Abstract—We present a framework for parameter and state estimation of personalized human kinematic models from motion capture data. These models can be used to optimize a variety of human–robot collaboration scenarios for the comfort or ergonomics of an individual human collaborator. Our approach offers two main advantages over prior approaches from the literature and commercial software: the kinematic models are estimated for a specific individual without a priori assumptions on limb dimensions or range of motion, and our kinematic formalism explicitly encodes the natural kinematic constraints of the human body. The personalized models are tested in a human–robot collaborative manipulation experiment. We find that human subjects with a restricted range of motion rotate their torso significantly less during bimanual object handoffs if the robot uses a personalized kinematic model to plan the handoff configuration, as compared to previous approaches using generic human kinematic models.

I. INTRODUCTION

Throughout daily life, humans collaborate with one another to manipulate objects in the world. Whether rearranging furniture or transferring tools, each person has a sense of their partner’s preferential limb and body configurations. We prefer to grasp objects: close to the body rather than out of reach; between waist and chest height rather than overhead or underfoot; and within view rather than out of sight. Endowing co–robots with this collaborative manipulation capability remains an obstacle to ubiquitous deployment of autonomous service robots.

One subproblem in this domain which has received attention recently is the selection of “handoff configurations” or “object transfer points” (Figure 1). This problem arises in tasks where a robot must pass an object to or from a human collaborator. Previous approaches construct a cost function over the possible handoff configurations with respect to the human collaborator, then select an optimal configuration with respect to this cost function from the set of all configurations which are feasible for the robot. Previous authors have designed these functions to capture desirable qualities of handoff configurations, such as safety, visibility, and comfort [1], or usability, naturalness, and appropriateness [2].

When computing these functions, many methods in the literature use a generic human kinematic model to capture the comfort or ergonomics of handoff configurations. These models typically use kinematic parameters determined a priori from average morphometrics [1]–[3]. Though appropriate for a segment of the population, it is clear that injury, disability, fatigue, and natural variation in body size and shape can make predictions from any particular kinematic model inapplicable to many individuals. Simple adjustments such as scaling kinematics with height may improve predictions, but a personalized model fit to an individual’s kinematics should give superior performance (at the expense of additional effort invested in model calibration and validation). Indeed, previous authors have noted that customization to individual human partners is needed [2].

Algorithms for fitting a kinematic “skeleton” to an individual using motion capture or vision data exist in the literature, but the resulting models often use only three degree of freedom spherical joints and fail to capture other kinematic constraints such as limb length and joint range of motion. The Kinect algorithm [9] for real-time skeletal pose estimation is a prominent example. The proprietary software included with many motion capture systems typically enforces fixed limb lengths, but does not model joints with limited degrees of freedom or restricted range of motion [10]–[12].

In this paper we present a system for estimating personalized human kinematic models from motion capture data. Our
work makes two main contributions: First, kinematic model parameters such as joint axes, joint ranges of motion, and limb lengths are estimated directly from a specific individual’s training data, without the use of a priori assumptions on joint and limb parameters. Second, our kinematic formalism explicitly encodes the natural kinematic constraints of the human skeleton, such as joints with limited degrees of freedom and range of motion, given minimal training data. We exploit a parsimonious twist representation [13] as in [14] to obtain a minimal intrinsic parameterization for joints. This work complements the existing literature, as the models it generates can be directly incorporated into frameworks for object handoff planning [1]–[3] and more general human modeling [15]–[17]. Specifically, our personalized kinematic models are a drop-in replacement for the generic human kinematic models used in [1]–[3] and our learned model parameters can be used to rescale and calibrate the detailed musculoskeletal models in [15]–[17] to a specific individual.

We expect that adapting robot behavior using personalized models will confer advantages including safer, more ergonomic interaction with humans of varying physical dimensions and more effective collaboration with humans whose capabilities are restricted by injury or disability. To demonstrate the utility of the proposed framework, we compared three schemes for generating bimanual object handoff locations from a robot (Baxter Research Robot, Rethink Robotics) to a human partner in a motion capture arena. The three handoff schemes differed in the data available at the moment of object transfer:

1) Constant: A handoff pose constant relative to the robot body frame. This was selected as a naive approach to serve as a control.

