A Data-Driven Method for Determining Natural
Human-Robot Motion Mappings in Teleoperation

Rebecca M. Pierce and Katherine J. Kuchenbecker

Abstract—Though many aspects of teleoperation systems have been fine-tuned through research, the mapping between the operator’s movement and the robot’s movement is almost always pre-defined as a simple scaling and offset. We believe that implementing nontraditional data-driven motion mappings has the potential to further improve the usability of teleoperation platforms, making it easier for a human to remotely complete challenging tasks. This paper presents a new paradigm for determining data-driven human-robot motion mappings for teleoperation: the human operator mimics the target robot as it autonomously moves its arm through a variety of trajectories. We analyze the resulting motion data to find the human’s chosen mapping combined with the systematic errors the human made when relying on proprioception to execute these arm movements. We report results from a study in which nine human subjects repeatedly mimicked the planar arm motions of a simulated PR2. We propose three nontraditional motion mappings (similarity, affine, and variable similarity), and we fit each of the proposed models to these data sets, verifying within and across trials and subjects. As hypothesized, the newly proposed mappings are found to outperform the traditional motion mapping model.

I. INTRODUCTION

In the wake of catastrophes such as the Great Hanshin earthquake and the Oklahoma City bombing in the mid-1990s, many researchers turned their efforts to creating machines that can safely explore disaster sites to find and assist survivors under the guidance of a remotely located operator. These new rescue robots were tested in action for the first time in the aftermath of the terrorist attacks on September 11, 2001, assisting with search and rescue at Ground Zero in New York City. Casper and Murphy reported these experiences and recommended a variety of improvements for rescue robotics technology, including significant changes to the human-machine interfaces [1]. These recommendations inspired us to take a fresh look at telerobotic control interfaces, hunting for ways to make them more natural for humans to use, especially in stressful situations.

First, Casper and Murphy state that rescue robots must be transportable and controllable by one person to minimize the number of people at risk during a mission. Furthermore, the danger of carrying large objects across hazardous environments requires all equipment to be transportable via wearable containers. Second, rescue workers usually do not sleep during the first forty-eight hours on the scene, and they sleep for no more than a few hours per day thereafter. Therefore, Casper and Murphy suggest that the control interfaces for rescue workers need to be made as intuitive as possible to account for the lower cognitive capacities that arise under extreme sleep deprivation. These guidelines strongly support investigation of wearable control interfaces that measure natural human motion through sensors such as inertial measurement units, as an alternative to the typical operator control boxes that rely on switches, buttons, and knobs. In addition, there is not yet one robot that can perform all the tasks needed in a search and rescue mission, so it would be beneficial for a single human-machine interface to be able to control several different robotic platforms.

Many attributes of a teleoperation interface affect its usability. One core aspect that has rarely been studied is the way in which the robot’s motion command is calculated from the operator’s arm movement. Though typically done with a simple scaling and offset, we believe that creating data-driven motion mappings from the operator to the target robot has the potential for improving the ease of use of wearable teleoperation control devices. Accordingly, we sought to create a new platform-independent paradigm for determining a natural mapping from the motion of an operator’s arm to the motion of a robotic manipulator.

After discussing relevant background material, we test the hypothesis that nontraditional human-robot motion mappings can be derived by having a subject mimic a robot moving its arm through a trajectory. Systematic differences between the human’s and robot’s trajectories can be modeled and identified from the data. The calculated data-driven mappings are the result of a coupling between the motion mapping chosen by the human and any systematic spatial distortions they made when moving their arm. We use within-trial, across-trial, and across-subject validation on data from nine individuals to demonstrate and test our approach.

II. BACKGROUND

Teleoperation predates nearly all achievements in modern robotics, starting in the early 1950s when Goertz and colleagues created mechanical and electromechanical systems to allow operators to safely handle radioactive materials [2],[3]. In these early teleoperators, the master and the slave were kinematically identical, and the natural arm motions of the
human were reproduced by the remote slave via joint-level mechanical or electrical connections. As the field of robotics began to expand, applications broadened, and teleoperation moved to using computational connections between the two devices. To work with highly complex robotic platforms, operator interfaces and telerobotic control architectures have grown increasingly sophisticated. However, one of the simplest teleoperation architectures, direct position mapping, is still widely used: the slave’s desired position is almost always calculated by applying a scaling and an offset to the measured position of the master device [4].

