Achieving Virtual Presence with a Semi-Autonomous Robot Through a Multi-Reality and Speech Control Interface

Kristian T Simsarian
kristian@sics.se
Jussi Karlgren
Jussi.Karlgren@sics.se
Lennart E. Fahlén
lef@sics.se
Ivan Bretan
ivan.bretan@sics.se
Emmanuel Frécon
emmanuel@sics.se
Tomas Axling
axling@sics.se
Niklas Frost
frost@sics.se
Lars Jonsson
jonsson@sics.se

Swedish Institute of Computer Science
S-16428 KISTA
Stockholm, Sweden
fax: +46 (8) 751-7230
tel: +46 (8) 752-1570

Abstract. This paper describes a model for a complex human-machine system where a human operator controls a remote robot through the mediation of a distributed virtual environment with a language interface. The system combines speech controlled graphical immersive environments with the live video from a robot working in a real environment. The worlds are synchronized and updated based on operator selections, commands and robot actions. This system allows the user to have a powerful tool with a high level of abstraction to create and control autonomous robots, thus making possible the realization of single and possibly multiple real-world autonomous robot applications.

1 Introduction

In this paper we describe our current work to construct a high-level remotely operated robot system. Control is acehieved via a high-level interface supplemented by language and gesture control within a graphical immersive environment containing live video of the remote space where the robot is situated.

The robot handles the *perception-action* of the human-machine system, the virtual environment is a model of the *world knowledge* of the system, and the interface, with both language and gesture interaction provides the tools for *interaction* thus manipulating the robot and the knowledge state of the entire system.

The virtual model is is able to display the current status of the system's awareness of objects and available actions and to specify high-level tasks, such as point-to-point navigation and pick-and-place manipulation, while the robot has the basic physical and perceptual skills to perform low level navigation in the form of path-planning and obstacle avoidance, and some vision processing. At the same time, the interaction between the user and the robot system is on a high level of abstraction facilitated by the combination of the use of a virtual environment, natural language and gesture interface. This releases the human operator from low-level tasks of robot control, and allows the operator to specify tasks in a high-level manner for possibly a number of robots.

The applications for completely autonomous robots are manifold; however, in the world today mobile robot high-level task planning is difficult. To have a robot perform complex tasks requires guidance or guided instruction from a human aide or controller. Designing a system to support a human guide for robot learning involves complex design decisions on several levels: firstly, the human guide needs information on the physical surroundings of the robot; secondly, the guide needs to be given a reasonable rendition of the robot's current awareness of those surroundings; thirdly, the guide needs to be given a useful and understandable mechanism to interact with the model and the robot's knowledge representation in order to be able to specify objects, entities, and tasks for the mobile robot. The above requirements are achieved by using the interface described in here.

This paper describes the framework we are using for the virtual and real world combination and demonstrates the principle which we are applying to performing remote tasks within a new immersive paradigm. This paradigm uses interaction mechanisms that will not limit the operator to low level manipulation.

2 Example Scenario

The human operator interacts with an immersive environment which represents a model of a remote real environment (see figure 2). In the virtual environment the operator has access to a mobile robot physically situated in the remote real environment. The robot is a vehicle that has the ability to carry the operator through both the virtual and remote-real worlds. The interface between user and robot in the virtual world consists of a real-time video view of the robot's real world environment accompanied by an interface control panel (figure in Appendix). The interface control panel consists of buttons, displays, and various data to aid the interaction. Some of the tools available can set robot speed, acceleration parameters, or command various image transforms. There are also additional displays that can give current real world robot and environment state, i.e. battery supply or radioactivity level. The robot has the ability to move through the world and can be controlled on a high level of abstraction by the operator through the interface. The operator gives simple instructions to navigate, such as “go there” accompanied by a pointing gesture in the virtual world. Or, alternatively, the user might give a more sophisticated command, such as “move toward the doorway” accompanied by the context of the robot's current view. Because the virtual model is at least roughly synchronized to the real world and information about specific doors are contained in the virtual environment model, this command can quickly be translated into navigational commands for the robot base.

