Assistive Robot Multi-Modal Interaction with Augmented 3D Vision and Dialogue

Juan G. Victores, Félix R. Cañadillas, Santiago Morante,
Alberto Jardón, and Carlos Balaguer

Robotics Lab research group within the Department of Systems Engineering and
Automation, Universidad Carlos III de Madrid (UC3M), `jcgvicto@ing.uc3m.es`

Abstract. This paper presents a multi-modal interface for interaction between people with physical disabilities and an assistive robot. This interaction is performed through a dialogue mechanism and augmented 3D vision glasses to provide visual assistance to an end user commanding an assistive robot to perform Daily Life Activities (DLAs). The augmented 3D vision glasses may provide augmented reality vision of menus and information dialogues over the view of the real world, or in a simulator environment for laboratory tests and user evaluation. The actual dialogue is implemented as a finite state machine, and includes possibilities of Automatic Speech Recognition (ASR), and a Text-to-Speech (TTS) converter. The final study focuses on studying the effectiveness of these visual and auditory aids for enabling the end user to command the assistive robot ASIBOT to perform a given task.

Keywords: assistive robotics, end-user development, human-robot interaction, multi-modal interaction, augmented reality, speech recognition

1 Introduction

During the past years, and due to the complexity of systems and robotic platforms to control, the importance of the developments in the field of Human-Robot Interaction (HRI) has been greatly increasing. This is most noticeable in systems developed for people with disabilities, such as assistive robotic systems, where HRI's must be designed taking the type of users who will use the system as well as their disabilities into account. In recent years, innovative breakthroughs have been developed in this field thanks to the development of multi-modal interfaces that may adapt to the needs of the users of these systems.

This paper presents the latest developments in multi-modal interfaces with the ASIBOT assistive robot [1]. The main components are an augmented reality 3D vision glasses system with inclinometer, and an interactive dialogue mechanism which has been implemented as a finite state machine. Figure 1 depicts an actual screenshot of the user's view of the developed augmented reality interface at run-time. The system as a whole is capable of guiding the user through the different options within the interactive dialogue, while presenting synchronized information to the user through the augmented reality 3D vision glasses interface. Through the interactive dialogue, the user is capable of commanding the robot to perform a set of actions, as well as controlling several different visual and functional aspects of the interface.

Fig. 1. Assistive Robot Multi-Modal Interaction Augmented 3D Vision screenshot.

Section 2 of this paper describes the State of the Art and several works related to the presented developments. Section 3 presents the Open System Architecture in its current implementation. Section 4 describes the experiments performed allowing users to test the system. Finally, Section 5 outlines several conclusions.

2 State of the Art

Augmented reality systems are being widely used for the development of multimodal interfaces, as they provide the possibility of multiple configurations. One of the main articles dealing with this kind of device is [2], which defines the types of configurations that augmented reality systems may have. In the article, systems called Head-Mounted Displays (HMD) are defined, a concept that refers to screens located near the eyes of a user's head, in addition to showing the advantages and disadvantages of using this kind of system. Subsequently, this study was updated in [3], introducing new concepts such as the called Head-Worn Displays (HWD), based on the use of small projectors that project onto a semi-transparent mirror to display information over the real world. Another relevant article is [4], where a survey is performed, which shows the main features of the displays used in augmented reality systems. In addition, in the article, the different techniques and positioning of the augmented reality systems based on the new technologies used in this field is described.

Before studying the different fields that rely on augmented reality technologies, it is important to review how the collaboration between humans and robots is when using augmented reality interfaces. A review on how HRI must be in the context of augmented reality interfaces can be found in [5]. This proposal is later evaluated in [6] by the same authors.

