MULTIMODALLY CONTROLLED INTELLIGENT ASSISTIVE ROBOT

Z. Kazi, M. Beitler, M. Salganicoff, S. Chen, D. Chester and R. Foulds
Applied Science and Engineering Laboratories, University of Delaware/A.I. duPont Institute
Wilmington, DE, USA

Abstract

The Multimodal User Supervised Interface and Intelligent Control (MUSIIC) project is working towards the development of an assistive robotic system which integrates human-computer interaction with reactive planning techniques borrowed from artificial intelligence. The MUSIIC system is intended to operate in an unstructured environment, rather than in a structured workcell, allowing users with physical disabilities considerable freedom and flexibility in terms of control and operating ease. This paper reports on the current status of the MUSIIC project.

Background

One of the most challenging problems in rehabilitation robotics is the design of an efficient control mechanism that allows users with motor disabilities to manipulate her environment in an unstructured domain. Rehabilitation robotics research literature describes many demonstrations of the use of robotic devices by individuals with disabilities [1, 2]

Prototype interfaces have taken two approaches to achieving effective use by individuals with disabilities. Some are command oriented where the users activate the robot to perform pre-programmed tasks [3, 4]. In contrast, there have been a number of projects in which the user directly controls all the movements of the manipulator much like a prosthesis [5, 6].

While direct control allows the user to operate in an unstructured environment, problems such as physical and cognitive load on the user, the requirement of good motor dexterity of the user and many other real-time perceptual and motor requirements preclude an efficient and useful assistive robot.

Command based systems also pose significant problems [7]. While modern speech recognizers provide access to large numbers of stored commands, effective command of a robot will require use of more commands than is reasonable for the user to remember. As the number of possible commands grows, the human/machine interface becomes increasingly unmanageable. Crangle and Suppes propose greatly expanding the capability of the robot to not only recognize spoken words, but also understand spoken English sentences [8].

A different approach to command-based robot operation was proposed by Harwin et al [9]. A vision system viewed the robot's workspace and was programmed to recognize barcodes that were affixed on each object and used the barcodes to determine the location and orientation of every item. While this was only successful within a limited and structured environment, it did demonstrate the dramatic reduction in *machine intelligence* that came by eliminating the need for the robot to perform object recognition and language understanding.

At the other extreme of robot control are the completely autonomous systems that perform with effectively no user supervision, the long elusive goal of AI, robotics and machine vision communities. Unfortunately, this goal seems far from practical at this point, although many important incremental advances have been forthcoming in the past decades. Furthermore, absolute automation poses a set of problems stemming from incomplete *a priori* knowledge about the environment, hazards, insufficient sensory information, inherent inaccuracy in the robotic devices and the mode of operation.

Objective

Therefore, what one should strive for is a synergistic integration of the best abilities of both *"humans"* and *"machines"*. Humans excel in creativity, use of heuristics, flexibility and "common sense', whereas machines excel in speed of computation, mechanical power and ability to persevere. While progress is being made in robotics in areas such as machine vision and sensor based control, there is much work that needs to be done in high level cognition and planning. We claim that the symbiosis of the high level cognitive abilities of the human, such as object recognition, high level planning, and event driven reactivity with the native skills of a robot can result in a human-robot system that will function better than both traditional robotic assistive systems and autonomous systems.

Our MUSIIC strategy overcomes the limitations of previous approaches by integrating a multimodal RUI (Robot User Interface) and a semi-autonomous reactive planner that will allow users with severe motor disabilities to manipulate objects in an unstructured domain. The multimodal user interface is a speech and deictic (pointing) gesture based control that guides the operation of a semi-autonomous planner controlling the assistive robot.

MUSIIC utilizes a stereo-vision system to determine the three-dimensional shape and pose of objects and surfaces which are in the environment, and provides an object-oriented knowledge base and planning system which superimposes information about common objects in the three-dimensional world [10, 11]. This approach allows the user to identify objects and tasks via a multimodal user interface which interprets her deictic gestures and speech inputs. The multimodal interface performs a critical disambiguation function by binding the spoken words to a locus in the physical work space. The spoken input is also used to supplant the need for general purpose object recognition. Instead, three-dimensional shape information is augmented by the user's spoken word, which may also invoke the appropriate inheritance of object properties using the adopted hierarchical object-oriented representation scheme.

