Improving Teleoperation: Reducing Mental Rotations and Translations

Brian P. DeJong, J. Edward Colgate, Michael A. Peshkin

Mechanical Engineering
Northwestern University
Evanston, IL
b-dejong@northwestern.edu

Abstract—**We consider teleoperation in which a slave manipulator, seen in one or more video images, is controlled by moving a master manipulandum. The operator must mentally transform (i.e. rotate, translate, scale, and/or deform) the desired motion of the slave image to determine the required motion at the master. Our goal is to make these mental transformations less taxing in order to decrease operator training time, improve task time/performance, and expand the pool of candidate operators. In this paper, we introduce a framework for describing the transformations required to use a particular teleoperation setup. We analyze in detail the mental transformations required in an interface consisting of one camera and display. We then expand our discussion to setups with multiple cameras/displays and discuss the results from an initial experiment.**

Keywords - teleoperation; mental transformation; mental rotation; mental translation; multiple views

I. INTRODUCTION

The teleoperation operator encounters many difficulties when attempting to complete a task. While some arise from the complexity of the task itself, much of the burden comes from controlling the robot slave via the restricted visual information in the teleoperation interface. Control consists of determining the desired motion of the slave as seen in the video images and then moving the master manipulandum to achieve that motion. Thus, the operator must mentally transform (i.e. rotate, translate, scale, and/or deform) the desired motion of the slave to determine the required input at the master, a task often mentally challenging and tiresome. These transformations are especially demanding with interfaces that involve multiple cameras and displays, as the operator must relearn them whenever he switches views.

A motivating example of this teleoperation difficulty is an application at Argonne National Laboratory (ANL). ANL has used teleoperation techniques to dismantle the inside of a retired nuclear reactor as part of an ongoing decontaminating and decommissioning project. The slave robot used was a Schilling Dual Arm Work Platform (DAWP), consisting of two six-degree-of-freedom robotic arms and several tilt/pan/zoom cameras. The control interface included two passive masters and several video monitors arrayed in an arc around the operator [1]. While teleoperation in the radioactive environment was cost-effective, ANL noted several problems. First, the operator selection and training process was very time-consuming and expensive; only 60% of the tested operators were skilled enough even to complete tasks. Second, they found that operators spent over 90% of their time setting up for rather than actually performing tasks. Finally, ANL's operators found the teleoperation tiring, especially when performing tasks that required switching between multiple camera views. Other work regarding the DAWP and subsequent designs can be found in [2] and [3].

The goal of the research reported here is to make the cognitive transformations inherent in teleoperation less mentally taxing. If we explore the mental transformations inherent in using a particular setup and can determine how cognitively challenging they are, we may find guidelines for how to make a setup less tiresome. We hope to reduce trial-and-error control, decrease training and task times, increase task performance, and expand the pool of candidate human operators.

Finding a teleoperation setup that requires no mental transformations may seem like the best objective, but doing so may be unnecessary. There are many transformations at which humans are very adept (such as using a computer mouse, driving a car, playing video games, or watching television from off-center). These transformations add such little mental cost that they may not be worth the extra effort or hardware required to remove them.

Therefore, instead of proposing a "zero transformation" setup, we investigate ways to reduce the teleoperation mental workload by optimizing component arrangement in an interface. We will limit our discussion primarily to monoscopic video since we are concerned with tasks that require more than one camera view, though most of our conclusions also apply to stereoscopic interfaces.

II. COORDINATE FRAMES

We define several coordinate frames for the key components in a teleoperation setup (Fig. 1). At the slave's site, we define

- The slave site's World coordinates, $Ws = (X,Y,Z)$

- The Slave's current control coordinates, $S = (P,Q,R)$
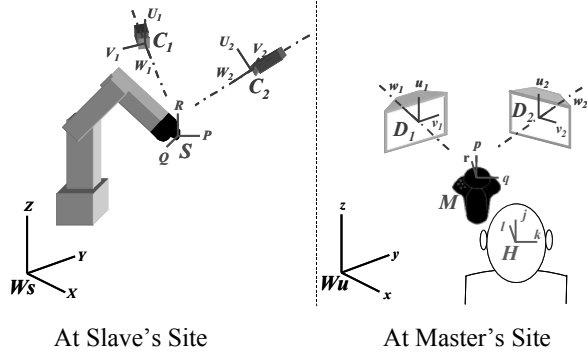
- Camera i's coordinates, $C_i = (U_i,V_i,W_i)$

Figure 1. Coordinate frames. The master shown is a 6-DOF force-sensing Spaceball device.

