AUT @Home 2014 Team Description Paper

Ameneh Sheikhjafari, Shaghayegh Gharghabi, Ismile Mehrabi, Marzieh Kariminoori, kourosh Sartipi, Alireza Sheikhjafari, Edwin Babaians, Sara Jahangiri, Ramin Gomari and Saeed Shiry Ghidary

Robotics Center, Amirkabir University of Technology (Tehran Polytechnic),
Tehran, Islamic Republic of Iran
http://autathome.ir

Abstract. In this paper we introduce the AUT@Home robot of Amirkabir University of Technology in Tehran, Iran for the 2014 RoboCup@Home competition. We describe the hardware characteristics and capabilities of this robot to perform tasks. In addition it covers our approaches for NI tasks such as speech and object recognition which has made AUT reliable for different tasks in home environments.

Keywords: AUT@home, @Home Robot, RoboCup, TDP.

1   Introduction:

AUT@home team consists of MSc and BSc students from different department of different universities. First participation of our team was in world RoboCup competition in 2008 in which we reached the finals and got the fifth place in the @home league. We participated in this competition for three consecutive years. At 2013 we restarted our work again by building a new robot. We participate in AutRobocup2013 competition with our young robot and ranked first in this competition. We consider to build a new robot with high performance. Also we want to improve our current robot to attend in world competition.

AUT@home robot is equipped with Kinect sensor and laser scanner for 3D perception and we use data gathered from these scanners for navigation, real time environment perception, object recognition and manipulation. This paper first gives a description of mechanical and electrical design of our robot then in the third section we describe its software platform and finally we end it with future work.

2   Hardware Design

2.1   Mechanical Part of AUT@Home Robot

The mechanical part of our robot consists of a mobile robot platform and manipulators. The upward of robot contains head and two arms which are attach to the shoulder. In this robot we tried to build arms to resemble human arms and properties of human body to enable us simulating human abilities. In each arm there is one griper that is used to grip, pick and place any object. Each arm without their wrist have two degrees of freedom. (See figure 1)

Fig. 1. AUT@Home

In order to compensate for missing torso pitch joint, shoulder of our robot can be adjusted with a prismatic joint that can move up and down in z axis. For shoulder movement, we used rack and pinion system that can make a safe and accurate movement.

Actuators.

For driving the joints and wheels of our robot we use Dynamixel motors [1] that are manufactured by Robotis Co. The Dynamixel motors have different types, sizes and specifications; in this robot we use Ax, Rx and Mx series in upward and wheels.

This motors have two modes (joint, wheel) therefore the operator can use them in any needed places. For more performance and energy efficiency, our elbow, shoulder pitch and waist motors are more powerful than the other joints. Our robot's head which contains a Kinect sensor will be held by 2 servo motors as a Pan-Tilt mechanism. We employ the strongest motor, one RX-64 (6.4Nm holding torque, 116 g) in elbow, shoulder, pitch and waist.

Motion System.

The motion system of our robot consists of three wheel Omni drive system. Because of the Holonomic advantage of Omni drive systems, they can be more useful in @Home robots field than other systems such as differential or other wheeled systems.

2.2   Electronic & Sensors

For more compatibility and sensor fusion, we used an AVR® microcontroller DX-IO board (DX-IO controller runs under 8bit AVR ATMega8) as a low level controller and Device Communication Manager (DCM) in our robot. Different instruments such as actuators and sensors communicate with main processor by this board. In low-level computation on this board, we drive 3 types of different sensors:

- A 3DOF magnetometer RM-G146 sensor.

- Integrated distance infrared based on GP2D120 sensors.

- Internal actuators for load, speed and absolute position sensors of Dynamixels.

