Mental Transformations in Human-Robot Interaction

B.P. DeJong[1], J.E. Colgate[2], and M.A. Peshkin[2]

[1] Central Michigan University, USA
[2] Northwestern University, USA

Abstract. Human-robot interfaces can be challenging and tiresome because of misalignments in the control and view relationships. The human user must mentally transform (e.g., rotate or translate) desired robot actions to required inputs at the interface. These mental transformations can increase task difficulty and decrease task performance. This chapter discusses how to improve task performance by decreasing the mental transformations in a human-robot interface. It presents a mathematical framework, reviews relevant background, analyzes both single and multiple camera-display interfaces, and presents the implementation of a mentally efficient interface.

Keywords: Mental transformation, control rotation, control translation, view rotation, teleoperation.

1 Introduction

In the summer of 1997, Argonne National Laboratory spent 2000 man-hours and $1.38 million dismantling their recently decommissioned nuclear reactor (Department of Energy, 1998). Rather than place humans in the radioactive environment, Argonne used a remotely controlled robotic system called the Dual Arm Work Platform (DAWP), consisting of two six-degree-of-freedom robotic arms and several tilt/pan/zoom cameras (Figure 1). Human operators sat at a console with several fixed video monitors and controlled the robots via two passive manipulanda[1] (Noakes et al., 2002).

The use of teleoperation[2] was cost-effective, but Argonne personnel noted several problems. First, training the operators was time-consuming and expensive; only 60% of the tested operators were skilled enough to complete tasks. Second, operators spent nearly 90% of the their time prepping rather than performing tasks. Finally, the teleoperation was mentally tiring, especially when performing complex tasks that required switching between multiple camera views (DeJong et al., 2004b).

---

[1] Manipulandum is a general term for the device that controls another, e.g., a joystick or kinematically-similar replica of the robot.
[2] Teleoperation is operation of a machine or robot at a distance.

X. Wang (Ed.): Mixed Reality and Human-Robot Interaction, ISCA 47, pp. 35–51.
springerlink.com

(a) (b)

Fig. 1. Argonne's Dual Arm Work Platform consisting of (a) remote robot and cameras, and (b) local user interface

Argonne's experience with teleoperation – and its difficulties – is not uncommon. Menchaca-Brandan and colleagues (2007) refer to similar struggles encountered during training of NASA astronauts. Teleoperated robots are now regularly used in space and underwater explorations (e.g., Menchaca-Brandan et al., 2007), military reconnaissance, and nuclear servicing (e.g., Park et al., 2002). These applications represent billions of dollars of funding and liability – for example, the liability for nuclear decommissioning work in the United States alone is around $30 billion (Park et al., 2004). Furthermore, teleoperation allows humans to reach into environments in ways that are otherwise impractical, such as in urban search and rescue (e.g., Murphy, 2004) or in minimally invasive surgery (e.g., Guthart and Salisbury, 2000).

Much of the mental burden in teleoperation, and human-robot interaction in general, is the result of misalignment in the hand-eye control of the robot. For example, to move a teleoperated robot an operator must determine the desired motion as seen in the camera views, and then impart the corresponding force or motion at the manipulandum. Thus, the operator must mentally transform (i.e., rotate, translate, scale, or deform) desired robot actions into necessary control inputs.

This chapter discusses the mental transformations found in human-robot interaction by investigating those inherent in teleoperation. It reviews relevant background, analyzes interfaces with only one camera and display, analyzes interfaces with multiple cameras and displays, and then examines the design of a low-mental-transformation interface.

The goal of this research is different from much of the literature. Rather than *adding* information into the interface to make the tasks easier (such as through augmented reality), it aims to reduce the complexity of the task by *removing* the need for mental transformations. If human-robot interfaces are made mentally efficient, then trial-and-error control may be reduced, training and task times decreased, and task performance improved.

2 Background

2.1 Components and Coordinate Frames

In general, teleoperation involves two worksites: a local worksite with the manipulandum, user interface, and human user, and a remote worksite with the robot (fixed or mobile) and several cameras (and possibly other sensors). The cameras may be on the robot or external to it, and may have tilt, pan, or zoom capabilities. The cameras' images are shown on displays in the user interface; for simplicity, assume each camera corresponds to one display.

Coordinate frames can be defined for each of the components – see Figure 2 and (DeJong et al., 2004). At the remote site there are:

- The site's world coordinates, $\boldsymbol{W}_R$
- The robot's control coordinates, $\boldsymbol{R}$
- Camera *i*'s view coordinates, $\boldsymbol{C}_i$

At the local site, there are:

- The site's global coordinates, $\boldsymbol{W}_L$
- The manipulandum's control coordinates, $\boldsymbol{M}$
- Display *i*'s view coordinates, $\boldsymbol{D}_i$
- The human operator's view coordinates, $\boldsymbol{H}$

