User Intentions Funneled Through A Human-Robot Interface

Michael T. Rosenstein **Andrew H.Fagg** **Shichao Ou** **Roderic A. Grupen**
Department of Computer Science, University of Massachusetts Amherst
140 Governors Drive, Amherst, MA 01003-9264
{mtr,fagg,chao,grupen}@cs.umass.edu

ABSTRACT

We describe a method for predicting user intentions as part of a human-robot interface. In particular, we show that *funnels*, i.e., geometric objects that partition the input space, provide a convenient means for discriminating individual objects and for clustering sets of objects for hierarchical tasks. One advantage of the proposed implementation is that very few parameters need tuning, and a simple heuristic for setting initial parameter values appears promising. We discuss the possibility of adapting the user interface with machine learning techniques, and we illustrate the approach with a humanoid robot performing a variation of a standard peg-insertion task.

Keywords

Human-robot interaction, telerobotics, predictive display

INTRODUCTION

In this paper, we describe a novel human-robot interface that supports both adjustable autonomy and hierarchical task selection. With adjustable autonomy, e.g., [6], a computer interface facilitates switching among a variety of control modes ranging from full supervision to full autonomy. With hierarchical task selection, the interface allows an operator to easily solve a high-level task autonomously or else to guide the robot through a sequence of lower-level subtasks that may or may not involve autonomous control.

For example, consider a robot capable of moving objects from one location to another. In an uncluttered environment, the operator may indicate the task by simply pointing to the desired object and then to the target location. But in other situations, the human supervisor may need to interrupt this high-level task and perform some lower-level maintenance operation, such as obstacle avoidance. In any case, the key challenge is that current forms of human-robot interaction suffer from the operator's inability to control a robot as easily as his or her own limbs. This leads to movements that are fatiguing for the operator and that fail to utilize the robot's full potential, especially in terms of speed. One goal of human-robot interface design, therefore, is to shift some responsibility or "cognitive load" from human to machine.

Our current approach is to predict operator intentions based on movement of the robot in the vicinity of key landmarks. As a first approximation, one could simply compute the distance from the robot to each candidate landmark and then predict that the operator's intention is movement toward the nearest landmark. With the proper feedback, the operator has two strategies for adjusting his or her commands *to elicit the desired prediction from the interface.* In particular, the operator can explicitly move the robot toward the desired landmark or take advantage of the autonomous system after moving to a region of space where the interface easily discriminates the target landmark from the other candidates. As described below, the novel part of this work is the use of funnels to enhance the discrimination capabilities of the interface. By *funnel*, we mean a geometric object that forms a decision surface in a suitable space, such as the user's motor space or the robot's state space. The intention is that the funnel plays the role of a virtual landmark which is easier to reach than its corresponding real landmark.

The use of funnels as described shortly offers a unique approach to human-robot interface design, although we are not the first to suggest that funnels are a useful metaphor when devising strategies for robot control. For example Burridge *et al.* [2] composed a sequence of controllers where the state of one controller converges to the basin of attraction for the next controller, eventually leading to a goal region. In the human-computer interaction community, spatial distortions [5] are sometimes used to provide a similar effect as a funnel. For instance, *fisheye views* [3] magnify a graphical user interface in the vicinity of a cursor and compress regions farther away. For instance, Gutwin [4] used a fisheye view to display numerous web pages simultaneously as a tree of small icons; when the user traversed the tree of web pages via a mouse, the icons nearest the cursor grew continuously toward full size.

FUNNELING USER INTENTIONS

One underlying assumption of this work is that landmark (or object) locations are known—in terms of some world coordinate frame as well as a relevant interface element. In this paper we assume that a computer vision system is able to recognize objects in a scene and make the correspondence between world and image coordinates. With this information in hand, we are then in a position to implement the funnel interface.

In general, "funnels" can have somewhat arbitrary shape, although for simplicity we consider a more prototypical hyperbolic profile. In particular, we represent a funnel in cylindrical coordinates as

$$\frac{r^2}{a^2} - \frac{(z-z_0)^2}{b^2} = 1, \tag{1}$$

where r is the radius at a distance z from the end of the "spout." In Eq. (1), the parameter a gives the minimum radius of the funnel and b determines the asymptotic radial slope relative to a. As a simple heuristic, one can set a to equal the object radius and b to reflect inter-object distances. In Eq. (1), z_0 is an offset parameter used to give the funnel a cylindrical shape near the object. Specifically, we set r equal to a whenever $z < z_0$ and use Eq. (1) to derive r otherwise. Figure 1 depicts the relationship among the various parameters.