One important difference in our model of robot control is that as the user

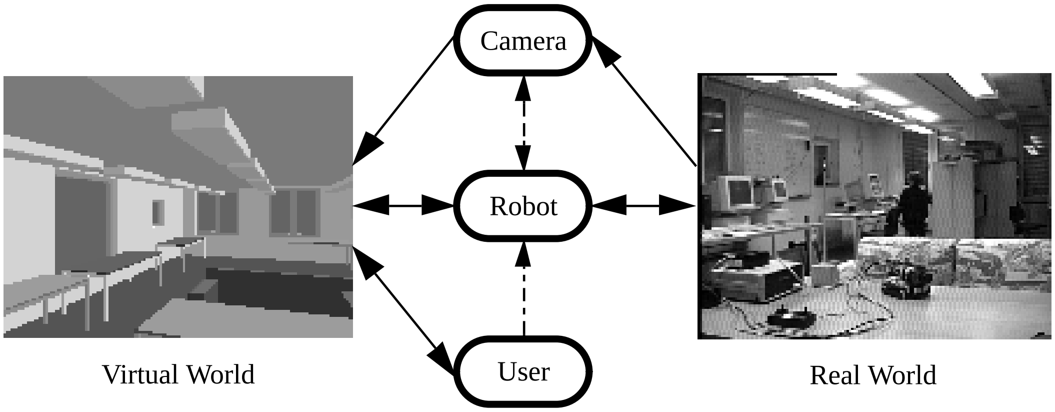


Figure 1: This figure shows the flow of information between the different system components. The solid lines indicate direct flow of information while the dotted lines indicate an indirect flow (see text).

interacts with this robot control interface, the machine is permitted to say "I don't know." The robot does not have to make high-level decisions, instead it performs as well as it can and always has the ability to return to the user with questions. Allowing this degree of relaxation in autonomy releases the system from many of the hardest problems in AI while simultaneously allowing us to build machines that can perform useful tasks and providing a novel platform for further research in autonomous robotics, sensing, man-machine interaction, and virtual environments.

For example, if our virtual model did not include a complete description for the real-world object that the user reference in the camera image, e.g. a book, then given a rough localization the robot could perform a vision process on the image to fit the right aspect ratios for a book at that distance and pose. When there is an ambiguity, the robot might ask "is this the book you mean?" while highlighting what it estimates to be the book boundary in the image. In some situations the robot might respond with a number of alternatives for the book including, among the alternatives, a box. These mistakes are, at least initially, allowed in the interface until visual recognition techniques have caught up with current needs. The selection of these alternatives would be part of the user-robot interaction. Thus the user would learn how to use the robot given its deficiencies. For future use, the user could segment out the book for future recognition and identification. The image texture of the book can also be used in the virtual world for aiding the identification by the operator and enriching the simulated environment, and also supplying visual features for the vision processes.

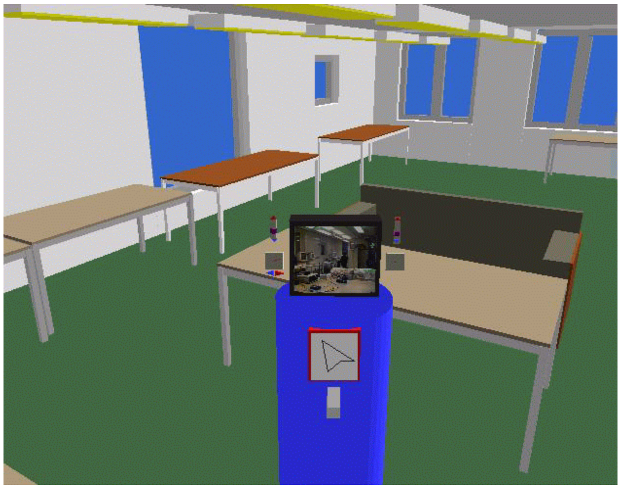


Figure 2: This figure displays a view of the immersive environment. It is a model of the laboratory space in which the robot is situated. In the center the virtual robot representation can be seen. On the robot sits a live camera view into the real world.

