WrightEagle@Home 2017 Team Description Paper

Jiangchuan Liu, Zekun Zhang, Bing Tang and Xiaoping Chen

Multi-Agent Systems Lab., Department of Computer Science and Technology,
University of Science and Technology of China, HeFei, 230027, China
{jkd, kzz, cgdsss,tb9527}@mail.ustc.edu.cn, xpchen@ustc.edu.cn,
http://ai.ustc.edu.cn/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

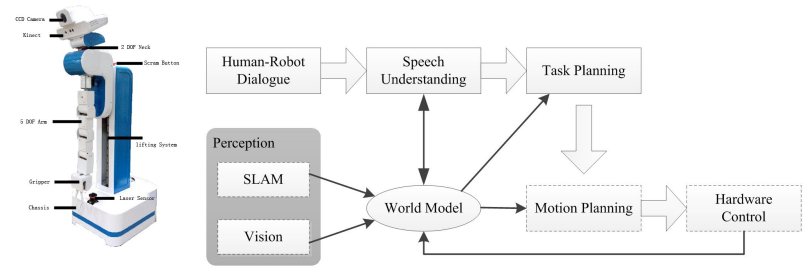
More and more researchers in Robotics and AI are showing their interest in intelligent robots. Research on intelligent service robots, which aims to fulfill a fundamental goal of Artificial Intelligence, is drawing much more attention than ever. Yet there are still challenges lying between the goal and 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[3] and knowledge acquisition[4]. 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 2016 competition, our robot KeJia got 3rd place. Other demo videos are available on our website[1].

[1] http://ai.ustc.edu.cn/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 4. Section 5 presents techniques for complicated task planning and Section 6 elaborates our approach to dialogue understanding. Finally we conclude in Section 7.

2 Hardware Design and Architecture



(a) The Robot KeJia E2

(b) Software architecture of KeJia

Fig. 1. The hardware and software architecture of KeJia

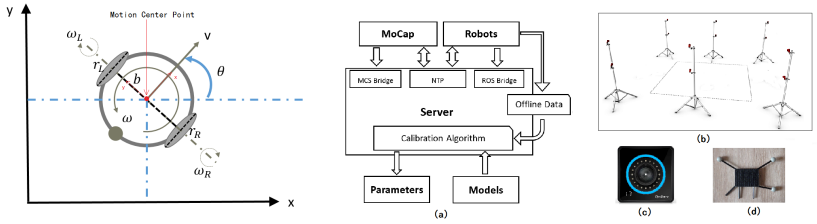
The KeJia service robot is designed to manipulate a variety of objects within an interior environment and has proved its stable performance since RoboCup@Home 2012. Our robot is based on a two-wheels driving chassis of 62*53*32 centimeters so that it can move across narrow passages. The lifting system is mounted on the chassis attached with the robots upper body. A six degrees-of-freedom (DOF) arm is assembled with the upper body. 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 robots 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[10] 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 Fig. 1(b). In case of simple tasks or pre-defined ones, a state machine is used instead of the Task Planning module.