While this method of calculating desired position has proven successful, it is important to remember that it represents just one of a wide variety of possible motion mappings between the human and the robot. Romano et al. compared the standard position mapping scheme to a rate controller, which maps the master’s position to the slave’s velocity, and to a mouse ballistics-inspired hybrid controller, which nonlinearly maps the master’s velocity to the slave’s velocity [5]. A user study showed that subjects were able to complete a targeting task using teleoperated steerable needles most accurately using the hybrid control law. This work provides evidence that humans may find nontraditional motion mappings to be more intuitive than the standard approach, depending on the needs of the task.

Many other researchers have created operator-adapted controllers to improve the fidelity, transparency, and robustness of teleoperation systems [6]. These researchers often look to mathematical models of human motion to create higher quality control schemes. One of the most commonly implemented models is Flash and Hogan’s minimum jerk criterion, which describes voluntary human arm motion as following the smoothest possible path [7]. For example, Maeda et al. [8] and Corteville et al. [9] successfully used the minimum jerk criterion to predict human motion for improved cooperative object transportation and manipulation. These methods could easily be adapted to teleoperative applications.

The minimum jerk criterion, however, does not describe how humans perceive space around their bodies when making voluntary arm motions. Gordon et al. showed that reaching movements are planned in a hand-centered coordinate frame with direction and magnitude as two independent parameters [10]. Ghez et al. later showed that healthy subjects forced to rely solely on proprioception make systematic directional errors in reaching movements if their hand is laterally displaced from their body centerline [11]. These directional errors were further characterized in [12] and [13] and are reprinted in Fig. 1. With the right arm, motions starting at the right are rotated clockwise, and motions starting at the left counter clockwise. This spatially variable directional error causes straight-forward reaching motions to tend to veer radially out from the body by up to 20° at the edges of the arm’s workspace; these errors are generally believed to stem from miscalibrations in proprioception. We note that an operator relies heavily upon proprioception to teleoperate a remote platform using natural arm motions, since the visual channel is focused on the target platform and the environment seen through a 2D camera view. Taking inspiration from Ghez et al., we hope to improve the naturalness of teleoperation by inverting the type of spatial distortions seen in Fig. 1 when calculating the desired position for the slave robot.

III. EXPERIMENTAL SETUP

We hypothesize that we can discover the distortions that exist in an operator’s spatial perception by having the individual mimic a robot moving its arm through a trajectory. Systematic differences in the trajectories should let us see the motion mapping the human chooses, which is presumably natural and comfortable. We can test this hypothesis by recording synchronized robot and human movements and finding the best mapping from the human’s movement to that of the robot. Willow Garage’s Robot Operating System (ROS) [14] is a natural choice for use in this project, since a major goal of our work is to make extensible algorithms to semi-automatically deduce motion mappings from an operator to a variety of robotic platforms. The algorithms developed in this paper are independent of the method used to capture the human’s arm movement; optical tracking, magnetic tracking, inertial measurement units, and sensors such as the Microsoft Kinect would all work.

We note that in graphics, retargeting recorded human motion to animate virtual characters has become a standard method for creating realistic movements [15]. Techniques from computer animation have also been adapted to animate humanoid robots, e.g., [16], [17], [18]. However, the goal of this body of prior research has been to create human-like robot motion to enhance human-robot interactions, while our work will use retargeting techniques to allow humans to intuitively teleoperate robotic platforms in real time.

A. ROS and the PR2

ROS’s modular, multi-lingual, and open-source packages facilitate the development of algorithms for use on several
different robotic platforms. Willow Garage’s humanoid, the PR2, is one of the best supported robots in ROS and is available in the University of Pennsylvania’s GRASP lab. We recorded the PR2 moving its arm through commanded trajectories using Gazebo, a three-dimensional multi-robot simulation environment supported by ROS [19]. Fig. 2 shows the recorded view of the simulated robot presented to the subject. Equivalent alternatives would have been to record the actual robot moving or to physically locate it with the operator during testing.

This work focuses on identifying transformations between human motion and robot motion for trajectories confined to a horizontal plane, since this is the plane in which Ghez et al. [13] found systematic distortions. Planar robot arm motion was produced using ROS’s real-time joint controller. The shoulder and elbow joints were controlled to follow pre-set trajectories over time, and the other joints were commanded to stay at fixed angles. Fig. 3 shows the trajectories of the PR2’s hand for the eight motions used. The robot’s hand moved with approximately constant speed in motions 1 through 5 and at varied speed in motions 6 and 7. The PR2 was recorded making each motion from six to ten times over approximately 90 seconds. The view point in the movie was overhead looking down, as in Fig. 1.