At the human operator's site, we define

- The operator interface's World coordinates, $Wu = (x,y,z)$
- The Master's current control coordinates, $M = (p,q,r)$
- Display i's coordinates, $D_i = (u_i,v_i,w_i)$
- The Human operator's view coordinates, $H = (j,k,l)$.

Note that all of our frames are right-handed.

In practice, slave frame S may be defined in a variety of ways, such as at the base of the slave, aligned with the tool, or with fixed orientation at the end of the tool, and it may be mobile or fixed. For our purposes, assume that S is located at the end-effector of the robot and translates with it. S is a control frame in that its axes correspond to the master's axes. That is, an inputted force/motion along M's p results in a motion of the robot along S's P, an inputted rotation about M's q causes the robot to rotate about S's Q, and so forth. Rotating or moving S does not affect M, and vice versa.

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 line represents the angle a camera/display is pointing. For our discussion, we represent the ith camera and display as three-dimensional coordinate frames C_i and D_i, respectively, with the frames' first two axes defining the planes and the third representing the centerlines. This places the origins of the frames at the intersection of their corresponding plane and centerline. Since C_i's positive W-axis points "forward" from the camera, D_i is defined so that positive w points "into" the image. Similar to S, C_i may be fixed or moving depending on whether a camera has tilt/pan/zoom abilities.

Note that the master frame M is the device's control coordinates and not necessarily its physical frame (e.g. in Fig. 1, the master may be physically oriented differently but have the same control frame M). Recall that the axes of M correspond to axes of slave frame S.

As for the human frame H, it is located not at the head, but rather at the eyes of the operator. The l axis should be considered the axis of where the eyes are looking; the point of

focus lies on this line. Axes j and k are in the peripheral directions and H moves with the eyes and/or head when the human changes where he is looking.

Using Special Euclidian notation (SE(3)), let the transformation from world frame to the slave control be

$$\begin{bmatrix} P \\ Q \\ R \\ 1 \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} & {}^{S}_{Ws}R & \end{pmatrix} & \begin{pmatrix} {}^{S}_{W}t \end{pmatrix} \\ 0\ \ 0\ \ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = {}^{S}_{Ws}T \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \qquad (1)$$

where ${}^{S}_{Ws}R$ and ${}^{S}_{W}t$ are the rotation and translation matrices from Ws to S. In the same fashion, define transformations ${}^{Ci}_{Ws}T$, ${}^{M}_{Wu}T$, ${}^{Di}_{Wu}T$, and ${}^{H}_{Wu}T$ as being from world to camera i, world to master, world to display i, and world to human, respectively. In most teleoperation situations, these matrices can be easily measured or calculated, and from these the transformations from component to component (e.g. from master to display i) can be calculated. For more information regarding rotational and translational matrices, see [4].

The transformation from camera i to display i is more difficult since C_i and D_i are physically two-dimensional finite planes with perpendicular centerlines. We will assume that the camera transformations are linear (an assumption that can be significantly violated by the distortions of wide-angle cameras [5]). Points in C_i are projected onto the camera's bounded plane (the image plane), and shown on D_i's bounded plane (the display's screen), frequently at a different scaling. This projection, shown in Fig. 2 where f_i is the camera's focal length, is essentially an image-depth-dependent scaling. That is, by similar triangles, a point at (U_i,V_i,W_i) is projected to