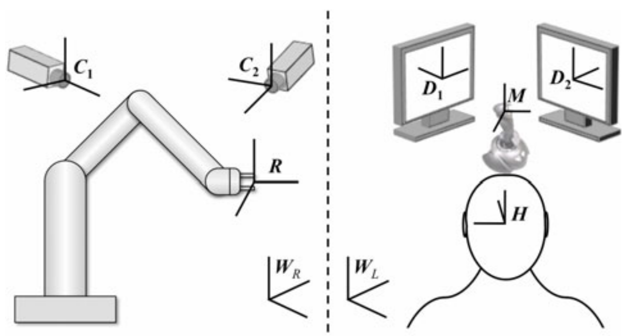

Fig. 2. Sample teleoperation components and their coordinate frames

Coordinate frames $\boldsymbol{M}$ and $\boldsymbol{R}$ represent the control of the robot – here, an input along one axis in $\boldsymbol{M}$ moves the robot along a corresponding axis in $\boldsymbol{R}$. In practice, $\boldsymbol{R}$ may be defined in a variety of ways, such as at the base of the robot or aligned with a tool, and it may be mobile or fixed. $\boldsymbol{M}$ may not correspond to the manipulandum's physical frame. In some situations, such as with joint-controlled robots, these frames are more complicated, although this analysis holds.

Each camera and display can be interpreted as a bounded plane (for the camera this is the image plane; for the display this is the screen's plane) and a perpendicular line through the center of that plane. This centerline represents the angle a camera or display is pointing. Since $\boldsymbol{C}_i$'s positive centerline points out of the camera, $\boldsymbol{D}_i$'s positive centerline points into the display. $\boldsymbol{C}_i$ may be fixed or moving if the camera has tilt/pan/zoom capabilities.

The human frame $\boldsymbol{H}$ is located at the eyes of the operator and moves with the eyes when the user looks around. One axis is where the eyes are looking; the point of focus lies on this line. The remaining two axes are in the peripheral directions.

In most teleoperation situations, the location transformations can be measured or calculated for each of the components. For example, using Special Euclidian notation (SE(3)), the transformation from remote world frame to the robot control is

$$\boldsymbol{R} = {}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{T} \cdot \boldsymbol{W}_R = \begin{bmatrix} \left(\quad {}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{R} \quad \right) & \left({}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{t} \right) \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix} \cdot \boldsymbol{W}_R, \tag{1}$$

where ${}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{R}$ and ${}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{t}$ are the rotation and translation matrices from $\boldsymbol{W}_R$ to $\boldsymbol{R}$. In the same fashion, matrices ${}^{\boldsymbol{C}_i}_{\boldsymbol{W}_R}\mathbf{T}$, ${}^{\boldsymbol{D}_i}_{\boldsymbol{W}_L}\mathbf{T}$, ${}^{\boldsymbol{M}}_{\boldsymbol{W}_L}\mathbf{T}$, and ${}^{\boldsymbol{H}}_{\boldsymbol{W}_L}\mathbf{T}$ are the transformations from the corresponding world frame to camera *i*, display *i*, manipulandum, and human. For more information regarding rotational and translational matrices, see (Mason, 2001).

These location matrices can be combined into more useful component-to-component transformations, such as from manipulandum to robot,

$${}^{\boldsymbol{R}}_{\boldsymbol{M}}\mathbf{T} = {}^{\boldsymbol{R}}_{\boldsymbol{W}_R}\mathbf{T} \cdot {}^{\boldsymbol{W}_R}_{\boldsymbol{W}_L}\mathbf{T} \cdot {}^{\boldsymbol{W}_L}_{\boldsymbol{M}}\mathbf{T}, \tag{2}$$

or from human to display,

$${}^{\boldsymbol{D}_i}_{\boldsymbol{H}}\mathbf{T} = {}^{\boldsymbol{D}_i}_{\boldsymbol{W}_L}\mathbf{T} \cdot {}^{\boldsymbol{W}_L}_{\boldsymbol{H}}\mathbf{T}. \tag{3}$$

However, the camera-to-display transformations, ${}^{\boldsymbol{D}_i}_{\boldsymbol{C}_i}\mathbf{T}$, cannot be directly calculated from the camera and display location matrices, because each display image is a two-dimensional projection of the three-dimensional workspace. Assume that theses projections are linear.[3] Points in $\boldsymbol{C}_i$ are projected onto the camera's bounded image plane, and shown on $\boldsymbol{D}_i$'s bounded screen plane, often at a different scaling. This projection, shown in Figure 3 where f_i is the camera *i*'s focal length, is a scaling dependent on distance from the camera. That is, by similar triangles, a point (u_i, v_i, w_i) in the camera's frame is flattened to

$$(u_i, v_i, w_i) \rightarrow \frac{f_i}{w_i + f_i} \cdot (u_i, v_i, 0) \tag{4}$$

[3] This assumption does not hold for wide-angle cameras (Weng et al., 1992).

on the image plane. The boundedness of the image plane and display mean that only a limited range of points are mapped from one frame to the other. Therefore, the transformation from camera to display can be interpreted as a camera projection scaling, a camera image plane cropping, a display scaling (for the display's resolution, pixel size, and aspect ratio), finally followed by a display cropping to fit the screen.

