“Intelligent” Telepresence: Introducing Virtual Reality In Advanced Robots

Riccardo Cassinis°, Mardy Terceros Rojas°°

° Dipartimento di Elettronica per l’Automazione, Università di Brescia.
°° Progetto di Intelligenza Artificiale e Robotica, Politecnico di Milano.

Abstract. The paper discusses some issues about a telepresence system to be used for remote driving of robots that operate in hazardous and/or hostile environments. The emphasis is on the problem of efficiently using and integrating information from different sensors, in order to provide the remote operator with readily understandable and usable data.

Solution hypotheses, which take advantage of virtual reality techniques, are presented.

Keywords

Robotics, Telemanipulation, Teleoperation, Virtual Reality.

1. Introduction

Virtual reality refers to realistic generation of a 3-D world in which the operator, using appropriate means, can interact with the virtual computer-generated objects as if he/she were in a real world.

On the other hand, telepresence is “a specialised form of teleoperation and telerobotics whereby the operator of a remote system is presented with workstation facilities capable of creating the illusion that he is actually present at a remote, hazardous worksite” [4].

The aim of this paper is to show how, combining telepresence with virtual reality techniques, the system can present the remote operator with more understandable data from the robot’s sensors, and in some cases use virtual data when real-world data are not available.

Telepresence has countless possible applications, among which, besides the obvious space, underwater and hostile environments, one may include handicapped people aid, civil and military defence, etc.

2. Problem's overview

Since 1965, when Sutherland proposed *The Ultimate Display*, in which the image should appear as if it were real [2], many advances have been made both in the fields of telepresence and virtual reality.

Today, special displays are available for showing remote-site images as if the operator were "inside" them. They are usually referred to as Head-Mounted Displays (HMD's).

Other devices provide the operator with simulated tactile feedback, stereophonic sound, etc. Usually, all these devices are integrated into a single telepresence system [3, 4].

The problem of remote reconstruction of sensory input, typical of telepresence applied to telemanipulation, has often been overlooked or at least underestimated. It is obvious that, if the remote operator were to be provided with enough information to give him/her complete sensory information about the environment the robot is in, countless sensors should be used, many of which are not even technologically available. Furthermore, the robot is often equipped with sensors that have no correspondence in the human body, such as ultrasonic or laser range sensors, metal proximity detectors, etc. For instance, the human body is fully covered with thermal and tactile sensors, that allow to determine contact among the body and foreign objects, and in many cases also their nature. Furthermore, humans can detect, with coarse approximation, the position of their limbs and the forces acting upon them. Clearly, if only a subset of such feelings is provided (e.g. eliminating temperature information, fundamental for determining the nature of objects being touched), or the amount of information is reduced (as when gloves or clothes are used, that reproduce some force and tactile feelings, but with a resolution that is much smaller than the resolution of the skin), the "presence" effect is unrecoverably lost.

In some cases, this may also eventually lead to a physical distress, as it happens in simple driving or flight simulators, where the lack of acceleration feelings often causes a sort of sea-sickness to the user.

Last, it should be kept in mind that manipulators are often provided with sensors that have no correspondence in the human body, such as optical or ultrasonic range detectors, that are nevertheless extremely efficient in measuring distances, and whose information should be somehow integrated with the optical, acoustic or tactile information the operator can readily understand.

2.1 Telepresence levels

From the preceding statements it should be clear that true telepresence cannot be achieved only by sending sensory data to a remote location, but requires, even at a low level, a complex data processing, according to the diagram shown in fig. 1.

This diagram however is not satisfactory, because human sensing is only seldom passive, but makes extensive use of actions, both on the sensors themselves (e.g. moving the head to change the viewpoint), and on the environment (touching objects, moving them, etc.). The diagram must then be expanded as shown in fig. 2.

