Grasping at Straws: An Intelligent Multimodal Assistive Robot

Zunaid Kazi, Shoupu Chen, Matthew Beitler, Daniel Chester and Richard Foulds
Applied Science and Engineering Laboratories
duPont Hospital for Children and the University of Delaware
http://www.asel.udel.edu/robotics/musiic/

Abstract

The goal of the Multimodal User Supervised Interface and Intelligent Control (MUSIIC) project is to develop an assistive manipulative robotic system for use in an unstructured environment by people with manipulatory disabilities. MUSIIC integrates a speech and pointing based human-machine interface with autonomous planning techniques from artificial intelligence. This paper describes the current status of MUSIIC.

Background

Several studies have demonstrated the need for the development of a general purpose manipulation aid for use by people with disabilities to perform everyday activities [1, 2]. Rehabilitation robotics research literature describes many demonstrations of the use of assistive robotic devices [3, 4, 5]. Prototype interfaces have generally taken two approaches; some are command oriented where the user activates the robot to perform preprogrammed tasks while others are control oriented where the user directly controls all the movements of the manipulator much like a prosthesis.

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 are hindrances towards the development of an efficient and useful assistive robot.

Command based systems also pose significant problems. 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.

At the other extreme of robot control are autonomous systems, the long elusive goal of the Artificial Intelligence (AI), robotics and machine vision communities. Although many important incremental advances have been made in the past decades, this goal seems far from being realizable at this point. Furthermore, absolute automation poses a set of problems such as incomplete *a priori* knowledge about the environment, hazards, insufficient sensory information, and inherent inaccuracy of the robotic devices themselves.

The MUSIIC Solution

The Multimodal User Supervised Interface and Intelligent Control (MUSIIC) project uses a novel multimodal (speech and gesture) human-machine interface built on top of a reactive and intelligent knowledge-driven planner that allows people with disabilities to perform a variety of novel manipulatory tasks on everyday objects in an unstructured environment.

MUSIIC demonstrates that by combining current state of the art in natural language processing, robotics, computer vision, planning, machine learning, and human-computer interaction. A practical robotic assistant can be built without having to solve the major problems in each of these fields, i.e., full text understanding, autonomous robot arm control, real-time object recognition in an unconstrained environment, planning for all contingencies and levels of problem difficulty, speedy supervised and unsupervised

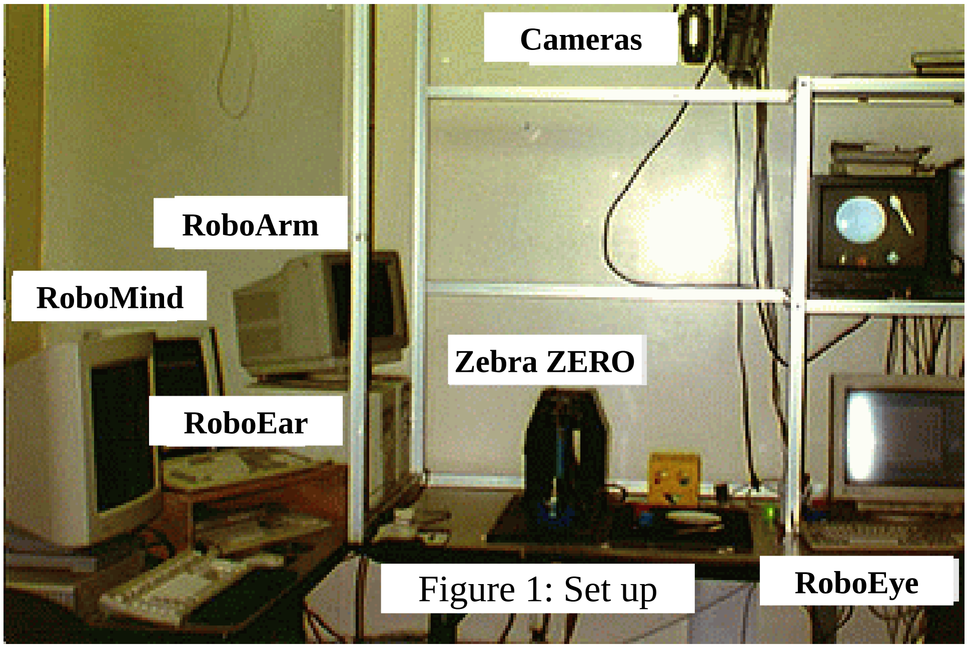

Figure 1: Set up

learning, and intelligent human-computer interfaces. Current solutions to these problems, when combined with each other and with the intelligence of the user, can compensate for the inadequacies that each solution has individually.

The MUSIIC Method

MUSIIC includes a knowledge driven planning subsystem in which objects are represented in an increasingly specialized sequence of object classes in an inheritance hierarchy [6]. The MUSIIC architecture is based on three knowledge bases: a hierarchical knowledge base of objects (WorldBase), a knowledge base of objects in the actual domain of operation (DomainBase), and a knowledge base of user extendible plans (PlanBase). Each object, depending on the degree of generalization, has a set of attributes such as shape, size, dimensions, weight, approach point, grasp points, constraints and plan fragments. The intelligent planner uses the three knowledge bases and user/sensor provided feedbacks, to synthesize robot plans.