2) Relative: A handoff pose constant relative to the human torso frame. This scheme represents configurations computed from the generic human kinematic models in [1]–[3], as discussed in Section II-E.

3) Personal: A preferred handoff configuration predicted using a personalized human kinematic model. This is the method developed in this paper, and is discussed in Section II-D.

To evaluate these approaches, we compared the rotation (Figure 1) in the human’s trunk at the moment of object handoff, since this statistic correlates with lower back injury [8, Table 1].

Each subject performed a randomized sequence of handoff experiments with the handoff location generated using each scheme. In addition, each subject repeated the process with two treatments: first unencumbered and subsequently with the dominant arm restricted by a strap. Restricting the dominant arm with a strap was intended to simulate loss of range of motion due to an acute injury, for instance if the limb is physically encumbered by a cast or sling. We expected to observe significantly more trunk rotation when the robot only had access to its own reference frame (the constant scheme), particularly for subjects whose morphology differed from the generic model used to choose the fixed handoff location. Furthermore, we expected that adjusting the handoff location to account for the human’s trunk position and orientation (the relative scheme) would nevertheless result in significant rotation when the dominant arm was restricted.

II. METHODS
A. Kinematic Model
Following conventions set by the International Society of Biomechanics [15], [16], we represent the topology of a human using a rooted tree composed of up to five kinematic chains. The tree is represented by \((J, E)\) where \(J\) is a set of joints and \(E \subset J \times J\) is a set of edges representing rigid links. Each \(j \in J\) has a single degree of freedom (DOF) specified by a twist \(\xi_j \in \text{se}(3)\). Thus displacing joint \(j\) by an amount \(\theta_j \in \mathbb{R}\) yields a rigid body transformation

\[
\exp \left( \xi_j \theta_j \right) \in \text{SE}(3)
\]

between the input and the output of \(j\), where the \(\exp\) operator maps a twist vector \(\xi = [\omega|v]^T \in \mathbb{R}^6\) to its equivalent homogeneous representation:

\[
\hat{\xi} = \begin{bmatrix} \hat{\omega} & v \end{bmatrix}, \quad \hat{\omega} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}
\]

Similarly, each edge \((i, j) \in E\) of the tree maps a twist vector \(\xi = [\omega|v]^T \in \mathbb{R}^6\) to its equivalent homogeneous representation:

\[
g_j(\xi, \theta) = \prod_{i \in (j)} \exp \left( \hat{\xi}_i \theta_i \right),
\]

where \(c(j)\) is the unique, ordered sequence of joints connecting \(r\) to \(j\). The world frame position of a feature \(p_i \in \mathbb{R}^3\) (e.g. a motion capture marker) rigidly affixed to the output of joint \(j\) is then given by

\[
g_j(\xi, \theta) \begin{bmatrix} p_i \\ 1 \end{bmatrix}.
\]

This compact representation can model revolute (rotational) in addition to prismatic (linear displacement) joints.

The twist formalism for kinematics has two main advantages for this application: First, its lack of singularities makes the parameter estimation cost function \(J(\xi, \theta, p)\) (in Section II-B) smooth with respect to the joint parameters \(\xi\). Second, with only six free parameters per joint, the twist parameterization is minimal, which minimizes the amount of training data necessary for the identification algorithm to achieve a specified accuracy.

To complete our model, we define the map \(\alpha : \{1, \ldots, m\} \to J\), which specifies which joint’s output each feature is rigidly attached to. We refer to the collection \(S = (J, E, \alpha, \xi, p)\) of a tree \((J, E)\) with a feature-to-joint mapping \(\alpha\), twists \(\xi \in \text{se}(3)^n\), and features \(p \in \mathbb{R}^{3 \times m}\) as a kinematic skeleton.
B. Parameter Identification

To model human motion using the kinematic skeleton developed in Section II-A, we assume the tree structure \((J, E)\) and feature-to-joint mapping \(\alpha\) are known but the twists \(\xi \in \text{SE}(3)^n\) and feature locations \(p \in \mathbb{R}^{3 \times m}\) are unknown. It is difficult to directly measure \(\xi\) and \(p\), therefore we estimate these quantities from a training dataset \(\eta_k \in \mathbb{R}^{3 \times m}\). This dataset consists of \(N\) (noisy) observations of the coordinates of \(m\) features in the world frame (provided, in our case, by a motion capture system). These are collected while the subject performs some sequence of training motions. For the upper body model used in our experiments, this sequence consisted of moving the shoulder and elbow joints through their full ranges of motion, as shown in Figure 2.