3 Robots, Operators, and Interaction

3.1 Autonomous Robotics

It is easy to see how having the capability to send autonomous robots into hazardous and remote environments (e.g. space, sea, volcanic, mine exploration, nuclear/chemical hazards), would be useful. Robotics can be built to stand higher environmental tolerance than humans, they can be built to perform specialized tasks efficiently and they are expendable. To this end there has been much research on fully autonomous robots that can navigate into an area, perform actions such as taking samples, performing manipulations, and return the information to base controllers without human assistance.

3.2 Teleoperated Robotics

Relatively independently from research in autonomous robotics, on the other end of mobile robot research, the field of teleoperated robotics has worked on the human-machine interface to enable an operator to control a robot remotely in space [24], and battlefield[4] applications and even used simulated environments for predictive task planning[20]. Some researchers have tried to bridge this gap from both directions. An autonomous robotics group has taken a schema-based reactive architecture and used this as a base-level for teleoperated control. In their architecture the mobile robot performs simple navigation while the operator's commands can be situated in the system either as another behavior that influences navigation or as a more global process that manipulates system parameters[2]. They have since extended this idea to allow the operator to control group behaviors in a multi-agent environment[3]. Other groups have recognized the need for tele-operators to move away from low-level robot movement control. One effort has created a multi-level architecture for robots to provide higher level navigation functions such as path-planning and obstacle avoidance with operator guidance[10].

3.3 Immersive Interfaces

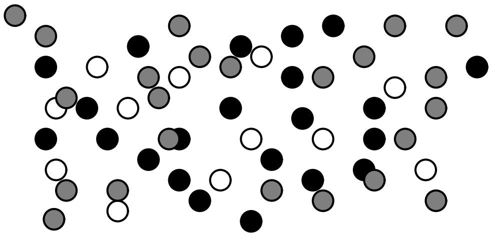
Meanwhile, recent research in immersive virtual environments and human-computer interaction at SICS has worked on building a framework for natural interaction. One aspect of this work has been the study of interaction between agents, human and others *in* a shared virtual environment [5]. Another aspect has been the building of mechanisms for human users to interact *with* the virtual environment [17]. We are using an immersive virtual environment as the interaction paradigm with the real world and the robot. Specifically our work is an application in the SICS Distributed Interactive Virtual Environment (DIVE) system[12].

3.4 Augmented Reality

We incorporate on-board video from the robot into the virtual world. The video can subsequently be enhanced and augmented to communicate information to the operator. This is quite similar to work in *Augmented Reality*, which at is base is the concept of combining real world images (e.g. video) with graphic overlays. Augmented reality techniques have been demonstrated in a variety of practical applications. Examples of these are displaying graphical information in printer repair operations[15]. Or displaying information from a 3-D model base on a real world artifact [26]. Other applications include virtual tape-measurement on real scenes as well as graphical vehicle guidance[23], and enhanced displays for teleoperated control[22]. All of these separate applications are relevant for our robot application. Additionally the reverse operation can be performed, the virtual environment can also be augmented in a similar fashion by real-world images.

4 Integrating Perception, Knowledge, and Interaction

There are four main distributed computational subsystems in the complete system. These are the robot system – the physical layer; the computer vision processing system – the perceptual layer; the graphical object database and rendering system – the knowledge model; the speech and language processing and the graphical object manipulation system – the interaction layer. The information that is passed around can also be viewed as flowing between the real and



"Select the grey marbles."

Figure 3: An example where natural language commands have significant advantage over "point and click."

"Where is the paper about virtual reality and robotics I sent to the workshop in Monte Carlo last fall?"

Figure 4: Try this with gestures.

virtual worlds via the camera, the robot and the user. This flow of information can be visualized in figure 1.

