Gaze tracking for robotic control in intelligent teleoperation and prosthetics

Claudio Castellini
(corresponding author)
LIRA-Lab, University of Genova
viale F. Causa 13, 16145 Genova (Italy)
claudio.castellini@unige.it
phone +39 010 353 2946
fax +39 010 353 2948

Giulio Sandini
Italian Institute of Technology, Genova (Italy)
and LIRA-Lab, University of Genova (Italy)
giulio.sandini@unige.it
phone +39 010 353 2779
fax +39 010 353 2948

Keywords
Gaze tracking, robotic control, teleoperation, prosthesis, machine learning

Introduction

In the rehabilitation of amputees it is crucial to be able to *control* a prosthesis in such a way to reduce as much as possible the gap between the user's intentions and the response of the prosthesis. In particular, as far as prosthetic *hands* are concerned, the situation is rather poor. Recent advances in building sophisticated artificial hands have made it possible to give the user an artifact which, although still far from a real human hand, could potentially allow for complex grasping actions to be performed. Moreover, advanced tactile sensors can be fitted on the hand, so that a great deal of sensory feedback can be sent to the user's Peripheric Nervous System (PNS in short). An example is the *CyberHand* project [1].

However, it is hard for the user to effectively *control* the prosthesis. Direct connection to the PNS will result in poor control abilities and sensory feedback – the technology of invasive PNS interfaces still does not allow for severed nerves to be connected one-to-one to artificial sensors and actuators. Other solutions, such as non-invasive interfaces (e.g., monitoring the electro-myographic signal) and high-level commands issued via voice or buttons are of no help in reducing the above mentioned gap. Good artifacts can be built, but there is no way for the disabled person to control them effectively.

In order to overcome this gap, we believe one must put *intelligence* in prostheses. The computer system controlling the prosthesis must be able to *learn* and *adapt* to the user's needs, capabilities and feelings. Picture, for instance, grabbing and holding a pen in order to hand-write: this type of grasping is extremely delicate and precise, but still the grip must be strong enough to allow the pen to be held against the paper. Moreover, although the shape of a pen is similar to that of the handle of a hammer, for example, what one can do with it is completely different.

In the framework of the Neurobotics project [2] we are trying to improve hybrid bionic systems this way by employing machine learning algorithms. As a prototypical experiment, we are working on a *teleoperation* setup which, after a period of training, will eventually *guess* the user's intentions and correctly grasp a series of different objects placed in front of the slave robot; the technology acquired will then be transferred to a prosthetic hand with high dexterity (for instance, that being developed at the German Aerospace Center, see http://www.dlr.de/rm/en/). Intelligent teleoperation bears more than a casual resemblance to intelligent prosthetics: driving a prosthesis is like teleoperating a robot, only the slave lies in the very same place as the master. We envision that the prosthesis will gather sensory data from cameras[2], tactile and pressure sensors, in order to gain insight on the shape and affordance [3] of objects lying around, and in order to guess what it is supposed to grasp.

In this paper we describe how the *gaze* of the user also can be used in order to direct the robot toward the correct object on a table and guess whether the user is actually willing to grasp that particular object.

Teleoperation

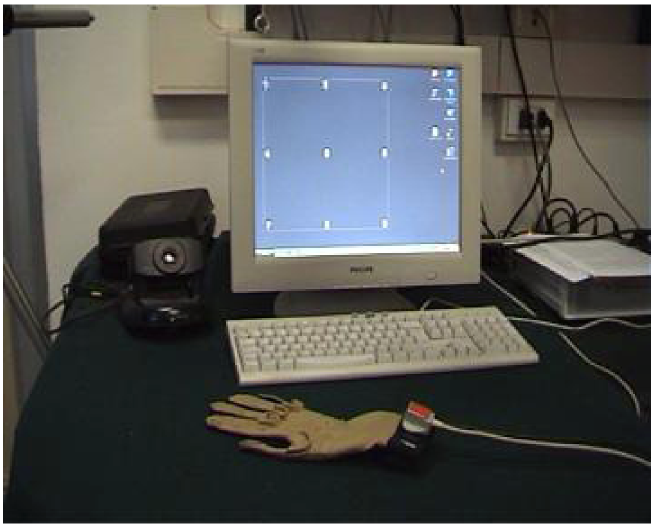
A basic teleoperation setup consists of a *master*, by means of which movement/sensory data are gathered off a human user, and a *slave*, a robot acting according to the intentions of the user. Figure 3 shows the setup. The master consists of an Immersion *DataGlove* with 22 sensors, describing in real-time the position of the fingers and wrist of the user; an Ascension *Flock-Of-Birds* magnetic tracker, which tells us where the wrist is, in absolute coordinates; and, lastly, an ASL *E504* gaze tracker, telling us where the user is looking at.