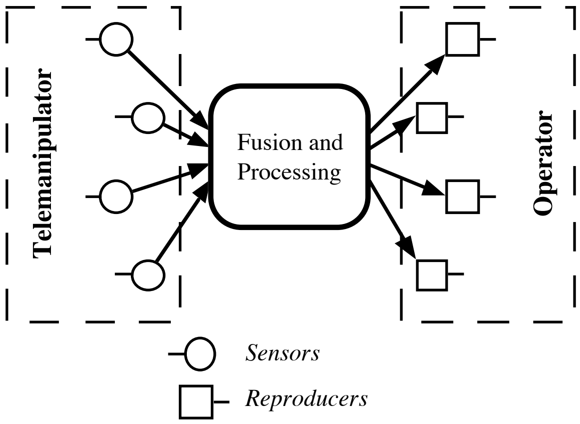

Fig. 1 First level of telepresence

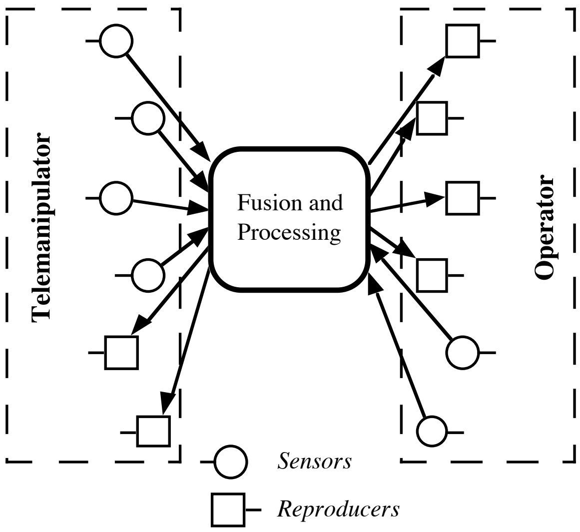

Fig. 2 Second level of telepresence

Although the system of fig 2 may be considered satisfactory for some applications [5, 6, 7, 8, 9], it does not solve several problems, the main being the impossibility of using sensors whose information cannot be directly reproduced and understood. Let us think for example to a scanner that may provide a range map: this map is obviously extremely useful, because it may indicate the proximity of dangerous objects, but it is also obvious that this information (a set of numbers) cannot be directly shown to the operator. A solution could be the integration of the "dangerous distance" information with the visual information coming from video cameras, using special colours for representing relevant objects.

To make this possible, the processing system must be capable of establishing a correspondence among range and visual information. This requires a world model, as shown in fig. 3.

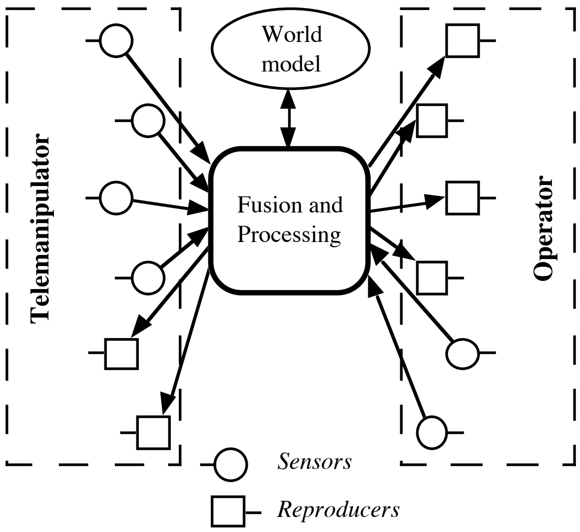

Fig. 3 Third level of telepresence

This world model can be either built in advance using a formal description of the environment in which the robot will operate, or in real time. It is obvious that both choices are extremely complex, and require several restrictions to be made feasible. It must be noted however that the purpose of this work is to indicate some strategies for improving the state-of-the-art telepresence systems, without necessarily achieving a complete presence effect, which, by the way, cannot be perfectly achieved even in simpler circumstances, such as Hi-Fi sound reproduction systems.

2.2 Sensor data fusion

Usually, current telepresence systems do not perform any kind of sensor data fusion, leaving this task to the human sensory system. In other words, information coming from each sensor is directly sent to the transducer that is coupled with the correspondent human sensor: visual data are sent to the eyes, tactile data to the fingers, sound data to the ears, etc. This strategy, shown in fig. 4, is correct only in a limited number of applications.

