Duckers 2017 @Home DSPL
Team Description Paper

Shota Isobe, Yuki Katsumata, Yoshiki Tabuchi, Akira Kinose, Tomoro Ishimine, Takashi Fukui, Hiroyoshi Kobayashi, Akira Taniguchi, Amir Aly, Yoshinobu Hagiwara, and Tadahiro Taniguchi

Ritsumeikan University, 1-1-1 Noji-Higashi, Kusatsu, Shiga 525-8577, Japan,
https://emlab.jimdo.com/

Abstract. This paper briefly describes our team's activity and background, and both the hardware and software architectures of Human Support Robot (HSR) in detail. Moreover, we shade on our research directions at our laboratory. Our team "Duckers" represents the Emergent Systems Laboratory at Ritsumeikan University in Japan, in the competition of RoboCup @Home Domestic Standard Platform League, 2017.

1 Introduction

RoboCup @Home competition was launched in 2006, and is now considered as the largest international annual competition of autonomous service robots. The purpose of this competition is to improve the integrated technology to home robots and to evaluate their ability to collaborate with human users in different tasks within a realistic non-standardized home environment setting like a kitchen or a living room.

2 Background

"Duckers" represents the team of the Emergent Systems Laboratory at Ritsumeikan University that participates in RoboCup @Home Domestic Standard Platform League, 2017. We participate at RoboCup Japan Open @Home Education, Open Platform, Simulation and Standard Platform leagues and Robo Cup @Home Domestic Standard Platform league. The aim of our team is to realize human-robot collaboration inspired by symbol emergence in robotics [1].

The motto of our team is "SER4U", which means Symbol Emergence in Robotics for you. Additionally, "4U" means Unique, Unpredictable, User-oriented, and Useful. The team's logo consists of a "wheel duck" robot, and "+R" which refers to Ritsumeikan University. The word "Ducker" represents the bird symbol of Shiga prefecture, where our university is located.

The team "Duckers" has undergraduate students, master's students, a PhD student, a posdoctoral research associate, and teaching fellows. "Duckers" has been participating in RoboCup Japan Open @Home since the 2015 tournament.

Fig. 1. RoboCup Japan Open 2015

Fig. 2. RoboCup Japan Open 2016

We won the 2nd place in RoboCup Japan Open @Home Education League and the 4th place in @Home Simulation League in 2015. In addition, we won the 3rd place in RoboCup Japan Open @Home Education League and the 7th place in @Home Open Platform League in 2016.

3 Hardware Architecture

Name of our HSR is Emma. Human Support Robot (HSR) developed by TOYOTA is a living support robot that assists handicapped and elderly people in everyday life. It has different advanced capabilities, such as: self-localization, recognition based failure avoidance, and motion planning. It can pick up objects placed on the floor and from shelves at different heights. HSR has four cameras and three sensors as shown in Fig. 3. The roles of cameras and sensors are described as follows:

- RGB-D sensor: It obtains color image and depth image. It can be used to recognize objects and environments.
- Wide-angle camera: It obtains a wide-angle view image of horizontal 135°.
- Stereo camera: It obtains the distance information of the object from the parallax of the left and right camera images.
- IMU sensor: It obtains acceleration and angular velocity of HSR.
- Force-Torque sensor: It obtains the force and torque acting on the wrist.
- Laser range sensor: It obtains the distance information of ambient environment with a horizontal scan angle 250°.

4 Software Architecture with Spatial Concept Model

4.1 Robot Software Description

Our system architecture is integrated by the Robot Operating System (ROS) as shown in Fig. 4. The robot can provide unique services by spatial concepts

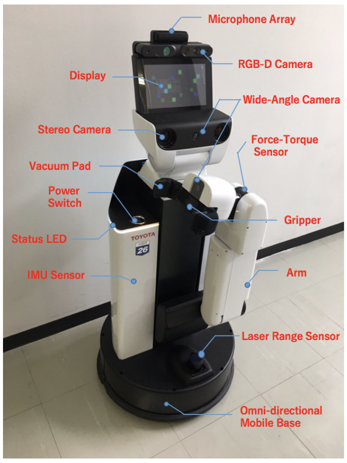

Fig. 3. Toyota Human Support Robot

formed from multimodal information obtained by the robot. Robot software consists of the following functions:

