Virtual Exoskeleton for Telemanipulation

Josep Amat
IRI. Robotics Institute (UPC/CSIC).
Campus Nord UPC
08034 Barcelona, Spain

Manel Frigola and Alícia Casals
Dep. of Automatic Control and Computer Engineering
Universitat Politècnica de Catalunya.
Pau Gargallo, nº 5, 08028 Barcelona, Spain.
Email: {frigola, casals}@esaii.upc.es

Abstract: The growing number of robotics application fields, mainly in services, has led to the increase of new needs as well as the development of new facilities for teleoperation. Research in the design of more efficient and easy to use human-machine interfaces has propitiated the development of friendly communication systems such as those based on voice or gesture recognition. This work describes a vision based human-machine communication system that allows a computer or a control unit to "see and track" the position of the hands of a human. Thus, the vision system can be used as a virtual exoskeleton for simple telemanipulation tasks.

1. Introduction

Teleoperation as a means to operate a robot using the intelligence of a human requires the availability of adequate human-machine interfaces. The use of communication means such as natural language or gestures enables us to expand teleoperation to new application fields, making it possible for any kind of user to operate a robot in different work environments. This is possible because human-robot interaction becomes much more comfortable and easier.

The operation of a robot by means of a joystick is very common in areas such as civil engineering, in applications for parts manipulation in construction, or in the guidance, from a van, of mobile robots within sewers, among others. When the number of degrees of freedom to control is high or the operation to be performed requires certain ability, it is convenient to use more sophisticated devices. Different hand-held devices have been designed to facilitate this human-robot interaction. Other structures, such as the well known phantom devices, that introduce the concept of haptics, provide augmented reality in the interaction of the robot with the environment. These kinds of devices are extremely useful in application fields ranging from space to surgery, areas in which perception is essential to understand the evolution of a teleoperated task. In all these areas, such physical interfaces enable a human operator to interact with real or virtual

environments, either for teleoperation works or for training applications, respectively.

The concept of exoskeleton, as the master device in teleoperation, started with the use of mechanical structures[1], like the Hardyman, a wearable articulated structure designed to amplify the human forces and movements in applications that require the manipulation of big or heavy loads. These devices have become lighter and they incorporate force feedback enabling us to perceive the effects of the performed actions by the robot, the slave, more effectively [2].

Electronic based devices aim to suppress the mechanical elements that, in some way, constrain the operator movements. Among such devices, data gloves are those that seem to be able to provide the best results.

With this same aim, to avoid the need of using mechanical structures, some efforts have been dedicated to designing computer vision systems to be used as the master, in a master-slave robotics configuration. Computer vision in industry and in robotic applications has progressed significantly, giving place to applications of detection of human movements and their gestures with the aim of interpreting signals or orders. Therefore, the human operator can avoid the need to wear a physical device over their body. Nevertheless, the detection and tracking of a human body in a natural environment presents certain difficulties if the background image is not homogeneous. Consequently, some of the developed systems use LED diodes or reflectors located in the body joints [3, 4, 5], or use colour information in applications where the user is forced to wear coloured clothes to facilitate the segmentation process [6, 7]. Other alternative systems are based either on magnetic position sensors or even on myoelectric sensors that convert the muscle's movements into signals, from which it is possible to detect the operator movements [8]. A mixture of them, magnetic sensors and cameras visualising some fiducial marks provide better performances since they combine the robustness of magnetic sensors with the precision of computer vision. Other researchers base their works on the analysis of the human movements, either through the use of optical flow [9, 10] or by means of the subtraction of successive images [11].

The present work is also based on the analysis of the human operator movements. In [12] the detection of movement is achieved working in highly contrasted environments. In our case, the system avoids the need to use specific *plateaus* with controlled lighting conditions or the need to wear special clothes or specific elements, or marks, on the body. This improvement on the working conditions is possible due to the fact that the system operates from the variations on the direction of the gradient between successive images. One such procedure notably improves the results obtained by the classical methods based on movement, enabling the availability of segmented images with very low noise level, even working in natural environments [13]. The computer vision system implements three basic functions: The detection of humans and segmentation of the hands of a person in a natural working environment, the tracking of the position of the upper limbs in 3D, and the control of the robotised arms, in accordance with the user's movements.