$$(U_i,V_i,W_i) \rightarrow \frac{f_i}{W_i + f_i} \cdot (U_i,V_i,0) \qquad (2)$$

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 C_i to D_i can be interpreted as a projection scaling, an image plane cropping, a scaling for the display's resolution, pixel size, and aspect ratio, finally followed by a cropping to fit the screen.
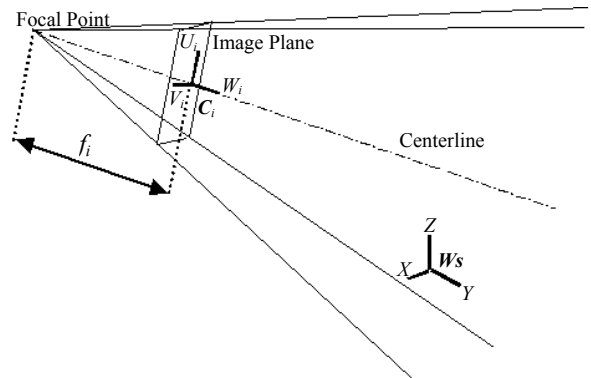

Figure 2. Projection of world points to camera i's image plane.

III. Reducing Mental Transformations: Single Display

We now consider ways to reduce mental workload by improving the components' arrangement. But what does mental workload entail? Sanders and McCormick define mental workload as "a measurable quantity of the information processing demands placed on an individual by a task" [6] and it can be measured in a variety of ways [7]. Mental workload is the cognitive effort required by the operator to achieve a goal. In teleoperation, completing a task requires mentally transforming desired outputs to needed inputs, so simplifying these relationships will reduce the mental effort expended. To see how, we begin with the case of a setup with one camera and display and drop the i subscript from our camera and display frames for convenience.

A. Control Translation

Ideally, the slave-to-camera spatial relationship would match the master-to-display spatial relationship. In the general case, however, the operator must perform the mental transformation to compensate any difference between slave-to-camera and master-to-display. This mental transformation, which we will call "control transformation", can include translation, rotation, and scaling.

We will consider first the translational part of the control transformation. Suppose the master and the view of the slave on the display are those in Fig. 3, where the master and slave frames are oriented the same, but the master is translated to the side and below. The operator need not mentally rotate control motions (if not rotating his head to look at the master when moving it) but he must mentally translate them from the slave to the master, i.e. to his hand. For example, if the desired motion of the slave is up-right-back, then the required motion of the master is a parallel line but translated to the master.

There is some research in the literature that is related to mental translations. Kosslyn [8] discusses results from several image scaling and scanning experiments showing that the greater the scaling or scanning, the longer the task time. Andersen [9] found that switching focus from one object on a mental map to another took slightly longer the farther away the objects were. Murray [10] tested operators' mental modeling abilities with three video images arranged on a screen either in a line or at map locations corresponding to where their cameras physically were. He found that operators performed slightly better with map locations, though not significantly. However, control translation is a transformation from screen to hand and not along a mental map, although research by Wexler [11] shows that there is a correspondence between mental and motor/physical rotation. Furthermore, there is plenty of anecdotal evidence that humans do not have much difficulty mentally translating motions to our hands. Examples include using a PC mouse or playing video/computer games. Mathematically, this means that having

$$ {}_S^C t \neq {}_M^D t \qquad (3) $$

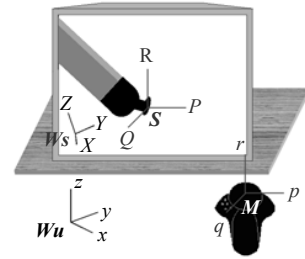adds mental translations of motions but is not very costly.

Figure 3.   Setup with the master translated from the perceived slave.

In fact, teleoperation almost always requires control translation along the display's depth axis. A setup that does not have this one-dimensional mental translation must have the master at the same location as the perceived slave. But at what distance along the display's depth is the perceived slave? The distance is not physical and depends on the operator's mental model of the slave and environment's size. That is, the operator may perceive the slave as far away if the image on the display is smaller than his mental model, or vice versa. Note that this translation is very similar to a scaling: as the perceived slave is moved along the display's depth axis, the scaling between master and perceived slave motions is changed. Therefore, this discussion on control translation is also a discussion on control scaling.

