WrightEagle@Home 2016 Team Description Paper

Wei Shuai, Jiangchuan Liu, Xiping Wang, Feng Zhou and Xiaoping Chen

Multi-Agent Systems Lab., Department of Computer Science and Technology,
University of Science and Technology of China, HeFei, 230027, China
{swwsag, jkd, wxiping, zf062}@mail.ustc.edu.cn, xpchen@ustc.edu.cn,
http://wrighteagle.org/en/robocup/atHome

Abstract. This paper aims at reporting the recent development and progress of our intelligent robot KeJia, whose long-term goal is to integrate intelligence into a domestic robot. The research issues range from hardware design, perception and high-level cognitive functions of service robots. All these techniques have been tested in former RoboCup@Home tests and other case studies.

1 Introduction

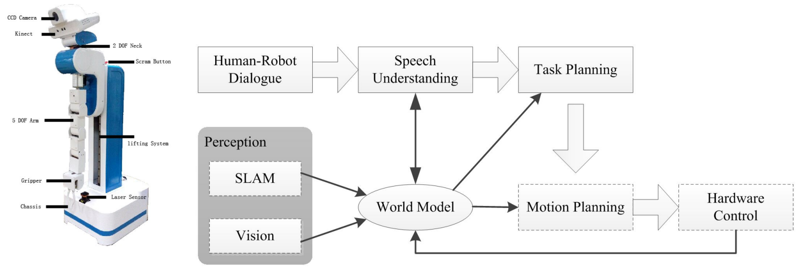
More and more researchers in Robotics and AI are showing their interest in intelligent robots. Research on intelligent service robots, which aims to fulfill a fundamental goal of Artificial Intelligence, is drawing much more attention than ever. Yet there are still challenges lying between the goal and reality. There are several essential abilities that a robot should have in order to make it intelligent and able to serve humans automatically. Although traditional robots which lacks intelligence and automation could serve human in some circumstances, robots with new characteristics which we would described later would do a much better job. Firstly, the robot should be able to perceive the environment through its onboard 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 a 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, hierarchical task planning[4] and knowledge acquisition[5]. Second, by participating in RoboCup@Home League, all these techniques could be tested in real-world like scenarios, which in return helps the development of such techniques. In the RoboCup@Home 2015 competition, our robot KeJia got 2nd place. Other demo videos are available on our website[1].

[1] http://wrighteagle.org/en/robocup/atHome

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 D2

(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 stable performance since RoboCup@Home 2012. Our robot is based on a two-wheels driving chassis of 62*53*32 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 six 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 computational needs. The image of our robot KeJia is shown in Fig. 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/

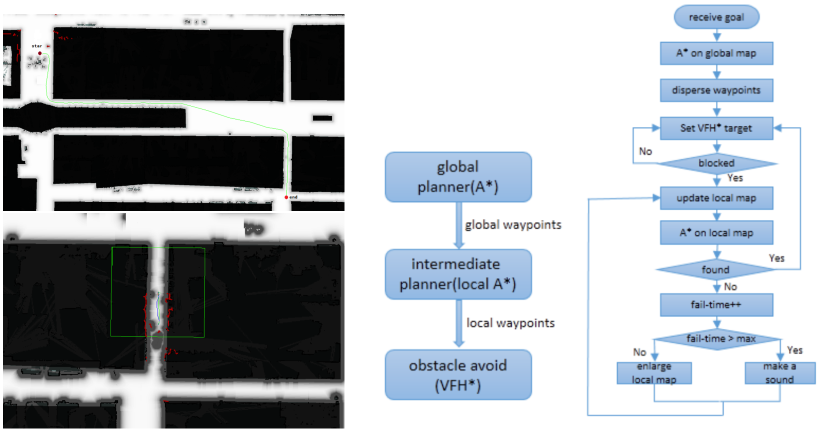
of Answer Set Programming (ASP) language[11] which is a Prolog-like logical language. An ASP

solver is employed in the Task Planning module to automati-cally 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 Fig. 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 2D occupancy grid map is generated first from the raw data collected by laser scanners through a round travel within the rooms aforehand[8]. 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, scan matching and probabilistic techniques are employed for localization. Besides the 2D grid map, we also create the 3D environment representation with Kinect using octree structure[10]. The system receives the point cloud information from



(a) Global path and path planned by in-termediate level

(b) Navigation structure and flow chart

Fig. 2. Navigation

the Kinect device and then process the data with the localization provided by 2D grid map. However, the drawbacks of the point cloud map are that the sensor noise and dynamic objects can't be handled directly and that it is not convenient to integrate with the navigation module. Fortunately, these two kinds of maps created by different techniques in the unified coordinate system, in normal conditions, are matched well with each other. Both maps describe the primary objects that rarely change in the environment, and the local map of navigation will be copied from the static maps and updated with the sensor data. The unified coordinate system with two maps can be used in avoiding obstacles in all height and motion planning.

