SUT Team Description Paper

Technical Report · November 2013

DOI: 10.13140/2.1.1441.3121

7 authors, including:

Azarakhsh Keipour
Carnegie Mellon University

Edwin Babaians
Amirkabir University of Technology

Hossein Bakhshi Golestani
RWTH Aachen University

Mohammadreza Mousaei
Shahid Beheshti University

4th Amir Kabir University of Technology Robotic Competitions (2013) - Service Delivery Robots League

SUT Team Description Paper

Azarakhsh Keipour[1], Edwin Babaians[2], Kourosh Sartipi[3], Sahand Sharifzadeh[4], Hossein Bakhshi Golestani[5], Faraz Foroughi[6], Mohammad Reza Mousaei[7]

[1] Sharif University of Technology, keipour@ieee.org
[2] Amir Kabir University of Technology, edwin.babaians@gmail.com
[3] Sharif University of Technology, kourosh.sartipi@gmail.com
[4] Technical University Munich, sahand.sharifzadeh@gmail.com
[5] Sharif University of Technology, h.b.golestani@gmail.com
[6] Sharif University of Technology, farazfo@yahoo.com
[7] Shahid Beheshti University, m.mousaei@gmail.com

Abstract. This paper provides an overview of the hardware and software of our RobX robot and presents the scientific achievements embodied in our AUTCup Service Delivery Robot entry. This robot has been built by the team SUT which competes on behalf of Sharif University of Technology. RobX is the champion of 2012 AUTCup and 4th National Khwarizmi Robotic Competition. It defines a Human Robot Interaction interface based on Speech, Face, Object and Gesture Recognition.

Keywords: Service Robots, Speech Recognition, Speech Synthesis, Face Recognition, Object Recognition, Gesture Recognition, Object Manipulation, Visual Odometry, SLAM.

1. Introduction

Service robots are hardware and software systems that can assist humans to perform daily tasks in complex environments. To achieve this, a service robot has to be capable of understanding commands from humans, avoiding static and dynamic obstacles while navigating in known and unknown environments, recognizing and manipulating objects and performing other several tasks that the human beings ask for [3].

Some of the capabilities we have included in RobX to fulfill the mentioned tasks include:

1. Speech Recognition (Persian and English).

2. Speech Synthesis (Persian and English).
3. Gesture Recognition
4. Face Recognition
5. Object Recognition
6. Object Manipulation
7. Weather Forecasting using online resources
8. Simultaneous localization and mapping using vision.
9. Accurate odometry using vision and wheel encoders.
10. Accurate human arm imitation by the robot arm

This Team Description Paper is part of the qualification package for 4th International AUT Robotic Competitions (2013) in Iran. First, the hardware and software of the RobX robot platform will be introduced, followed by a description of our different methods for localization, and face, object, gesture and speech recognition.

2. RobX' Hardware

2.1. Body

The body of our robot is made from wood, to be environmental friendly and also to decrease the costs. The base of it is in the shape of a polygon with 60cm diameter. It weighs 10kg and the design is so that it has three major sections, each devoted to a part of robot; the first one for the mechanical part including motors and batteries, second for the processing unit and third is for sensing and manipulation. The height of each section is designed so that it bares the weight of the upper sections perfectly. The vision system is placed at a height of about 60cm to have a perfect view to ground for Arrow Following application, a perfect view to standard table size for Recognition and Object Manipulation and an upper view for human robot interaction including face and gesture recognition.

RobX has four wheels, two of which are connected to motors. There is a 24v motor fully equipped with encoders and a 49:1 reduction gearbox. It has a maximum speed of 100-rpm. It is ideal for medium to large robotic applications, providing cost effective drive and feedback for the user. It also includes a standard noise suppression capacitor across the motor windings [2]. The wheels have a diameter of 12.5cm.The motors are controlled by a closed loop system using the feedback received through the driver. This provides a very high odometer accuracy helping the navigation system.

The main processing unit of the robot is a windows-based laptop with 2.8 GHz Intel Core-i7 processor. The network-based architecture of the software allows the remote access to the main processor which enables us to do distribute computing.

The vision and sensory system of the robot is supported by Microsoft Kinect which enables us to access depth data of the surroundings and eliminates the need to use an external laser scanner.

The power of the whole systems is supplied by two 12volts lead-acid batteries with 7.2Ah.