4.1 Physical competence: Robotics

The robot exists in the real world. The robot is endowed with a basic model of the environment from an architectural-type drawing of the basic physical world structure and artifacts, and through movement and exploration, the robot has the ability to augment this model, with new objects, methods and descriptions that are more useful for comparison to its sensor data.

The robot that we are using for this system is a Real World Interface B21 robot with on-board processing and sensing. In this system this robot can perform basic tasks such as navigate around obstacles, recognize objects to the best of its ability and take its high level commands from a human, thus displacing the artificial high-level planner with a human one.

The robot can perform basic point-to-point navigation tasks and avoid basic obstacles while negotiating the indoor structured environment in which it is situated.

Using the robot encoders and periodic self-localization such as described in [28, 21] to account for drift, the robot can synchronize with the model represented in the virtual reality. Having access to a model of the environment as well as access to the operator's knowledge and assistance gives great leverage on the harder problems of robot navigation, such as localized sensor pitfall situations.

This type of hybrid human-machine system will give the operator something we are calling *virtual presence*, where the operator is present by way of a machine proxy, or by way of a virtual rendition of the surroundings, dependent on the perspective we take.

4.2 Perceptual processing: Vision

The robot sends the data it perceives to the vision processing system. The output from the vision processing system is sent further to the knowledge level of the system to be encorporated the virtual model of the world.

General visual segmentation of real world images continues to be a hard problem. However in this situation we are assisted in this tasks by both having some knowledge of appearance (through saved images of the real world such as texture maps), approximate location and the user. In many instances this harder problem of visual segmentation breaks down to an easier problem of verification and fine localization. The user can also interact with the vision processing to aid in the segmentation, identification and localization. Using graphical interaction tools such as "snakes" (an interaction method that allows a user to roughly identify a region which then shrinks to surround the nearest edge) the user can be facilitated in performing hand segmentation of the image. These images that are cut out the environment are used for both identification by the vision processor on the video images as well as identification by the user within the virtual environment. This part of the system is work in progress.

4.3 Knowledge model: The Augmented Virtual World

The video from the camera flows from the real world to the virtual world. These images represent the real-world from a robot-centered perspective. The user sends commands to the robot via the interface, thus these commands are made by the operator interacting with the virtual world. The commands may be as simple as updating velocity and orientation or may also be higher-level and more complex involving path specification, navigational targets, and grasping tasks.

The present interface takes video from the world and brings it into the virtual world which allows the possibility to superimpose graphics on video. Such an *augmented reality* interface can display information that may not be visible, but useful in the real world, thus augmenting the information present in the video channel with navigational aid information or nonvisual sensory data. It is also a convenient way of displaying to the operator what the current state of the system is: items in the environment could be graphically emphasized or deemphasized (dimmed), based on the needs of the operator, robot and task at hand. For example a book in the physical environment that is recognized and localized by the vision processor could be colored and emphasized for the operator to know that this object is known and could possibly be interacted with. If a object is not recognized, it is chance for the user to advise the system what the object in question is.

In addition to this standard notion of augmented reality, we also have the power to perform the complement operation, embellishing the virtual environment with real-world images. With the proper, possibly user-guided, feature extraction and image warping we can decorate the world with much of the richness of the real world. Thus, objects that the operator needs to interact with can be visually more informative than the pure virtual reality system would allow them to be.

Thusly the virtual world represents the knowledge state of the robot. It is not intended to be perfect or complete. Within this framework many of these methods for real and virtual world interaction for both the user and robot are embedded and distributed in the world itself (see section 4.4.2).

4.4 Interaction with the Operator

The virtual world serves as the communication medium between the robot and the user. It is through the interaction of the robot with the virtual environment and the operator with the virtual environment that interaction between the operator and the robot can take place. Thus bi-directional communication and command specification is achieved via the virtual world. In complement to operator commands, the robot can make queries of the operator regarding task direction as well as update the environment with objects and model features discovered in the course of exploration.