- Spatial concept formation
- Speech recognition and synthesizing
- Object recognition
- Object manipulation
- Navigation, localization and mapping

4.2 Spatial Concept Formation for @Home Tasks

Our main research focuses on spatial concept formation using autonomous robots through unsupervised machine learning [2–5]. In particular, we use a nonparametric Bayesian generative model. The spatial concept is a place category constructed on the basis of the robot's experiences. The robot learns spatial concepts from multimodal information such as self-positions, human speech signals, and image features as shown in Fig. 5.

Ishibushi et al. [2] proposed a computational model to estimate the spatial region of a place based on spatial distance and the distance of visual features. The study revealed that the accuracy of the self-localization is improved by classifying the object recognition result by using a Convolutional Neural Network (CNN) [6] and the self-position information by Monte Carlo localization (MCL). The experimental results showed that the method was able to converge particles for self-localization and to reduce estimation errors in the global self-localization.

Taniguchi et al. [3] proposed a method for simultaneously estimating self-position and words from noisy sensory information and utterances. Their method integrated ambiguous speech-recognition results with the self-localization method

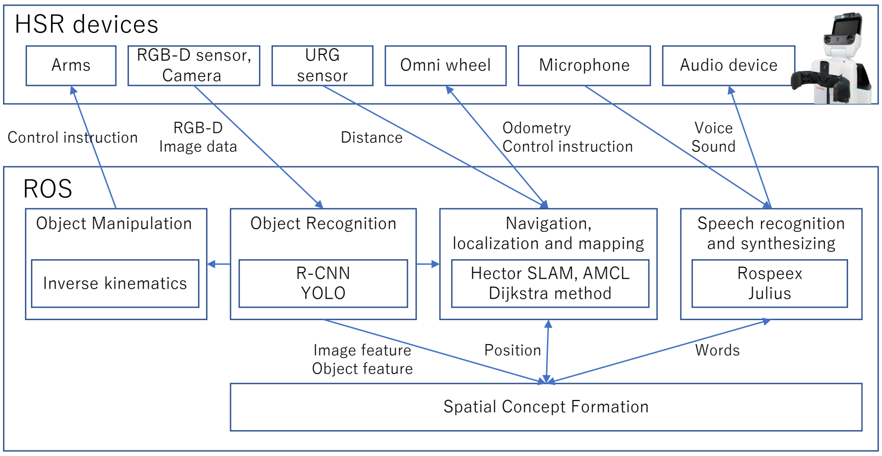

Fig. 4. System Architecture

for learning spatial concepts. Furthermore, Taniguchi et al. [4] proposed nonparametric Bayesian spatial concept acquisition method (SpCoA) based on an unsupervised word-segmentation method known as latticelm[1]. This method enables word segmentation with consideration of phoneme errors in speech recognition more efficiently than the nested Pitman–Yor language model (NPYLM) does.

Hagiwara et al. [5] proposed a model to infer the bottom-up hierarchical structure of places based on multimodal information such as position and visual information with hierarchical multimodal latent Dirichlet allocation (hMLDA). The experimental results demonstrated the formation of hierarchical place concepts by hMLDA.

As results of our research in @Home tasks, we succeeded that the robot acquires the spatial concept in the environment of RoboCup Japan Open 2016. Fig. 6 shows the example of formed spatial concepts that consist of images, occurrence probabilities of objects, and word probabilities. Based on this spatial concept, we demonstrated navigation to move the robot to the "Living room" by voice command. The goal of our research is to enables a robot to naturally communicate with human and provide services to support human daily life through the spatial concept.