2. System structure

The aim of the vision system is the recognition and tracking of the arms and hands of a human operator to remotely control one or more robots. From such recognition, the postures and gestures of an operator have to be interpreted. This interpretation is based on the detection of an operator postures and the interpretation of the gestures derived from their movements are based on the location in a tridimensional space of the body more relevant parts for this application, using multiple views of the operating scenario. The location process is based on a first segmentation phase and a second one that validates, over a simplified model, the detection of the human figure.

Since the human detection and the image segmentation are based on the movement in the scene, it is necessary to distinguish a moving human body from other possible moving objects in the scene. Consequently, we use a 3D adaptable geometrical model of the human arms to make this detection more reliable, not only considering the target dimensions, but also its shape. The model is simple enough to be applicable in real time, but also complete enough to enable the description of the arms and hands position. The model is polycylindrical, articulated and tridimensional, and it is adaptable to changing shapes to fit in real time with the operator moving body.

The first step of the gesture human-machine communication interface is the detection and tracking of the arms and hands by estimating, over time, their position at every instant. The estimated position sequences will describe the body movement indicating the actions the person desires to express. The complete system, as shown in fig. 1, consists of the following tasks:

- Dedicated low level processing for movement detection
- Features extraction and detection of the arms singular points
- 3D position measurement, from stereo, and arms posture data validation using a simple geometrical model
- Filter and, operator-robotic arms, frames transformation

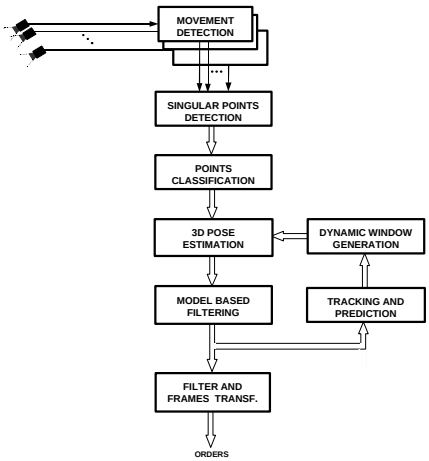


Fig. 1 Functional system structure.

3. Arms detection and tracking

The process of movement detection to interpret the human body orders to control the robot requires us to follow different steps. The vision process extracts the person silhouette from the image, for its analysis. Then from the human upper limbs silhouette the features that characterise their posture are obtained providing the data necessary to fit the human figure to a predefined adaptable human model. The process can thus be summarised in the three following steps: background extraction, features extraction and 3D model fitting.

3.1 Background extraction

Image segmentation is one of the main problems to face up to in computer vision. In complex scenes where it is not possible to find features discriminating enough to use satisfactorily any common segmentation process to extract the desired objects, it is possible to resort to the analysis of image sequences and to analyse the images' temporal variation, provided that the objects to be segmented are in movement.

Since the detection of a person's movements for the interpretation of her gestures requires the detection of the human figure and no more information from the scene is necessary, the level of image segmentation can be reduced to objects extraction from their background. The extraction of human figures in natural scenarios, either indoor or outdoor, should rely on segmentation techniques not dependent neither on the heterogeneity of the possible elements in the scene and its lighting conditions, nor on the person movement itself.

In this work, we use images subtraction, but to improve the system performances in complex scenarios the comparison pixel by pixel is carried out from the estimated gradient vector instead of using only its absolute value. In this case, images subtraction is performed as follows:

$$\left| \vec{G}_t(x,y) - \vec{G}_{t-1}(x,y) \right| \qquad \text{(Eq. 1)}$$

Where $\vec{G}_t(x,y)$ represents the measurement of the gradient vector computed at position x,y of the grey image $I(x,y)$, taken at instant t. The advantage of comparing images using the gradient vector instead of the gradient module is the increase from one to two dimensions in the pixels description, thus enhancing their characterisation.