3 Calibration



(a) The structure of differential-driven wheels

(b) Mocap system

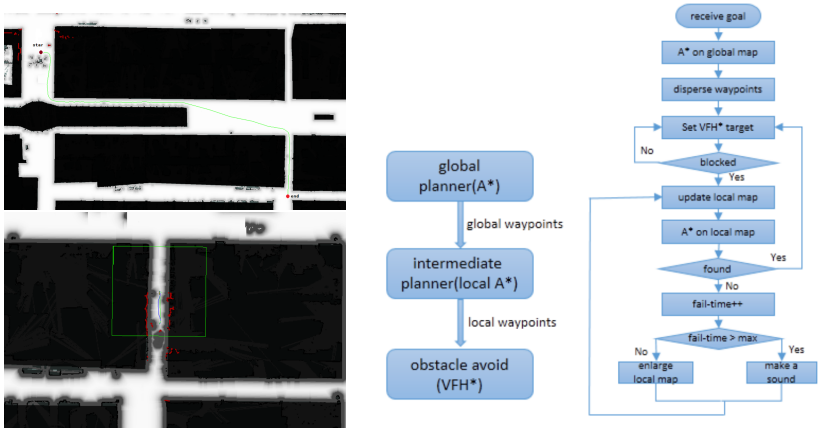
Fig. 2. Calibration

A good calibration is a prerequisite for robot KeJia to perform precise tasks. For example, it is helpful for Kejia to have an accurate odometry model(raddi and distance between the wheels) and a precise pose of the laser in self-localization and navigation task. KeJia need a good hand eye calibration to perform the manipulation task. For calibration of odomety and sensor parameters we follow the approach as proposed in [2]. This method does not require the robot to move along particular trajectories and experimental results show the accuracy of the method is very close to the attainable limit given by the Cramér-Raobound. For calibration of camera pose relative to the base-link of robot. we use MCS(motion capture system) to calculate the motion center point as shown in Fig. 2. Firstly, we command the robot spin on the spot, assuming that the robot's center point is fixed during the operation, thus we could get the radius and the circle center of the trajectory, then we drive the robot forward along the direction of its x axis, and we can determine its base-link axis. Then we use the method proposed in [12] to calculate the pose of camera relative to base-link.

4 Perception

4.1 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[7]. 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



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

(b) Navigation structure and flow chart

Fig. 3. Navigation

2D grid map, we also create the 3D environment representation with Kinect using octree structure[9]. The system receives the point cloud information from 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[5]. 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. 3).

4.2 Vision

In our robot vision system, a Microsoft Kinect and a high resolution (1920×1440) CCD camera are used simultaneously to obtain both aligned RGB-D images and high quality RGB images. Two cameras are calibrated to establish transformation between the corresponding points. By combining the information from both kinds of images, our vision system is capable of human detecting and tracking, as well as recognition of different categories of objects.

People Awareness We developed a point cloud based method to efficiently detect both still and fast-moving humans. Since each human occupies a continuous and almost fixed-size space, we segment the corresponding point cloud into multiple connected components. Then we can measure the shape of each component according to the relative distance between pixels. Candidates are transferred into a pre-trained HOD[13] upper body detector to determine whether they are humans or not. We use HAAR[14] face detector from OpenCV[1] to detect human faces. After faces are detected, VeriLook SDK and Microsoft Cognitive Services are used to identify each of them.

Object Detection and Recognition We follow a proposed approach[11] to detect and locate tabletop objects such as bottles, cups, boxes, etc. First the largest horizontal plane is extracted from point cloud using Point Cloud Library (PCL)[11]. Then point clouds above the plane are clustered into different pieces as candidates of objects. To improve the detection performance and reduce false positive rate, clusters with abnormal sizes are filtered out.Three kinds of features are used in this system: gradients feature in object contours, HSV color histogram, and SURF feature. First, we use LINEMOD[8] to detect the contours of object candidates in RGB-D images with multi-modal templates. Matching based on this feature is unreliable, since objects with similar shapes have similar contour gradient features. Thus HSV color histograms are used to reduce false recognitions. After that SURF feature matching is applied to the rest of the candidates to get the final recognition results. One example of objects recognition is shown in Fig. 4.

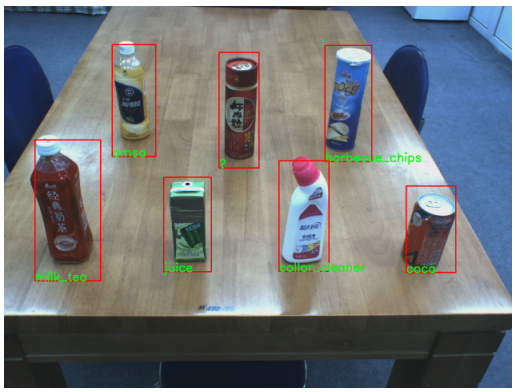


Fig. 4. Object recognition

We also use a joint model[13] to deal with occlusion and small objects when detect objects. The joint model integrates scene classification, object and visual phrase detection, as well as scene parsing together. By encoding them into a Conditional Random Field model, occlusion and small objects could be solved jointly.