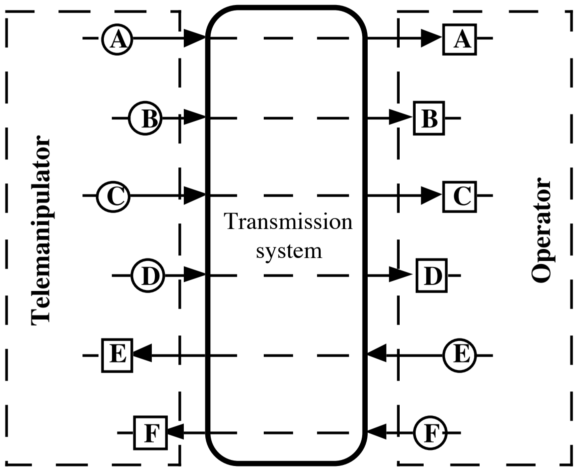

Fig. 4 "One-to-one" correspondence among sensors and reproducers

The proposal presented here, on the other hand, implies full separation among the acquisition of sensory data and their reproduction, in the sense that reproduction can be deferred, and that sensory data can be presented to the operator in a form that has little to do with their original structure.

This kind of approach is complex, but offers several advantages that can be summarised as follows:

1. **Data acquisition time**. Many sensors require a long time for data acquisition. However, one can assume that in a quasi-static environment such data have little changes in time. This means that previously acquired data can be used at a later time. As an example, let us consider the possibility of acquiring and storing a "panoramic" video image, obtained by combining several smaller images. If the environment is static, a view from the same point, but with any orientation, can be readily obtained by choosing a subset of the stored image, while, with traditional methods, one should at least move the cameras to the new viewing position.

2. **Complex processing**. Data processing (e.g. for combining a 2D video image and the corresponding range scan) may require long times, and it is unacceptable that the operator, after having directed his/her attention to a particular area, has to wait several seconds before being able to see the objects it contains. But, if acquisition and processing have been performed at a previous time, the display can be instantaneously updated, provided that enough computing power is available.

3. **Bandwidth reduction**. In many cases a constraint is placed by the narrow bandwidth of the communication channel being used. A partial solution to this problem is to use all available time to transmit and to store information in a place close to the operator.

A distinction should be made among sensor information that must be acquired and transmitted in real time[1], and information that can be acquired in advance.

The first category obviously includes visual information about the area in which the robot causes physical changes, manipulating or moving objects, and sound information thus produced.

The second category includes all data that vary slowly: symbolic and geometric maps of the environment, range maps from sonar and laser scanners, global views of the environment[2].

3. Sensors for telepresence

3.1 Introduction

The needs of telepresence require using several kinds of sensors, some being the same sensors that are used for autonomous robot operation, and some, both for their nature and for the way they are used, typical of this application.

Although measurement and signal transfer methods are well assessed, the particular needs of robotics have caused in the last years deep changes in the way sensory data are used. Initially, sensors were mainly used as “measurement devices” of physical quantities. Nowadays, the accent is mainly placed on the usage of sensory information, and this usage can be completely different from the purpose sensors were originally built for. For instance, using a temperature sensor for detecting the presence of human beings is not a “classical” way of employing a temperature sensor.

Moreover, data fusion techniques are being widely used. Data from heterogeneous sensors are combined, often in a variable and state-dependent way, in order to obtain new data whose nature is sometimes completely different from the nature of original data.

Some considerations follow on the sensors that have special application in telepresence systems.

---

[1]Unless predictive techniques are used, which are not covered by this paper.

[2]Conditions under which these assumptions are realistic obviously depend from the characteristics of the environment: a road with traffic on it varies much more than an unhabited building, but the general map of the road does not change, while the internal map of the building, for instance in the case of a fire, may vary unpredictably and quickly.