Given this training dataset, the skeleton parameters are estimated as in [14] using nonlinear least-squares prediction error minimization [18] on a collection of error vectors with the form

\[
\varepsilon(\xi, \theta_k, p_i) = g_{\alpha(i)}(\xi, \theta_k) \begin{bmatrix} p_i \\ 1 \end{bmatrix} - \begin{bmatrix} \eta_k \end{bmatrix}.
\]

(5)

Note that in addition to the skeletal parameters \(\xi\) and \(p\), the joint displacements \(\theta\) must also be estimated for each frame in the training dataset to completely specify the prediction error. Thus, the error function which is minimized is

\[
J(\xi, \theta, p) = \sum_{k=0}^{N} \sum_{i=0}^{m} \| \varepsilon(\xi, \theta_k, p_i) \|^2.
\]

(6)

For all joints \(j \in c(\alpha(i))\) that precede \(\alpha(i)\), the derivatives of \(\varepsilon\) with respect to \(\xi_j, \theta_j\), and \(p_i\) are given by (see also (33) in [19])

\[
\begin{align*}
D_{\xi_j} \varepsilon(\xi, \theta, p_i) &= \hat{A}_j p_i, \\
D_{\theta_j} \varepsilon(\xi, \theta, p_i) &= (A \hat{g}_j(\xi, \theta)) \cdot p_i, \\
D_{p_i} \varepsilon(\xi, \theta, p_i) &= R_j(\xi, \theta),
\end{align*}
\]

where the matrix \(\hat{A}_j\) is given in [19, Eqn. 14] and \(R_j\) is the rotational component of \(g_j(\xi, \theta)\). (For joints \(j \notin c(\alpha(i))\) that do not precede \(\alpha(i)\), the derivatives \(D_{\xi_j} \varepsilon, D_{\theta_j} \varepsilon, D_{p_i} \varepsilon = 0\).) Prediction error minimization was performed with the SciPy [20] interface to the \texttt{lmder} routine in MINPACK [21]. Though we have not derived conditions ensuring formal identifiability of the model parameters or asymptotic consistency of the parameter estimates [18], we anecdotal report reliably obtaining good fits using datasets that scale linearly with the number of joints, i.e. \(N \approx C |J|\), with \(C\) as small as \(C = 3\) observations for every joint in the kinematic tree.

C. State Estimation

After estimating the geometric parameters of a kinematic skeleton \(S\) for a specific individual as in Section II-B, we can estimate the state of the model (the joint displacements \(\theta\)) online from a sequence of motion capture feature position measurements. This estimation was performed using an Unscented Kalman Filter (UKF) [22]. The filter is applied to discrete–time stochastic processes of the form

\[
\begin{align*}
x_{k+1} &= f(x_k) + u_k, \quad u_k \sim \mathcal{N}(0, U_k), \\
y_k &= h(x_k) + v_k, \quad v_k \sim \mathcal{N}(0, V_k),
\end{align*}
\]

(7)

where \(f : \mathbb{R}^n \to \mathbb{R}^n\) specifies the deterministic dynamics, \(h : \mathbb{R}^n \to \mathbb{R}^m\) the observation function, and \(u_k, v_k\) are independent and (respectively) identically distributed Gaussian random variables. The initial state distribution is assumed Gaussian and denoted by \(\mathcal{N}(\bar{x}_0, P_0)\).

For the kinematic model of Section II-A the state is given by the generalized joint coordinates \(x = \theta\), the dynamics are driven entirely by the process noise \(u\) (i.e. \(f \equiv \text{id}_{\mathbb{R}^n}\)), while the observation of a feature affixed to the output of joint \(j\) at a position \(p\) is given in (4), and \(h(\theta)\) is therefore obtained by vertically concatenating the three-dimensional position vectors from all features.