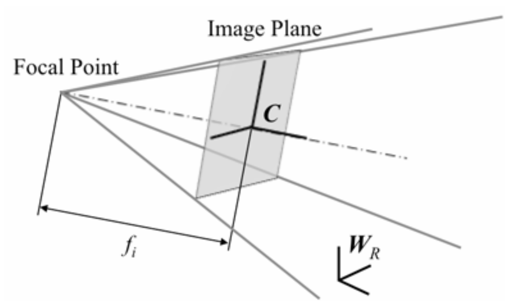

Fig. 3. Projection of points in the camera coordinate frame

2.2 The Cost of Mental Tranformations

Human-robot interfaces frequently have a misalignment between inputs at the manipulandum and corresponding robot actions, as seen in the displays. These misalignments force the user to mentally transform between control outputs and inputs.

We encounter such control misalignment quite frequently in our daily lives, although it is usually simple and easy to learn. For example, controlling a computer mouse consists of sliding a horizontal mouse to move a vertical pointer – the task is (initially) more difficult when switching to a new computer with a different mouse scaling or if the mouse frame is rotated with respect to the screen frame. Driving a car consists of rotating a vertical steering wheel to steer our horizontal path – the task becomes more difficult when driving in reverse. First-person and third-person video and computer games often have inverted pitch or yaw – in third-person games, the task is temporarily harder when the character is facing sideways or backwards in the camera view. Remote controlled cars, boats, and planes are steered in their local coordinates that may be rotated from the operator's frame – the task is noticeably difficult when the vehicle is driving towards the operator.

While we learn simple transformations quite readily, research shows that they can be mentally taxing, especially when first learning.

In the cognitive science realm, studies show that all mental transformations can be challenging. Shepard and Metzler (1971) found that the time to mentally rotate objects is linearly related to the angle of rotation. Wexler et al. (1998) studied the relationship between mental and motor rotation. Anderson (1995) and Kosslyn (1980) performed experiments in mental image scanning showing that the time to mentally translate is related to distance. The results from these sources and others (e.g., Pylyshyn, 1979; Barfield et al., 1988) show that mental rotation is by far the most costly.

Also in the cognitive science literature, there is much discussion on mental workload[4] (Hancock and Chignell, 1988; Schvaneveldt et al., 1998; Johnson et al., 2003; DiDomenico, 2003). Most of the research has focused on reducing the mental workload of multiple sensor feedback. Since it has been shown in literature that mental transformations are costly, it follows that reducing them will decrease mental workload and improve task performance.

Furthermore, there has been research on hand-eye coordination and adaptation to errors in it. Miyake (1982) tested rotational misalignment for various arm positions and found that motions were identified in egocentric, rather than external, coordinates. Many studies have addressed visual distortions, such as nonlinear mappings by Flanagan and Rao (1995) and rotations by Krakauer et al. (2000).

Even in teleoperation literature, proper hand-eye coordination is not a new topic. For example, Sheridan (1992) discusses it in terms of a misalignment of proprioception, or "sense of self". Menchaca-Brandan et al. (2007) ran an experiment showing that users' spatial abilities correlate to task performances. A different teleoperation experiment (DeJong et al., 2004) shows that interface arrangement can affect task time and accuracy. Other experiments show that task performance depends on the definition of robot's control frame (Hiatt and Simmons, 2006; Lamb and Owen, 2005). For fixed-base robots with a single camera, control alignment is often achieved by computationally rotating the manipulandum's frame, as done by Thompson (2001), or in the well-known DaVinci surgical system (Guthart and Salisbury, 2000). Regarding multi-camera systems, Chiruvolu et al. (1992) show that hand-eye misalignments are even more critical when the operator is simultaneously using multiple video displays.

Many human-robot interfaces do not attempt to reduce mental transformations, but rather try to make them easier by adding augmented reality overlays. Cao (2000) presents the use of augmented reality navigational and orientational cues in endoscopy, to improve navigation and help surgeons "effectively perform mental rotations and mappings". Nawab et al. (2007) overlay their robot's frame onto the displays, to help the user learn the transformations. Similarly, Nielsen and his colleagues (2007) use augmented reality to fuse video, map, and robot-pose information, to help the user "mentally correlate the sets of information".

3 Single Camera/Display Interface

Given the previous terminology, this section examines the mental transformations found in a teleoperation setup with only one camera and display. In general, a mental transformation is required any time the perceived robot[5] frame is misaligned with the human operator's internal frame. There are three primary sources of

[4] Sanders and McCormick (1993) define mental workload as "a measurable quantity of the information processing demands placed on an individual by a task."

[5] The perceived robot is the robot as seen in the video displays.

misalignment: rotation between perceived robot and manipulandum (called control rotation), translation between perceived robot and manipulandum (called control translation), and rotation between human and display (called view rotation).