In the previous researches, the navigation module combined global planner and local planner was introduced, global path replanning will be triggered when robot traps in local dilemma. This idea is not fully applicable for us since it may lead robot repeatedly alter its route in vain. Closer analysis, however, the main trouble with this method is that all things may influent robot's motion are treated as barrier, which means robot can only try to avoid obstacles but not communicate with them while the obstacles are human or other agents. In reality, it would not be the best choice for robot to replan every time for the following reasons.

- The robot may move back and forth between two blocked alleys frequently without progress.
- Refinding a global path on the whole map is time-consuming.
- Making a long detour sometimes is expensive than just waiting for a while.

In order to eliminate this disharmony between global and local planner, a intermediate layer is employed[6]. Once a goal is received, Firstly, the path from the robot's position to goal is computed. Next, a serial of ordered way points are generated from the global route, then the way points will be sequentially dispatched to the local planner which will find a local path for the well-tuned VFH* module to track. If the local planner fails to find a suitable path, the local map would continue enlarging until a maximum limit is reached. After several failures, robot will demand the crowd to give way, if all these attempts fail, a global replan happens. This approach endows the robot ability of adapting environments, meanwhile, reduces the unnecessary global path plan (shown in Fig. 2).

3.2 Vision

In our recognition system, two cameras are used, a high resolution CCD camera (1920*1440) and a Microsoft Kinect, to obtain aligned RGB-D images as well as high quality 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 an 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 robots reference frame. Since human will occupy a continuous and almost fixed-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[13] upper body detector to decide whether it is human or not. We use HAAR[14] face detector from OpenCV[3] to detect and localize human face. If present, the VeriLook SDK is used to identify each face.

Object Detection and Recognition We follow the approach as proposed in [12] to detect and localize table-top objects including bottles, cups, etc. The depth image is first transformed, then the largest horizontal plane is extracted using Point Cloud Library (PCL)[12], and point clouds above it are clustered into different pieces. At last, to further enhance the detection performance and decrease FP rate, we check each detection cluster and filter out those vary too much in size. A traditional resolution for object recognition is to extract SURF feature[1] from the detected region. The SURF feature matching against the stored features are applied to each region. The one with the highest match above a certain threshold is considered as recognition. However, there are several problems:

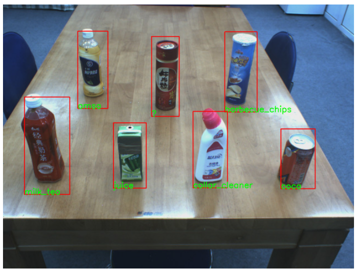


Fig. 3. Object recognition

- This method cannot satisfy real-time requirements. The detection method in [12] is categories independent, which cannot provide any information for instances recognition. As a result, the SURF feature extracted from each detected region need to be matched with every stored feature of every object. This will consume much time. So we need an example-based object detection method.
- There are limitations for one single feature. For example, SURF feature cannot deal with texture-less objects, or illumination change. So we need combine multi-features.
- There are limitations for depth image based detection method. For example, transparent objects in depth images are missing data.

To address these above problems. We developed a real-time combining multi-features objects recognition system. Three kinds of features are used in this system, which including gradients feature in object contours, HSV histogram, SURF feature. At first, line-mod[9] detects these possible object candidates contour in RGB-D images with multimodal templates. As objects with similar shape have similar contour gradient feature, there are many error detection results generated by the line-mod detector. HSV color histogram is a simple but useful feature to distinguish some of them. At second step , HSV histogram are used to exclude some error candidates. At last, SURF feature matching against the rest of these candidates are applied to the detected region. Recognition result is shown in Fig. 3.

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). ASP provides a unified mechanism of handling commonsense reasoning and planning with a solution to the frame problem. There are many ASP program solver tools which can produce a solution to an ASP program. The solver KeJia uses is iclingo.