The small cost associated with control translations is also of interest because it means zooming a camera is not very harmful. Recall that the projection operation between camera and display is a scaling dependent on depth and focal length (2). Zooming a camera, i.e. changing the focal length, changes this scaling. Points are moved in or out radially from the camera's/display's center; the perceived slave is translated along the display's depth axis and most likely also translated in display's plane. Thus zooming a camera (ignoring any cropping effects) is similar to a three-dimensional translation of the slave, camera, master, or display. If we desired to reduce control translations, then when the camera is zoomed, the master or display must be translated correspondingly to avoid the inequality in (3).

Panning or tilting the camera also adds control translation. Since we defined our camera frame at the camera's image plane, tilting or panning the camera results in an arc motion of that frame. (The camera frame rotates about a center of rotation, i.e. it translates and rotates with respect to world coordinates. If the center of rotation of the camera is at the same location as the camera frame's origin, then the translation component is zero.) Thus, the location of the slave in the video image changes; if we wish to minimize control translation when moving our camera, we must correspondingly translate the master or display.

B. Control Rotation

The relationship that proves the most critical in our setup is the rotation component of control transformation, i.e. control rotation. Suppose the master and view of the slave on the display are instead those shown in Fig. 4 where the image of the slave frame is rotated from the master frame (ignore the axial translation from master to the perceived slave). If the
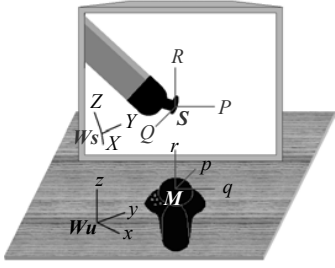
Figure 4. Setup with the master rotated from the perceived slave.

operator wishes to move the slave in a direction, he may need to push or move the master 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. Having a rotation between master and perceived slave means that

$$_D^M R \cdot {}_C^D R \cdot {}_S^C R \neq I \qquad (4)$$

where I is the identity matrix. Recall that the transformation from camera to display is a scaling and cropping (i.e. no rotation), so $_C^D R = I$. Therefore, the issue is that the rotation from slave to camera is different than that from master to display:

$$_S^C R \neq {}_M^D R \ . \qquad (5)$$

Mental rotation is in fact a very taxing transformation, and there has been much research regarding mental rotations of objects. For example, time to mentally rotate a three-dimensional object has been shown to increase with rotation angle, whether linearly [12] or not [13, 14]. While most studies in mental rotation deal with simple objects rotated about a vertical or horizontal axis, control rotation involves rotation of the desired motion often about a more complicated axis. To remove this very taxing mental transformation, $_S^C R$ must equal $_M^D R$.

Restricting the rotation between the master and perceived slave presents a problem when panning or tilting the camera. Recall from our control translation discussion that panning or tilting the camera usually results in an arc motion of the camera frame. To avoid the rotation matrix inequality in (5) when panning or tilting, we must rotate the slave, master, or display correspondingly. Rotating the display is dangerous because it alters the image presented to the operator since it affects the human/display relationship, as we will explore later, and has limited range before the image is physically unviewable. Rotating the slave frame is allowable only if it does not have an important, intuitive control relationship that will be changed. For example, if the slave frame was chosen such that one of the axes is aligned with an arm of the slave, the rotated frame may alter this relationship. Therefore, rotating the master control frame, either computationally or by physically rotating the master, is usually the best choice although it has dangers. Computational rotation will create harmful kinematic rotation if the slave and master are kinematically similar, but is very simple to implement and requires no additional hardware.

Both computational and physical rotation of the master frame will likely lead to erroneous motion if performed while the operator is still pushing or moving the master, though physical rotation gives the operator helpful haptic information of the change. (The rotation of the master frame while the operator is performing motions usually is not a problem, though, since panning or tilting of the camera tends to occur when setting up for a task or when readjusting the camera to get a better view, rather than during actual motion.)