3.1 Control Rotation

The relationship that proves the most critical is the rotation between manipulandum and the perceived robot, called *control rotation*. Suppose the manipulandum and perceived robot are those shown in Figure 4: the perceived $\boldsymbol{R}$ is rotated with respect to $\boldsymbol{M}$. To move the robot in one direction, the operator may need to push or move the manipulandum in an entirely different direction. This mental rotation can be confusing, especially if it is about an axis that is neither vertical nor horizontal, or if it is a large rotation. To have no control rotation, the transformation from manipulandum to robot must be identity, $\boldsymbol{I}$:[6]

$$ {}^{R}_{M}\mathbf{R} = {}^{R}_{C}\mathbf{R} \cdot {}^{C}_{D}\mathbf{R} \cdot {}^{D}_{M}\mathbf{R} = \boldsymbol{I} \ . \tag{5} $$

Recall that the transformation from camera to display is a scaling and cropping (i.e., no rotation), so ${}^{C}_{D}\mathbf{R} = \boldsymbol{I}$. Therefore, to eliminate control rotation, the rotation from robot to camera must be the same as that from manipulandum to display:

$$ {}^{C}_{R}\mathbf{R} = {}^{D}_{M}\mathbf{R} \cdot \tag{6} $$

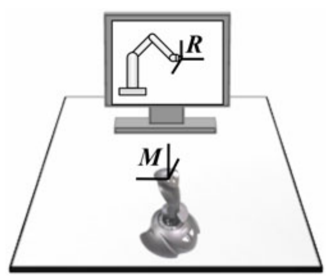

Fig. 4. Single camera and display interface with control rotation

Often, a camera used in teleoperation can pan or tilt, meaning that its orientation matrix, ${}^{C}_{R}\mathbf{R}$, changes. To avoid control rotation when panning or tilting the camera, the robot control frame, the manipulandum control frame, or the display itself must be rotated in unison. Rotating the display affects the human/display relationship, as shown later, and has limited range before the image is physically unviewable. Rotating the robot frame may destroy an intuitive control relationship with regards to the robot's kinematics. For example, if the robot frame was chosen

[6] The subscript *i* has been dropped because there is only one camera and display.

such that one of the axes is aligned with an arm of the robot, the rotated frame may alter this relationship.

Therefore, computationally or physically rotating the manipulandum control frame is usually the best choice, although it still has dangers. If the user is holding the manipulandum while panning or tilting the camera, then rotating the manipulandum's frame can cause erroneous input. Computational rotation is simple to implement and requires no additional hardware, but it may destroy kinematic mapping if the manipulandum is kinematically similar to the robot. On the other hand, physical rotation requires additional hardware, but it offers visual and haptic feedback of the rotation.

3.2 Control Translation

Even if the control is rotationally aligned, it may still require mental translations, called *control translation*. Suppose the manipulandum and the perceived robot are those in Figure 5, where the frames are oriented the same, but the manipulandum is translated to the side. The user needs to mentally translate control inputs from the perceived robot to the manipulandum. For example, if the desired motion of the robot is up-right-back, then the required input is on a parallel line but translated to the manipulandum. To eliminate control translation, the translation matrices must be identical:

$$ {}^{C}_{R}\mathbf{t} = {}^{D}_{M}\mathbf{t}\,. \tag{7} $$

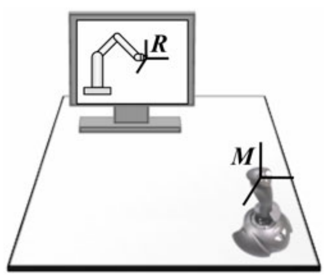

Fig. 5. Single camera and display interface with control translation

As shown in the literature, mental translation is nontrivial, but not as burdensome as mental rotation. In fact, almost all human-robot interfaces involve some control translation, such as along the depth axis of the display (i.e., the perceived robot in a physical display cannot be in the exact same location as a physical manipulandum). For that matter, depth in the display is hard to quantify, because the image is a projection onto the image plane. The user may perceive the robot as near (or far) if it appears larger (or smaller) than expected.

Panning or tilting the camera affects control translation as it did control rotation. Panning and tilting move the camera about a center of rotation that is usually not the center of the image plane, so the image plane is rotated and translated with respect to world coordinates.[7] Thus, the perceived robot translates and rotates in the video image.

In addition, camera zoom affects the control relationship, although it is less significant. Zooming a camera (i.e., changing its focal length) changes the scaling in its projection: points in space move radially to and from the image's center. That is, the perceived robot changes size and location in the display's image. Therefore, camera zoom (ignoring any cropping effects) is similar to a three-dimensional translation of the robot, camera, display, or manipulandum.

3.3 View Rotation

The third transformation of interest is the view transformation from human user to display, ${}^{D}_{H}\mathbf{T}$. The human coordinate frame is defined as always pointing in the direction the user is looking, such as at the display. This definition restricts human motions to three dimensions: longitudinal and lateral rotation about the display (as on a sphere), and translation towards and away from the display (radially).