In addition, we also use USBzDXL (is a high speed USB to half-duplex serial) as a direct motor controller for Dynamixels actuators from main processor. This enables more than 50Hz update rate for high resolution data of internal sensors. Fig.3 shows low-level controller and main controller and peripheral connected device.
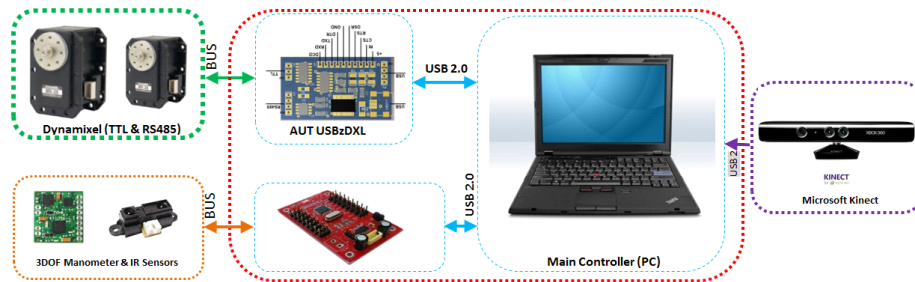

Fig. 2. DX-IO low-level controller & main PC and peripheral connected device

3   Software

The robot's operating system architect is based on ROS (Robot operating system) which runs under Ubuntu operating system. It has a core node which the other parts connect to it. Core node manages the requests of other parts. The main duty of the core is to manage communications between different nodes.

3.1   Human Robot Interaction (HRI)

One of the most important subjects that should be implemented in Service Robots is Human Robot Interaction (HRI). Human detection and following is the main concept of HRI. This should be done with minimum risk for both the human and the environment. @Home robots are designed to work in a real world indoor dynamic environments and robot should follow a safe and collision free path. A robot with this ability should ignore the messy background and focus on people.

By using Kinect sensor we detect skeletons in raw image frames with OpenNI [2]. The main problem is detection of only specific person. For this we first crop detected person's contour form RGB frames and then extracts features from the skeleton and cropped image data by using SURF [3] algorithm. In addition we want to use SIR filter [4] for model tracking. The face is the main identifier for a special person so if the face is visible to the camera we crop the face and track it with a tracking algorithms implemented in OpenTLD library and OpenCV [5] for face detection and classification algorithms.

We can increase performance by increasing the detection solutions, but our resources are limited. So with only these three approaches (skeleton detection, human image contour features and face detection) we can successfully detect and track defined persons. Finally for better data fusion a Kalman filter is used. Next step after getting the person's local position (x, y, z) is controlling robot actuators. By implementing a PID controller robot can successfully control the distance between the person and rotate smoothly to face the person. In these three approaches we don't need scanner data. Sensor fusion with KINECT and laser scanner data's[6] are another approach that can increase accuracy and solve KINECT near mode problem with tracking and detecting user legs that is our future work.

3.2   Person Re-identification

Most of person re-identification approaches are appearance-based in which it is necessary individual does not change one's clothing. For accurate person re-identification, we use 3D information which is computed by using Kinect [7]. In this approach we considered a set of 3D soft-biometric cues as features. Based on these features we can re-identify individuals independent of their appearance.

3.3   Voice interaction

Voice interaction unit includes speech recognition and speech synthesis. It is one of most important abilities of a robot. During competition, robot waits to receive a command from user. For better interaction, if any of the commands are not properly detected by the robot, the robot will ask to repeat it again. We use Microsoft Speech SDK [8] for speech recognition task in our robot. It uses categorized words in dictionary in order to increase the recognition accuracy. Dictionary selection is based on recognized vocabulary (keywords) during match.

For reducing environmental noise effect, we employ a preprocessing stage on data received by microphone before recognition stage. It is a speaker independent system, so we don't need to learn during match. Also, we have been used a source separation technique. By using this method, robot changes its direction to user. In speech synthesis

task, we use Festival[9]. We will try to identify gender and age of user to select appropriate speaker. Also, we will adjust synthesized speech volume based on distance from the user.