2.2. Arm

RobX is equipped with a robotic arm to manipulate the objects. The planning of RobX's arms is done by the V-rep inverse kinematic solver. V-rep is Virtual Robot Experimentation Platform based on Lua scripting language. RobX manipulator is a 6-DOF arm that can handle 0.5 Kg objects. The first step of processing is

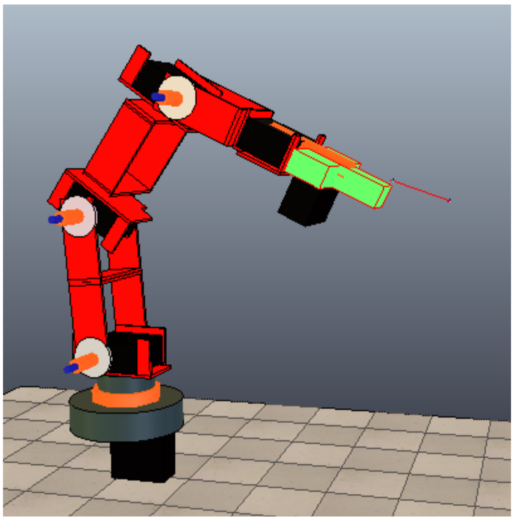

Figure 1. RobX' arm simulation.

to gather information from Microsoft Kinect sensor with creating image depth map, then the Kinect application side program communicate with v-rep via socket, and set IK coordinate target. Arm controller driver is based on Atmel AVR microcontroller that controlled with pc software that is linked to v-rep invers kinematic system via socket.

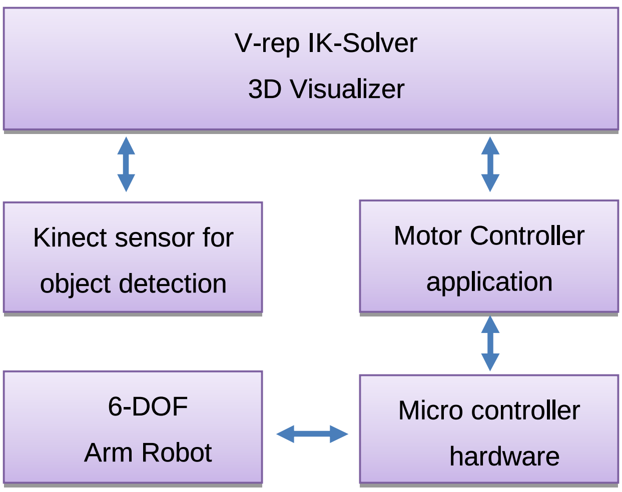

Figure 2. RobX' arm controller using Kinect data.

3. RobX's Software

The software on RobX is developed using Microsoft Visual Studio. We developed our software on a network-based architecture which allows distributable computing.

There are different services available on the software which is explained briefly as follows.

3.1. Speech

In this robot, the "Julius" speech recognition engine is employed. Julius is a high performance, open source speech recognition decoder used for academic research. It used latest speech recognition achievements and can perform large vocabulary continuous speech recognition (LVCSR) in real-time processing. The Julius incorporates a language model (LM) and an acoustic model (AM) for reliable recognition. The language model consists of a word pronunciation dictionary and a syntactic constraint. Word N-gram model, rule-based grammar and a simple word list (for isolated word recognition) are some types of supported LM. Acoustic models should be HMM defined for sub word units [10]. To prepare acoustic model, the "Hidden Markov Model toolkit (HMM)" is used. The HTK that was originally developed at Cambridge University is a portable toolkit for building and manipulating hidden Markov models. This tool provides sophisticated facilities for speech analysis, HMM training, testing and results analysis [11].

Julius has been developed as research software for Japanese LVCSR and it has been applied for various languages such as English, French, Mandarin Chinese, Thai, Estonian, Slovenian, and Korean. In this work, the required database for "Persian speech recognition" is prepared and the Julius is optimized and trained by employing the HTK toolkit. The results show the excellent accuracy of our Persian speech recognition system when a free-noise microphone is employed.

3.2. Face Detection and Recognition

Our robot's real-time face recognition system involves two stages:

1. *Face Detection*: video frame is searched to find faces and to narrow regions of interest.
2. *Face Recognition*: features of the detected face are extracted and are compared to a database of trained faces.