Since the objects the master wants to grasp are in front of the slave setup (which is exactly the situation one would be presented with in intelligent prosthetics), we place a monitor in front of the user, showing the robot's point of view. The slave consists of two colour cameras mounted on a five degrees-of-freedom (DOFs) robotic head, a *PUMA200* six DOFs robotic arm, and a custom built dexterous humanoid hand, also having six DOFs. The communication is realised via YARP [4], a modular, abstract robotic control environment, which allows distributed computation and fast transmission of data through a standard network.

The control loop starts with the data gathered off the master: hand position (using data the magnetic tracker), shape of the grasp (using data from the DataGlove) and direction of gaze, coming from the gaze tracker. The user looks at the monitor and contextually moves his/her arm and hand, performing reaching and grasping actions. The master's data is sent to the slave, where it is interpreted according to the robot's geometry and kinematics, and immediately executed. (A simple inverse kinematics algorithm is applied to evaluate the robot joints positions.) The loop is then closed showing the user what the slave is doing in real time on the monitor.

The robotic control is, at this stage, still performed *in position*, that is, position data is translated to position commands to the robotic joints. This means the user has no control over the *speed* of the robot movements, which in some case can be awkward. However, thanks to a smart evaluation of the velocity profiles onboard the robot, the delay (of about half a second) is tolerable, giving the user a reasonable feeling of tele-presence.

2 Light microcameras can be embedded in a pair of glasses, which would make the disabled person able to carry the system along.



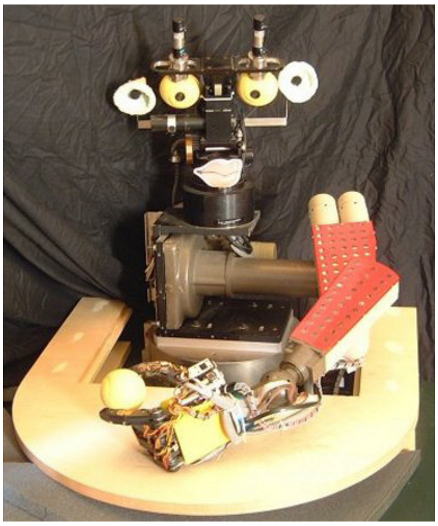


Figure 3: the master (left) and the slave (right) parts of the teleoperation setup.

Intelligent teleoperation

Basic teleoperation (what the master commands, the slave blindly executes) has several problems: mainly, it requires a high bandwidth, and it impossible for the user to correct every problem which might occur in the slave's setup. As an example, imagine the slave holding a mine while carrying it to a safe place, when suddenly a hole in the ground is met. In this case the user cannot react fast enough to prevent the mine to shake and possibly blow up.

In order to overcome these problems, we envision a learning machine to be put in the control system, able to react in place fast enough to compensate for such problems[3]. Even more interesting in this context is the ability of the slave to learn, from the real-time data coming from the user, *what the user wants to do*. For instance, the system could monitor a reasonable time-frame of data coming from the hand position and the gaze of the master; in most cases, *gazing* at an object and *moving the hand* towards that object means: *I want to grasp that object*. After a supervised training phase, the system would then learn to associate a certain speed of reaching, associated with the gaze fixation upon an object, with the *action* of grasping that object. In a second phase then, the system would be instructed, upon recognising a "grasp-that-thing" sequence, to de-activate basic teleoperation, grasp the object and then release the control to the user. The user would then have the feeling of the machine "having read his/her thoughts".