3.2 Visual sensors

Sensors that are most widely used in telepresence systems obviously are visual sensors (mainly video cameras). The signal they provide can be directly transmitted to the human eye trough CRT or liquid crystal displays. Usually, couples of cameras are employed, whose signals are separately sent to couples of displays, each one placed in front of one of the operator's eyes, with the interposition of special purpose lenses. It is then possible to re-create the stereoscopic effect of binocular vision, but, to make it realistic, cameras must be movable, and must repeat exactly the movements of the operator's head.

This is because stereoscopic vision is not only due to the availability of a couple of images the brain processes and coupes together, but also to the relative displacement of objects that can be seen when the point of view is moved. This displacement, that also causes objects to be occluded or to re-appear, explains how men may have a good perception of the relative position of objects, even when looking at them with a single eye.

Stereoscopic effect is very poor if this possibility of moving the cameras synchronously with the operator's head is missing. This movement is currently included in all advanced telepresence systems.

3.3 Non-visual sensors

The other two categories of sensors that may have useful applications in telepresence systems are range and contact sensors.

Generally speaking, range sensors measure, using several different methods, the distance of obstacles placed in front of the sensor. This measurement can be done on the sensor's axis, or in an area in which the space in front of the sensor is partitioned into a matrix, and range data are provided for each element of the matrix. It must be noted that, independently from the working principle of such sensors and of the errors that may affect readings, there is no possibility of directly presenting these data to the operator, because the human body does not have equivalent sensors. This problem will be further discussed in the sequel.

Touch sensors can provide information about their contact with a solid surface. This information can be binary or analog: in the latter case there exists an equivalent in the human body, that is tactile feeling. Systems have been built that are capable of reproducing with some accuracy tactile feelings on fingertips, hands or on the whole body. This reproduction is anyway quite coarse, although it can be considered sufficient for some purposes.

4. Problems of reconstruction and of representation of the environment

4.1 Building the world model

From what has been previously said, it should be clear that the availability of a CAD model of the world in which the robot is operating would allow using sensors' data in a much more efficient way. Generally, the world model is not available at all, or it is available in an approximate and difficult to use form. Let us consider as an example the case of an industrial environment: usually maps are available that, although describing with some accuracy the shape of the environment, don't take into account tooling, machines, furniture, etc. The latter are described in layout drawings, but always in an incomplete and approximate form, and, what is more important, a form that is not suitable for direct utilisation. Manual building of a model of the environment is an extremely long and expensive operation. The solution probably lies in a semi-automatic acquisition of the model, done using the following method:

- Several images of the robot's working environment are acquired, either directly from an operator, or through the robot's cameras. Each image includes some *reference points*, whose spatial co-ordinates must be known.
- Each image is digitised and related to other images, taken from different points of view, but containing all or some of the same reference points. The operator manually indicates these points to the computer, together with the edges of relevant objects that should be included in the world model.
- The result of the operation is an accurate 3-D model, which can be augmented introducing physical attributes of the objects, and is the base of the aforementioned world model.

Such systems are already commercially available. An example are the KOH-I-NOOR PhOX cards.

4.2 Using the world model

In order to conveniently use data contained in the system, which are partly real (acquired in real or deferred time) and partly synthetic or semi-synthetic, they must be integrated using a graphic system with suitable characteristics. Today, one can build high performance graphic systems using CAD systems running on high-power PC-type computers or workstations, and integrating them with software packages for image synthesis enhancement (animators, ray tracing, etc.), and with transputers for increased processing speed. It is thus possible to produce quite realistic synthetic images at a rate of at leas 25 frames/sec.

The need for a CAD system mainly derives from the fact that each set of sensory data is gathered from a single point of view. If the same data have to be used for representing the situation seen from a different point of view, it is usually necessary that the first acquisition be "understood" and suitably modelled. Let us consider as an example the case of a classical "horizon sonar scan", a diagram that provides, for any direction on a horizontal plane, the distance of the nearest obstacle from the sensor. In this case, moving the observation point only requires a simple geometric transformation to obtain a representation of distances from the new viewpoint. Instead, if we think at a stereoscopic image from two video cameras, in order to obtain the image of what would be seen moving the cameras, an image interpretation is essential, at least to compute depth data. Therefore, the classical sequence (segmentation – iconic representation – modelling) becomes essential. It must also be reminded that usually information coming from heterogeneous sensors has to be associated to each object.