The UKF recursively estimates the first two moments of the state \(\mathcal{N}(x_{k+1}, P_{k+1})\) at step \(k + 1\) using the estimate \(\mathcal{N}(x_k, P_k)\) from the previous step and observations \(y_{k+1}\); for details we refer the interested reader to [22]. First, an array of sigma points [22] are assembled. The empirical process and observation covariances and cross-covariances are estimated by propagating these points through the model (7). Finally, the innovation step of the classical Kalman filter [23] is performed using the estimated covariances.
Note that while it would have been possible to repeat the kinematic parameter estimation step (Section II-C) on each subsequent dataset from a given person to obtain minimal prediction error, in practice we found the parameter estimation to be remarkably consistent across a range of specific training datasets and initializations, given the amount and quality of data available in our experiments. Because of this, we simply fixed each test subject’s kinematic parameters after the initial estimation step.

D. Pose Prediction

Given a kinematic tree \((J, E)\) calibrated to a subject as in the preceding section, we now consider the problem of predicting the behavior of the human subject during a collaborative manipulation task. We begin by reviewing the rich scientific literature that aims to address this problem before describing how some of the most popular theories can be incorporated into our framework.

At present, one of the most popular and fruitful theories of motor control is optimal feedback control theory, where it is posited that the central nervous system synthesizes motion by minimizing a cost \(C \in \mathbb{R}\) that varies as a function of the joint angle \(\theta\), torque \(\tau\), or end effector \(x\) trajectory (and, possibly, their derivatives) over a time interval \([0, T] \subset \mathbb{R}\), either in open–loop or through receding–horizon feedback [24], [25]. For trajectory generation, one of the earliest proposed and oft–cited forms for the cost function is minimum jerk [26],

\[
C_x = \int_0^T \| \ddot{x}(t) \|^2 dt. \tag{8}
\]

However, subsequent studies have shown that other statistics such as minimum torque change [27],

\[
C_\tau = \int_0^T \| \dot{\tau}(t) \|^2 dt. \tag{9}
\]

or minimum motor command [25] produce better predictions.

In the present setting, we are more concerned with the final pose of the subject than the trajectory adopted to reach that pose. For static posture prediction, a classical law (alternately attributed to Donder or Listing [28]) posits that attributed to each hand pose there exists a unique preferred limb posture. Though appropriate for some experimental settings, [28] found that this “law” yields poor predictions for limb posture in a reaching task, and demonstrated that minimum work,

\[
C_W = \int_0^T \tau^T(t) \dot{\theta}(t) dt \tag{10}
\]

provides superior predictions.

We conclude that, depending on the collaborative manipulation task under consideration, a cost function with the form given in either (8), (9), or (10) may provide superior predictions of human behavior. Note that our framework is applicable to any cost function that varies smoothly with respect to joint angle \(\theta\), torque \(\tau\), or end effector \(x\) trajectory, including but not limited to (8–10). For the handoff experiments in this paper, we elected to choose the simplest cost function that is consistent with our kinematic skeleton model. Specifically, we select a posture \(\theta^*\) that is merely feasible based on the subject–specific joint limits computed in the calibration step, then minimize the preferred posture cost

\[
C_\theta = \| \theta - \theta^* \| \tag{11}
\]

using the choice \(\theta^* = \frac{1}{2}(\theta_m + \theta_m)\) i.e. halfway between the upper and lower joint limits. We emphasize that the cost (11) was chosen for its simplicity, and that we expect choosing a more biologically–plausible cost will yield strictly superior results relative to those obtained with (11). In particular, any ergonomic benefit conferred by employing the preferred posture cost (11) should be enhanced by instead minimizing jerk (8), torque change (9), or work (10). Note that because the dynamic cost functions (8)-(10) assign a cost to complete trajectories instead of a static posture as in (11), the handoff posture \(\theta_T\) with the lowest cost will, in general, be dependent upon the person’s initial posture \(\theta_0\) before the handoff.