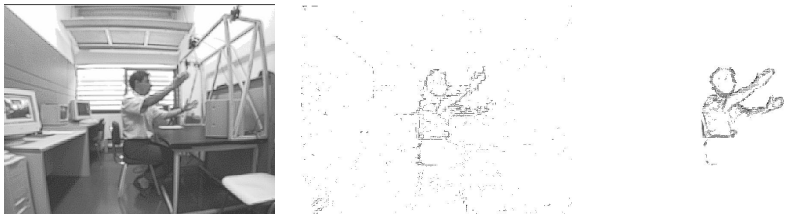
In spite of this advantage, images subtraction is still sensitive to lighting variations. In natural environments, or in environments with fairly controlled lighting conditions it is necessary to use a more robust comparison. The new expression will include the gradient direction that does not depend on variations of lighting intensity, as follows:

$$\left| Atan2(\vec{G}_t(x,y)) - Atan2(\vec{G}_{t-k}(x,y)) \right| \qquad \text{(Eq. 2)}$$

where Atan2 is the extension of the atan function to two dimensions.

Since equation (1) is simpler than (2) the former is normally used in this system when lighting conditions are fixed, and only when lighting variations

could decrease reliability it is necessary to rely on (2). Fig. 2 shows the subtraction resulting from a sequence grey level images (a), using the gradient module (b), and using the gradient vector (c).



a) b) c)

Fig 2 Results of the segmentation operator, a) original image, b) segmented image using the gradient module, and c) using the gradient vector.

3.2 Features extraction for posture characterisation

From the human arms silhouette detected, the following step is the extraction of the features that characterise each operator's posture. The process followed to extract the required data, operating in different kind of environments and without imposing strict operating restrictions, is achieved by splitting the problem into three steps: features extraction, singular points detection and singular points classification.

The features selected to detect and to locate singular points are the clusters of pixels that verify some pre-established heuristic conditions. First, the clusters considered are the areas of the image whose distance among pixels are less than a maximum value, this value being chosen according to a compromise between efficiency and computing cost. A second parameter is the size of the clusters. The next step is to detect from these clusters the body singular points. The singular points considered are the most prominent ones of the silhouette. Fig. 3 shows some candidate clusters of pixels (a) and the set of points considered as arms singular points (b), corresponding to the scene in fig. 2.

a) b)

Fig. 3 Features extraction. a) pixels clustering, b) singular points candidates

According to the geometry of the singular points distribution, they have to be identified, by means of a model, as the corresponding arm or hand parts, such as the elbow or a finger tip.

3.3 Tridimensional human model

With the aim of robustly detecting and recognizing a human upper limbs configuration or the arms posture and to avoid false detections, a model is defined to validate the extracted silhouettes. The model was defined based on a compromise between simplicity and speed on one hand and efficiency on the other. It was designed according to the human body structure, its shape and moving capability. The articulated body state will be defined by a set of variables that, at a given time instant, defines position, speed and acceleration of the different model constituting parts. Therefore, the model has been designed as an articulated structure composed of geometrical primitives.

The imposition of some anthropomorphic constraints and the availability of some dimensional measures make it possible to reject wrong detections without the need to apply the model, thus reducing the operation time. Therefore, it is possible to reject the shapes that do not fit to an adequate profile. The person model adopted is constituted by a set of cylinders that fit to the moving parts profile. The model consists of two coaxial cylinders that are adjusted to the head and body, and also a set of up to four cylindrical surfaces per arm, that are adjusted to the body overhanging elements, that can correspond to the arms and hands (Fig. 4).