Once the world model has been built, it can be used for representing visual and non-visual information in a comparatively simple way: for instance, simulating object views from different viewpoints is no more a problem, and particular characteristics (colour, brightness, surface texture, etc.) can be assigned to each object according to specific parameters (distance, danger, temperature, etc.).

A proposal for the described structure is partially shown in fig. 5.

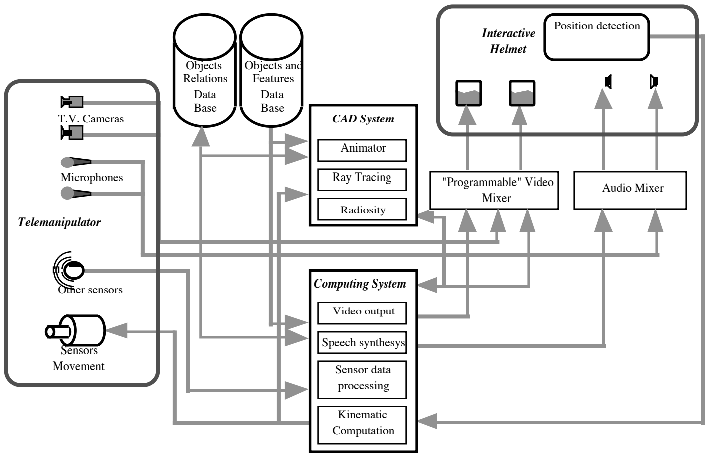

Fig. 5 A partial structure of the proposed system

5. A practical realisation

To carry on practical experiments, a system was built which includes an interactive helmet, a sensor system mounted on an industrial robot (ROBOSOFT GT-6A), interfaces and control software (fig. 6).

Remote system

Two high sensitivity colour CCD cameras were mounted on the manipulator's arm, their axes being parallel and their distance being of 70 mm, which resembles the placement of the eyes in the human body.

This placement allows getting realistic stereoscopic images.

The cameras are mounted on the arm, because one of the main features of the system is to be able to move them according to the operator's head position.

Interactive helmet

The helmet was built using two colour LCD displays and LEEP (Large Expanse Extra Perspective) lenses directly placed in front of the wearer's eyes. It provides the illusion of really being present at the remote worksite where the robot is placed.

A 3Space-IsoTrack position sensor was placed on the helmet to allow real time acquisition of the position and orientation of the operator's head. The software system processes IsoTrack signals and accordingly moves the robot arm that carries the cameras.

Additional sensors

An ultrasonic range sensor was mounted on the robot, its axis being parallel to the camera's axes, which provides the distance of the closest object the cameras are pointing to.

The computing system

In the actual configuration we use a personal computer for data processing and for communication handling.

The main task of the developed software is the transformation of IsoTrack data into data suitable for manipulator control. This processing includes co-ordinate transformation for mapping the operator's head position to the cameras' position, and inverse kinematic transformation for driving the manipulator's joints. Computing speed is critical at this point, and efforts have been made to keep processing time to a minimum, in order to achieve the fastest possible response from the robot when the operator moves his or her head.