Door and Handle Detection Using cloud data and color image of Kinect, we designed a method to locate a door. We use hough transform to get the lines which may belong to a door, and filter out lines which are too short or two long. The line segments are combined into rectangular regions which may contain a door handle.

5 Task Planning

One of the most challenging tests in the RoboCup@Home competition is GPSR(General Purpose Service Robot). In this test the robot needs to perform the task according to the user's request. To meet this requirement, we have developed a set of techniques. And one of the most important modules is task planning.

For the purpose of allowing an autonomous robot to use task instructions for task planning, we present a formalization for specifying structured task instructions and provide an approach for integrating these instructions with

robot?s built-in knowledge to compute plans for open-ended tasks. Our planning approach is based on McCain and Turner's causal theories(McCain and Turner, 1997). The main idea is to specify the action domain and the planning problem in causal theories such that a causal model corresponds to a planning solution, then use sophisticated AI tools or solvers to compute robot plans for tasks. And we choose a logic programming language named Answer Set Programming(ASP)(Baral, 2003) for the calculation of causal theories and an efficient ASP solver iclingo (Gebser et al., 2008) is used for computing task plans. Because they provide a unified mechanism of handling commonsense reasoning and planning with a solution to the frame problem.

In the task planing module, an extended task planning problem Δ is firstly converted to the causal theory *tr(Δ)*, then *tr(Δ)* is converted to its corresponding ASP program whose answer sets are computed by a sophisticated ASP solver iclingo. In our implementation, the Dialogue Understanding module of our robot extracts a new piece of knowledge, then it will be transformed into ASP-rules and added into the corresponding part of ASP program. At last, plans of Δ are extracted from the computed answer sets.

6 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 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 [6] to obtain the syntax tree of the speech. For the semantic interpretation, the lambda-calculus is applied on the syntax tree to construct the semantics.Fig. 5 shows an example of semantic interpretation.

7 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

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

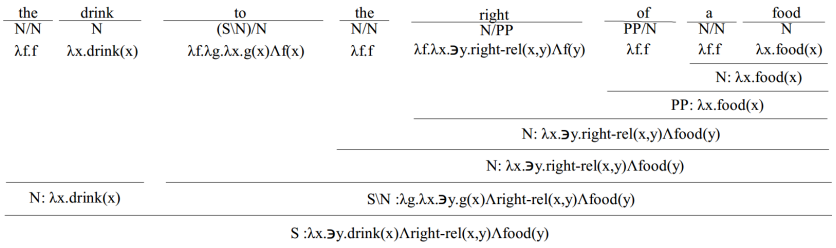


Fig. 5. Example parse of "the drink to the right of a food." The first row of the derivation retrieves lexical categories from the lexicon, while the remaining rows represent applications of CCG combinators.

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: Guangda Chen, Zhao Zhang, Zhongxiao Jin, Peichen Wu, and Lan Lin.

Table 1. Hardware overview of the robots

	C2	E2
Name	C2 For KeJia Series	E2 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	Sound Localization Modules
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

Table 2. Software overview of the robots

Operating system	Ubuntu 14.04 LTS Desktop
Middleware	ROS Indigo
SLAM	Gmapping
	http://wiki.ros.org/gmapping
Face recognition	VeriLook SDK
	http://www.neurotechnology.com
	Microsoft Face API
	https://www.microsoft.com/cognitive-services/en-us/face-api
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*

Bibliography

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

[2] A. Censi, A. Franchi, L. Marchionni, and G. Oriolo. Simultaneous calibration of odometry and sensor parameters for mobile robots. *IEEE Transactions on Robotics*, 29(2):475–492, 2013.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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*.

[9] 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.

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

[11] 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.

[12] Y. C. Shiu and S. Ahmad. Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax= xb. *ieee Transactions on Robotics and Automation*, 5(1):16–29, 1989.

[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.