KeJia: The Intelligent Domestic Robot for RoboCup@Home 2015

Xiaoping Chen, Wei Shuai, Jiangchuan Liu, Song Liu, Xiping Wang, Ningyang Wang, Dongcai Lu, Yingfeng Chen and Keke Tang

Multi-Agent Systems Lab., Department of Computer Science and Technology, University of Science and Technology of China, HeFei, 230027, China
xpchen@ustc.edu.cn, {swwsag, jkd, lszq, wxiping, wny257, ludc, chyf, kktang}@mail.ustc.edu.cn
http://wrighteagle.org/en/robocup/atHome

Abstract. This paper aims at reporting the recent progress of our intelligent robot KeJia, whose long-term goal is to integrate intelligence into a domestic robot. It covers the content ranging from hardware design, perception and high-level cognitive functions. All these techniques have been tested in former RoboCup@Home tests and other open demostrations.

1 Introduction

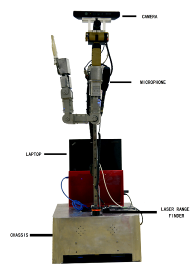
Remarkable progress has been made on research into intelligent robots, in particular, service robots. Also in recent years, there has been an increasing interest in integrating techniques drawn from areas of AI and Robotics [5, 6, 23, 21, 22, 8, 11, 13, 18]. Yet there are still challenges lying between the goal and reality. There are several essential abilities that a robot should have in order to make it intelligent and able to serve humans. Firstly, the robot should be able to perceive the environment through on-board sensors. Secondly, the robot has to independently plan what to do under different scenarios. Thirdly and most importantly, the robot is expected to be able to communicate with humans through natural languages, which is the core difference between service robots and traditional robots. As a result, developing an intelligent service robot requires huge amount of work in both advancing each aspect of abilities, and system integration of all such techniques.

The motivation of developing our robot KeJia is twofold. First, we want to build an intelligent robot integrated with advanced AI techniques, such as natural language processing [23], hierarchical task planning and knowledge acquisition [22, 21]. Second, by participating RoboCup@Home League, all these techniques could be tested in real-world like scenarios, which in return helps the development of such techniques. In previous RoboCup@Home competitions, our robot KeJia got two 2nd places in 2013 and 2011, respectively. Other demo videos are available on our website [1].

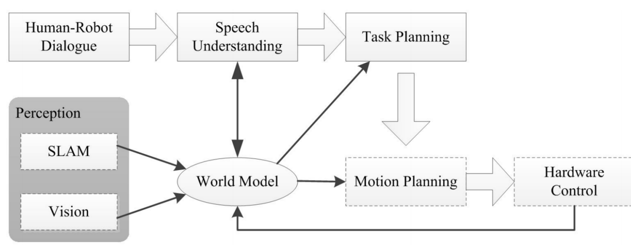
[1] http://wrighteagle.org/en/demo/index.php

In this paper, we present our latest research progress with our robot KeJia. Section 2 gives an overview of our robot's hardware and software system. The low-level functions for the robot are described in Section 3. Section 4 presents techniques for complicated task planning and Section 5 elaborates our approach to dialogue understanding. Finally we conclude in Section 6.

2 Hardware Design and Architecture



(a) The Robot KeJia



(b) Software architecture of KeJia

Fig. 1. The hardware and software architecture of KeJia

The KeJia service robot is designed to manipulate a wide range of objects within an indoor environment, and has shown its outstanding performance since RoboCup@Home 2012. Our robot is based on a two-wheels driving chassis of 5050 centimeters in order to move across narrow passages. A lifting system is mounted on the chassis, attached with the robot's upper body. Assembled with the upper body is a five degrees-of-freedom (DOF) arm. It is able to reach objects over 83 centimeters far from mounting point and the maximum payload is about 500 grams when fully stretched. The robot's power is supplied by a 20Ah battery, which guarantees the robot a continuous running of at least one hour. As for real-time perception needs, our robot is equipped with a Kinect camera, a high-resolution CCD camera, two laser sensors and a microphone. A working station laptop is used to meet the computation needs. The image of our robot KeJia is shown in Figure 1(a).