Translation of the human frame towards and away from the display (radially), without changing the control relationship, does not require any additional mental transformations. Consideration should be given, of course, to the viewable distance of the display and its resolution. This also means that the size of the display does not affect mental transformations, as long as the display is not so large that the user cannot view all of it without turning his head.

On the other hand, rotation of the human frame about the view (longitudinally and laterally) is important. This rotation is called *view rotation*. Suppose the human and display are situated as in Figure 6, with the human frame rotated from the display's centerline. Since the image on the display is two dimensional, the user sees the same information regardless of angle. The angled image is not of the perceived robot from a new angle as it would be if the user moved about the real robot. Because of the user's mental model of the monitor and the depth into the image, the image is interpreted (correctly) as a rotated two-dimensional plane. To move the robot, the user now must determine the desired robot action in this rotated plane, requiring mental rotation of the image[8] so they are perpendicular. That is, the mental model of the planar image is mentally rotated to determine control inputs. To eliminate view rotation, the human's view axis must be perpendicular to the displays plane, *at the area of interest*. This constraint is clearly impractical because it means that as the user changes his area of interest, he must translate. Thus, a more reasonable constraint is to place the human on the display's centerline to reduced view rotation. This constraint can be represented mathematically by the conditions

---

[7] If the center of rotation is the center of the image plane, then the translation component is zero.

[8] Or the user, though this appears to be harder (Wraga et al., 1999).

$$ {}^{D}_{H}\mathbf{R} = I \text{ and } {}^{D}_{H}\mathbf{t} = \begin{pmatrix} 0 \\ 0 \\ l_0 \end{pmatrix}, \tag{8} $$

where l_0 is the radial distance from human to display.

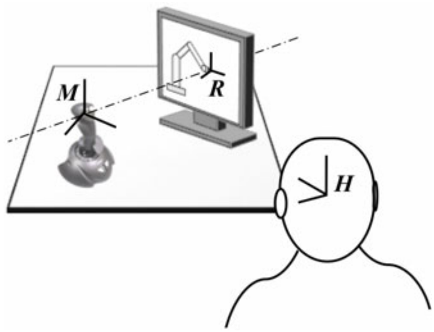

Fig. 6. Single camera and display interface with view rotation

For view rotation, using stereovision is more harmful than using monoscopic vision because of "induced stereo movement" (Tyler, 1974; Arthur; 1993). If the display is stereoscopic, the image appears three-dimensional when the user is looking perpendicular at the screen. Imagine that the image on the screen is a hand pointing at the user. When the user is rotated to the side, the two stereoscopic images are the same, so the hand still appears to be pointing at the user rather than in the original direction. With teleoperation, this means that the perceived robot has rotated with respect to the manipulandum. Thus, with stereovision, it is necessary that the human remains on the display's centerline. Systems like the DaVinci telesurgical robot that rely heavily on stereovision restrict the human's head to a fixed location and orientation (Guthart and Salisbury, 2000).

3.4 How to Design a Single Camera/Display Interface

Therefore, there are three relationships of interest, in order of importance: control rotation, view rotation, and control translation. The cognitive science literature shows that these transformations are mentally taxing, and the human-robot literature shows that reducing them can increase task performance (DeJong, 2003).

In designing an interface with only one camera and one display, the designer may wish to position the camera, manipulandum, and video display properly so as to eliminate or minimize the control transformations. Given the locations of any two of these three components, the designer can calculate the location for the third component that follows Eqs. 6-7. For example, given the camera and display locations, there is a specific position and orientation for the manipulandum that minimizes the mental transformations.

Once the control is aligned, the user should be situated on the display's centerline, to follow Eq. 8.

4 Multiple Camera/Display Interfaces

Often, human-robot interfaces incorporate more than one camera and display to give the user additional visual feedback of the remote worksite. The previous single camera/display methods are easily extended to interfaces with multiple cameras and displays.

4.1 Multiple Control and View Relationships

When an interface involves more than one camera-display pair, the control and view alignments are especially important. If the user is using one of the displays and it involves mental transformations, he may eventually learn the mapping such that control is relatively easy. However, once the user switches attention to a different display, the mapping has changed, and he must learn a new mapping. Even when switching back to the original display, the user must relearn the relationships. This switching and relearning has been shown to be mentally taxing (Chiruvolu et al., 1992).

For example, suppose the interface consists of two displays, as shown in Figure 7. This interface has properly aligned manipulandum and perceived robot frames for the camera-display on the left. When the user is using the left display, control is straightforward: an input on the manipulandum moves the perceived robot in the same way. However, when the user attempts to use the display on the right, the control is misaligned: an input moves the perceived robot in an unintuitive way.

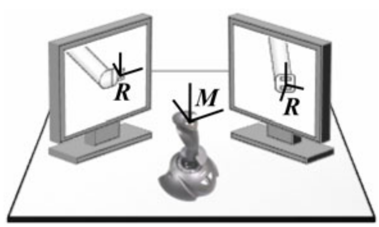

Fig. 7. An example of a mentally inefficient interface with two cameras and displays

4.2 How to Design a Multiple Camera/Display Interface

Designing a mentally efficient interface is more complicated when it involves multiple control and view mappings. Each camera/display should independently satisfy the control and view constraints mentioned previously.