As is shown in [7], technologies based on augmented reality are being used in many fields thanks to several possibilities that these system allow. In the industrial robotic field, augmented reality systems are used for the tele-operation of industrial robots. The use of an augmented reality interface for the control of a manipulator robot in unstructured environments is described in [8]. Another example is the system proposed in [9], where the positioning of a robot and generation of its trajectories is obtained through the use of augmented reality. A different area where augmented reality systems are being used is in the field of robotics oriented medical applications. In [10], the authors suggest the use of augmented reality system for controlling a robot as support in surgery tasks. Another interesting application in this field is proposed in [11], where they use augmented reality as support for performing laparoscopic surgery. In rehabilitation robotics, studies are underway, also based on augmented reality with the aim to help patients with mobility disabilities. In [12], the authors use an augmented reality system to support the rehabilitation of the hand following a cerebrovascular condition. A more limited number of studies can be found in the field of assistive robotics. Among them is the system proposed in [13], which uses an augmented reality setup for control and interaction between a wheelchair and the user.

The most common interfaces that are being used in assistive and social robotics are based on voice recognition, where a person can interact with a robot through voice commands, sending orders or requesting information. An example of interactive dialogue mechanism between a human and a social robot can be found in [14]. Examples of robots that may perform tasks, actions, or exchange information with a user can be found in [15] or [16]. An example of a voice recognition system to control a wireless assistive environment can be found in [17].

3 Open System Architecture

The ASIBOT Open System Architecture is provided through the use of the YARP [18] robotics platform. It acts as glue between the components, which are simultaneously decoupled and asynchronously updated with the flow of user and environmental information. Figure 2 depicts a basic connection diagram between the different components that compose the architecture. Two types of connections are used: streaming data flow connections for information that should be updated quickly, and remote procedure port connections that return acknowledgements of reception that may also be used for information on the degree of accomplishment of a certain task.

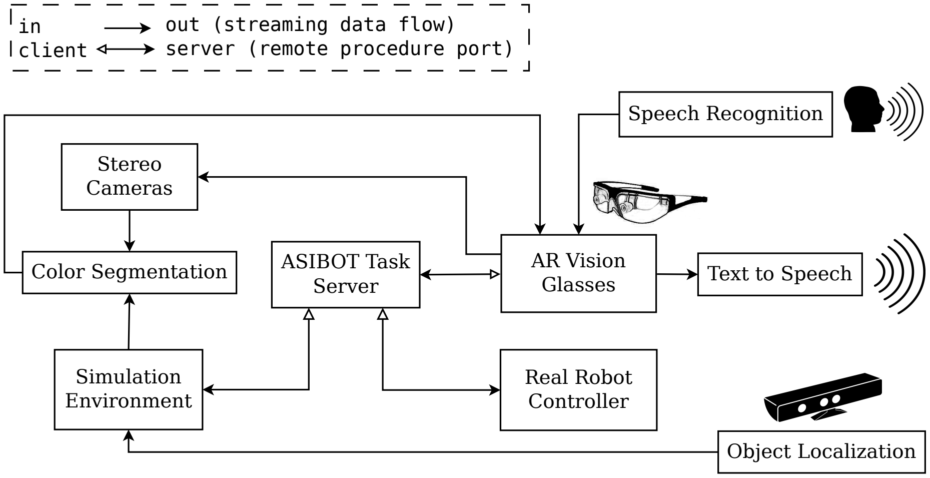

Fig. 2. Assistive Robot Multi-Modal Interaction Augmented 3D Vision scheme.

The following subsections will describe some of the main characteristics of the different components that compose the system and are depicted in the diagram.

3.1 Augmented Reality 3D Vision Glasses

The Augmented Reality 3D Vision Glasses used in this system are the Vuzix Wrap 920AR Glasses, which are shown in Figure 3. This device provides a Head-Mounted Display with two frontal cameras for the stereo video capture. In addition, it has a 6-degree of freedom head inclination tracker, and a high fidelity stereo audio output.

Fig. 3. Vuxiz Wrap 920AR Augmented Reality System.

The development of the interface that is shown on the display is based on the Open Source OpenGL libraries. To develop the augmented reality graphical interface, text and geometrical elements with textures are overlapped upon the real world or simulator images that are incoming from the cameras.