As for the software system, Robot Operating System (ROS)[2] has been employed as the infrastructure supporting the communication between modules in our KeJia robot. In general service scenarios, our robot is driven by human speech orders, as input of the robot's Human-Robot Dialogue module. Through the Speech Understanding module, the utterances from users are translated into the internal representations of the robot. These representations are in the form

[2] http://www.ros.org/wiki

of Answer Set Programming (ASP) language [12] which is a Prolog-like logical language. An ASP solver is employed in the Task Planning module to automatically make decisions given the translated results. The Task Planning module then generates the high-level plans for users' tasks. The generated course of actions is fed into the Motion Planning module. Each action is designed as a primitive for KeJia's Task Planning module and could be carried out by the Motion Planning module and then autonomously executed by the Hardware Control module. A figure describing the architecture is shown in Figure 1(b). In case of simple tasks or pre-defined ones, a state machine is used instead of the Task Planning module.

3 Perception

3.1 Self-Localization and Navigation

For self-localization and navigation, a probabilistic quadtree map is generated first from the raw data collected by laser scanners through a round travel within the rooms aforehand[9].

The probabilistic quadtree map is presented instead of traditional grids map which is used widely in robot mapping and localization filed. A quadtree is a wellknown data structure capable of achieving compact and efficient representation of large two-dimensional environments. The simulation experiments are conduct in *gazebo* [2], which is a well-designed simulator makes it possible to rapidly test algorithms, design robots, and perform regression testing using realistic scenarios. A manually created family environment is showed as Fig.2(a) ,and the furniture models are imported form *Google SketchUp* [1] and been adjusted to make them better suited to the gazebo.The remain pictures in Fig.2 are presented to validate the feasibility and reliability of the proposed quadtree mapping approach.

Then the map is manually annotated with the approximate location and area of rooms, doors, furniture and other interested objects. Finally, a topological map is automatically generated, which will be used by the global path planner and imported as a part of prior world model.With such map, scanning match and probabilistic techniques are employed for localization.

Moreover, VFH+ [19] is adopted to avoid local obstacles while the robot is navigating in the rooms. Frontier-based exploration strategy[24] and Gmapping[9] algorithm are used to explore unknown environment.

We also create the 3D environment representation using octo-tree structure[10], the system receive the point cloud information from the Kinect device, and then process the data with the localization provided by 2D grid map, eventually we get an effective and efficient 3d map, the map can be used in avoiding obstacles in all height and motion planning.

3.2 Vision

In our recognition system,two cameras are used, a high resolution CCD camera and a Microsoft Kinect,to obtain aligned RGB-D images as well as high quality

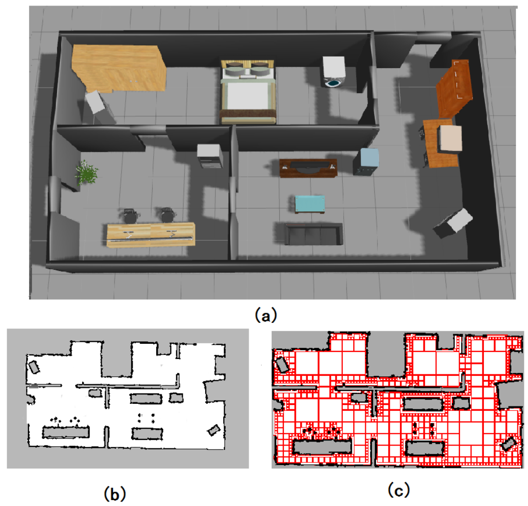


Fig. 2. (a) Simulate home scenarios in gazebo (17m*10m) (b) Grids map of the scenarios(resolution 0.05m) (c) Quadtree map of the scenarios

RGB images. Both cameras are calibrated so we can directly get the correspondence between the images. We obtain an aligned RGB-D image by combining the RGB image with the depth image. With such aligned RGB-D image, our vision module is capable of detecting, tracking people and recognizing different kinds of objects.

People Awareness We developed a fast walking people detection method to efficiently detect standing or walking people.The depth image is transformed into the robot's reference frame. Since human will occupy a continuous and almost fxed-size space, we segment the point cloud into multiple connected-components,and analyze the shape of each component based on the relative distance between pixels. Each candidate is then passed into a pre-trained HOD [17] upper body detector to decide whether it is human or not.We use HAAR [20] face detector from OpenCV [4] to detect and localize human face. If present, the VeriLook SDK is used to identify each face.