Our interface design is multimodal – meaning that it makes use of live video, 3-D graphics, gestures, menu choice, speech, and text as input and output channels. Language and graphics (or, indeed, any abstract and any analog manipulable representation) complement each other, in the sense that tasks of different types require different modalities [14], and that users have varying preferences for modalities or differing capacity to make use of them [16].

DIVERSE (DIVE Robust Speech Enhancement) is a speech interface to the virtual reality platform DIVE. DIVERSE is developed at SICS for use as a test system to experiment with multimodal interaction[17]. DIVERSE allows for spoken language control of operations that are normally carried out through direct manipulation in DIVE, such as transportation of objects, change of view, object creation, deletion, colouring etc, while still retaining the possibility to perform actions through direct manipulation whenever that is more suitable [16].

Interaction in DIVERSE is mediated through an animated agent to allow explicit modeling of the linguistic competence of the system, both in terms of output language and in terms of the gestures.

4.4.1 Spatial model

Inside the virtual environment that the DIVE system implements, there is a strong model of spatial interaction[7],[6]. This model provides a method of interaction for the operator, the robot, and the objects within the virtual and real worlds. In this section this spatial interaction model and the methods it suggests are described.

Here we summarize key concepts which constitute the DIVE and DIVERSE model of interaction, the details for this model can be found in [6] and [17]. The goal of the spatial model is to provide a small but powerful set of mechanisms for supporting the negotiation of interaction across shared virtual space. The spatial model, as its name suggests, uses the properties of space as the basis for mediating interaction. We briefly introduce the key abstractions of *space,* *objects, aura, awareness, focus, nimbus*, and *boundaries* which define part of the spatial model, and the concepts of *interlocutor* and *discourse compost* which are central to the linguistic interaction.

Aura is defined to be a sub-space which effectively bounds the presence of an object within a given medium and which acts as an enabler of potential interaction. Objects carry their auras with them when they move through space and when two auras collide, interaction between the objects in the medium becomes a possibility. It is the surrounding environment that monitors for aura collisions between objects.

Once aura has been used to determine the potential for object interactions, the objects themselves are subsequently responsible for controlling these interactions. This is achieved on the basis of quantifiable levels of **awareness** between them. Awareness between objects in a given medium is manipulated via **focus** and **nimbus**, further subspaces within which an object chooses to direct either its presence or its attention. More specifically, if you are an object in space the following examples help define the concept:

focus -the more another object is within your focus, the more aware you are of it;

nimbus -the more another object is within your nimbus, the more aware it is of you.

This notion of spatial focus as a way of directing attention and hence filtering information is intuitively familiar from our everyday experience (e.g. the concept of a visual focus). The notion of nimbus requires a little more explanation. In general terms, a nimbus is a sub-space in which an object makes some aspect of itself available to others. This could be its presence, identity, activity or some combination of these. Nimbus allows objects to try to influence others (i.e. to be heard or seen). Nimbus is the necessary converse of focus required to achieve a power balance in interaction.

Awareness levels are calculated from a combination of nimbus and focus. Aura, focus and nimbus may most often be implicitly manipulated through fundamental spatial actions such as movement and orientation. Additionally, aura, focus and nimbus may be manipulated through boundaries in space. Boundaries divide space into different areas and regions and provide mechanisms for marking territory, controlling movement and for influencing the interactional properties of space.

Language usage adds some complexity to the interface. Using language presupposes a *counterpart*, and the design of the dialog will hinge on how the counterpart is conceptualized. There are several conceivable models of the interlocutor [16, 17]. For the present application the robot itself is a natural counterpart for most tasks – the experiments in DIVERSE so far have involved a separately rendered **Agent** to anchor the discourse communicative competence of the system.

