Proceedings of the 6th Győr Symposium and 3rd Hungarian-Polish
and 1st Hungarian-Romanian Joint Conference on Computational Intelligence

Point-and-command Paradigm in Human - Robot Interaction

Răzvan Gabriel Boboc[1], Doru Talabă[1]

[1]Transilvania University of Bra□ov, Robotics and Virtual Reality Research Lab 29 Bd. Eroilor, 500036 – Bra□ov, Romania
E-mail: razvan.boboc@unitbv.ro

Abstract: This paper presents an interaction concept aimed to provide an easy-to use and intuitive way of interaction with humanoid robots in domestic environments. The interaction is performed trough gesture commands and dialogue mechanism to provide 'natural' means for user to command assistive robots which performs some given tasks. "Point-and-command" concept was evaluated using two Nao robotic platforms. A DSS based on Fuzzy Logic was implemented for selecting between two robots when one of them becomes unusable.

Keywords: human-robot interaction, pointing gesture, humanoid robot, assistive robotics

1. Introduction

With the advent in development of assistive mobile robots, has become increasingly evident the need to find new and more intuitive ways for people to interact with them, so that any user, whether experienced or not, can use them without having too many preliminary instructions. In this sense, simple and easy to use interfaces are required within Human-Robot Interaction (HRI).

In this paper we present a specific approach of multimodal interaction with humanoid mobile robots that assist people in domestic or office environments. The aim of our research is to develop an assistive robot that help humans in daily activities and naturally interact with them, providing communicative capabilities. The implemented interface allow user to express his intentions using both verbal and non-verbal expressions. In this way, the interaction is more natural and more human-centered.

2. Related Work

Despite the multitude of ways to interact with a mobile robot, gesture and speech channels remains the most suitable. When dealing with assistive robots, the user should be able to inform them about an object's location. This interaction is crucial in home environments because most of the tasks performed by the robot require manipulation of objects. In this regard, pointing gesture is recognized as an essential way of interaction with robots.

There have been many research works on interaction with robots using deictic communication. In some of them a robot is used as testbed to investigate specific issues related to social development of humans [1], other works try to enable a robot to generate its pointing gestures [2], or both human and robot use pointing gestures in interaction [3], but most papers focus on the ability of robots to respond to pointing gestures performed by human operators. Some papers are focusing on recognition methodology and estimating pointing direction [4], on the control part of the system [5], or on the process of achieving a natural deictic communication between robots and humans [6].

We propose a simple solution of deictic communication that combines vision and speech recognition using a single device and we name this paradigm "Point-and-Comand" (PaC). Beside this, we implement a Decision Support System (DSS) for selecting between two robots to be assistants in home environments. The DSS is based on Fuzzy logic and inform user when a robot become unusable. DSS systems are a class of computer-based information systems with the main aim to help decision makers to make the best decision when dealing with complex situations and information [7]. There are few studies that make use of DSS for robotic applications [8], [9], [10]. To our knowledge, DSSs were little or no used in HRI applications.

3. System architecture

The system proposed in this paper is designed for interaction with personal mobile assistive robots, operating at home or in offices for tedious, repetitive tasks. In fig. 1 is depicted the system architecture.

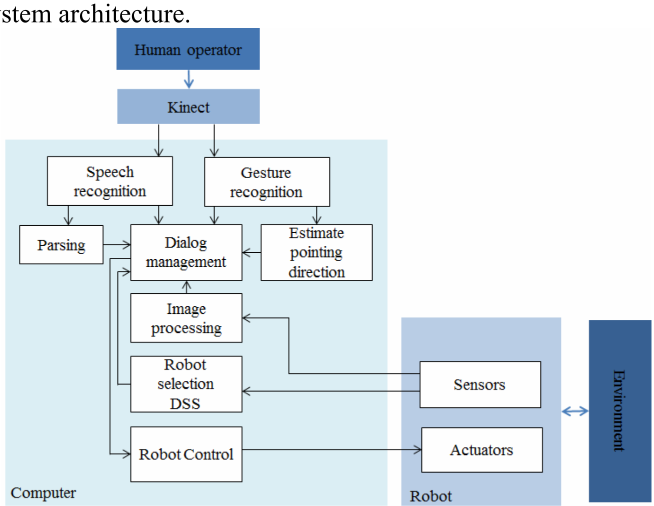

Figure 1. System overview

Human operator sends gestural and voice commands captured by a Kinect device to a mobile robot. Speech and gesture recognition are two separate processes, but they are combined in Speech/gesture fusion block in order to provide more complex commands. A Robot selection DSS is responsible for choosing between two robots the 'active' one, which is prepared for interaction. A Control unit sends the commands to robot, which are generated in Dialog management block.

4. Point-and-command paradigm

People use pointing gesture in communication as a mechanism to indicate a reference object or to inform listener about its location. Pointing gestures can be used also by humans to provide direction information to robots [11].