MUSIIC also includes a vision subsystem for the determination of the three-dimensional shape, pose and location of objects in the domain [7]. *No object recognition is performed.* A human-machine interface subsystem uses a multimodal input schema where users of the system use deictic gestures (pointing, achieved by a head mounted laser pointer) to indicate locations, and spoken commands to identify objects and specific actions.

Illustration

As shown in Figure 1, the actual hardware setup includes a vision subsystem (RoboEye), containing a pair of color cameras, an SGI workstation, and associated vision software, a six degree of freedom Zebra ZERO robot and controller (RoboArm), a speech recognition subsystem (RoboEar), and the planning and knowledge base system (RoboMind). These reside in different computing platforms and communicate with each other through Remote Procedure Calls (RPC) components.

The operation of MUSIIC is illustrated through two annotated task scenarios. These scenarios involves the task of inserting a straw into a cup and bringing the cup to the user. The workspace contains a cup and a straw and the WorldBase contains entries for “straws” and “cups”.

Scenario 1

Instruct the system to load in the plan library.

User: "Load plans"
MUSIIC: "Plan loading complete"

Instruct the system to synchronize the various system components.

User: "Synchronize"
MUSIIC: "Synchronization complete"
User: "Home"

The robot then moves to it's home configuration.

MUSIIC: "Home successful"
User: "Scan"

The vision system generates object size, position, orientation and color information.

MUSIIC: "Scanning complete"

Instruct the vision system to transfer the information to the planning subsystem to build up the DomainBase.

User: "Load domain"
MUSIIC: "Domain loading complete"

The user points to the straw while simultaneously saying the word "straw".

User: "That's a straw"
MUSIIC: "Looking for the straw"

MUSIIC searches the WorldBase for the "straw".

MUSIIC: "I found the straw"

User points to the cup and identifies it to the system.

User: "That's a cup"
MUSIIC: "Looking for the cup"
MUSIIC: "I found the cup"

Instruct the robot to insert the straw into the cup.

User: "Insert the straw into the cup"

MUSIIC inserts the straw into the cup. On success:

MUSIIC: "I am ready"

Instruct the robot to bring the cup to the user.

User: "Bring the cup"

The arm approaches the cup and grasps it by the rim. It then brings the cup to a predetermined position that is accessible to the user.

Scenario 2

User: "Bring that"

Here we present an alternate scenario where the user did not explicitly identify the cup. The user points to the cup while simultaneously saying the verbal deictic "that". RoboEye continuously records any identified spot along with time-stamp values that mark the time when the spot was recorded. The speech system also time-stamps utterances and these values are used to determine the location of the spot that was generated when a verbal deictic such as "that" is spoken. The system then finds the object that is in that specific location and invokes the "bring" task. Since the object has not been identified specifically, planning is based on general principles. Instead of grasping it by the rim the arm grasps the cup along its width and brings it to the user.

Conclusion

Human intervention as well as an intelligent planning mechanism are essential features of a practical assistive robotic system. We believe our multimodal user interface is an intuitive one for interaction with a three-dimensional unstructured world that also allows the human-machine synergy that is necessary for practical manipulation in a real world environment. The hierarchical object oriented knowledge base allows the planner to synthesize plans for object manipulation tasks solely from shape, pose and location information obtained from the vision system. By engaging in dialogue with the user in such a way that natural deictic gestures and voice input can be used to carry out a task, MUSIIC makes general purpose object recognition unnecessary.

Since MUSIIC is “instructable” [8], the user is capable of giving verbal and gestural instructions to the robot and teaching it new tasks in a familiar environment. MUSIIC is also capable of adapting previously learned plans and tasks to new situations and objects.

Reference

[1] Stanger CA, Anglin C, Harwin WS, Romilly DP. Devices for assisting manipulation: A summary of user task priorities. IEEE Trans. Rehabilitation Engineering 1994;2(4):256–265.

[2] Batavia A, Hammer G. Consumer criteria for evaluating assistive devices: Implications for technology transfer. In: 12th annual Conference on Rehabilitation Technology. Washington DC: RESNA Press, 1989;194–195.

[3] Foulds RA, ed. Interactive Robotics Aids-One Option for Independent Living: an International Perspective, Monograph 37. World Rehabilitation Fund, 1986.

[4] 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.

[5] der Loos MV, 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.

[6] 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.

[7] Chen S. World information through color images in task planning for a rehabilitation robot. Technical Report ROB9605. A.I. duPont Institute / University of Delaware, Wilmington, DE 19899: Applied Science and Engineering Laboratories, 1996.

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

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 of the United States Department of Education and Nemours Research Programs.

Author address

Zunaid Kazi
Applied Science and Engineering Laboratories
duPont Hospital for Children / University of Delaware
1600 Rockland Road,
Wilmington, DE 19899
Email: kazi@asel.udel.edu
Phone: (302) 651-6830
Fax: (302) 651-6895