For face detection, we used OpenCV face detector, which is fairly reliable. OpenCV's Face Detector, working in roughly 90-95% of clear frames of a person looking forward at the camera. It is usually harder to detect a person's face when they are viewed from the side or at an angle, and sometimes this requires 3D Head Pose Estimation. It can also be very difficult to detect a person's face if the photo is not very bright or if part of the face is brighter than another or has shadows or is blurry or wearing glasses, etc. Our robot detects a frontal face in a video frame using its Haar Cascade Face Detector (also known as the Viola-Jones method).

We used Eigenfaces (also called "Principal Component Analysis" or PCA), a popular method of 2D Face Recognition from a photo or video frame. For higher accuracy, we initially convert color images to gray scale, and then apply Histogram Equalization as a method of automatically standardizing the brightness and contrast of facial images. Also, we resize images to a standard size while keeping its aspect ratio.

After preprocessing facial image, we perform PCA for Face Recognition using OpenCV Library. We also use a database (training set) of trained frames to recognize each of our teammates.
You can see the first 32 eigenfaces in the image below:

Figure 3. First 32 eigenfaces of a face image.

Our Face Detection and Recognition system is shown in action in figure below:

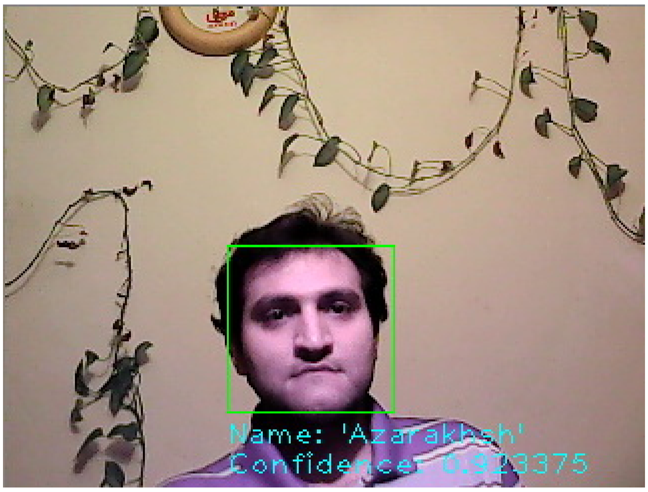

Figure 4. Robx' face recognizer in action.

3.3. Object Recognition

For object recognition we use SURF features. A picture is taken from the object we want to recognize and feature points are extracted from that image. Later in recognition phase, features extracted from each object, are compared against the features extracted from the image taken from the scene to decide if there is any of those objects are in there. Also because SURF does not consider color information, a histogram comparison is performed. For histogram comparison Chi-square distance is used. For chi-square, a low score represents a better match than a high score. A perfect match is 0 and a total mismatch is unbounded (depending on the size of the histogram).

$$d_{chi-square}(H_1, H_2) = \sum_i \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)}$$

Another challenge is to decide when to take an image of the scene and look for the objects in it. To overcome this problem we use an ultra-sonic distance sensor and frame difference information. Ultra sonic sensor gives information about the distance of the object (if there is any) from the camera. When an object is closer than a pre-defined distance from the robot, Camera starts taking pictures continuously. Every picture is subtracted from its previous one to calculate the difference between the two. If the difference is negligible the object recognition phase is triggered.

A sample object detection and recognition using SURF descriptors is shown in the figure below:

Figure 5. Robx’ object recognizer in action.

3.4. Road sign and direction detection

In robot navigation, we need to detect direction using signs drawn on the road. Each sign consists of a circle with an arrow in it. We use a simple, but fast and reliable method to find arrows direction.

At first, to find the circle on the road we use Hough Transform. After finding the circle, we narrow region of interest to only that circle. We use template matching techniques on the narrowed region to find arrows direction. Using obtained direction, our robot changes its moving direction.

3.5. Gesture Recognition

RobX has a gesture recognition engine which uses 3D depth data from Microsoft Kinect. We can recognize different gestures performed by users and do a task based on the recognition. On the current version of the RobX, the user can order the robot to move to different directions using hand gestures with a high accuracy of recognition.

Gesture Recognition enables a novel human robot interface which can be used for many purposes on service robots.

Figure 6. An example of Kinect RGB and Depth Data.

3.6. Weather Forecasting

Weather forecasting and news reading are examples of Human Robot Interaction applications which are important for service robots. As a task for AUTCup competition, RobX has a weather forecasting service which receives its data from online resources. It gets the data whenever there is connectivity and by saving them it also works on offline mode.