C. View Rotation

So far we have discussed possible control transformations. However, there is one more transformation that deserves attention: that between the human operator and the display, which we will call "view transformation". Recall that we defined the human frame so it is always pointing at the area of interest on the display. Therefore we cannot purely rotate the frame; we must also translate it to satisfy our definition. Nor can we purely translate the origin of the human frame perpendicular to the display's centerline (for example, by moving the operator to the right or by lowering the height of his chair); by definition his frame also rotates. Thus, movement of the human frame involves two motions: rotation arcs around the display frame (translation and rotation with respect to the world) and axial translation along the human frame's view axis, l.

The rotation component of the arc operation is of concern, called view rotation. Suppose the human and display are situated as in Fig. 5, with the human frame translated and rotated from the display's centerline. Since the image on the display is two-dimensional, the operator sees the same information regardless of angle (ignoring monitor viewing angle effects). The new image is not of the slave from a new angle as it would be if the operator rotated around the real slave. Because of the operator's mental model of the monitor and the varying focus depth of the image, he sees this "squashed" image not as a thinner image, but as a two-dimensional plane rotated. To move the slave, the operator now must determine the desired motion from this rotated plane, requiring mental rotation of the image (or operator, though this appears to be harder [15]) so they are perpendicular. That is, the mental model of the planar image is mentally rotated (recall the discussion regarding control rotation) to determine direction. To completely remove this rotational difference, the human's view axis must be perpendicular to the displays plane, *at the area of interest*. This constraint is clearly impractical because, as the operator changes his area of interest, he must
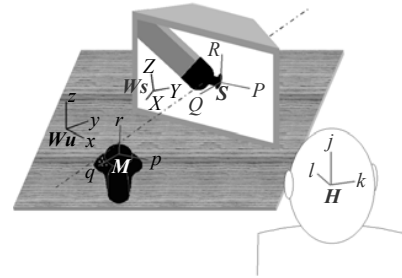

Figure 5. Setup with the human translated and rotated.

translate. Thus, a better constraint is to place the human on the display's axis to minimize the maximum possible view rotation. This constraint can be represented mathematically by the condition that

$$_H^D R = I \tag{6}$$

and

$$_H^D t = \begin{pmatrix} 0 \\ 0 \\ l_0 \end{pmatrix}. \tag{7}$$

For view rotation, using stereovision is more harmful than using monoscopic vision because of "induced stereo movement" [16, 17]. If the display is stereoscopic, the image appears three-dimensional when the operator is looking perpendicular at the screen, with the depth dimension of the image along the human's view axis, l. Imagine that the image on the screen is a hand pointing at the operator. When the operator is moved and rotated to the side, the two stereoscopic images are the same and the depth dimension is still along l. So the hand still appears to be pointing at the operator, rather than in the original direction. With teleoperation, this means that the perceived slave has rotated with respect to the human! Thus, with stereovision, it is critical that the human remains on the display's centerline. Systems like the Da Vinci telesurgical robot that rely heavily on stereovision actually force the human's head to adopt a fixed location and orientation [18].

Translation of the human frame along its view axis, l, on the other hand, costs next to nothing. Except for changing the maximum possible angle between the display's centerline and l, or making the image harder to interpret when it is too far away or too close, translating the frame along l adds no mental difficulty. This also means the size of the display is insignificant as long as the image is viewable by the operator and is not so large that the operator must turn far to see into the corners.

IV. REDUCING MENTAL TRANSFORMATIONS: MULTIPLE DISPLAYS

The results from our single camera/display exploration are easily expanded to setups with multiple cameras and displays. In fact, the significant relationships are more critical now that the operator has more than one source of visual feedback [19]. When using a mentally inefficient setup, the operator must relearn the master / perceived slave relationship each time he switches which display he is looking at. So not only is there a potential for mental transformations, there may also be learning curves occurring with every switch of attention.