3.2 Simulation Environment

Our basic setup usually involves a three-layered structure: the simulator class (which uses the OpenRAVE-core libraries for graphical and physical aspects), a robot kinematic solver class, and a robot kinematic controller class. Figure 4 depicts the default loaded simulated environment.

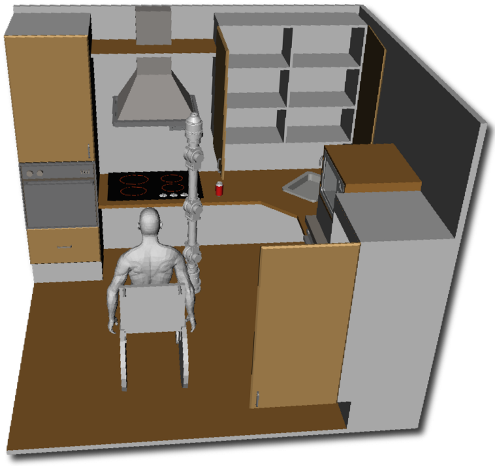

Fig. 4. ASIBOT assistive kitchen simulation environment.

As a new feature for this application, an OpenRAVE plugin called externObj has been developed. This plugin enables a Simulation Environment input port that receives streaming data from the Object Localization module, synchronizing the position of an object in the simulated environment with the actual real position of an object provided by the Object Localization module.

3.3 Real Robot Controller

The Real Robot Controller is also three-layered, with the real robot controller class at the motor-sensor level. This class manages the movements of the real robot using CAN-bus messaging with the robot's drivers. The bus is internally treated as a shared resource protected by software semaphores to avoid the possibility of different threads attempting to access the bus simultaneously.

3.4 ASIBOT Task Server

A Task Server has been developed, which allows tasks to be implemented as classes that inherit from a same Task base class and be instanced through remote procedure port calls. The specific tasks used for this application have been those implemented in the TaskGrabCan and the TaskDrink classes. TaskGrabCan is capable of making the robot to grab a can in the environment, taking the current position of the robot and the located object into account. TaskDrink moves a grabbed object near the user's lips to allow the user to drink.

3.5 Color Segmentation

An Open Source library that wraps around OpenCV has been released, namely Travis. Travis (which stands for Tracking and Vision library) is a small library for computer vision in robots [19]. The parametric image segmentation Travis provides has been used within the Color Segmentation module, which processes an incoming stream of images within a periodical thread. The Color Segmentation module is used as a bypass module that outputs to the Augmented Reality 3D Vision Glasses display. Its input may be switched between the 3D Vision Glasses' cameras or their simulated environment analogous, providing object segmentation information to the user in either of the two cases. Figure 5 depicts the Color Segmentation of the red can of the simulated environment. The visual output includes the object contour (pink contour), centroid (blue dot), and rotated bounding box (blue rectangle).

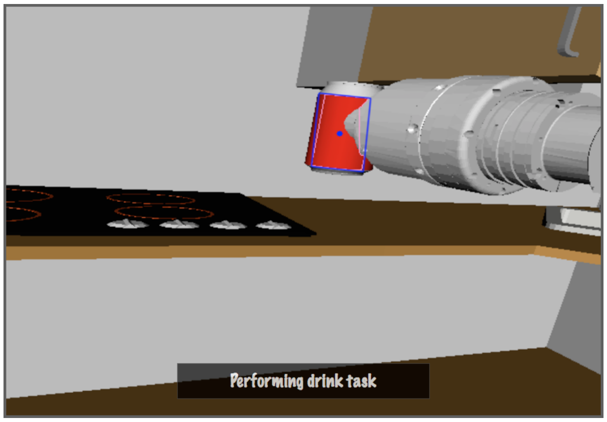

Fig. 5. User's view of the Color Segmentation module bypassing the simulator image.