For simplicity, assume that the cameras and robot frames are fixed and known, but the displays, manipulandum, and human frames need to be arranged. This is often the case. The robot frame is often intelligently chosen for the given application and robot kinematics. Similarly, the cameras are often chosen based on where they physically can be placed and where they are most beneficial for the given tasks.

The best solution is to align the manipulandum with every perceived robot frame simultaneously, by placing the manipulandum and then arranging the displays around it such that Eq. 6 is satisfied for each display. If this can be done, then the control is aligned for whichever display the user is currently using – more importantly, there is no need to learn a new mapping when switching displays. Figure 8 shows a two-display teleoperation setup that achieves this. This method arranges the displays on a sphere around the user and manipulandum, such as that in Figure 9.

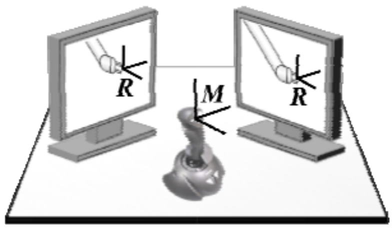

Fig. 8. An example of an aligned interface with two cameras and displays

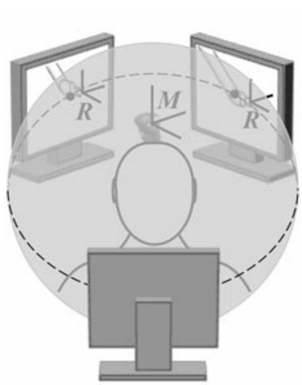

Fig. 9. Method of placing displays on sphere around the user

A constraint that follows from simultaneous alignment is that the relative orientation of the displays must be the same as the relative orientation of the cameras. For simultaneous rotational alignment of any two displays with the same robot and manipulandum (recall Eq. 6),

$$ {}^{C_1}_{R}\mathbf{R} = {}^{D_1}_{M}\mathbf{R} \text{ and } {}^{C_2}_{R}\mathbf{R} = {}^{D_2}_{M}\mathbf{R}\,, \tag{9} $$

which leads to:

$$ \begin{gathered} {}^{C_1}_{R}\mathbf{R} \cdot {}^{R}_{C_2}\mathbf{R} = {}^{D_1}_{M}\mathbf{R} \cdot {}^{M}_{D_2}\mathbf{R} \\ {}^{C_1}_{C_2}\mathbf{R} = {}^{D_1}_{D_2}\mathbf{R} \end{gathered} \tag{10} $$

This makes sense: if the two cameras are perpendicular to each other, the displays must be perpendicular as well. The camera on the left as viewed from the robot (facing the front of the camera) corresponds to the display on the left as viewed by the user (facing the front of the display).

This ideal solution is not always feasible, such as when the aligned location for a display is not physically achievable due to space conflicts. When the displays cannot be aligned simultaneously, the issue becomes how to enforce alignment for the display the operator is currently using.

One way to accomplish alignment is to perform a transition of the manipulandum frame when the user is switching views. In Figure 7, this means that when the user switches attention to the display on the right, the manipulandum's frame must be rotated (computationally or physically) to align it with the new perceived robot frame. The transition can be initiated from various sources, such as eye tracking or the user pressing a key corresponding to the new display.

Unfortunately, transitioning the manipulandum frame has drawbacks. First, rotating it dynamically can cause erroneous motion. Conversely, forcing the user to stop inputs during transition increases task time. Second, passive measures of attention switching, such as eye tracking, may initiate transition when the user is merely glancing between displays. On the other hand, requiring the user to press a button when switching adds time and may increase overall mental workload for an interface. Third, if the task requires close attention to both views simultaneously (e.g., when trying not to collide with multiple objects seen in separate views), transition becomes completely infeasible.

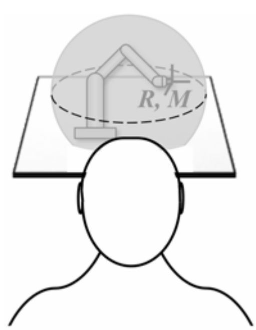

Fig. 10. Method of placing displays facing out from a common point – the ideal of which is a hologram

An alternative method to placing the displays around the user is to place the user and manipulandum around the displays. If the displays can be placed facing out from a central point, then the user and manipulandum can be moved around the displays while keeping control and view alignment. This method applies if the sphere of displays is replaced with a hologram of the robot, in which case a kinematically similar manipulandum can be placed physically coincident with the perceived robot (see Figure 10).

4.3 Ergonomics and Camera Motion

One disadvantage to the arranging the displays as presented in the previous section is that the location and orientation for a display that minimizes mental transformations may be in an ergonomically poor location. It is possible that a display should be placed above or even behind the operator to minimize transformations. For example, if two cameras face in opposite directions (e.g., front and back or left and right of the robot), then the displays should also do so (as in Figure 9).