A. View Transition is the Key

To apply these lessons to multiple displays, we must be sure each display independently satisfies the constraints mentioned previously. That is, when the operator is using a display, the master must not be rotated from the perceived slave (no control rotations) and the operator should be on the centerline of the display (reduced view rotations). Let us assume that the slave frame is the same for all the displays and is chosen intuitively for the specific teleoperation application and therefore given. Let us also assume the every camera frame is also given, since they are typically chosen by where cameras can be mounted and by which views are the most beneficial for the tasks being performed. So the slave-to-camera relationships are already known but there is flexibility in the master, displays, and human arrangement. How the displays are situated effects what movements of the master, display, and/or human are required when switching views.

B. Examples

The following four examples (also shown in Fig. 6) illustrate the impact of the displays' setup. Each example uses movement of the human and possibly the master, although it is possible to instead move the displays and possibly the master to achieve the same effect. The four examples are:

- Example (a): The displays are lined up side-to-side. When switching views, the operator must translate parallel to this line and the master must rotate (and possibly translate so the operator can still reach it) to match the transformation constraint of the new camera/display.

- Example (b): The displays are instead situated in a tight 3-dimensional cluster, all facing out from the same point, such that one orientation of the master frame simultaneously satisfies all their Control Rotation constraints. Now the operator must rotate on a sphere around them while pointing at the cluster. The master might need to translate so the operator can still reach it.

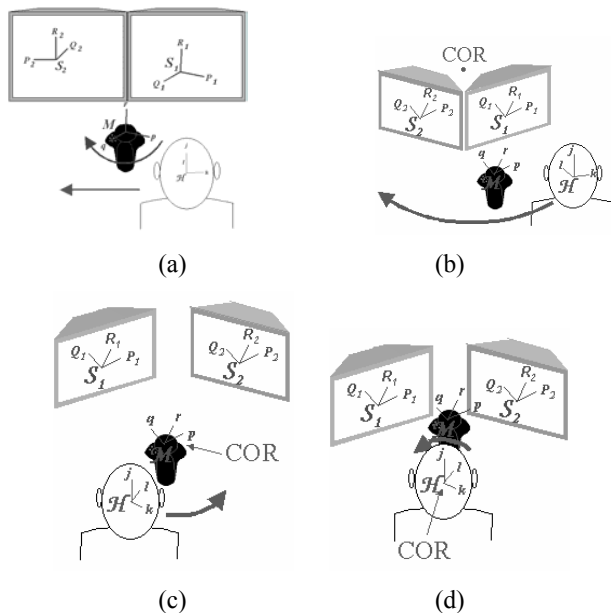(a)   (b)

(c)   (d)

Figure 6.   Motions required to eliminate control rotations and reduce view rotations for different example interfaces. Note that though each interface shows motion of the human and possibly the master, the same result can be achieved by instead moving the displays and possibly the master.

- Example (c): The displays are arranged in a cave of sorts such that they all face the master and their control rotation constraints are all satisfied. Here, the operator must again rotate on a sphere around the master, within that cave and with master between it and the current display.

- Example (d): The displays are again arranged in a cave such that their control rotation constraints are simultaneously satisfied for one orientation of the master. This time their centerlines meet at the origin of the human frame and the master is situated somewhere within reach. To switch view while maintaining control rotation, the human need only turn his head. The master need not translate since it does not get out of reach.

If we want to also restrict the master to be between the operator and the display of interest to reduce control translations, examples (a), (b), and (d) require motion of the master similar to that of the human (or of the display if the operator instead remains stationary). In example (c), the master can remain stationary since its location is always on the displays' centerlines.

V. PRELIMINARY EXPERIMENT

A. Description and Results

A preliminary teleoperation experiment was performed that adds validity to this analysis of mental transformations [20]. The key components consisted of a 6-DOF PUMA robot, a 6-DOF force-sensing Spaceball master, two video cameras, and two CRT monitors. This experiment tested teleoperation via three two-display setups, each with the master fixed and in the same location; see Fig. 7 and Table I for drawings and descriptions. For simplicity, the setups are titled *ALR* (Angled, with views looking Left and Right at robot), *FLR* (Flat, with views looking Left and Right at robot), and *ARL* (Angled, with views looking Right and Left at robot). Twelve subjects were tested on the three setups, evenly spanning the six possible orders of setup presentation. For each subject, the setups were tested with several hours between to minimize effects of presentation order. Subjects were naïve to the goal of the experiment, familiar with robots, although none had previously used a Spaceball extensively.