3.4   Simultaneous-Localization and Mapping

For localization and motion estimation we fuse data from both a laser scanner and a Kinect sensor, although during certain tasks visual odometry using Kinect might be disabled to reduce processing. Localization using laser scanner is done using standard AMCL [10]. Visual odometry is more complex and is done using Fovis [11] algorithm but will fall back to GICP [12] if Fovis failed to provide a motion estimate.

We used a keyframe approach for SLAM and place recognition. Keyframes are defined when the robot moves a certain distance or rotates for a certain angle. If a keyframe was detected, the current image captured by the Kinect sensor would be sent to place recognition module which uses RTAB-Map [13] algorithm. In the case that the image was previously seen, the place recognition algorithm will produce a loop-closure message for the SLAM system. Our SLAM system uses g2o [14] algorithm to sparsely represent robot movements.

Although we mostly rely on laser scanner data for navigation using GMapping and ROS navigation stack, for more complex environments we can also use OctoMap [15] to produce a 3D map with Kinect point clouds, but we currently map the 3D result to two dimensions since path planning is less complex and satisfies our current needs.

3.5   Object Recognition

As a robot has to be able to carry out tasks with the objects, the ability to recognize objects is very important. For object recognition, we used an object model which is given as input then we identify an object in an image scene. The algorithm should recognize the objects in conditions where the object has been rotated/translated or when the object is viewed from different viewpoints.

Our object recognition process uses 3D descriptor matching. It obtains a set of point to point correspondences between stored models and objects in the current scene. Then for clustering this set we use a correspondence grouping algorithm.

SURF is used to extract features of the objects. The correspondence clustering algorithm is based on the 3D-Hough Voting scheme [15]. These algorithms are implemented using Point Cloud Library (PCL) [16]. The features of the RGB-D camera image are matched to features of trained image based on Euclidean distance. In the last stage the previously found correspondences are clustered by using 3D-Hough Grouping.

3.6   Object Manipulation

For performing manipulation, coordinates of the object are obtained from object recognition stage. The robot arm is placed in a pre-grasp pose. This pose is obtained by object

recognition stage that includes height of tables and object coordinates. Distance between gripper and object is measured by the angle of the RGB-D camera and then robot will change its position so that the coordinates of the gripper and object are matched and grasping is performed.

Currently we are working on a system that uses an inverse kinematics algorithm of 6DOF arm to move gripper to object coordination and to avoid obstacles throughout the path. The object coordinates are provided by visual object tracking algorithm.
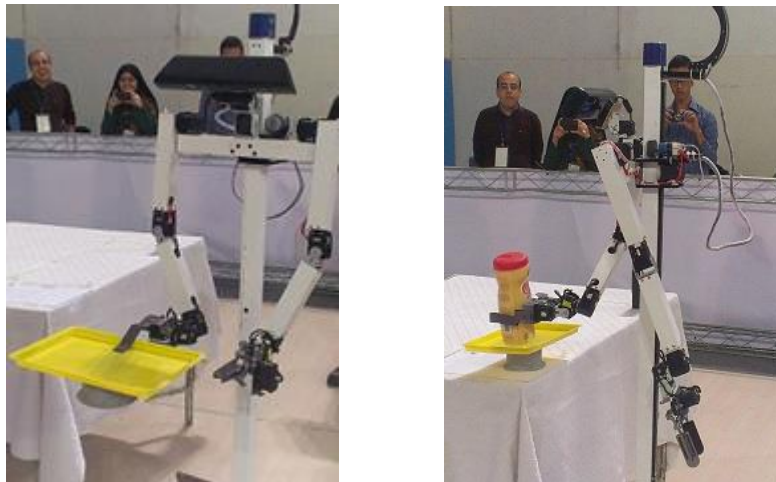

Fig. 3. Object recognition and grasping during the AUTRoboCup 2013 in Iran

4   Conclusion

In this paper we described our robot system and a part of its abilities. We participated with our young robot in AutRobocup2013 which we got first place. We hope by building a stronger robot, improving current robot and developing our software programs have a good demonstration in RoboCup 2014.