Finally, we note that each of the costs above (8–11) only require a personalized kinematic model for the subject. It is conceptually straightforward to extend the framework in this paper to accommodate personalized dynamic models, but doing so requires conducting a more elaborate set of calibration experiments to estimate inertial parameters [29], [30]. A dynamic model is unnecessary in the present paper; since we focus on demonstrating the value of personalized models using the simplest quasistatic cost (11), the inertial parameters of a dynamic model would have no effect on the pose prediction.

E. Handoff Experiments

We compared three schemes for generating bimanual object handoff locations from a robot (Baxter Research Robot, Rethink Robotics) to a human partner in a motion capture arena. The three handoff schemes differed in the data available at the moment of object transfer:

1) Constant: Generates a handoff pose in a fixed location relative to the robot body frame. This scheme does not make use of any data about the human’s location or kinematic structure. It is a naive scheme which we include as a control.

2) Relative: Generates a handoff pose that is constant relative to the human torso frame. This scheme represents the handoff locations generated using the generic human kinematic models in prior approaches. For example, in [2], [3], the authors use a generic human kinematic model to evaluate one component (denoted \(f_{\text{take}}\)) of their handoff cost function, which is maximized at the object pose with the largest number of possible “take” configurations. Similarly, in [1], the authors optimize a handoff cost function which includes an “arm comfort” term. This term is itself a sum of the squared displacement of the human’s joints from a resting configuration, plus the gravitational potential energy of the arm’s current configuration, and is also evaluated using a generic human kinematic model. Because these approaches use generic human models, they will produce the same optimal handoff configurations with respect to the human’s body frame, regardless of variations in the limb dimensions
or range of motion of a particular human partner. The relative scheme represents these approaches by performing the object handoff at a configuration which is constant relative to a frame attached to the human collaborator’s torso.

3) Personal: Predicts a preferred handoff configuration using a personalized kinematic model. This scheme is our approach. It generates a handoff configuration which is optimal with respect to the chosen ergonomic cost function (Section II-D), for the limb dimensions and range of motion of an individual human collaborator, identified as described in Section II-B.

Each subject performed a randomized sequence of $3 \times 10$ handoff experiments with the handoff location generated 10 times using each scheme. In addition, each subject repeated the process with two treatments: first unencumbered and then with the dominant arm restricted by a strap (Figure 3). Restricting the dominant arm with a strap was intended to simulate loss of range of motion due to acute injury. As an evaluation criteria, we compared the rotation in the human’s trunk at the moment of object handoff. We expected the following outcome:

$H_0$ The constant and relative schemes generate significantly more rotation than the personal scheme with the subject’s arm restricted.

Fig. 3: The strap configuration used to restrict the subject’s arm motion. The resulting range of motion was roughly equivalent to that which would be expected with one’s arm in a sling following an acute injury.

F. Human Subjects Protocol

Each test subject was first outfitted in an upper body motion capture suit. A total of 24 markers were attached to the suit, with four on the subject’s chest, four on the back, and eight distributed along the length of each arm and hand (Figure 5). An active marker motion capture system was used so that each marker was uniquely identifiable in the resulting dataset. Data was collected at 50 Hz.

Before beginning the handoff trials, the test subject performed a calibration sequence to fit a personalized kinematic model as described in Section II-B, and to estimate the subject’s preferred handoff pose as in Section II-D. The kinematic model had two degrees of freedom per arm, for a total of four degrees of freedom. The subject performed three sets of each of the calibration motions shown in Figure 2. The subject was then instructed to stand with their feet at marked locations on the floor and the following test procedure was performed for a total of 30 trials:

(i) robot picks up a single cable from a table while displaying “Please wait” on its head display;
(ii) robot chooses a handoff pose, hands the cable to the subject (Figure 5), and displays “Ready?” on its display;
(iii) after two seconds, robot’s grippers open and arms retract;
(iv) subject takes the cable and plugs it into a piece of network hardware, then waits for the next handoff.

The handoff pose in each trial was randomly chosen as one of constant, relative, or personalized, as defined in Section II-E and illustrated in Figure 4. The handoff poses were chosen such that 10 handoffs of each type were performed in one session. During each trial, motion capture was used to record the pose of the subject’s torso both before the beginning of the trial and at the moment the handoff occurs.