Method

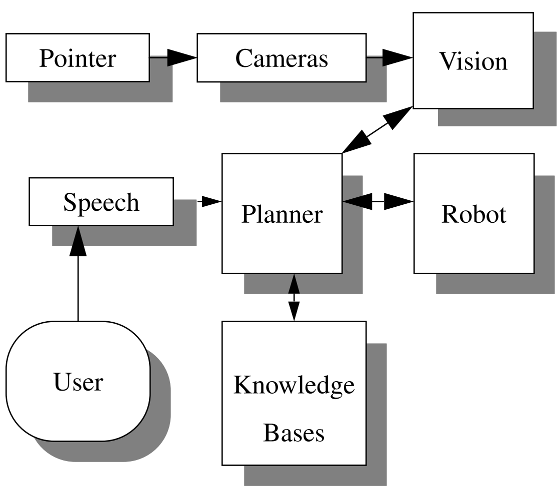

Figure 1: System Configuration

The previous sections lead naturally to a description of the essential components of the MUSIIC system [Figure 1]. We require a ***planner*** that will interpret and satisfy user intentions. The planner is built upon ***object oriented knowledge bases*** that allow the users to manipulate objects that are either known or unknown to the system. A ***speech input*** system is needed for user inputs, and a ***gesture identification*** mechanism is necessary to obtain the user's deictic gesture inputs. An ***active stereo-vision*** system is necessary to provide a snap-shot of the domain; it returns object shapes, poses and location information without performing any object recognition. The vision system is also used to identify the focus of the user's deictic gesture, currently implemented by a laser light pointer, returning information about either an object or a location. The planner extracts user intentions from the combined speech and gesture input. It then develops a plan for execution on the world model built up from the *a priori* information contained in the knowledge bases, the real-time information obtained from the vision system, the sensory information obtained from the robot arm, as well as information previously extracted from the user dialog. Prior to execution, the system allows the user to preview and validate the planner's interpretation of user intentions via a 3-D graphically ***simulated environment*** [12]. Figure 2 shows the actual system set-up.

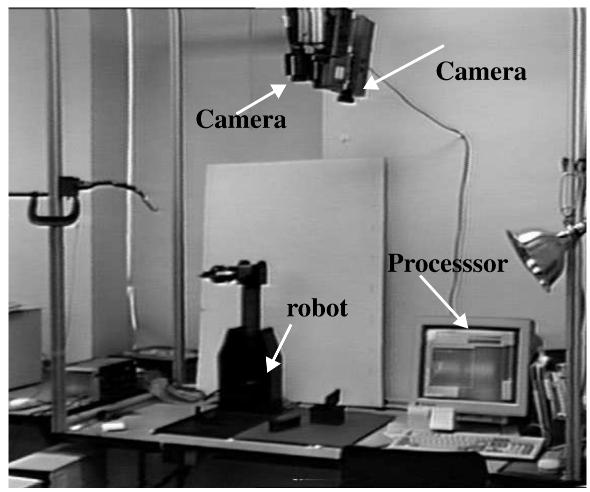

Figure 2: Physical set-up

Result and Illustration

The current operational implementation of MUSIIC is able to manipulate objects of generic shapes at arbitrary locations. A set of robot control primitives are used to build up higher level task commands with which the user instructs the assistive robot. The robot primitives include approaching, grasping and moving an object amongst others. The vision system first takes a snap shot of the domain and returns to the planner object sizes, shapes and locations. This information is then combined with the knowledge base of objects to model the workspace in question. The user then points to objects using a laser light pointer while verbally instructing the robot to manipulate an object.

For example, the user may say "`Put that here`", while pointing at an object as she says "`that`" and pointing to a location as she says "`here`". First, the combined gesture and verbal deictic is interpreted by the planner based on information extracted from the vision system as well as the object knowledge base. The planner then uses the plan knowledge base to approach and grasp the object and then move the object to the desired location.

In addition to high level commands as illustrated above, the user is also able to instruct the robot at a lower level, by commands such as "`move there`", "`open gripper`", "`move down`", "`close gripper`", "`move here`" to obtain the same functionality as the "`move that here`" instruction.