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 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

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

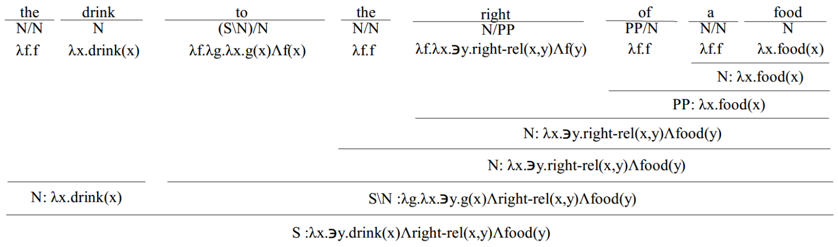

the
N/N
λf.f
drink
N
λx.drink(x)
to
(S\N)/N
λf.λg.λx.g(x)∧f(x)
the
N/N
λf.f
right
N/PP
λf.λx.∃y.right-rel(x,y)∧f(y)
of
PP/N
λf.f
a
N/N
λf.f
food
N
λx.food(x)
N: λx.food(x)
PP: λx.food(x)
N: λx.∃y.right-rel(x,y)∧food(y)
N: λx.∃y.right-rel(x,y)∧food(y)
N: λx.drink(x)
S\N :λg.λx.∃y.g(x)∧right-rel(x,y)∧food(y)
S :λx.∃y.drink(x)∧right-rel(x,y)∧food(y)

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.

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 [2] is applied on the syntax tree to construct the semantics.Fig. 4 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 bases and solve tasks it has not met before.

Acknowledgement

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, USTC 985 project and the core direction project of USTC. Other team members beside the authors are: Zekun Zhang, Guangda Chen, Guowei Cui, Bing Tang, and Lan Lin.

Bibliography

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

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

[3] G. Bradski. The opencv library. In *Dr. Dobb's Journal of Software Tools*, 2000.

[4] X. Chen, J. Ji, J. Jiang, G. Jin, F. Wang, 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, 2010.

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

[6] Y. Chen, F. Wang, W. Shuai, and X. Chen. Kejia robot-an attractive shopping mall guider. In *ICSR*, 2015.

[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] 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.

[9] S. Hinterstoisser, C. Cagniart, S. Holzer, S. Ilic, K. Konolige, N. Navab, and V. Lepetit. Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes. In *Proc. IEEE Int'l Conf. Computer Vision.*

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

[12] 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.

[13] I. Ulrich and J. Borenstein. People detection in rgb-d data. In *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, page 38383843, 2011.

[14] P. Viola and M. Jones. People detection in rgb-d data. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 511518, 2001.

Table 1. Haedware overview of the robots

	C2	D2
Name	C2 For KeJia Series	D2 For KeJia Series
Base	Two-wheels driving chassis	Two-wheels driving chassis
Manipulators	5 degrees-of-freedom (DOF) arm	5 degrees-of-freedom (DOF) arm
Neck	2 degrees-of-freedom (DOF) neck	2 degrees-of-freedom (DOF) neck
Head	PointGrey HD Camera Kinect for XBox 360 Kinect2.0 for XBox 360	PointGrey HD Camera Kinect for XBox 360
Additional sensors	Sound Localization Modules by iFLYTEK	Sound Localization Modules by iFLYTEK
Dimensions	Base: 0.5m x 0.5m Height: 1.7m	Base: 0.45m x 0.45m Height: 1.7m
Weight	80kg	75kg
Microphone	MAKAD EN-8800 SUPER	MAKAD EN-8800 SUPER
Batteries	1x Lithium battery 24 V, 20 Ah 1x Lithium battery 20 V, 20 Ah	1x Lithium battery 24 V, 20 Ah 1x Lithium battery 20 V, 20 Ah
Computers	ThinkPad w530 PC	ThinkPad w530 PC

Table 2. Software overview of the robots

Operating system	Ubuntu 14.04 LTS Desktop
Middleware	ROS Indigo
Localization	Monte Carlo using ED Scan matching
SLAM	Gmapping *http://wiki.ros.org/gmapping*
Navigation	Global: A* planner Local: vfh*
Object Recognition	HSV Linemod SURF
People detection and Face detection	OpenCV *http://opencv.org/* PCL *http://pointclouds.org/*
Face recognition	VeriLook SDK *http://www.neurotechnology.com*
Natrual Language Understanding	The Stanford Parser *http://nlp.stanford.edu/software/lex-parser.shtml*
Task planner	Answer Set Programming(ASP)
Speech recognition	iFLYTEK
Speech recognition	*http://www.iflytek.com/en/audioengine/list_3.html*
Speech synthesis	iFLYTEK
Speech synthesis	*http://www.iflytek.com/en/audioengine/list_4.html*