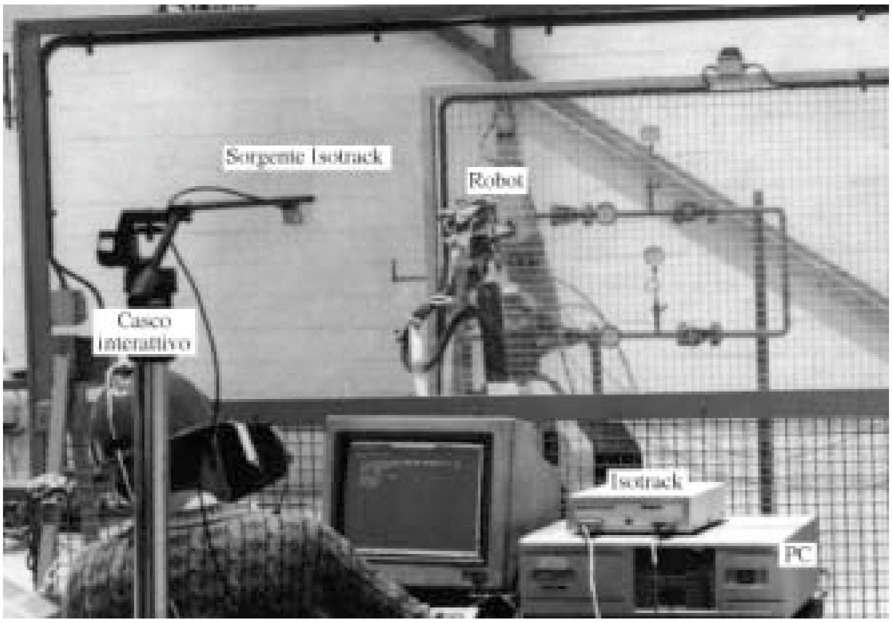

Fig. 6 The implemented experimental system

6. Conclusions and future developments

Although the experimental system just described is not suitable for implementing the world model mentioned in the paper, due to its obvious limitations as far as the processing power is concerned, some results were determinant for evaluating the possibility of implementing a more complex system [10]. Furthermore, the existing system can already be used for some applications, improving its characteristics with limited effort.

As far as the proposal of sensory data integration and modelling is concerned, a deep experimentation is obviously needed, in order to establish matching problems among real world and synthetic sensory data.

The possibility of physically realising the proposed system, given the constant improvement of computing systems and of generation and representation of virtual reality, is very high. Availability of countless software packages for CAD images treatment makes software development quite fast.

7. Bibliography

[1] Chung, J. C., et al.: *Exploring Virtual Worlds with Head-Mounted Displays*, Proc. SPIE, Vol. 1083, Los Angeles, Ca., 1989.

[2] Sutherland, I.E.: *The Ultimate Display*, Proc. IFIP Congress 2, 1965.

[3] Engelberger, J.: *Robotics in service*, Kogan Page Ltd, London, 1989.

[4] Stone, R. J.: *"The Best Of Both Worlds": A Combined Virtual-Real Human-Computer Interface For Telepresence And Remote Driving*, proc. of '91 ISART - International Symposium On Advanced Robotics Technology, Tokyo, 1991.

[5] Stone, R. J.: *The UK VERDEX Project, 2 Years On: Virtual Reality, Telepresence and the Human Factor*, Proc. Imagina '92, Monte Carlo, 1992.

[6] McGreevy M. W., Stoker C. R.: *Telepresence for Planetary Exploration*, SPIE The International Society for Optical Engineering, Cooperative Intelligent Robotics in Space, R. J. de Figueiredo and W. E. Stoney Editors, vol. 1387 pp. 110-123, Boston, Massachusetts, November, 1990.

[7] Noseworthy J. R., Gerhardt L. A.: *Three dimensional vision-requirements and applications in a space environment*, SPIE The International Society for Optical Engineering, Cooperative Intelligent Robotics in Space, R. J. de Figueiredo and W. E. Stoney Editors, vol. 1387 pp. 26-37, Boston, Massachusetts, November, 1990.

[8] McDonnell J. R., Solorzano M. R., Martin Stephen W. M., Umeda A. Y.: *A Head Coupled Sensor Platform for Teleoperated Ground Vehicles*, Unmanned Systems, Vol. 8, N. 4, pp. 33-38, Fall Editore S.p.A., 1990.

[9] Dimitriadis B. Papanikolau K.: *Stereoscopic Vision for AGV Guidance*, NATO Asi Series, vol 171, Expert Systems and Robotics, Springer-Verlag, 1991.

[10] Terceros, M.: *Telepresenza intelligente*, Tesi di laurea, Università degli Studi di Milano, 1992.

Acknowledgements

The experimental part of the present work has been carried on with the contribution of CESI SpA - Milano.