3.6 Object Localization

The Object Localization module uses data from a Microsoft Kinect sensor placed in the environment. The color segmentation of the object is based on the Travis library, described in the previous subsection. It obtains object centroids, that are then matched with the depth image provided by the official YARP wrapper of the Kinect OpenNI drivers. The real world coordinates of the object are computed through homogeneous transformation matrices provided by the developed ASIBOT TinyMath library.

3.7 Automatic Speech Recognition

The Automatic Speech Recognition module uses the CMU Pocketsphinx and GStreamer packages in a Python environment. Natural language parsing is not used, and instead a small vocabulary corpus of specific words is compiled for our applications.

3.8 Text-to-Speech

The Text-to-Speech module used is the iSpeak module that can be directly found within the iCub Software repository. It is actually a wrapper around the Festival and eSpeak packages for speech synthesis.

4 Experiments

The experiments where performed with ten healthy robotics-related people using the system in the ASIBOT kitchen environment[1] in conjunction with its simulated representation. The range of ages of the participants was between 25 and 35 years old. After a brief description of the system and its components (approximately 5 minutes), they were allowed to use the robotic system freely.

The following is an actual transcription of a dialogue performed between one of the human users and the ASIBOT assistive robot developed system.

ASIBOT→ I'm ready. Ask me to show tasks, or to perform an action.
User→ *Show tasks.*
ASIBOT→ I understood SHOW TASKS. Is that correct?
User→ *Yes.*
ASIBOT→ How do you want to see the tasks? Text, speech, icons?
User→ *Text.*
ASIBOT→ I understood TEXT. Is that correct?
User→ *Yes.*
-Text with task names appear in the interface-
ASIBOT→ Okay, perfect. I'm showing the text. I'm ready. Ask me to show tasks, or to perform an action.
User→ *Grab.*
ASIBOT→ Performing grab task.
-The robot performs the task of grabbing a red can object-
ASIBOT→ Finished grab task.

---

[1] It is important to notice, however, that the system is ubiquitous and has been designed to work in any part of a living environment.

In order to measure the satisfaction of the users with the robotic system, we provided them with SUS tests (System Usability Scale). As a summary of the results:

- The average punctuation was 84 ± 10.88 over 100 (where 100 is the best score). This is higher than the 70.5 ± 9.5 score achieved using a web-browsable multi-modal system, recently published by the authors [20].
- The best results were obtained in the items "I though the system was easy to use" and "I would imagine that most people would learn to use this system very quickly", both with an average of 4.7 ± 0.48 (where 5 is the best score).
- On the other hand, the worst results were obtained in: "I think that I would need the support of a technical person to be able to use this system", with an average of 2.3 ± 1.16 (where 5 is the best score).

5 Conclusions

In this paper, the creation of an Assistive Robot Multi-Modal Interface system based on augmented 3D vision and interactive dialogue has been proposed. To this end, different systems based on augmented reality and interactive dialogue mechanisms have been studied. A complete system has been developed under our Open System Architecture, which has been tested with users in the form of a closed user-ready system.

As a result, we have understood the capabilities and functionalities of our system. The limitations of our current developments detected by the users in the tests and our own subjective and objective appreciations will lead to future developments with increased accessibility, usability, and end user satisfaction.

Acknowledgments

The research leading to these results has received funding from the ARCADIA project DPI2010-21047-C02-01, funded by CICYT project grant on behalf of Spanish Ministry of Economy and Competitiveness, and from the RoboCity2030-II-CM project (S2009/DPI-1559), funded by Programas de Actividades I+D en la Comunidad de Madrid and co-funded by Structural Funds of the EU.

References

1. Alberto Jardón Huete, Juan G Victores, Santiago Martinez, Antonio Giménez, and Carlos Balaguer. Personal autonomy rehabilitation in home environments by a portable assistive robot. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 42(4):561–570, 2012.
2. R. T. Azuma. A survey of augmented reality. *Presence*, 6(4):355–385, 1997.
3. Ronald T. Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, 21(6):34–47, 2001.