B. Motion Capture

We recorded human movement using the Vicon motion capture system in the Penn SIG Center. The subject wears a full-body suit covered with 53 passive retroreflective fiducial markers. These markers are individually placed adjacent to all major joints in the body, to provide a stationary reference frame is directed from the left shoulder to the right shoulder, the $X$-axis of this reference frame is directed from the left shoulder to the right shoulder, the $Y$-axis points out from the chest, and the $Z$-axis points up. A visual examination of the captured data showed that there were often very large discrepancies between the human and robot motions at the beginning of each data set, during the time when the human was learning the periodic motion of the robot, as would be expected. Additionally, subjects often stopped mimicking the robot just before the end of each video. Therefore, the first twenty seconds and the last two seconds of each data set were excluded from analysis.

At the end of the study, each subject completed a short questionnaire to explain the strategies and methods they used to accurately reproduce the robot’s motions. Subjects also completed a NASA Task Load Index (NASA-TLX) [20] to rate the difficulty of the task. Subjects also indicated

IV. FOUR MOTION MAPPING MODELS

Once synchronized in time and space, the motion data obtained during the study was analyzed to look for trends in the differences between each subject’s motion and that of the robot. These differences can be attributed to the motion mapping chosen by the subject, as well as unintentional spatial distortions made by the subject while mimicking the
The best fit affine transformation from the human trajectory to the robot trajectory can be solved for in the same manner as the best fit similarity transformation, relaxing the constraint that $T$ be an orthogonal matrix.

D. Variable Similarity

The final and highest dimensional model analyzed in this paper is a position-based warping of space around the human. This model was inspired by the data presented by Ghez et al. in [13]. As depicted in Fig. 1, these prior results show that directional errors strongly depend on position of the subject’s hand. To look for similar trends, we partitioned the human and robot data into half-second segments and found the best fit similarity transformation for each matched pair. The rotation, scale, $X$ offset and $Y$ offset ($\theta, c, \gamma_1$, and $\gamma_2$ from (2) and (3)) were plotted against the $X$ and $Y$ position of the first data point in the time segment. To maintain a reasonable model dimensionality, we found the best-fit plane for each parameter of the similarity fitting, resulting in a model with twelve independent parameters:

$$
\begin{align*}
\gamma(x, y) &= \begin{bmatrix} \gamma_1(x, y) & \gamma_2(x, y) \end{bmatrix} \\
c(x, y) &= a_c x + b_c y + c_c \\
\theta(x, y) &= a_\theta x + b_\theta y + c_\theta \\
\gamma_1(x, y) &= a_{\gamma_1} x + b_{\gamma_1} y + c_{\gamma_1} \\
\gamma_2(x, y) &= a_{\gamma_2} x + b_{\gamma_2} y + c_{\gamma_2}
\end{align*}
$$

While studying pilot data, we also tested a motion mapping scheme that transforms the direction and magnitude of the user’s velocity as a function of the position of the user’s hand. This mapping in the velocity domain creates robot motions that depend on the path of the subject. Thus, certain human motions could cause the mapping in the velocity domain to command desired positions beyond the robot’s workspace. Additionally, only rotation angle and scaling could be fit in the velocity domain, while the warping presented in this paper can also fit the $X$ offset and the $Y$ offset. In the variable similarity transformation, the $X$ offset and $Y$ offset describe both global translation and differential scaling.

E. Summary of Mappings

All of the proposed data-driven motion mappings will be used to find the best fit transformation from the human motion data to the robot data. Applying the fitted mapping to human motion will calculate the desired robot trajectory for each data set. The four motion mappings investigated in this paper are summarized in Table I. This table includes a visualization of how the space around the human will be morphed to command a desired robot pose. The traditional position mapping uniformly scales and offsets human motion to map it to a desired robot position. The similarity mapping uniformly scales, offsets, and rotates the human motion,
while the affine transformation adds differential scaling and shear. Finally, the variable similarity motion mapping smoothly warps, scales, and offsets the human position to calculate desired robot motion.

<table>
<thead>
<tr>
<th>Model</th>
<th>Fitted Parameters</th>
<th>Visual Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional</td>
<td>$S$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td></td>
</tr>
<tr>
<td>Similarity</td>
<td>$c$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\theta$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td></td>
</tr>
<tr>
<td>Affine</td>
<td>$a$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$d$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td></td>
</tr>
<tr>
<td>Variable Similarity</td>
<td>$a_{c1}, b_{c1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_{c1} \gamma_1$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{c2}, b_{c2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_{c2} \gamma_2$</td>
<td></td>
</tr>
</tbody>
</table>

TABLE I
THE FOUR MOTION MAPPINGS CONSIDERED IN THIS WORK.