As a basis for predicting operator intentions with multiple objects, funnels yield relatively discriminating results near the target object but lead to aliasing at greater distances. This property is actually convenient for hierarchical tasks because it provides a way to group nearby objects into clusters that may be acted on as a unit. More specifically, for each object the robot is either within the corresponding funnel or not, and we use the terms *active* and *inactive* to describe these two cases. In some situations several funnels may be active at the same time and the corresponding objects form a "meta-object" at a higher level of abstraction. The computer interface can then indicate to the operator the choice of acting on this meta-object or of picking a specific member of the cluster to interact with. This scenario is made clear by the following example.

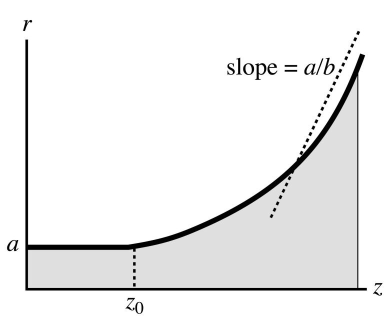

Figure 1: Schematic of a funnel as defined in Eq. (1).

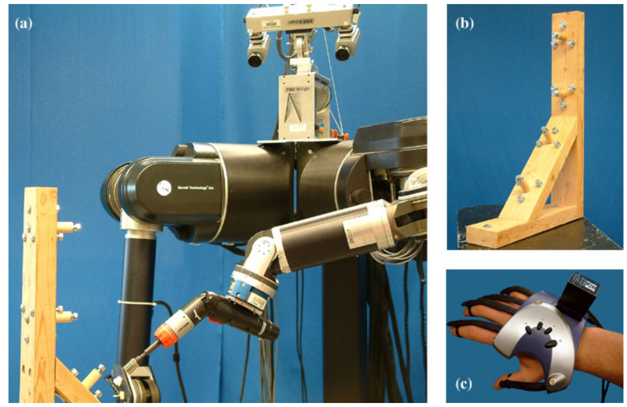

Figure 2: Teleoperation setup with (a) Dexter, the UMass Amherst humanoid robot, (b) the workpiece with 16 bolts arranged in four clusters, and (c) the glove input device and orientation sensor.

EXAMPLE

As a preliminary demonstration of how funnels might be used to predict operator intentions, we devised a variation of a canonical peg-insertion task. Figure 2 shows the setup for a teleoperation experiment with Dexter, the UMass Amherst humanoid robot. The task is to engage and then tighten each bolt fixed to a wooden stand. A glove-like input device (model P5 from Essential Reality, Mineola, NY) was used to specify translations of the end-effector (a cordless drill) and an inertial orientation sensor (InertiaCube from InterSense, Burlington, MA) was used to specify orientations. For the particular example shown below, the operator moved toward one of the bolts with no change in orientation.

Figure 3 shows several images from one of two cameras that provide the operator with a three-dimensional video interface. When the end-effector is far from the target bolt (upper panels) all 16 funnels are active and the interface makes the trivial prediction that the user intends to interact with the workpiece. This prediction is indicated by a circular overlay centered on the mean location of the bolts in the image plane; the radius of the circle is four times the standard deviation of the vertical pixel locations. The middle panels in Figure 3 illustrate that at closer distances to the target many funnels become inactive and then the interface predicts that the user intends to work on the upper bolt cluster. And near the target (lower panels) only one funnel remains active and so the interface is able to discriminate the desired bolt from the others.

DISCUSSION

The example in Figure 3 illustrates that a hierarchical human-robot interface is easily constructed by partitioning Cartesian space with funnels. Although funnels as represented in Eq. (1) work quite well with this example, we anticipate the need for greater flexibility when specifying the funnel shape with other tasks. One possible variation is to