5 Argonne Redesigned

Motivated by Argonne's past experience with teleoperation, we (the authors and our colleagues, including Argonne researchers) established a collaborative testbed for improving teleoperation. The remote worksite was located at Argonne and included a six-degree-of-freedom robotic arm, two video cameras, and a structured light system (Park et al., 2004). The local worksite was located thirty miles away at Northwestern University and consisted of a cobotic manipulandum (Faulring et al., 2006) and two displays with augmented reality. The manipulandum, robot, and user interface communicated via Ethernet. The testbed successfully implemented reduced mental transformations, structured light sensing, virtual surfaces, cobotic technology, and augmented reality aids (DeJong et al., 2006).

In designing the interface, we carefully arranged the components to minimize mental transformations. We positioned the displays around the user with simultaneous control alignment using the following steps (see Figure 11).

First, we placed the cameras at the remote site to provide useful views of the robot's workspace.

Second, we placed the two displays in front of the user such that the user was at the intersection of their centerlines. Doing so minimized view rotation without transition needed when shifting from one view to the other. Furthermore, we angled the displays to match the angle between the cameras (following Eq. 10).

Third, we needed to properly position the manipulandum to minimize control rotation and translation. Our manipulandum was not kinematically similar to the robot, so we had freedom in the computational definition of the manipulandum frame. Because of the size and shape of our cobotic manipulandum, we placed it on a table between our two displays. Doing so also minimized translational misalignment between manipulandum and perceived robot, i.e., it minimized control translation.

Finally, we needed to orientate the manipulandum frame to eliminate rotational misalignment between manipulandum and perceived robot, i.e., eliminate control rotation. The manipulandum had an internal reference frame for recording inputs, $\boldsymbol{Int}$, and a computational definition of the manipulandum control frame, ${}^{M}_{Int}\mathbf{R}$. Using one camera and display, we calculated this definition from Eq. 6, knowing ${}^{R}_{C}\mathbf{R}$ and ${}^{D}_{Int}\mathbf{R}$:

$$ {}^{M}_{Int}\mathbf{R} = {}^{M}_{D}\mathbf{R} \cdot {}^{D}_{Int}\mathbf{R} = {}^{R}_{C}\mathbf{R} \cdot {}^{D}_{Int}\mathbf{R} . \tag{11} $$

The manipulandum recorded inputs, multiplied them by this matrix, and then commanded them to the robot.

The teleoperation testbed showed significant improvement in task performance (DeJong et al., 2006). When arbitrary and minimized-transformation interfaces were both used by Argonne personnel, they were surprised by how much the reduced transformations made human-robot interaction easier. Figure 11 shows the mentally efficient interface.

Fig. 11. Mentally efficient interface for Northwestern-Argonne teleoperation

6 Conclusion

Clearly, mental transformations in human-robot interfaces degrade task performance, increase the skill required by users, and are mentally tiresome for the user. These transformations come from misalignment between manipulandum and perceived robot (control rotation and translation), and user and display (view rotation).

Ideally, human-robot interfaces should be carefully designed to minimize the mental transformations. This can be accomplished through intelligent arrangement of the interface's components, as shown in the sample implementation. Doing so makes tasks inherently easier, and can increase task performance.

Acknowledgements

The authors would like to thank their Argonne colleagues Young Park, Hyosig Kang, and Thomas Ewing, as well as Eric Faulring for use of the cobotic manipulandum. The original research was funded by the Department of Energy.

References

Anderson, J.R.: Review: perception-based knowledge representations. In: Cognitive Psychology and Its Implications, 4th edn. Freeman, New York (1995)

Arthur, K.W.: 3D task performance using head-coupled stereo displays. M.S. thesis, University of British Columbia (1993)

Barfield, W., Sandford, J., Foley, J.: The mental rotation and perceived realism of computer-generated three-dimensional images. International Journal of Man-Machine Studies 29(6), 669–684 (1988)

Cao, C.G.L.: Augmented reality displays for endoscopic orientation and navigation. In: CHI 2000 extended abstracts on Human Factors in Computing Systems (2000)

Chiruvolu, R., Hwang, V.S., Sheridan, T.B.: Virtual display aids for teleoperation. In: SPIE – The International Society for Optical Engineering (1992)

DeJong, B.P.: Improving Teleoperation Interfaces: Creation of a Testbed and Thinking Inside the Box. M.S. thesis, Northwestern University (2003)

DeJong, B.P., Colgate, J.E., Peshkin, M.A.: Improving teleoperation: reducing mental rotations and translations. In: IEEE International Conference on Robotics and Automation, New Orleans (2004)

DeJong, B.P., Colgate, J.E., Peshkin, M.A.: Improving teleoperation: reducing mental rotations and translations. American Nuclear Society, Gainesville, FL (2004)

