to learn spatial concepts, a lexicon, and an environmental map sequentially based on interactions with an environment or human, even in an environment in which it has no prior knowledge. We propose a novel unsupervised Bayesian generative model and an online learning algorithm that can perform simultaneous learning of the spatial concepts and an environmental map from multimodal information. The proposed method can automatically and sequentially perform place categorization and learn unknown words without prior knowledge.

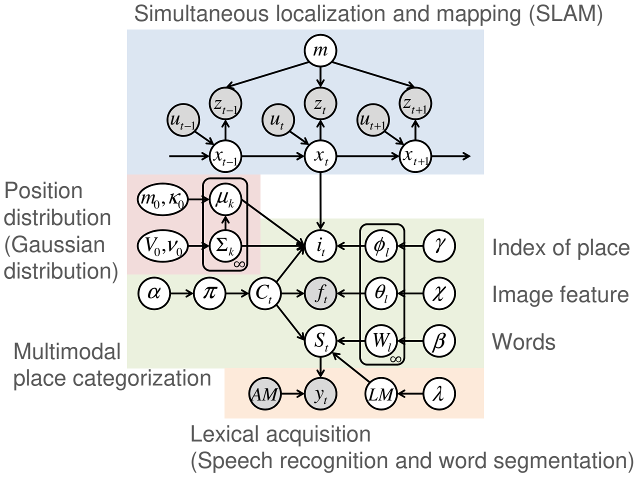

Fig. 1. Graphical model representation of SpCoSLAM; This model integrates multimodal place categorization, lexical acquisition, and SLAM as the Bayesian generative model. Gray nodes indicate observation variables.

II. SpCoSLAM

A. Overview

The proposed method is an online spatial concept acquisition and simultaneous localization and mapping (SpCoSLAM). This method can learn sequential spatial concepts for unknown environments and unexplored regions without maps. It can also learn many-to-many correspondences between names and places via spatial concepts, and can mutually complement the uncertainty of information by using multimodal information. Figure 1 shows the graphical model of SpCoSLAM. The steps in the procedure of SpCoSLAM are described as follows. (a) A robot obtains the weighted finite-state transducer (WFST) speech recognition results of the user's speech signals using a current language model. (b) The robot obtains the observation likelihood by performing a sample motion model and a measurement model of FastSLAM [2]. (c) The robot performs unsupervised word segmentation latticelm [3] using WFST speech recognition results. (d) The robot obtains latent variables of spatial concepts by sampling. (e) The robot obtains the marginal likelihood of observation data as the importance weight. (f) The robot updates an environmental map. (g) The robot estimates the set of parameters of spatial concepts from data and sampled values. (h) The robot selects a language model of the maximum weight for the next step. (i) Particles are re-sampled according to weights. (b) – (g) are performed for each particle. The details of this procedure are described in [1].

*This paper is an extended abstract version of the accepted paper [1] in IROS2017.

*This work was partially supported by JST CREST and JSPS KAKENHI Grant Number JP17J07842.

[1] Akira Taniguchi, Yoshinobu Hagiwara, and Tadahiro Taniguchi are with Ritsumeikan University, 1-1-1 Noji-Higashi, Kusatsu, Shiga 525-8577, Japan {a.taniguchi, yhagiwara, taniguchi} @em.ci.ritsumei.ac.jp

[2] Tetsunari Inamura is with the National Institute of Informatics / SOKENDAI (The Graduate University for Advanced Studies), 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan inamura@nii.ac.jp

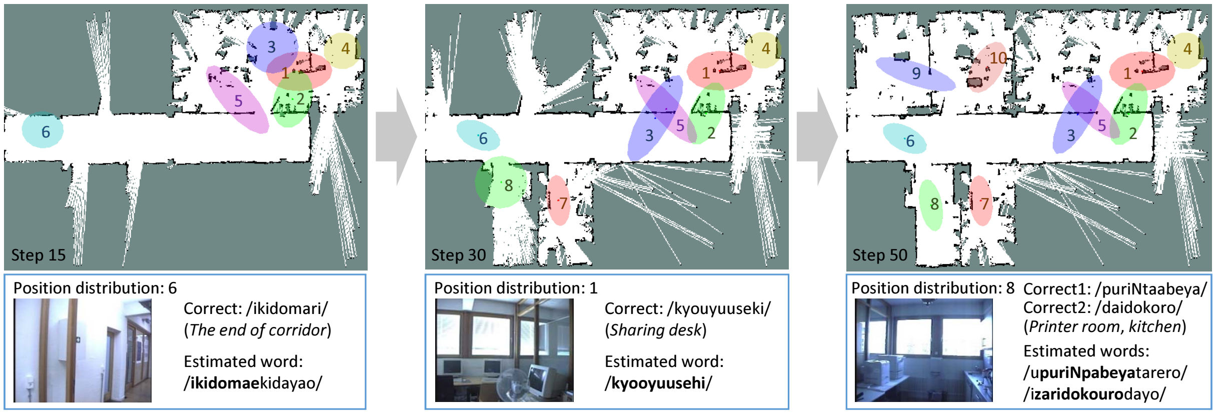

Fig. 2. Learning results of spatial concepts; Ellipses denoting the position distributions are drawn on the map at steps 15, 30, and 50. The colors of the ellipses were determined randomly. Each index number is denoted as $i_t = k$.