Table II shows results for this experiment. A single task "step" consisted of moving a pointer on the end of the robot from one of six hanging PVC pipes to another, as prompted by the computer, and as quickly as possible. Control was in world coordinates. For each setup, subjects performed 5 sets of 10 of these goal steps with the first 20 steps used to learn the
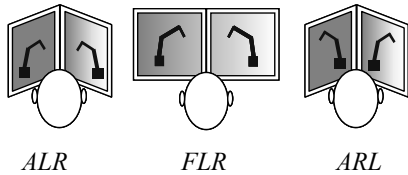

Figure 7. Setups tested in the preliminary experiment.

setup/master. These fifty steps from goal to goal were the same for each setup and for each subject, included every possible step, and were long enough that the subjects did not learn or remember step order from setup to setup. Student t-tests on the results found that the distinction between the setups was significant (all p's <.001).

B. Analysis

These results support some of our transformation conclusions. First, our proposal that control rotations are mentally taxing is supported by the *ARL* times (with control rotations) being slower than *ALR* times (without control rotations). Second, by changing a setup from *ARL* to *FLR*, i.e. by adding slight control translations but subtracting large control rotations, one can greatly decrease task time. Therefore, our view that control rotations are more costly than control translations is also supported. Note that this experiment does not test the effect of view rotation.

The subjects were also asked questions after each setup and after completing the test. Responses from these questions and their t-tests showed

- Subjects found the *ARL* setup, i.e. control rotation, harder to learn, harder when already learned, and more tiring than the *ALR* and *FLR* setups.

- Subjects preferred *ALR* to the others and *FLR* to *ARL*. They found the *ALR* setup the easiest and they thought they did the best with it.

- Almost every subject thought the *ALR* setup was intuitive, most thought the *FLR* setup was, and none thought the *ARL* setup was.

These results further support our transformation conclusions.

It should be noted that the subjects in the preliminary experiment were not taught to move when changing views.

TABLE I.   DESCRIPTIONS OF SETUPS TESTED IN THE PRELIMINARY EXPERIMENT.

Title	Additional Description	Control Rotation	Control Translation	View Rotation
ALR	Displays located so that the master was properly rotated and on the intersection of their centerlines.	None	None	A little
FLR	*ALR* setup except that the displays were "opened" into a plane	A little	A little	A little
ARL	*ALR* setup but with the two displays switched.	A lot	None	A little

TABLE II.   STEP TIME DISTRIBUTION VALUES FOR EACH SETUP (IN SECONDS).

	ALR	*FLR*	*ARL*
Max	92	98	310
Mean	27.5	29.9	43.0
Min	10	9	9
Std. Dev.	12.5	15.1	32.5

It is probable that task times would improve in all of the setups if the following transition rules were enforced:

- In *ALR* - the human arcs left/right so that he remains on the display-of-interest's centerline to reduce view rotation

- In *FLR* - the human slides left/right to reduce view rotation; the master rotates to eliminate control rotation and possibly slides to minimize control translation

- In *ARL* - the human arcs left/right to minimize view rotation; the master rotates much farther than in *FLR* to eliminate control rotation.

VI. CONCLUSIONS AND FUTURE WORK

The framework introduced in Section II is a useful foundation for exploring the mental transformations found in teleoperation control. From the discussion of an interface with a single camera/display, it is clear that mental workload during teleoperation can be decreased (and therefore performance can be increased) by eliminating control rotations, reducing view rotations, and possibly reducing control translations. For interfaces with multiple cameras/displays, eliminating/reducing these transformations may be more critical, as the operator may have to relearn them every time he switches his attention between displays. To improve teleoperation, a setup should minimize control and view rotations and possibly reduce control translations for each display.