DeJong, B.P., Faulring, E.L., Colgate, J.E., Peshkin, M.A.: Lessons learned from a novel teleoperation testbed. Industrial Robot 33(3), 187–193 (2006)

Department of Energy: Dual Arm Work Platform Teleoperated Robotics System, DOE/EM-0389 (1998)

DiDomenico, A.T.: An investigation on subjective assessments of workload and postural stability under conditions of joint mental and physical demands. Ph.D. thesis, Virginia Polytechnic Institute and State University (2003)

Faulring, E.L., Colgate, J.E., Peshkin, M.A.: The cobotic hand controller: design, control and performance of a novel haptic display. The International Journal of Robotics Research 25, 1099–1119 (2006)

Flanagan, J.R., Rao, A.K.: Trajectory adaption to a nonlinear visuomotor transformation: evidence of motion planning in visually perceived space. Journal of Neurophysiology 74(5), 2174–2178 (1995)

Guthart, G.S., Salisbury, J.K.: The intuitive telesurgery system: overview and application. In: IEEE International Conference on Robotics and Automation (2000)

Hancock, P.A., Chignell, M.H.: Mental workload dynamics in adaptive interface design. IEEE Transactions on Systems, Man, and Cybernetics 18(4), 647–658 (1988)

Hiatt, L.M., Simmons, R.: Coordinate frames in robotic teleoperation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (2006)

Johnson, C.A., Adams, J.A., Kawamura, K.: Evaluation of an enhanced human-robot interface. In: IEEE International Conference on Systems, Man, and Cybernetics (2003)

Kosslyn, S.M.: Image and Mind. Harvard University Press, Cambridge (1980)

Krakauer, J.W., Pine, Z.M., Ghilardi, M.F., Ghez, C.: Learning of visuomotor transformations for vectorial planning of reaching trajectories. Journal of Neuroscience 20(23), 8916–8924 (2000)

Lamb, P., Owen, D.: Human performance in space telerobotic manipulation. In: ACM Symposium on Virtual Reality Software and Technology (2005)

Mason, M.T.: Mechanics of Robotic Manipulation. MIT Press, Cambridge (2001)

Menchaca-Brandan, M.A., Liu, A.M., Oman, C.M., Natapoff, A.: Influence of perspective-taking and mental rotation abilities in space teleoperation. In: ACM/IEEE International Conference on Human-Robot Interaction (2007)

Miyake, Y.: Coordinate systems for hand control. Ph.D. thesis, University of California (1982)

Murphy, R.R.: Human-robot interaction in rescue robotics. IEEE Transactions on Systems, Man, and Cybernetics 34(2) (2004)

Nawab, A., Chintamani, K., Ellis, D., Auner, G., Pandya, A.: Joystick mapped augmented reality cues for end-effector controlled tele-operated robots. In: Virtual Reality Conference (2007)

Nielsen, C.W., Goodrich, M.A., Ricks, R.W.: Ecological interfaces for improving mobile robot teleoperation. IEEE Transactions of Robotics and Automation 23(5) (2007)

Noakes, M.W., Love, L.J., Lloyd, P.D.: Telerobotic planning and control for DOE D and D operations. In: IEEE International Conference on Robotics and Automation (2002)

Park, Y.S., Ewing, T.F., Yule, T.J., Colgate, E., Peshkin, M.: Enhanced teleoperation exhibiting tele-autonomy and tele-collaboration. Department of Energy, ANL/TD/CP-107730 (2002)

Park, Y.S., Kang, H., Ewing, T.F., Faulring, E., Colgate, J.E., Peshkin, M.: Enhanced teleoperation for D&D. In: IEEE International Conference on Robotics and Automation, New Orleans (2004)

Pylyshyn, Z.W.: The rate of 'mental rotation' of images: a test of a holistic analogue hypothesis. Memory and Cognition 7, 19–28 (1979)

Sanders, M.S., McCormick, E.J.: Human Factors in Engineering and Design, 7th edn. McGraw-Hill, New York (1993)

Schvaneveldt, R.W., Gomez, R.L., Reid, G.B.: Modeling mental workload. Cognitive Technology 3(1) (1998)

Shepard, R.N., Metzler, J.: Mental rotation of three-dimensional objects. Science 171, 701–703 (1971)

Sheridan, T.B.: Telerobotics, Automation, and Human Supervisory Control. MIT Press, Cambridge (1992)

Thompson, R.L.: Integration of visual and haptic feedback for teleoperation. Ph.D. thesis, University of Oxford (2001)

Tyler, W.: Induced stereo movement. Vision Research 14, 609–613 (1974)

Weng, J., Cohen, P., Herniou, M.: Camera calibration with distortion models and accuracy evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(10) (1992)

Wexler, M., Kosslyn, S.M., Berthoz, A.: Motor processes in mental rotation. Cognition 68, 77–94 (1998)

Wraga, M., Creem, S.H., Proffitt, D.R.: The influence of spatial reference frames on imagined object and viewer rotations. Acta Psychologica 102, 247–264 (1999)