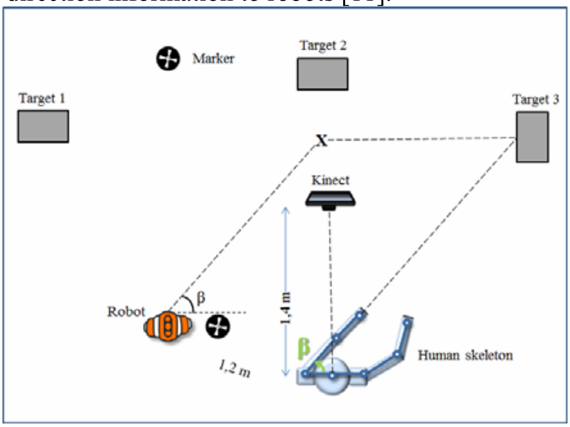

Figure 2. Experimental setup

In our work, we exploit the idea of combining human pointing gestures with speech for communicate the intentions to a robot, in a user-centered approach, named Point-and-Command (PaC). In simple terms, the idea of the interaction process is the following: a person points at an object and says something such us "grab that object", and then the robot will navigate in the indicated direction and will perform the task (identify object and grab it).

4.1. Gesture detection and recognition

In this work a Kinect sensor is used to track human motions. The Kinect software API provides positions of the body joints of the user in real time. In our case just 7 are relevant, which are depicted in fig. 2 (blue circles on human skeleton). The pointing direction is given by computing the angle between the operator's arm and his body. The vector from the shoulder to the wrist indicate the direction where operator pointing at.

The robot has a fixed position, where the power supply is located. It initially has the same orientation like human operator. Thus, it will start to move on the direction indicated by the human arm (given by β angle), but from its position. Knowing his initial orientation towards operator, it can rectify its orientation by periodically rotating the head and looking for objects by color. Thus, after a certain distance (e.g., to the X point on figure 2), it change the moving direction.

For gesture recognition, the frames obtained from Kinect sensor are converted into feature vectors that contain the positions of different joint of the body. The motion frame is expressed as a 24-elment feature vector, containing the x,y,z positions for all joints of each arm (shoulder center, shoulder, elbow and wrist) expressed in meters in the coordinate system of the Kinect camera. So, a gesture is represented by a sequence of arm postures over time and the recognition process involves the comparing of current sequence with some previously recorded sequences using Dynamic Time Warping (DTW) algorithm.

4.2. Speech recognition

For capturing the human voice the same device (Kinect) was used. Kinect for Windows SDK allows the use of Microsoft.Speech recognition API. The streaming audio is captured from the Kinect microphone array using a speech recognition engine provided by SDK. Microsoft.Speech was used to create grammars, which can recognize a single world or a short phrase. Depending on the interaction context, there are different grammars implemented for each situation. Each grammar has several rules that define a pattern or a sequence of words. If a word is detected, but has a very low confidence level, the robot will notify human, asking him to repeat the command.

4.3. Object detection

After the system estimates the moving direction, the image processing unit comes into play. The robot moves forward in that direction, searching for the potential objects. In this process data from robot's sensors (video camera, sonar sensor) are processed for detecting objects or obstacles. Emgu CV was used for image processing, which is a .NET wrapper to OpenCV library [12]. In this paper, an OpenCV implementation of Canny algorithm [13] for edge detection was used. The objects used for experiments are Styrofoam balls and wood block colored in different colors.

5. Decision Support System

One of the main limitations for mobile robots is represented by their energetic autonomy. In the case of the mobile platform used in this research, it does not surpass 45 min of autonomy in operational mode. We choose to use two robots for interaction. When the battery of the first is going flat, the second swings into action. But the decision of selecting the other robot belongs to human operator. For this purpose an Intelligent Decision Support System (IDSS) was developed for helping him to decide when to stop the interaction with a robot and to call the other one.

The DSS based on Fuzzy Logic (FL) is designed to deal with the uncertainty for aiding the operator to make the decision on which robot to interact with considering 4 input variables: percentage of remaining power, the distance between robot and the power supply, a variable that show the internal state of the robot, and operating time. These variables are obtained from different sensor readings and are fuzzified into meaningful fuzzy sets. Membership function of all these fuzzy are trapezoidal.

The output of the Fuzzy system is a value from 1 to 4 representing the operating state of the robot: "good", "normal", "low", "very low". The rule base of the system contains some rules that determine the state used to inform user when the robot is fully operable and when not. When the state is “very low” or even “low” the user should ask robot to return in the home position and then he will call the other robot.

6. Experiments

The platform selected for this research is humanoid robot NAO, created by the French company Aldebaran Robotics to be a true daily companion [14]. A simple testing scenario was proposed in order to examine the functionality of the entire system. The experiment was performed in an indoor environment. Three boxes with objects of various colors and shapes were placed in the environment in different positions as illustrated in fig. 3. Two NAO robots sit on the left of the user at a certain distance from him, besides the power supply.