4.3 Speech Recognition and Synthesizing

In the speech recognition task, the robot is required to hear, recognize, and respond to the human voice. We use the rospeex[2] for speech recognition. It is a

[1] latticelm: an unsupervised word-segmentation tool
`http://www.phontron.com/latticelm/`

[2] Rospeex: a cloud-based speech communication toolkit
`http://rospeex.org/top/`

Fig. 5. Overview of spatial concept formation based on multimodal information

cloud-based multilingual communication package for ROS. It has high recognition accuracy and can be easily implemented using its APIs in Python or C++. Rospeex receives a speech signal from the accompanying waveform monitor application. After noise reduction and speech segment detection, the speech signal is converted into text through the speech recognition engine on the cloud.

However, rospeex does not work without a network environment. If the communication line (WiFi-connection) is not stable, the recognition precision will degrade. Therefore, we will also use another speech recognition system Julius[3]. Julius is a high-performance open-source software of large vocabulary continuous speech recognition (LVCSR) for speech-related researchers and developers. We can perform speech recognition without network connection by using Julius, to prepare an acoustic model, a word dictionary, and a language model. These two types of speech recognition systems are used on a case-by-case basis.

4.4 Object Recognition

In order to perform the manipulation task, it is necessary to perform object recognition by using a camera attached to the robot. You Only Look Once (YOLO)[4] is used as object recognition algorithm in our team. YOLO is one of the CNN based object recognition methods. It simultaneously estimates bounding boxes and classes of objects from captured images as shown in Fig. 7. In the previous method such as R-CNN, the object classification was performed after

[3] Julius: Open-Source Large Vocabulary Continuous Speech Recognition Engine `https://github.com/julius-speech/julius`

[4] YOLO: a state-of-the-art, real-time object detection system `https://pjreddie.com/darknet/yolo/`

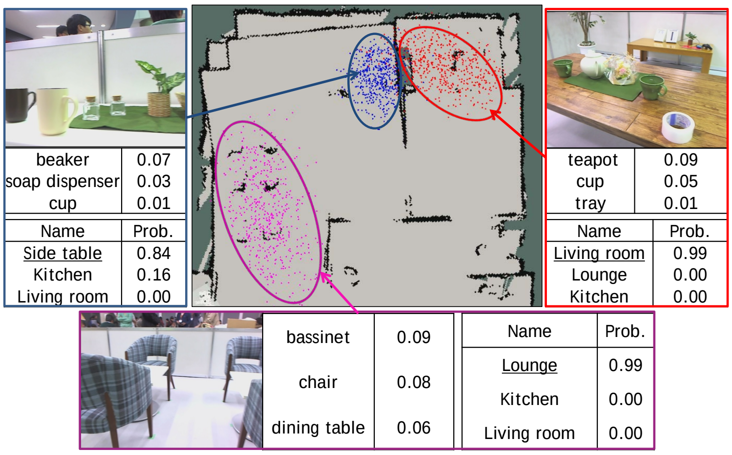

Fig. 6. Spatial concepts in RoboCup Japan Open 2016 @Home environment

Fig. 7. Object recognition by YOLO

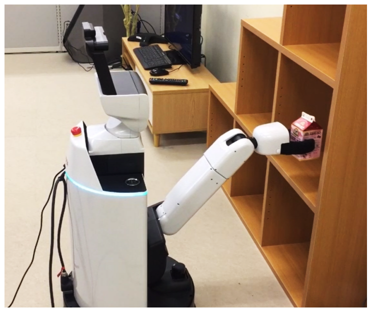

Fig. 8. Manipulation

detection of region proposal. R-CNN need to search region proposals of several thousand times on a single image. In contrast, YOLO can obtain classes and bounding boxes of objects by segmenting the entire image for each grid region. YOLO is sufficient to perform a single search on single image since the entire image is seen at the time of testing. Therefore, the detection accuracy of YOLO is slightly inferior to Faster R-CNN, but it is possible to achieve the detection speed that can be used for robots.

4.5 Object manipulation

In the manipulation task, the robot has to detects and recognizes objects placed on shelfs and tables, and to manipulate these objects. Our robot measures the three-dimensional position of the object based on the distance information obtained by the depth sensor and the object recognition result obtained by YOLO.

Then, the robot grasps the target object by inverse kinematics. In addition, the robot specifies the movement destination of the object to be grasped from the depth image and moves the object.