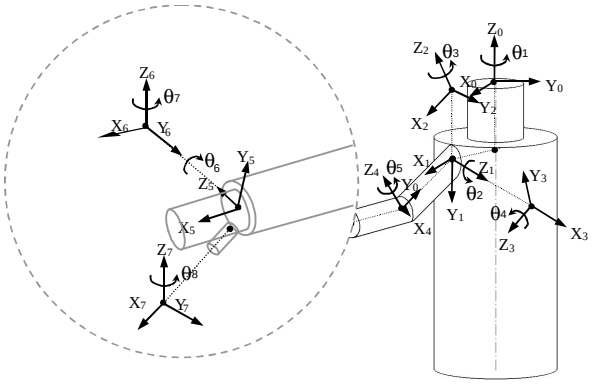


Fig. 4 The polycylindrical model with its joint reference frames.

From the pair of singular points in the two stereoscopic images, classified as belonging to the head, its corresponding 3D position is determined by triangulation. This head singular point defines the central axis of the two main cylinders of the body model (those corresponding to the head and to the trunk), and consequently the body position.

Based on this estimation of the body position, all the singular points that have also been detected, either those located at a distance too far or too close from the main axis, compared to a previously defined cylinder radius (R_c), are eliminated. In this way, we avoid the ambiguity derived from considering arm configurations that imply that the arm is too close or even in contact with the body.

Consequently, all the singular points located at a reasonable distance from the main axis form the set of points that will be used to generate the different

hypothesis about the operator's gesture. Every point considered is again validated in the Cartesian world. Fig. 5 shows, part by part, the cylinders that fit with the relevant body parts for telemanipulation: the trunk, head, arms and hands.

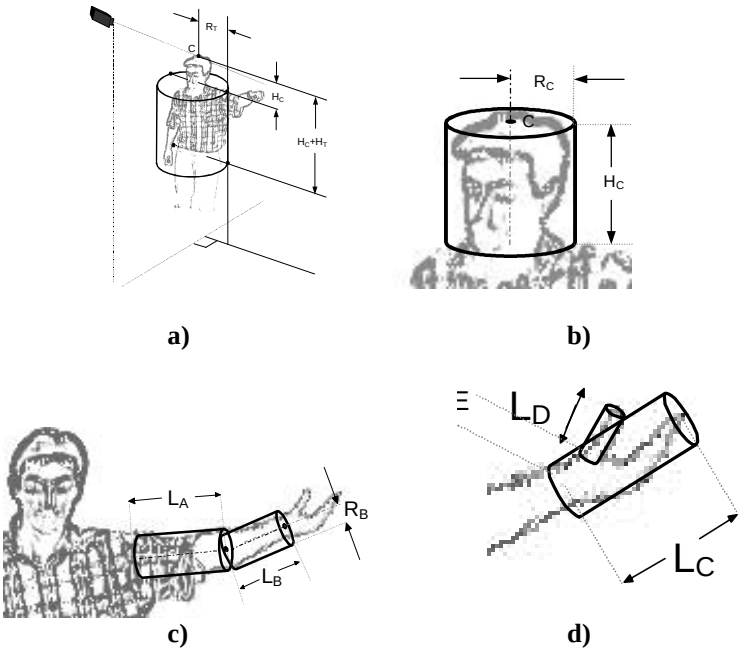


Fig. 5 Cylinders parameters: a) and b) cylinders associated to main axis, trunk and head, c) arm cylinders and d) hand model.

4. Model fitting and frames transform

Once the position and orientation of the different model elements that fits with the operators posture has been obtained, it is necessary to transform the posture, or configuration of the operator's arms and hands, to the reference frames corresponding to the robot arms. This transform is necessary since the operator and the robot arms have different geometry, and it is set basically as an inverse kinematics problem.

With respect to these transforms there are two facts to consider. First, the changes of the head position indicate the potential user's displacements within the environment. Second, the elbows' orientation of the operator's model moves in such a way as to make them fit with the robot elbows' rotation, or that of the robotic platform supporting the teleoperated arms.

Later on, the inverse kinematics that provides the joint angles of the teleoperated arms is computed. This computation has to maintain the extreme points of the teleoperated arms at the same distance (with an adjustable scale factor) as the distances between the operator's arms extreme points and their

shoulders. From the multiple possible configurations that enable us to achieve the same point, those that minimise the configuration changes are chosen.

