RobX – AUTCUP 2012 & Khwarizmi Robotic Competition @Home League

Team Description Paper

Azarakhsh Keipour, Sahand Sharifzadeh, Edwin Babaians, Farid Taherpour, Amin Malekpour, Hossein Bakhshi Golestani
Sharif University of Technology
Iran
keipour@ieee.org, s.sharifzadeh@acm.org, edwin.babaians@gmail.com,
{farid.taherpour, amin.malekpour.6083, h.b.golestani}@gmail.com

Abstract. This Paper provides an overview of the hardware and software of the RobX robot and presents the scientific achievements embodied in our AUTCup @Home entry. This robot has been built by the team SUT which competes on behalf of Sharif University of Technology. RobX defines a Human Robot Interaction interface based on Speech, Face, Object and Gesture Recognition.

1. Introduction

The 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
2. Gesture Recognition
3. Face Recognition
4. Object Recognition
5. Object Manipulation
6. Weather Forecasting using online resources
7. Following Arrow signs on the ground by recognition of their direction using accurate vision system and odometer.
8. Accurate human arm imitation by the robot arm
This Team Description Paper is part of the qualification package for AUTCup 2012 and Khwarizmi Robotic Competition 2012 in Iran. First, the hardware and software of the RobX robot platform will be introduced followed by a description of our different recognition methods for Face, Object, Gesture and Speech.

2. RobX’s 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 that so 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 are 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 Corei7 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 12volt lead acid batteries with 7Ah.

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. First step of processing is 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.
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 data-base 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:

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

For object detection and recognition, we used SURF (Speeded Up Robust Feature) descriptors implemented using OpenCV. SURF is a robust local feature detector, first presented by Herbert Bay et al. in 2006, that can be used in computer vision tasks like object recognition or 3D reconstruction. It is partly inspired by the SIFT descriptor. SURF descriptors are invariant to rotation, translation and other affine transformations and also they can greatly deal with overlap, which make them suitable for our object recognition task.

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

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 can ask the robot about 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.

4. Conclusions

We introduced the software and hardware specifications of RobX. Other than the capacities needed for the AUTCup2012 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.

5. References

3. Pumas @Home 2012 Team Description Paper
11. http://htk.eng.cam.ac.uk