Determining what object an utterance refers to is non-trivial in the general case. Using the agent we model the system's conception of the world and the saliency of various objects by displaying a list of referents. This list – the **discourse compost** – is composed by the system giving each object present in the discourse a saliency grade, based on recent mention, highlightedness, gestural manipulation by the user, and above all, visual awareness. So, primarily, if the agent has a high degree of *awareness* of an object, it is a candidate for reference. This effect declines rapidly when the agent becomes less aware of it. Secondly, users can *manipulate* or *point at* an object. An object which the user points at gets a high saliency grade, with a rapid decline after the pointing gesture has been completed. Thirdly, we keep track of which objects have been mentioned. Objects in the recent *dialog history* are likely to be referred to again. The evidence from these different sources is compiled in the compost to determine which objects are likely future referents.

4.4.2 Methods

We use the interaction model to create an interactive and informationally rich immersive environment that stores the methods to aid the robots interaction in the real world. The concepts of aura, nimbus, focus are key to the way the robot interacts with the virtual and real worlds. Using the concepts of spatial boundaries and auras we can define interaction mechanisms and methods for sharing information between the robot and the environment.

For example using the concept of object aura we can define a means of transferring information for navigation and object identification. If the robot's aura collides with an object's aura that object may then open up a channel, i.e. the robot focuses and the object projects nimbus, thus enabling the object to pass information to the robot that would be pertinent to the mutual interaction. In this way each object stores information and methods about itself. This information can include: 1) object identification, 2) object function, 3) navigational approach method, 4) grasping method, 5) recognition method.

These last three types of information deserve special mention. An object may store the actual methods in which to perform a local interaction such as recognition. Given that the position of the object and the position of the robot are well known these methods can be rather specific.

Likewise, using the boundaries in space, various locations in the environment may store information and methods regarding navigation. For example there may be certain areas of the environment where great care must be taken, so crossing a boundary could then act like entering a 'speed control zone' and thus negotiate control for the robot's velocity. Similarly there could also be areas in the environment where certain configurations or specific paths should be avoided or taken. Crossing a boundary into such an area would open up a channel to transfer specific navigational commands to the robot.

Using this model of interaction unweights the robot control process from the need to have knowledge about the entire environment at all times. Using this spatial model we are distributing the processes and specific information throughout the environment. Also using the virtual world as a knowledge model in this way it makes it less necessary for a robot to have much knowledge about a new environment before actually entering it. Thus when the robot crosses the boundary into a new environment the new environment would contain all the necessary global information regarding that world.

4.5 How Do We Actually *Tell* The Robot What To Do?

Hitherto, controlling and manipulating a virtual or augmented reality has mainly been through *direct manipulation*, an interaction paradigm based on immediacy of control and tightly linked feedback between action and effect. Direct manipulation interfaces are generally share three main characteristics of 1. continuous representation of the object of interest, 2. physical actions or labeled button presses instead of complex syntax, and 3. rapid incremental reversible operations, whose impact on the object of interest is immediately visible [27].

These characteristics have usually been seen as standing in contrast to command based interfaces that build on more expressive forms of input such as formal command languages or human languages. While Shneiderman's points certainly have been understood as a justification for a completely analog interface representation such as pictures and graphs, and analog operations such as gestures, points one and three do not in fact in any way contradict the possibility of using text or other language input – indeed, any interface at all, be it language based or point-and-click based would do well to follow the principles. We will use language, in our case speech or typewritten text, as one of the mechanisms of interaction, thus relaxing the constraints posed by Shneiderman's second point, but continuing to observe points one and three. Language, as we will show below, is necessary to manage the level of complexity following from instructing a robot.

4.5.1 Why Charades Are Difficult

Virtual reality offers the user intuitively useful means of selecting and manipulating objects in the vicinity, much as gestures do in real life. Cognitive concepts like "this" and "that" are easily defined and formalized in virtual reality. Human languages are by design a step beyond deixis or the simple acts of ostentation behind "this" and "that". They allow the user to refer to entities other than concrete objects, using arbitrary conventions: abstract concepts ("air", "battery charge", "algorithm"), actions ("running", "picking up"), objects that are not present ("the tool kit in the other room"), objects that are no longer present ("my December salary"), objects that will be present ("Summer"), and objects and conditions that are impossible ("unicorn", "perpetuum mobile"), or objects with some specified property ("slow things").