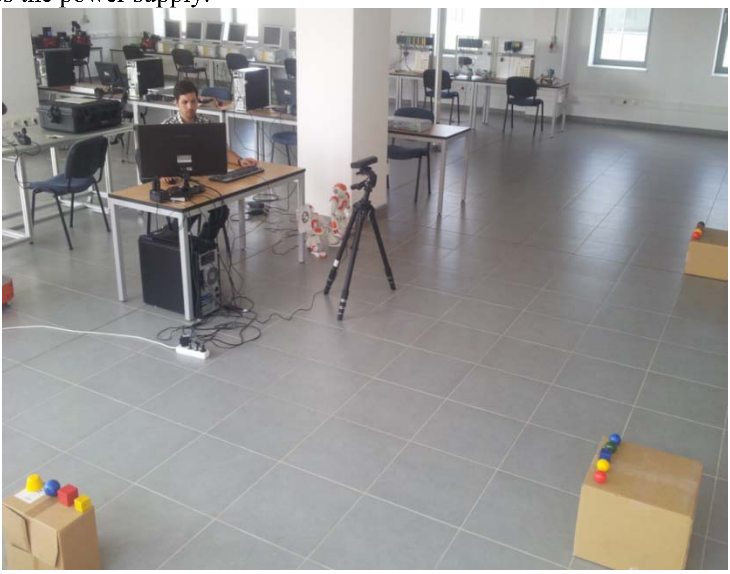

Figure 3. Layout of the experimental room

The tests were performed by 3 persons, each of them for three times. In the first round of tests, the user ask NAO to bring him the following objects: a red ball from the left box, a green ball from the box placed in the middle, and yellow cube from the right box relating to human position.

The accuracy of pointing gesture recognition was 91 % and for speech recognition 95 % was achieved. Regarding the vision processing algorithms, the accuracy of object detection and recognition was also acceptable. Each experiment requires the recognition of 9 objects: 3 (for each task) x 3 (for each user). So, in total 27 objects had to be recognized throughout all three experiments and of these only twice the robot misidentified an object, but after the user gave it another instruction, it found the correct one. Robot always ask user before performing a grasping action.

7. Conclusion

The purpose of this work was to develop an interface for assisting people in their everyday tasks, using two humanoid robots that navigate in the human space, being able to detect and to fetch objects. The human-robot interaction is based on pointing gesture combined with vocal commands. A “point-and-command” interaction concept was proposed aimed to provide user an easy-to-use and intuitive interface and to improve the efficiency of human-robot collaboration. An Intelligent DSS was also implemented in order to allow user to decide when to ‘dismiss’ a robot because his battery become out of charge and to call the other one to continue the task performed by the first.

Further research includes improvements on these areas: human-robot communication, pointing angle estimation, robot localization and obstacle avoidance, more tests with more users. Also, the robot’s learning ability should be investigated. It must learn and ‘keep in mind’ the identified objects.

Acknowledgement

This paper is supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), ID134378 financed from the European Social Fund and by the Romanian Government.

References

[1] Nagai, Y.: Learning to Comprehend Deictic Gestures in Robots and Human Infants, in the 2005 IEEE International Workshop on Robot and Human Interactive Communication, Nashville, USA, 2005.

[2] Sauppé, A., Mutlu, B.: Robot deictics: how gesture and context shape referential communication, presented at the Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, Bielefeld, Germany, 2014.

[3] Beuter, N., Spexard, T., Lutkebohle, I., Peltason, J., Kummert, F.: Where is this? - gesture based multimodal interaction with an anthropomorphic robot, in th IEEE-RAS International Conference on Humanoid Robots, 2008.

[4] Breuer, T., Ploeger, P.G., Kraetzschmar, G. K.: Precise pointing target recognition for human-robot interaction, presented at the Workshop on Domestic Service Robots in the Real World, SIMPAR, Darmstadt, Germany, 2010.

[5] Schiffer, S., Ferrein, A., Lakemeyer, G.:Caesar: an intelligent domestic service robot, Intelligent Service Robotics, vol. 5, pp. 259-273, 2012.

[6] Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H., Hagita, N.: Natural deictic communication with humanoid robots, in the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2007), 2007.

[7] Cebi, S., Kahraman, C.: Developing a group decision support system based on fuzzy information axiom, Knowledge-Based Systems, vol. 23, pp. 3-16, 2010.

[8] Mouaddib, A. I.:Controlling and Sharing Authority in a Multi-Robot System, in first conference on Humans Operating Unmanned Systems (HUMOUS'08), Brest, France, 2008.

[9] X. C. Ding, X. C., Powers, M., Egersteddt, M., Shih-yih, Y., Balch, T.:Executive decision support, IEEE Robotics & Automation Magazine, vol. 16, pp. 73-81, 2009.
[10] Pernalete, N., Gottipati, R., Grantner, J., Edwards, S., Janiak, D., Haskin, J. Dubey, R. V.: Integration of an Intelligent Decision Support System and a Robotic Haptic Device for Eye-Hand Coordination Therapy, in the IEEE 10th International Conference on Rehabilitation Robotics (ICORR 2007), 2007.
[11] Bauer, A., Wollherr, D., Buss, M.: Information retrieval system for human-robot communication - Asking for directions, in IEEE International Conference on Robotics and Automation (ICRA '09), 2009.
[12] http://www.emgu.com/wiki/index.php/Main_Page.
[13] S. Shin, *Emgu CV Essentials*, Birmingham Packt Publishing, 2013.
[14] http://www.aldebaran.com/en/humanoid-robot/nao-robot