We have not considered up to now other strategies, as for instance those that would avoid a collision between the robot elements, those that form the teleoperated structure, with the rigid objects in the scene. This is due mainly to the computing requirements, that would introduce an excessive delay that make the teleoperation difficult.

5. Results

The virtual exoskeleton has been tested on two different robots in our laboratory: a Cartesian robot and Garbí, an underwater vehicle provided with two arms.

Garbí is a low cost underwater robot, designed to carry out some simple manipulation tasks such as to collect some samples from the sea bed for applications in biology. To cope with such specifications its arms were designed to have uniquely three degrees of freedom each, plus the open-close movement of the gripper. Fig. 6

Fig. 6 Garbí, underwater robot.

Garbí has been the main test-bed for the experimentation of teleoperation tasks using the virtual exoskeleton. The Garbí exoskeleton itself, two simple articulated arms mounted on a chair and provided with potentiometers, in their joints, and push-buttons to order the open-close movements of the gripper, were replaced by the described vision based virtual exoskeleton.

The tasks more feasible to carry out, with the constraints derived from the simplicity of the structure of the arms that are not provided with any rotation in the wrists, can be classified into two types:

- Collecting samples of small parts, sea-weeds, stones, etc.
- Hoisting of objects using ropes and hooks steered from the assistant ship.

The system has proved to be efficient enough for sample collection, and the work carried out, fig. 7, has been always more comfortable than using the mechanical exoskeleton.

Other more complex tasks, such as the recovery of objects from the sea bed, have indeed been carried out. The experimentation environment has been a test

laboratory scenario. The tasks experimented were oriented mainly to carry out operations for hoisting an object with ropes provided with a hook, fig. 8. The Garbí robot arms can be the means of transmitting the ability of the operator to the working place, but not the force required for hoisting the part. Consequently, the ropes arrive from the surface, from the support ship, and the forces required to proceed to the extraction of the hooked object are provided by the crane.

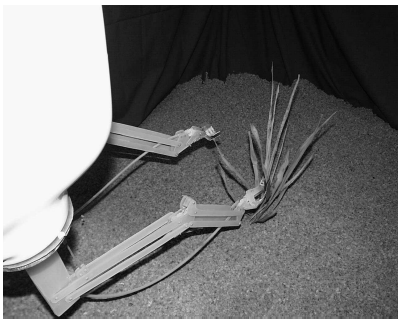


Fig. 7 Samples collection.

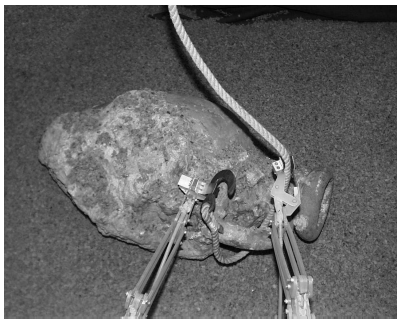


Fig. 8 Hoisting a part from the sea bottom.

In order to evaluate quantitatively the location and tracking precision some trials have been performed. The trials consisted of placing a finger tip over a reference stick, on the working table, and measurimg the error with which the robot the robot arms position themselves in the corresponding homologue points in their working space.

The robot positioning and repeatability observed over the vertical axis, can be seen in fig. 10. These errors, that in the worst case are within a radius of about 2 cm, are mainly due to the hysteresis of the robot arms movements. Nevertheless, these errors do not have a special effect in teleoperation since the users themselves continuously correct the operation visually.

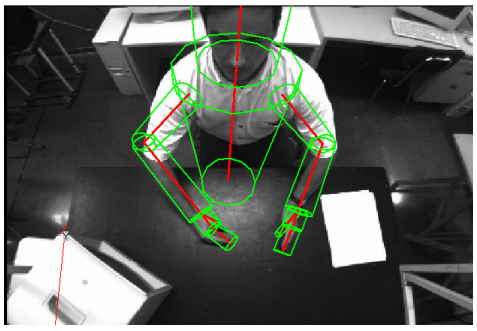


Fig. 9 The virtual exoskeleton.