Object Recognition We follow the approach as proposed in [16] to detect and localize table-top objects including bottles, cups,etc. The depth image is first transformed and segmented, then the largest horizontal plane is extracted using Point Cloud Library(PCL) [15], and point clouds above it are clustered into different pieces. After that the SURF feature matching against the stored features are applied to each piece [3]. The one with the highest match above a certain threshold is considered as a recognition. At last, to further enhance the detection performance and decrease FP rate, we check each recognized cluster and filter out those vary too much in size. Recognition result is shown in Figure 3.

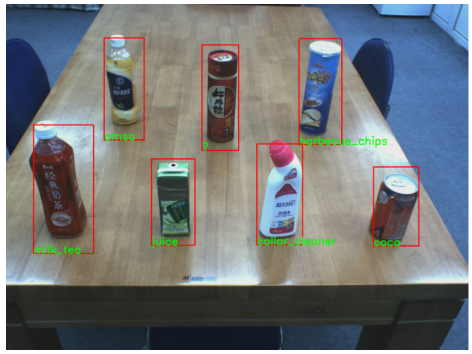


Fig. 3. Object Recognition of object instance

4 Task Planning

One of the most challenging tests in the RoboCup@Home competition is GPSR, where a robot is asked to fulfill multiple requests from an open-ended set of user tasks. This ability is generally required for real-world applications of service robots. We are trying to meet this requirement by developing a set of techniques that can make use of open knowledge, i.e., knowledge from open-source knowledge resources, including the Open Mind Indoor Common Sense (OMICS) database, whose knowledge was input by Internet users in semi-structured English. This section provides a brief report on this effort.

In the KeJia project, the Task Planning module is implemented using Answer Set Programming (ASP), a logic programming language with Prolog-like syntax under stable model semantics originally proposed by Gelfond & Lifschitz (1988). The module implements a growing model $M =< A; C^*; P^*; F^* >$, the integrated decision-making mechanism, and some auxiliary mechanisms as an ASP program M^{Π}. The integrated decision making in M is then reduced to computing answer sets of M^{Π} through an ASP solver. When the robots Dialogue Understanding module extracts a new piece of knowledge and stores it into M, it will be transformed further into ASP-rules and added into the corresponding part of M^{Π}.

5 Dialogue Understanding

The robot's Dialogue Understanding module for Human-Robot Interaction contains Speech Recognition module and Natural Language Understanding module,it provides the interface for communication between users and the robot.

For speech synthesis and recognition, we use a software from iFlyTek [3]. It is able to synthesis different languages including Chinese, English, Spanish etc. As for recognition, a configuration represented by BNF grammar is required. Since

[3] http://www.iflytek.com/en/index.html

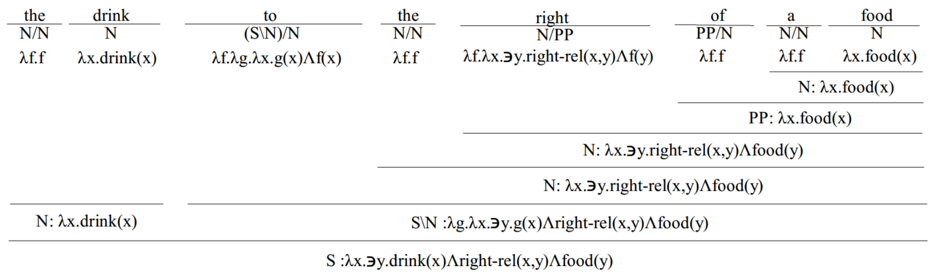


Fig. 4. Example parse of "the drink to the right of a food." The first row of the derivation retrieves lexical categories from the lexicon, while the remaining rows represent applications of CCG combinators.

each test has its own set of possible speech commands, we pre-build several configurations to include all the possible commands for each test.

The Natural Language Understanding module is used for the translation to its semantic representation.With the results of Speech Recognition module and the semantic information of the speech,the Natural Language Understanding module is able to update the World Model,which contains the information from the perceptual model of the robot's internal state,and/or to invoke the Task Planning module for fulfilling the task.

The translation from the results of the Speech Recognition module to semantic representation consists of the syntactic parsing and the semantic interpretation. For the syntactic parsing,we use the Stanford parser [7] to obtain the syntax tree of the speech.For the semantic interpretation ,the lambda-calculus [14]is applied on the syntax tree to construct the semantics.Fig.1 shows an example of semantic interpretation.