Typical virtual reality tools constrain their users to the *here* and *now*, even if "here" and "now" may be defined differently than in physical reality. In figure 3, the reference to the grey marbles would be very difficult without the use of natural language. The idea that someone might want to refer to grey marbles if they are represented as in the picture ought not to be surprising: the concept of the set of grey marbles is not inherently complex. In figure 4, the reference to an object which is not actually present will pose a difficulty, unless there is a way of referring to objects that are not visible by their temporal location or their content. Referring to "virtual reality", as in the example, without using human language will of course be a considerable challenge.

The motive for including language in a robot-control interface is to add a level of abstraction to the system: to be able to specify goals on a higher-level than pointing at visible objects. This, of course, presupposes a level of representation abstract enough for symbolic reasoning: we have achieved this through the explicit model of the robot's real world knowledge in the virtual world.

5 Conclusions and Future Work

This paper describes the framework for including a human operator and a real robot in a remote environment in a virtual presence system. The virtual environment layer gives a natural level of representation for the world knowledge of the system; the robot is a natural repository for physical competence; the vision system for perceptual processing; and the multimodal interaction is an intuitive tool for control.

Besides consolidating the framework into a complete system and improving the various modules in it, there are some natural openings to continue development. The perceptual module (grounded on the remote physical layer) should accommodate higher level processing of other sensory data, both human-like speech or sound recognition and placement and non-human, such as temperature and other measurement interpretation; including renditions of information about various other actors and objects in the virtual environment; the interface must be capable of modeling more sophisticated knowledge that the robot learns – physical competence, among other things.

In addition to having tight temporal and causal coupling between virtual and real environments, the operator could move around the virtual environment freely, without involving the robot, and specify tasks for the robot to perform. As the operator navigates through the virtual environment the operator can specify point-to-point navigational tasks as well as pick-and-drop manipulation tasks. These then turn into batch-like higher-level goals – or in essence, programs by high-level example – to be submitted at some later time for the robot's navigational path-finding and grasping systems.

currsize

References

[1] Magnus Andersson, Lennart E. Fahlén, and Torleif Söderlund. A virtual environment user interface for a robotic assistive device. In *Proceedings of the second European Conference on the Advancement of Rehabilitation Technology*, pages 33–57, Stockholm, May 1993.

[2] Ronald Arkin. Reactive control as a substrate for telerobotic systems. *IEEE AES Systems Magazine*, pages 24–31, June 1991.

[3] Ronald Arkin and Khaled S. Ali. Integration of reactive and telerobotic control in multi-agent robotic systems. In *Proceedings of Third International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, Brighten, UK, 1994.

[4] W.A. Aviles, T.W. Hughes, H.R. Everett, A.Y. Martin, and A.H. Koyamatsu. Issues in mobile robotics: The unmanned ground vehicle program teleoperated vehicle (tov). In *SPIE Vol. 1388 Mobile Robots V*, pages 587–597, 1990.

[5] S Benford, J. Bowers, L. Fahlén, and C. Greenhalg. Managing mutual awareness in collaborative virtual environments. In *Proceedings of VRST'94*, ACM, Singapore, 1994.

[6] S. Benford and L. Fahlen. A spatial model of interaction in large virtual environments. In *Third European Conference on Computer-Supported Cooperative Work*, pages 109–124. Kluwer Academic Publishers, 1993.

[7] Steve Benford, John Bowers, Lennart E. Fahlen, Chris Greenhalgh, John Mariani, and Tom Rodden. Networked virtual reality and co-operative work. *To appear in Presence*, 1995.

[8] Alan W. Biermann, Bruce W. Ballard, , and Anne H. Sigmon. An experimental study of natural language programming. *International journal of man-machine studies*, 18:71–87, 1983.

[9] Edwin Bos, Carla Huls, , and Wim Claassen. Edward: full integration of language and action in a multimodal user interface. *International Journal of Human-Computer Studies*, 40:473–495, 1994.