Discussion

While MUSIIC is still very much a work in progress, the current test-bed implementation has amply demonstrated the flexibility in use of an assistive robot achievable by our multimodal RUI built on top of an intelligent planner. Work is continuing in fleshing out the complete object hierarchy that will allow the planner to plan tasks at any level of specialization, from objects about which nothing is known except what the vision system returns to objects which are well known, such as a cup often used by the user. The reactive component is also nearing completion. Reactivity will be achieved in two ways: An autonomous run-time reactivity will be obtained through sensor fusion and a human centered reactivity will be used where the user can take over the planning process when the planner fails to make correct plans as a consequence of incomplete information or catastrophic failures. The user will engage in a dialog with the system, either to update the knowledge bases or to perform plan correction or editing.

Conclusion

Human intervention as well as an intelligent planning mechanism are essential features of a practical assistive robotic system. We believe our multimodal RUI is not only an intuitive interface for interaction with a three-dimensional unstructured world, but it also allows the human-machine synergy that is necessary for practical manipulation in a real world environment. Our novel approach of gesture-speech based human-machine interfacing enables our system to make realistic plans in a domain where we have to deal with uncertainty and incomplete information.

References

[1] Foulds RA. Interactive Robotics Aids-One Option for Independent Living: an International Perspective. World Rehabilitation Fund, 1986.

[2] Bacon DC, Rahman T, Harwin WS, eds. Fourth International Conference on Rehabilation Robotics: AI DuPont Institute, Wilmington, Delaware, USA, 1994. Applied Science and Engineering Laboratories.

[3] Fu C. An independent vocational workstation for a quadriplegic. In: Foulds R, ed. Interactive Robotics Aids-One Option for Independent Living: an International Perspective, volume 1 of 37. World Rehabilitation Fund, 1986;42.

[4] van der loos M, Hammel J, Lees D, Chang D, Schwant D. Design of a vocational assistant robot workstation. Annual report. Palo Alto VA Medical Center, Palo Alto, CA: Rehabilitation Research and Development Center, 1990.

[5] Zeelenberg A. Domestic use of a training robot-manipulator by children with muscular dystrophy. In: Foulds R, ed. Interactive Robotic Aids-One Option for Independent Living: An International Perspective, volume Monograph 37. World Rehabilitation Fund, 1986;29–33.

[6] Kwee H. Spartacus and manus: Telethesis developments in france and the netherlands. In: Foulds R, ed. Interactive Robotic Aids-One Option for Independent Living: An International Perspective, volume Monograph 37. World Rehabilitation Fund, 1986;7–17.

[7] Michalowski S, Crangle C, Liang L. Experimental study of a natural language interface to an instructable robotic aid for the severely disabl. In Proc. of the 10th Annual Conf. on Rehabilitation Technology 1987;466–467.

[8] Crangle C, Suppes P. Language and Learning for Robots. Stanford, CA: CSLI Publications, 1994.

[9] Harwin W, Ginige A, Jackson R. A potential application in early education and a possible role for a vision system in a workstation based robotic aid for physically disabled persons. In: Foulds R, ed. Interactive Robotic Aids-One Option for Independent Living: An International Perspective, volume Monograph 37. World Rehabilitation Fund, 1986;18–23.

[10] Kazi Z, Beitler M, Salganicoff M, Chen S, Chester D, Foulds R. Multimodal user supervised interface and intelligent control (MUSIIC) for assistive robots. In: 1995 IJCAI workshop on Developing AI Applications for the Disabled. 1995;47–58.

[11] Kazi Z, Beitler M, Salganicoff M, Chen S, Chester D, Foulds R. Intelligent telerobotic assistant for people with disabilities. In: SPIE's International Symposium on Intelligent Systems: Telemanipulator and Telepresence Technologies II: SPIE, 1995.

[12] Beitler M, Foulds R, Kazi Z, Chester D, Chen S, Salganicoff M. A simulated environment of a multimodal user interface for a robot. In: RESNA 1995 Annual Conference. Vancouver, Canada: RESNA press, 1995;490–492.

Acknowledgments

Work on this project is supported by the Rehabilitation Engineering Research Center on Rehabilitation Robotics, National Institute on Disabilities and Rehabilitation Research Grant #H133E30013, Rehabilitation Services Administration Grant #H129E20006 and Nemours Research Programs.

Author Address

Zunaid Kazi
Applied Science and Enginnering Laboratories
University of Delaware/A.I. duPont Institute
1600 Rockland Road, Wilmington, DE 19899
Phone: (302) 651-6830 / Fax: (302) 651-6895
WWW URL: http://www.asel.udel.edu/~kazi/