Notice that this schema overcomes the bandwidth limitation detailed above, since all calculations and learning would happen onboard the slave. The schema can also be extended toward more elaborate forms of learning, e.g., learning to grasp by recording the shape of the master's hand during the grasp, and associating it with the visual appearance of the object. This would aid the slave in grasping the object the right way automatically. Not incidentally, this is exactly what is needed in intelligent prosthetics.

[3] By *learning machine* we mean a software artifact based upon a machine learning algorithm, such as, e.g., Support Vector Machines [5].

Gaze tracking for saccades control

Gaze is a quintessentially voluntarily driven motion, and can be actively used to infer cognitive processes (see, e.g., [8] for a survey). Our slave's robotic head is currently able to *saccade* to a certain point in its viewfield. We have linked the gaze tracking device to the saccading mechanism in order to control the head's position by gazing. Having the slave gaze accordingly to the master has the beneficial effect of placing the objects of interest at the center of the viewfield: this enables us to employ *log-polar* vision [6], which greatly reduces the bandwidth needed by the transmission of images – in most cases, the bottleneck of the system. The system works in real time: we evaluate the mean and variance of the gaze signal over a carefully chosen time-window; once we find that the variance has remained "small" for the whole duration of the time frame, we guess that the master is *fixating* a point in the slave's view field, and therefore command a saccade toward that point. The coordinates of the saccade are gathered by considering the mean of the gaze signal, essentially the *center* of the "cloud" of the gaze data.

The above mentioned time-window is currently set at 400 milliseconds: in an extensive series of experiments on human adults, Johansson et al. [7] have shown that, during ordinary reaching and grasping tasks, (a) we always gaze at the *reaching and grasping points*, and never at our own hand; (b) we fixate the objects to be grasped for about 350-450 milliseconds, and then direct our hand toward the object. Therefore, it seems reasonable to instruct our system to do the same. The data reported in [7] is actually an average over nine human subjects; indeed, the time-window necessary for deciding when to saccade varies from person to person. Therefore, we plan to extend the learning mechanism in this sense too: to adapt the time-window to the needs and will of the user.

Conclusion

The Neurobotics project is due at the end of 2007. By that time, we plan to have accomplished the two phases sketched above: (a) setting up and testing the described learning machine on the intelligent teleoperation setup, and then (b) migrating the system to a dexterous robotic hand, which will then be worn by an amputee. Gaze will therefore be extensively used to understand how a disabled person can realise a better control of his/her artificial hand, and reduce the frustration gap induced by the poor chances of a direct control.

Acknowledgments

This work is supported by the Neurobotics project, FP6-IST-001917. We wish to thank Claes Von Hofsten for lending us the gaze tracker, and Giorgio Metta and Francesco Orabona for invaluable discussions and support.

References

[1] The CyberHand project, IST-FET-2001-35094. See *http://www.cyberhand.org*
[2] The Neurobotics project, FP6-IST-001917. See *http://www.neurobotics.org*
[3] Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston.
[4] The YARP project. See *http://yarp0.sourceforge.net*.
[5] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.

[6] *Anthropomorphic Visual Sensors*. F. Berton, G. Sandini, G. Metta. In the Encyclopedia of Sensors. Ed. C.A. Grimes, E.C. Dickey, M. V. Pishko, American Scientific Publishers. Volume X, pp. 1-16.

[7] Roland S. Johansson, Göran Westling, Anders Bäckström, and J. Randall Flanagan, *Eye-Hand Coordination in Object Manipulation*, The Journal of Neuroscience, September 1, 2001, 21(17):6917-6932

[8] Mary Hayhoe and Dana Ballard, *Eye movements in natural behavior*, TRENDS in Cognitive Sciences, Vol. 9 No. 4, 2005.