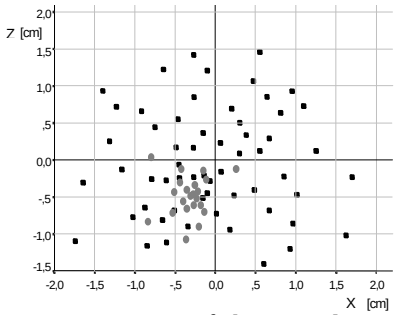


Fig. 10 Dispersion of the arrival position of the hand, in the vertical plane, ● from a fixed position, and ■ from random origin.

6. Conclusions

The described vision system, operating as an arms and hands position sensor, has permitted us to substitute the mechanical exoskeleton of the teleoperation working station of the underwater robot Garbí, by a remote, non invasive, sensor. This change makes the teleoperation tasks easier. The underwater robot, with its two teleoperated arms controlled from the surface, can grasp small samples of the sea bed, move objects, or manipulate a rope dedicated to hoisting objects and their retrieval.

Due to the simplicity of the upper-limbs human model used, the system developed can be used as a virtual exoskeleton in teleoperating tasks, in which the end-effector has up to 3 + 2 degrees of freedom, three for the wrist position and two for the end-effector orientation. The use of such friendly master-slave teleoperated structure opens the possibility of using teleoperation in environments or situations where it is not desirable or convenient to wear specific devices, or when a very intuitive way of user-robot communication is necessary, such as controlling a crane remotely. Thus, this work provides some advances towards "natural" human- machine interfaces.

References

[1] R. C. Goertz (1964) Manipulator systems development at ANL. In proc. Of the 12th Int. RSTD Conference.

[2] M. Bergamasco (1995). Force replication to the human operator: The development of arm and hand exoskeleton as haptic devices. In The seventh ISRR, Germany, pp. 173-182

[3] R. Azuma, G. Bishop (1994). Improved Static and Dynamic Registration in an Optical See-through HMD. In Proc. SIGGRAPH, Orlando.

[4] R. Rasid (1979). LIGHTS: A study in motion. In Proc. DARPA Image Understanding Workshop, pp. 57-68, Nov.

[5] M. Ward, R. Azuma, R. Bennett, S. Gottscalk, H. Fuchs (1992). A Demonstrated Optical Tracker with Scalable Work Area for Head-Mounted Display Systems. In Proc. of the Symposium on Interactive 3D Graphics, pp. 43-52.

[6] M. Yachida and Y. Iwai(1998). Looking at Human Gestures. Computer Vision for Human-Machine Interaction. Ed. by R. Cipolla and A. Pentland. Cambridge University Press.

[7] D.M. Gavrila and L.S. Davis (1996). 3-D model-based tracking of humans in action: a multi-view approach. IEEE Computer Vision and Pattern Recognition.

[8] O.A. Alsayegh and D.P. Brazakovic (1998). Guidance of Video Data Acquisition by Myoelectric Signals for Smart Human-Robot Interfaces. In Proc. of ICRA'98.

[9] A. Pentland and B. Horowitz (1991). Recovery of nonrigid motion and structure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(7): 730-742.

[10] Y. Yacoob and L. Davis. Learned Temporal Models of Image Motion. ICCV'98, pp. 446-453, 1998

[11] H. Nugroho, J. Hwang and S. Ozawa (1994). Tracking Human Motion in a Complex Scene Using Textural Analysis. IECON 94, pp. 727-732.

[12] J.M. Buades, R. Mas & F.J. Perales (2000). Matching Human Walking Sequence with a VRML Synthetic Model. Works. Articulated Motion and Deformable Objects, pp. 145-158.

[13] J. Amat, A. Casals, M. Frigola. Stereoscopic System for Human Body Tracking in Natural Scenes. ICCV Workshop on Modelling People, MPEOPLE'99, 1999.