4.6 Navigation, localization and mapping

Many technologies such as map generation, localization, path planning and obstacle avoidance are required for the robot to autonomously move in the environment. Our team uses hector SLAM for map generation. Hector SLAM is one of the SLAM method without odometry information. Adaptive MCL (AMCL) is used for the localization method [7]. Dijkstra method is used for global planning and Dynamic Window Approach (DWA) algorithm is used for local planning such as dynamic obstacle avoidance [8].

5 Multimedia

5.1 Web Site

- RoboCup @Home Our Team Web Site
  `https://emlab.jimdo.com/`
- Emergent Systems Laboratory
  `http://www.em.ci.ritsumei.ac.jp/en`

5.2 Video

- Object recognition and Manipulation
  `https://youtu.be/EQHpGJbn40k`
- Navigation test
  `https://youtu.be/jJ-0CTEMbhY`
- Object manipulation based on spacial concept
  `https://youtu.be/5fblpisxLr4`
- Human robot interaction based on hierarchical spatial concept
  `https://youtu.be/0-xsfaaRZtY`
- Human life support based on spatial concept
  `https://youtu.be/dz3mncPVfH0`
- Transfer learning of spatial concept
  `https://youtu.be/bK77nOKYGW4`
- Service example based on spacial concept
  `https://youtu.be/nuUnWWu6Y28`
- Spacial concept learning
  `https://youtu.be/SHg_mDi4QrQ`

5.3 Download contents

We have released the following open datasets. These datasets were collected for learning the spatial concept. The datasets contain the robot's position and image information.

- Emergent Systems Lab. (Office environment)
  `https://github.com/hirokoba/Datasets-EmergentSystemLab`
- Experimental room 10 (Home environment)
  `https://github.com/hirokoba/Datasets-Experimental-room-10-`

5.4 Open Source Software

- Spacial Concept Formation [2]
  `https://github.com/EmergentSystemLabStudent/Spatial_Concept_Formation`
- Nonparametric Bayesian Double Articulation Analyzer (NPB-DAA) [9]
  `https://github.com/EmergentSystemLabStudent/NPB_DAA`

References

1. T. Taniguchi, T. Nagai, T. Nakamura, N. Iwahashi, T. Ogata, and H. Asoh, "Symbol emergence in robotics: A survey," *Advanced Robotics*, vol. 30, pp. 706–728, 2016.
2. S. Ishibushi, A. Taniguchi, T. Takano, Y. Hagiwara, and T. Taniguchi, "Statistical localization exploiting convolutional neural network for an autonomous vehicle," in *Industrial Electronics Society, IECON 2015-41st Annual Conference of the IEEE.* IEEE, 2015, pp. 1369–1375.
3. A. Taniguchi, T. Taniguchi, and T. Inamura, "Simultaneous estimation of self-position and word from noisy utterances and sensory information," in *13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*, 2016.
4. ——, "Spatial concept acquisition for a mobile robot that integrates self-localization and unsupervised word discovery from spoken sentences," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 4, pp. 285–297, 2016.
5. Y. Hagiwara, I. Masakazu, and T. Taniguchi, "Place concept learning by hmlda based on position and vision information," in *13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*, 2016.
6. A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
7. S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics.* MIT Press, 2005.
8. D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.
9. T. Taniguchi, S. Nagasaka, and R. Nakashima, "Nonparametric bayesian double articulation analyzer for direct language acquisition from continuous speech signals," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 3, pp. 171–185, 2016.

Addendum

– Team name : Duckers
– Contact information : isobe.shota@em.ci.ritsumei.ac.jp
– Website url : https://emlab.jimdo.com/
– Team members' names : Shota Isobe, Yuki Katsumata, Akira Kinose, Tomoro Ishimine, Yoshiki Tabuchi, Hiroyoshi Kobayashi, Takashi Fukui, Akira Taniguchi, Amir Aly, Yoshinobu Hagiwara and Tadahiro Taniguchi.
– Description of the hardware used : Toyota HSR (Human Support Robot)
– Brief, compact list of 3rd party reused software packages : ROS, rospeex and Julius.