V. RESULTS

Preliminary analysis of the data showed that all of the motion mapping schemes were capable of transforming human motion to the general shape of the robot motion, but they were unable to fit the global position. For each of the three data-driven models, more than 20% of the error in initial within-trial tests was explained by the offset centers of the transformed human and robot motion. This effect greatly worsened when human motion was transformed using a model trained on different data sets, rising to more than 35% for all four motion mappings. This discrepancy is due to the fact that one’s proprioceptive estimation of arm position drifts significantly over time [23]. However, even with a large drift in proprioception, the direction and extent of motion remain relatively constant [24]. This effect is clearly evident when humans blindly draw repeated shapes: subjects render several nearly identical shapes with offset centers [25], [26]. In the post-study survey, subjects were asked to estimate the percentage of time that they focused the center of their vision on their arms. The mean response to this question was 9.2% with a standard deviation of 13%, indicating that subjects relied heavily upon proprioception to complete the task. Thus, it was not surprising that a similar drift was found in our data set. For this reason, we chose to evaluate all fittings by translating the center of the transformed human motion to the center of the robot motion.

![Fig. 4. Average errors for tests 1 through 4 on the four tested motion mappings.](image)

A. Validation of Models

The quality of each motion mapping model was determined using four tests that differed in their choice of training and validation data. In each test, the parameters of the three data-driven motion mappings were determined by fitting human motion from a training set to the corresponding robot motion. The identified mappings were then used to transform the validation set’s human motion to a predicted robot motion for comparison to the actual robot motion. The scale factor in the traditional fittings was always taken as the ratio of the length of the PR2’s arm to the length of the validation subject’s arm. In the first test, the models were trained and tested on the same data; each trial represented one of the fourteen recordings for one subject. Second, the training data was set to be the first half of each recording, and the second half of the recording was used as the validation set. Third, we performed a standard leave-one-out cross validation test: for every subject, each of the fourteen data sets was used as the validation data for models trained on the combined data of the remaining thirteen sets. Fourth, a standard leave-one-out cross validation test was performed across all subjects; the combined data of each of the nine subjects was used as the validation set for motion mappings trained on the combined data for the other eight subjects. The median of the Cartesian distance from the transformed human position to the robot’s actual position was used as the metric to evaluate the goodness of each fit.

The errors yielded by the four validation tests for each of the four motion mapping schemes are shown in Fig 4. A two-way analysis of variance (ANOVA) was implemented on the average values of the motion mapping errors for each test, using the fixed factor of mapping type and the random factor of subject number. These ANOVAs enable us to determine whether the motion mapping models yielded significantly different errors in each test, taking $\alpha = 0.05$. If model errors were found to differ significantly, a Tukey-Kramer post-hoc multiple comparison test was conducted at a confidence level of $\alpha = 0.05$ to determine which models produced significantly different errors. When the training and validation sets were the same (Test 1), the similarity, affine, and variable similarity transformations produced significantly lower error than the traditional fitting ($F(1,35) = 26.96$, $p_1 < 0.0001$, $\eta^2_1 = 0.2717$), as one would expect from...
When performing longer data captures, it will be important to accurately model a human’s entire workspace. This is likely caused by the fact that they were using proprioception to estimate their arm positions, which was better than the similarity and affine transformation. All three data-driven mappings yielded higher dimensionality of these fittings. When a mapping trained on data from a given subject was tested on previously unseen motion data recorded from the same trial (Test 2) or a different trial (Test 3), the three data-driven mappings yielded similar errors, with the variable similarity performing slightly better than the similarity and affine transformation. All three data-driven mappings were again statistically significant improvements over the traditional fitting ($F_2(3,35) = 17.33$, $p_2 < 0.0001$, $\eta_2^2 = 0.1462$; $F_3(3,35) = 4.76$, $p_3 = 0.0096$, $\eta_3^2 = 0.0168$). The errors yielded by the cross-subject validation (Test 4) did not differ significantly from each other ($F_4(3,35) = 1.58$, $p_4 = 0.2201$). Though not significantly better, the variable similarity mapping is the only one that performs better than the traditional fitting in Test 4.