6 Conclusion

In this paper we present our recent progress with our intelligent service robot KeJia. Our robot is not only capable of perceiving the environment, but also equipped with advanced AI techniques which make it able to understand human speech orders and solve complex tasks. Furthermore, through automated knowledge acquisition, KeJia is able to fetch knowledge from open source knowledge base and solve tasks it has not met before

Acknowledgements

This work is supported by the National Hi-Tech Project of China under grant 2008AA01Z150, the Natural Science Foundations of China under grant 60745002, 61175057, and USTC 985 project.

Bibliography

[1] 3D models sources. http://www.sketchup.com/. [Online; accessed 19-Jan-2015].

[2] Simulation tool. http://gazebosim.org/. [Online; accessed 19-Jan-2015].

[3] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features. 110(3):346–359, 2008.

[4] G. Bradski. The opencv library. 2000.

[5] W. Burgard, A. Cremers, D. Fox, D. Hahnel, G. Lakemeyer, and Schulz. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1):3–55, 1999.

[6] R. Cantrell, M. Scheutz, P. Schermerhorn, and X. Wu. Robust spoken instruction understanding for HRI. In *Proceedings of the 5th ACM/IEEE International Conference on Robot Interaction*, 2010.

[7] D.Klein and C.Manning. Accurate unlexicalized parsing. *In Proceedings of the 41st Annual Meeting on Association for Computational Linguistics(ACL-03)*, pages 423–430, 2003.

[8] J. Dzifcak, M. Scheutz, C. Baral, and P. Schermerhorn. What to do and how to do it: Translating natural language directives into temporal and dynamic logic representation for goal management and action execution. In *IEEE International Conference on Robotics and Automation*, pages 4163–4168. ICRA, 2009.

[9] G. Grisetti, C. Stachniss, and W. Burgard. Improved techniques for grid mapping with rao-blackwellized particle filters. *Robotics, IEEE Transactions on*, 23(1):34–46, 2007.

[10] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard. Octomap: an efficient probabilistic 3d mapping framework based on octrees. *Auton. Robots*, 34(3):189–206, 2013.

[11] T. Kollar, S. Tellex, D. Roy, and N. Roy. Toward understanding natural language directions. In *5th ACM/IEEE International Conference on Human-Robot Interaction*, 2010.

[12] V. Lifschitz. Answer set planning. In *Proceedings of the 1999 International Conference on Logic Programming (ICLP-99)*, pages 23–37, 1999.

[13] D. Nyga and M. Beetz. Everything robots always wanted to know about housework (but were afraid to ask). In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

[14] P.Blackburn and J.Bos. Representation and inference for natural language: A first course in computational semantics. *CSLI Publications*, 2005.

[15] R. B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2011.

[16] R. B. Rusu, A. Holzbach, G. Bradski, and M. Beetz. Detecting and segmenting objects for mobile manipulation. In *Proceedings of the 12th IEEE*

International Conference on Computer Vision: Workshop on Search in 3D and Video, 2009.

[17] L. Spinello and K. O. Arras. People detection in rgb-d data. In *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3838–3843, 2011.

[18] S. Tellex, T. Kollar, S. Dickerson, M. Walter, A. Banerjee, S. Teller, and N. Roy. Understanding natural language commands for robotic navigation and mobile manipulation. In *Proceedings of National Conference on Articial Intelligence*, 2011.

[19] I. Ulrich and J. Borenstein. Vfh+: reliable obstacle avoidance for fast mobile robots. In *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, volume 2, pages 1572–1577 vol.2, 1998.

[20] P. Viola and M. Jones. Rapid object detection using a boosted cascade of sim-ple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 511–518 vol.1, 2001.

[21] J. J. X. Chen, J. Xie and Z. Sui. Toward open knowledge enabling for humanrobot interaction. *Journal of Human-Robot Interaction*, 2(1):100–117, 2012.

[22] J. J. G. J. X. Chen, J. Jiang and F. Wang. Integrating nlp with reasoning about actions for autonomous agents communicating with humans. In *Proceedings of the 2009 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pages 137–140, Hefei, China, 2009.

[23] J. J. G. J. F. W. Xiaoping Chen, Jianmin Ji and J. Xie. Developing high-level cognitive functions for service robots. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, pages 989–996, Hefei, China, 2010. AAMAS.

[24] B. Yamauchi. A frontier-based approach for autonomous exploration. In *Computational Intelligence in Robotics and Automation, 1997. CIRA'97., Proceedings., 1997 IEEE International Symposium on*, pages 146–151, 1997.