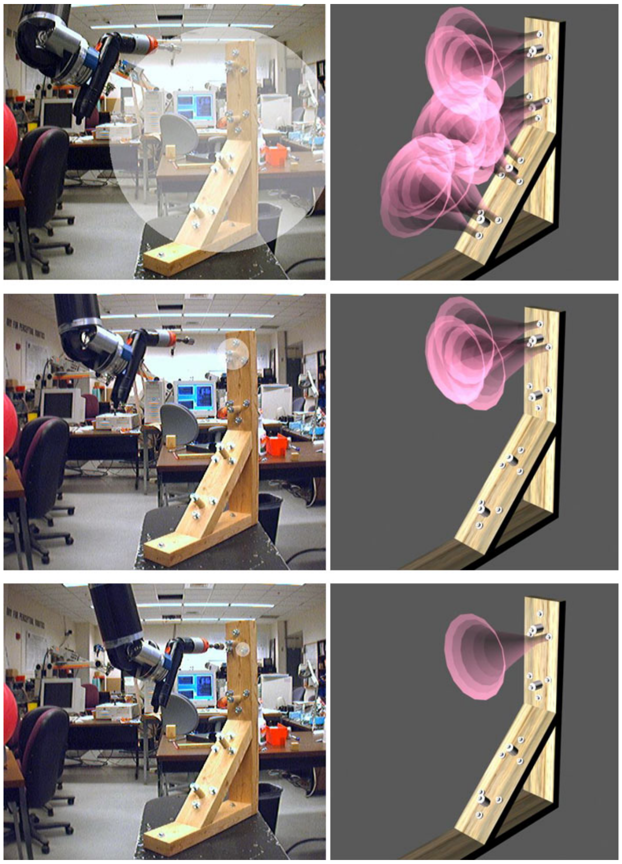

Figure 3: Sequence of images from the user interface as the operator guides the end-effector toward a target. The white, circular overlay depicts the hierarchical prediction, and the right-hand panels show the active funnels at the corresponding times. Funnel parameters were set to $a = 0.02$m, $b = 0.05$m, and $z_0 = 0.15$m

construct other kinds of funnels by union of geometric primitives (e.g., spheres, polyhedra).

Another possible variation of this work is to use machine learning techniques that adapt the interface to fit the operator's idiosyncrasies as well as the particulars of the task. Learning of this kind is similar to *learning from demonstration*, e.g., [1], although one key difference is that learning by demonstration typically involves a small set of trajectories demonstrated before the robot makes an attempt at solving the task. This contrasts with the ongoing involvement of a human supervisor who adds a degree of robustness to the learning process as well as the opportunity to learn from new situations as they arise. Another possible use of machine learning is to consolidate an operator's control "knowledge" into a machine compatible form. For instance, if expert operators always approach an object from a particular direction (relative to the object) then this may be useful evidence to bias a machine vision system searching for salient image features.

CONCLUSIONS AND FUTURE WORK

Perhaps more important than the use of learning methods, is future work to assess the benefits of the proposed interface technique. In this paper we sketched out one particular implementation and demonstrated the graphical display with the robot under full supervision by the human operator. The next logical step is to enable selection of autonomous movement through the object hierarchy. A subsequent user study will then allow us to quantify improvements in terms of execution speed, error, or progress on a parallel task.

Another key part of future work is to augment the funnels to include additional information. As described above, funnels are simply geometric objects used to carve up space into useful regions. For an effective human-robot interface, one must design (or learn) not only the funnel's shape but also the space over which it is defined. With some applications, the space may need to account for orientation in addition to position. Moreover, there may be circumstances where it is helpful to mix real-valued quantities with discrete sets of values. For example, one extra bit of memory to specify if a bolt has already been tightened can help the interface predict whether the operator intends to run the drill in Figure 2 forward or reverse. In any case, the approach outlined in this paper is a promising one as a general design strategy for human-robot interaction. Moreover, we expect the many possible variations of the basic theme to lead to opportunities for varied applications.

REFERENCES

1. C. G. Atkeson and S. Schaal. Robot learning from demonstration. In D. H. Fisher, editor, *Proceedings of the Fourteenth International Conference on Machine Learning*, pages 12–20, San Francisco, CA, 1997. Morgan Kaufmann.
2. R. R. Burridge, A. A. Rizzi, and D. E. Koditschek. Sequential composition of dynamically dexterous robot behaviors. *International Journal of Robotics Research*, 18(6):534–555, 1999.
3. G. W. Furnas. Generalized fisheye views. In *Proceedings of CHI '86*, pages 16–23, New York, 1986. ACM Press.
4. C. Gutwin. Improving focus targeting in interactive fisheye views. In *Proceedings of CHI '02*, pages 267–274, New York, 2002. ACM Press.
5. Y. K. Leung and M. D. Apperley. A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction*, 1(2):126–160, 1994.
6. P. Scerri, D. V. Pynadath, and M. Tambe. Towards adjustable autonomy for the real world. *Journal of Artificial Intelligence Research*, 17:171–228, 2002.