Figs. 5–8 show how the model trained in each one of the validation methods transforms a sample movement by Subject 6 to that of the robot. These plots make it clear that the variable similarity fitting can better match the features of human motion data to those of the robot motion data. In the same vein, Figs. 9–11 visually displays the distortion of space around each of the nine subjects’ bodies when mapping human motion to robot motion under the data-driven similarity, affine, and variable similarity transformations trained on the combined data of each subject. These figures show that the similarity and affine transformations can be viewed as local approximations to the total spatial warping around the subject’s body. Thus, if the similarity and affine transformations are trained on data recorded when the subject’s hand is in a certain region, the resulting mappings may not be able to map human motion to robot motion when the subject’s hand moves to another region. Since the global position of the subject’s hand drifted over time, which was likely caused by the fact that they were using proprioception to estimate their arm positions, it will be important to accurately model a human’s entire workspace when performing longer data captures.
The fact that the variable similarity fitting performed slightly better than the traditional mapping in the cross-subject leave-one-out validation study means that some parameters of this transformation are consistent across subjects. Though the variable similarity fittings shown in Fig. 11 clearly differ across subjects, there are some striking similarities among all of the identified mappings. Notably, these similarities are in general agreement with the systematic rotational errors described by Ghez et al. [13] (Fig. 1). Fig. 11 shows how the human’s motion would have to be transformed to best match the robot’s motion. Thus, if the findings from our study are consistent with those described by Ghez et al., our transformation will appear to be an inversion of the distortions they described. Looking again at Fig. 11, we see that the angle of rotation error of the subject is consistent with the previously reported results: the magnitude of the direction error grows with lateral displacement and is increasingly counterclockwise to the left and clockwise to the right. Furthermore, the rotational errors made by subjects when the hand was between the body midline and the right shoulder were fairly small. With the exception of points very near to the body for subjects 1 and 3, the lateral location where the directional errors change from clockwise to counterclockwise is within 30 cm of the shoulder at all points over all nine subjects, which is consistent with the findings of Ghez et al. Additionally, much like their described distortions, our variable similarity fittings show a large dependence on the lateral position of the subject’s hand and a much smaller dependence on the extension of the subject’s hand.

VI. CONCLUSION AND FUTURE WORK

This paper is the first work to consider non-traditional data-driven motion mappings for teleoperation with the goal of intuitiveness. We have created a standard paradigm that allows us to determine motion mappings from the recordings of a human mimicking a target robot autonomously moving through a trajectory. Simultaneously measuring the motions of the human and the robot allows us to characterize how a subject systematically distorts space around his or her body while imitating the robot. We believe these models can be used to allow the subject to naturally teleoperate the robot.

Three new motion mapping models were discussed in this paper: similarity, affine, and variable similarity. To validate these models, we trained and validated each model in four tests: the same data set, portions of the same data, data from other trials by the same subject, and finally data recorded from other subjects. In the tests where mappings were trained and validated on motion data from the same subject, the three data-driven motion mappings all yielded significantly lower errors than the traditional motion mapping. The variable similarity fitting was the only motion mapping that yielded a lower error than the traditional mapping for the cross-subject validation study, though the difference was not significant.

The validation tests used for this study were an appropriate starting point, but to see the true effectiveness of each motion mapping model, they will need to be implemented on a teleoperation platform and used by human operators. Therefore, we plan to run a user study in which subjects will mimic the PR2 as it moves through planar motions to allow us to find data-driven motion mappings for each individual, as done in this study. We will then implement these mappings on the PR2 and ask each subject to perform tasks via teleoperation. During this study we will be interested in determining how constant the data-driven motion mappings are across individuals, and whether the new motion mappings affect task performance or cognitive load.

We also plan on extending our data-driven motion mapping models to three dimensions. All of the motion mappings developed in this paper can easily be extended to the three dimensional case. Both the similarity and affine transformations require no additional work to find the best three-dimensional fitting. The variable similarity model can be fit in 3D by finding the best fit linear function for rotation, scaling, X, Y, and Z offset as a function of human X, Y, and Z position. Because Ghez et al. [13] reported spatial distortions only in the horizontal plane, we do not expect significant dependence on the third dimension.

Finally, we are interested in applying similar data-driven
motion mappings to assistive technology. Specifically, we plan on investigating how data-driven motion mappings would allow a person with limited or deteriorated motion, such as a patient suffering from cerebral palsy or post-stroke paresis, to more easily control a target robot.

REFERENCES