B. Online learning based on RBPF

The online learning algorithm of the proposed method can be derived by introducing sequential update equations for the estimation of the model parameters of the spatial concepts into the formulation of the FastSLAM based on RBPF. The proposed method applies the grid-based FastSLAM 2.0 algorithm. In the formulation of SpCoSLAM, the joint posterior distribution can be factorized to the probability distributions of a language model LM, a map m, the set of model parameters of spatial concepts $\Theta = \{\mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \theta, \phi, \pi\}$, and the joint distribution of the trajectory of self-position $x_{0:t}$ and the set of latent variables $\mathbf{C}_{1:t} = \{C_{1:t}, i_{1:t}, S_{1:t}\}$. C_t is an index of spatial concepts and i_t is an index of position distributions. We describe the joint posterior distribution of SpCoSLAM as follows:

$$\begin{aligned} & p(x_{0:t}, \mathbf{C}_{1:t}, LM, \Theta, m \mid u_{1:t}, z_{1:t}, y_{1:t}, f_{1:t}, AM, \mathbf{h}) \\ & = p(LM \mid S_{1:t}, \lambda) p(\Theta \mid x_{0:t}, \mathbf{C}_{1:t}, f_{1:t}, \mathbf{h}) p(m \mid x_{0:t}, z_{1:t}) \\ & \quad \cdot \underbrace{p(x_{0:t}, \mathbf{C}_{1:t} \mid u_{1:t}, z_{1:t}, y_{1:t}, f_{1:t}, AM, \mathbf{h})}_{\text{Particle filter}} \end{aligned} \tag{1}$$

where the set of hyperparameters is denoted as $\mathbf{h} = \{\alpha, \beta, \gamma, \chi, \lambda, m_0, \kappa_0, V_0, \nu_0\}$. Note that the speech signal y_t is not observed at all times. In this study, the proposed method is equivalent to FastSLAM at the time when speech signal y_t is not observed, i.e., the speech signal is a trigger for the place categorization. The particle filter algorithm uses sampling importance resampling (SIR). The details of online learning are described in [1].

III. Experiment

In this study, we performed an experiment for online spatial concept acquisition in a real environment. We extended the gmapping package, implementing the FastSLAM in the robot operating system (ROS). We used an open dataset (albert-b-laser-vision) containing a rosbag file in which the odometry, laser range data, and vision data were recorded. This dataset was obtained from the Robotics Data Set Repository (Radish) [4], and the authors thank Cyrill Stachniss for providing this data. We prepared Japanese speech signal data that correspond to the movement of the robot of the above dataset. The speech recognition system used Julius. The initial word dictionary contains 115 Japanese syllables, and the unsupervised word segmentation system used latticelm [3]. We used a pre-trained CNN Places205-AlexNet [5] as an image feature extractor. The number of particles was $R = 30$. The hyperparameters were set as follows: $\alpha = 20$, $\gamma = 10$, $\beta = 0.2$, $\chi = 0.2$, $m_0 = [0, 0]^{\mathrm{T}}$, $\kappa_0 = 0.001$, $V_0 = \mathrm{diag}(2, 2)$, and $\nu_0 = 3$. Fig. 2 shows the position distributions and estimated words in the environmental maps[1]. Based on the results, we showed that the spatial concepts are acquired while sequentially mapping.

IV. Conclusion

This paper described online learning methods for spatial concepts, a lexicon, and an environmental map using a mobile robot. The proposed method integrated the spatial concept acquisition into FastSLAM based on the RBPF approach. In the experiments, we conducted online learning in a novel environment using the robot without a pre-existing lexicon or map. We consider that SpCoSLAM can further improve the estimation accuracy by incorporating forgetting [6] and rejuvenation [7].

References

[1] A. Taniguchi, Y. Hagiwara, T. Taniguchi, and T. Inamura, "Online spatial concept and lexical acquisition with simultaneous localization and mapping," in *Proceedings of IROS*, 2017.

[2] G. Grisetti, C. Stachniss, and W. Burgard, "Improved techniques for grid mapping with Rao-Blackwellized particle filters," *IEEE Transactions on Robotics*, vol. 23, pp. 34–46, 2007.

[3] G. Neubig, M. Mimura, and T. Kawahara, "Bayesian learning of a language model from continuous speech," *IEICE Transactions on Information and Systems*, vol. 95, no. 2, pp. 614–625, 2012.

[4] A. Howard and N. Roy, "The robotics data set repository (radish)," 2003. [Online]. Available: http://radish.sourceforge.net/

[5] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Proceedings of NIPS*, 2014, pp. 487–495.

[6] T. Araki, T. Nakamura, T. Nagai, S. Nagasaka, T. Taniguchi, and N. Iwahashi, "Online learning of concepts and words using multimodal LDA and hierarchical Pitman-Yor Language Model," in *Proceedings of IROS*, 2012, pp. 1623–1630.

[7] K. R. Canini, L. Shi, and T. L. Griffiths, "Online inference of topics with latent Dirichlet allocation." in *Proceedings of AISTATS*, vol. 9, 2009, pp. 65–72.

[1] Video: `https://youtu.be/hVKQCdbRQVM`