4. DWF Van Krevelen and R Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9(2):1, 2010.
5. Scott A Green, Mark Billinghurst, XiaoQi Chen, and GJ Chase. Human-robot collaboration: A literature review and augmented reality approach in design. *International Journal of Advanced Robotic Systems*, 5(1):1–18, 2008.
6. Scott A Green, J Geoffrey Chase, XiaoQi Chen, and Mark Billinghurst. Evaluating the augmented reality human-robot collaboration system. *International journal of intelligent systems technologies and applications*, 8(1):130–143, 2010.
7. Julie Carmigniani, Borko Furht, Marco Anisetti, Paolo Ceravolo, Ernesto Damiani, and Misa Ivkovic. Augmented reality technologies, systems and applications. *Multimedia Tools and Applications*, 51(1):341–377, 2011.
8. Paul Milgram, Shi Yin, and Julius J Grodski. An augmented reality based teleoperation interface for unstructured environments. In *ANS 7th Meeting on Robotics and Remote Systems, Augusta*, pages 101–123. Citeseer, 1997.
9. HC Fang, SK Ong, and AYC Nee. Interactive robot trajectory planning and simulation using augmented reality. *Robotics and Computer-Integrated Manufacturing*, 28(2):227–237, 2012.
10. Son-Lik Tang, Chee-Keong Kwoh, Ming-Yeong Teo, Ng Wan Sing, and Keck-Voon Ling. Augmented reality systems for medical applications. *Engineering in Medicine and Biology Magazine, IEEE*, 17(3):49–58, 1998.
11. Li-Ming Su, Balazs P Vagvolgyi, Rahul Agarwal, Carol E Reiley, Russell H Taylor, and Gregory D Hager. Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3d-ct to stereoscopic video registration. *Urology*, 73(4):896–900, 2009.
12. Xun Luo, Tiffany Kline, Heidi C Fischer, Kathy A Stubblefield, Robert V Kenyon, and Derek G Kamper. Integration of augmented reality and assistive devices for post-stroke hand opening rehabilitation. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 6855–6858. IEEE, 2005.
13. Rodrigo AM Braga, Marcelo Petry, Antonio Paulo Moreira, and Luis Paulo Reis. A development platform for intelligent wheelchairs for disabled people. In *International Conference on Informatics in Control, Automation and Robotics*, pages 115–121, 2008.
14. F Alonso-Martin and Miguel A Salichs. Integration of a voice recognition system in a social robot. *Cybernetics and Systems: An International Journal*, 42(4):215–245, 2011.
15. Mary Ellen Foster, Tomas By, Markus Rickert, and Alois Knoll. Human-robot dialogue for joint construction tasks. In *Proceedings of the 8th international conference on Multimodal interfaces*, pages 68–71. ACM, 2006.
16. L Seabra Lopes and A Teixeira. Human-robot interaction through spoken language dialogue. In *Intelligent Robots and Systems, 2000.(IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on*, volume 1, pages 528–534. IEEE, 2000.
17. Eric Becker, Zhengyi Le, Kyungseo Park, Yong Lin, and Fillia Makedon. Event-based experiments in an assistive environment using wireless sensor networks and voice recognition. In *Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments*, page 17. ACM, 2009.
18. P. Fitzpatrick, G. Metta, and L. Natale. Towards long-lived robot genes. *Robotics and Autonomous Systems*, 56(1):29–45, 2008.
19. Santiago Morante. Interfaz y librera para visión artificial, navegación y seguimiento en robótica. Undergraduate's honors thesis, Universidad Carlos III de Madrid, Dpto. Ing. Sistemas y Automática, June 2012.

20. Juan G. Victores, Santiago Morante, Alberto Jardón, and Carlos Balaguer. Give me the red can: Assistive robot task creation through multi-modal interaction. In *V Congreso Internacional de Diseño, Redes de Investigación y Tecnología para todos (DRT4ALL)*, Sept. 2013.