Future work

There are many items which we consider to use in our future robot for improving the performance of it. In case of electronic and control system:

- Using Cubieboard instead of laptop.
- Using accurate IMU for estimating movement angle of robot.
- Using Battery management system (BMS) to control and monitor the battery charge and discharge. BMS system has the capability of estimating the remaining charge, optimal charging and battery consuming.

In case of mechanical system:

- Modifying robot structure in order to improvement appearance of robot.
- Modifying robot arm in order to increasing DOFs and improving performance.

In addition in software part we developed a novel approach in accurate person recognition which we want to publish it.

Team Members and their contribution

-Ameneh sheikhjafari: team leader,

-Shaghayegh gharghabi: person recognition,

-Ismaiel mehrabi: object manipulation, object recognition,

-Marzie kariminoori: object manipulation, object recognition,

-Sara jahangiri: speech recognition,

-Edwin babaians: human tracking,

-Ramin gomari: frame work, path planning,

-Kourosh sartipi: Simultaneous-Localization and Mapping,

-Amir badamchi: speech recognition,

-Pantea habibi: programming group,

-Elahe khani: programming group,

-Alireza sheikhjafar: electronic designer, mechanical designer

-Mohammadreza mousaei: electronic designer,

-Pooya parsa: frame work.

References

1. ROBOTIS. (2013). Dynamixel Actuator. Available: http://www.robotis.com/xe/dynamixel_en. Last accessed 25th January 2014.

2. OpenNI, OpenNI. http://openni.org/ (Last Accessed: April 2011).

3. Bay, Herbert, et al. "Speeded-up robust features (SURF)." *Computer vision and image understanding* 110.3 (2008): 346-359.

4. Klein, Dominik Alexander, et al. "Adaptive real-time video-tracking for arbitrary objects." *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010.

5. Bradski, Gary, and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. O'reilly, 2008.

6. Alvarez-Santos, Víctor, et al. "Feature analysis for human recognition and discrimination: Application to a person-following behaviour in a mobile robot." *Robotics and Autonomous Systems* 60.8 (2012): 1021-1036.

7. Barbosa, Igor Barros, et al. "Re-identification with rgb-d sensors." *Computer Vision–ECCV 2012. Workshops and Demonstrations*. Springer Berlin Heidelberg, 2012.

8. Esposito, Dino. "The Microsoft Speech SDK." *Microsoft Internet Developer, Feb*(1999).

9. Black, Alan, et al. "The Festival Speech Synthesis System, Version 1.4. 2."Unpublished document available via http://www. cstr. ed. ac. uk/projects/festival. html (2001).

10. Fox, Dieter, et al. "Monte carlo localization: Efficient position estimation for mobile robots." *AAAI/IAAI* 1999 (1999): 343-349

11. Huang, Albert S., et al. "Visual odometry and mapping for autonomous flight using an RGB-D camera." *Int. Symposium on Robotics Research (ISRR),(Flagstaff, Arizona, USA)*. 2011.

12. Segal, Aleksandr, Dirk Haehnel, and Sebastian Thrun. "Generalized-ICP." *Robotics: Science and Systems*. Vol. 2. 2009.

13. Labbé, Mathieu, and François Michaud. "Appearance-Based Loop Closure Detection for Online Large-Scale and Long-Term Operation." 1-12.

14. Kuemmerle, Rainer, et al. "g2o: A general framework for graph optimization." *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2011.

15. Wurm, Kai M., et al. "OctoMap: A probabilistic, flexible, and compact 3D map representation for robotic systems." *Proc. of the ICRA 2010 workshop on best practice in 3D perception and modeling for mobile manipulation*. Vol. 2. 2010.

16. Tombari, Federico, and Luigi Di Stefano. "Object recognition in 3D scenes with occlusions and clutter by Hough voting." *Image and Video Technology (PSIVT), 2010 Fourth Pacific-Rim Symposium on*. IEEE, 2010.

17. Rusu, Radu Bogdan, and Steve Cousins. "3d is here: Point cloud library (pcl)." *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011.