The preliminary experiment performed shows that teleoperation setup does affect teleoperation performance and supports several of our conclusions. Future experiments should better isolate the three transformation types and test quantitatively the effects of motion during view-transition. A next step is to assign metrics to the mental transformations and quantify their relative cost both between types and possibly between axes of transformation (e.g. mental rotation about one axis may not be as costly as about another axis).

REFERENCES

[1] Argonne National Laboratory-East, "Dual arm work platform teleoperated systems," U.S. Department of Energy, Rep. DOE/EM-0389, December 1998.

[2] M. W. Noakes, L. J. Love, and P. D. Lloyd, "Telerobotic planning and control for DOE D and D operations," *Proc. Of IEEE Int. Conf. on Robotics and Automation*, Washington, D.C., pp. 3485-3492, 2002.

[3] J. V. Draper and L. M. Blair, "Dual arm work package performance estimates and telerobot task network simulation," *American Nuclear Society Seventh Topical Meeting on Robotics and Remote Handling*, Savannah, Georgia, 1997.

[4] M. T. Mason, *Mechanics of Robotic Manipulation*. Cambridge, MA: MIT Press, 2001.

[5] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 10, pp. 965-980, October 1992.

[6] M. S. Sanders and E. J. McCormick, *Human Factors in Engineering and Design*, 7th ed. New York: McGraw-Hill, 1993.

[7] N. Meshkati, P. A. Hancock, M. Rahimi, and M. S. Dawes, "Techniques in mental workload assessment," in *Evaluation of human work: A practical ergonomics methodology*, J. R. Wilson and E. N. Corlett, Eds. Philadelphia: Taylor & Francis, 1995, pp. 749-782.

[8] S. M. Kosslyn, *Image and Mind*. Cambridge, MA: Harvard University Press, 1980.

[9] J.R. Anderson, "Review: perception-based knowledge representations," in *Cognitive Psychology and Its Implications*, 4th ed. Freeman, 1995, p73-10.

[10] S. A. Murray, "Human performance studies for control of multiple remote systems," U.S. Navy, Rep. NraD TR 1658, July 1994.

[11] M. Wexler, S. M. Kosslyn, and A. Berthoz, "Motor processes in mental rotation," *Cognition*, vol. 68, pp. 77-94, 1998.

[12] R. N. Shepard, and J. Metzler, "Mental rotation of three-dimensional objects," *Science*, vol. 171, pp. 701-703, February 1971.

[13] Z. W. Pylyshyn, "The rate of 'mental rotation' of images: a test of a holistic analogue hypothesis," *Memory and Cognition*, vol. 7, pp. 19-28, 1979.

[14] W. Barfield, J. Sandford, and J. Foley, "The mental rotation and perceived realism of computer-generated three-dimensional images," *International Journal of Man-Machine Studies*, vol. 29, no. 6, pp. 669-684, December 1988.

[15] M. Wraga, S. H. Creem, and D. R. Proffitt, "The influence of spatial reference frames on imagined object- and viewer rotations," *Acta Psychologica*, vol. 102, pp. 247-264, 1999.

[16] W. Tyler, "Induced stereo movement," *Vision Research*, vol. 14, pp. 609-613, 1974.

[17] K. W. Arthur, "3D task performance using head-coupled stereo displays," Masters Thesis, University of British Columbia, July 1993.

[18] G. S. Guthart and J. K. Salisbury, "The intuitive telesurgery system: overview and application," *Proc. of the IEEE Int. Conf. on Robotics and Automation*, San Francisco, CA, pp. 618-622, 2000.

[19] R. Chiruvolu, V. S. Hwang, T. B. Sheridan, "Virtual display aids for teleoperation," *Proc. of SPIE – The International Society for Optical Engineering*, vol. 1612, pp. 299-310, 1992.

[20] B. P. DeJong, "Improving teleoperation interfaces: creation of a testbed and thinking inside the box," Masters Thesis, Northwestern University, May 2003.