[10] S. Bouffouix and M. Bogaert. Real time navigation and obstacle avoidance for teleoperated vehicles. In *SPIE Vol. 1831 Mobile Robots VII*, pages 265–275, 1992.

[11] Ivan Bretan, Niklas Frost, and Jussi Karlgren. Using surface syntax in interactive interfaces. In *The 10th Nordic Conference of Computational Linguistics*, University of Helsinki, 1995.

[12] Christer Carlsson and Olof Hagsand. DIVE – a platform for multi-user virtual environments. *Computers and Graphics*, 17(6), 1993.

[13] R. Chandrasekar and S. Ramani. Interactive communication of sentential structure and content: an alternative approach to man-machine communication. *International Journal of Man-Machine Studies*, 30:121–148, 1989.

[14] Philip R. Cohen. The role of natural language in a multimodal interface. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pages pp. 143–150, Monterey, CA, 1992.

[15] Steven Feiner, Blair MacIntyre, and Doree Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):52–62, July 1993.

[16] Bretan I. *Natural Language in Model World Interfaces.* Licentiate Thesis, Department of Computer and Systems Sciences. The Royal Institute of Technology and Stockholm University, Stockholm Sweden, 1995.

[17] Jussi Karlgren, Ivan Bretan, Niklas Frost, and Lars Jonsson. Interaction models, reference, and interactivity for speech interfaces to virtual environments. *Proceedings of Second Eurographics Workshop on Virtual Environments – Realism and Real Time*, 1995.

[18] Fred Karlsson. Constraint grammar for parsing running text. In Karlgren, editor, *Thirteenth International Conference On Computational Linguistics (COLING - 90)*, University of Helsinki, Helsinki, 1990.

[19] Fred Karlsson, Atro Voutilainen, Juha Heikkila, and Arto Anttila (eds.). *Constraint Grammar.* Mouton de Gruyter, Berlin, 1995.

[20] Jacqueline H. Kim, Richard J. Weidner, and Allan L. Sacks. Using virtual reality for science mission planning: A mars pathfinder case. In *ISMCR 1994: Topical Workshop on Virtual Reality*, pages 37–42, Houston, 1994. NASA Conference publication 10163.

[21] Eric Krotkov. Mobile robot localization using a single image. In *IEEE Proceedings of Robotics and Automation*, pages 978–983, 1989.

[22] M. Mallem, F. Chavand, and E. Colle. Computer-assisted visual perception in teleoperated robotics. *Robotica (10)*, pages 99–103, 1992.

[23] Paul Milgram and David Drascic. Enhancement of 3-d video displays by means of superimposed stereo-graphics. In *Proceedings of the Human Factors Society 35th Annual Meeting*, pages 1457–1461, 1991.

[24] NASA. *Proceedings of the NASA Conference on Space Telerobotics.* JPL Publication 89-7, Vol 1-5, Pasadena, Ca, 1989.

[25] Jane Robinson. Dependency structures and transformational rules. *Language*, 46:259 – 285, 1970.

[26] Eric Rose, David Breen, Klaus H. Ahlers, Chris Compton, Mihran Tuceryan, Ross Whitaker, and Douglas Greer. Annotating real-world objects using augmented vision. Technical report, European Computer-Industry Research Center GmbH, Arabellastrasse 17 D-81925 Munich, 1994.

[27] Ben Shneiderman. Natural vs. precise concise languages for human operation of computers: Research issues and experimental approaches. In *Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, 1980.

[28] K. T. Simsarian, N. Nandhakumar, and T. J. Olson. Mobile robot self-localization from range data using view-invariant regions. In *IEEE Proceedings of the 5th International Symposium on Intelligent Control*, pages 1038–1043, 1990.

[29] P.C Woodland, J.J. Odell V. Valtchev, and S.J. Young. Large vocabulary continuous speech recognition using htk. In *Proceedings of ICASSP'94, Adelaide*, 1994.