After completing the first set of 30 handoffs with normal arm movement, the test subject’s dominant arm was affixed to their torso with a strap to restrict its movement. This was intended to simulate the range of motion observed with one’s arm in a sling after an injury. The test subject then completed another calibration sequence. This sequence was used to identify a new kinematic model and predict a new handoff pose given the newly restricted range of motion. A new session of 30 trials was then run using the same protocol as before.

III. RESULTS

A. Personalized Kinematic Models

After fitting a kinematic model to a test subject’s calibration sequence, the accuracy with which the model’s rigid kinematic structure captured the subject’s actual motion was evaluated by computing the reprojection error for each motion capture marker in each frame of the calibration sequence. Across all four test subjects and both the restricted and unrestricted motion trials, the median reprojection error for the 16 arm markers ranged from 4.86 cm to 0.29 cm, with
a mean of 1.49 cm. This relatively low error suggests that the
kinematic model identified by the parameter fitting algorithm
accurately captured the kinematic constraints observed in the
subject’s actual motion.

Figure 6 shows the feasible workspaces computed from
the personalized kinematic models of two test subjects.
Note that though the workspaces are qualitatively similar,
there are significant differences between the two subjects. In
particular, the workspaces of the two subjects with restricted
arm motion have only a small area of overlap, indicating that
a single, generic kinematic model would have had difficulty
capturing both subjects’ physical constraints simultaneously.
Even in the unconstrained case, the portion of the reachable
workspace with the person’s arm extended rearward is
significantly larger for Subject 2 than for Subject 1 (Figure 6).

B. Handoff Ergonomics

We applied a pairwise t-test [31] to assess whether the
choice of the constant, relative, or personal handoff schemes
produced a statistically significant change in the torso ro-
tation angles measured at the moment of the handoff. This
analysis was performed both individually for each subject
and on the pooled dataset from all four subjects. The t–
test implicitly assumes that the true distributions of rotation
angles for a given subject are Gaussian and have identical
variance across all three handoff schemes.

For trials with unrestricted arm movement, there was
no statistically significant difference ($p < 0.05$) between
the torso rotation angles measured with the three different
handoff schemes when tested across the pooled dataset
(Figure 7b). However, the analysis on the pooled dataset
for trials with restricted arm movement produced significant
differences in rotation angles between the personal and
constant as well as the personal and relative handoff schemes
(Figure 7d).

This analysis suggests that the choice of handoff scheme
produced a significant difference in torso rotation angles in
subjects with a restricted range of motion.

IV. DISCUSSION

The similarity of the torso rotation angles measured for
all three handoff schemes in the unrestricted case suggests
that handoff planning methods based on generic kinematic
models perform well when interaction partners have “typical”
body dimensions and range of motion (Figures 7a and 7b).
However, the performance of these methods degrades when
subjects’ range of motion is restricted, since this necessitates
significant compensatory motion in the torso to adapt to the
robot’s chosen handoff configurations (Figures 7c and 7d). The use of handoff configurations generated using a personalized kinematic model significantly improved performance by allowing test subjects to maintain a neutral body posture at the moment of object handoff.

We believe the use of personalized kinematic models shows promise not only for the specific case of object handoffs, but also more broadly for human–robot collaboration. Our present implementation utilizes motion capture data to provide a proof-of-concept demonstration of the personalized kinematic model framework. The framework is easily extensible to real-world applications by using inertial measurements and other data streams such as camera data that can be obtained from devices such as the Kinect [32]. For simplicity we employed the personalized kinematic model only to estimate a feasible handoff location; it is straightforward to extend our framework to incorporate other constraints and objective functions [33] to predict human posture and motion...
from energetic [27], [34] or dynamic [30] principles. In future work we hope to investigate the use of these classes of objective functions as well as other task-specific ergonomic metrics such as manipulability, gravity compensation torques, and related concepts. This work complements the existing literature, as the models it generates can be easily incorporated into frameworks for object handoff [1]–[3] and more general human modeling [15]–[17]. More broadly, it provides a foundation for employing personalized kinematic models in human–robot collaboration.

V. ACKNOWLEDGMENTS

We thank Gregorij Kurillo for helpful consultations on experiment design, and Ethan Schaler for assistance with photography.

REFERENCES