The weather system is working in connection with speech service. User might ask the robot the weather of a city and RobX would respond to it by showing it on the screen or talking to the user. It can recognize names of the cities in Iran or other countries.

3.7. Human Arm Imitation

We have built a novel human arm imitation ability to RobX. It can accurately perform the exact movements made by user's arm in 3D Space.

3.8. Visual Odometry

As Kinect sensor can produce both RGB and depth images, according to [15] the relative motion between two robot poses can be calculated using only two captured frames.

The main algorithm used for calculating robot motion is "Fast Odometry from Vision" [16]. In this algorithm the FAST features of RGB images are extracted and matched, and then by minimizing re-projection error motion between two frames is calculated. The main problem of this method is that it is dependent on finding good features in RGB images and in environments with very few available features the algorithm might fail. In case of failure, we use GICP [17] algorithm. As applying GICP on data captured by Kinect is too time consuming, we have developed a technique to down-sample Kinect data in such a way that important features of the environment are not lost, which guarantees fast convergence and accurate results, and is also not time consuming. If motion estimation using GICP also fails, we can be almost certain that the two frames do not hold enough information to estimate motion and should be discarded, fortunately this rarely happens.

4. Conclusion

We introduced the software and hardware specifications of RobX. Other than the capacities needed for the AUTCup2013 we have got other achievements during the research and development phase which will be introduced in the Open Show of the competitions. Currently we are working more on the interaction systems of RobX and in the future this robot will appear with improved capabilities including better navigation and localization.

Figure 7. RobX robot (without the manipulator arm).

5. References

1. Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. Int. J. Digit. Libr. 1 (1997) 108–121
2. http://www.robot-electronics.co.uk/htm/emg49.htm
3. Pumas @Home 2012 Team Description Paper
4. P. Chang and J. Krumm. Object recognition with color cooccurrence histogram. In Proceedings of CVPR '99, 1999.
5. S. M. LaValle. Planning Algorithms. Cambridge University Press, Cambridge, U.K., 2006. Available at http://planning.cs.uiuc.edu/.
6. S. Caraian and N. Kirchner, \Robust Manipulability-Centric Object Detection in Time-of-Flight Camera Point Clouds," in *Proc. of the 2010 Australasian Conference on Robotics and Automation (ACRA-10)*, 2010, pp. 1{8.
7. Lowe, D.: Distinctive image features from scale-invariant keypoints. International journal of computer vision 60(2) (2004) 91–110
8. Bradski, G., Kaehler A.: Learning OpenCV. O'reilly Press (2008).
9. R.Kelly, V.Santibanez and A.Loria:Cotrol of Robot Manipulators in Joint Space. Springer (2005)
10. A. Lee and T. Kawahara, "Recent Development of Open-Source Speech Recognition Engine Julius," In Proc. Asia-Pacific Signal and Information Processing Association (APSIPA2009), pp. 131-137, 2009.
11. http://htk.eng.cam.ac.uk
12. H. Bay, A. Ess, T. Tuytelaars and L. van Gool, "Speeded-up robust features (SURF)", Computer Vision and Image Understanding, 110:3, 2008, pages 346–359.
13. B. Schiele and J. L. Crowley "Recognition without correspondence using multidimensional receptive field histograms", International Journal of Computer Vision, 36:1, 31-50, 2000.
14. M. J. Swain and D. H. Ballard "Colour indexing", International Journal of Computer Vision, 7:1, 11-32, 1991.
15. Y. Ma, S. Soatto, J. Kosecka and S. S. Sastry, An Invitation to 3-D Vision: From Images to Geometric Models, Springer, 2005.
16. A. S. Huang, A. Bachrach, P. Henry, S. Prentice, R. He, M. Krainin, D. Maturana, D. Fox and N. Roy, "Visual Odometry and Mapping for Autonomous Flight Using an RGB-D Camera," International Symposium on Robotics Research (ISRR), 2011.
17. Segal, Aleksandr, D. Haehnel and S. Thrun, "Generalized-ICP," Robotics: Science and Systems, vol. 2, 2009.
18. A. Keipour, S. Sharifzadeh, E. Babaians, F. Taherpour, A. Malekpour, H. Bakhshi Golestani, "RobX–AUTCUP 2012 & Khwarizmi Robotic Competition@ Home League Team Description Paper," AUTCUP 2012 & 4th Khwarizmi Robotic Competitions, 2012.