2019-04-26

Object Transfer Point Estimation for Prompt Human to Robot Handovers

Heramb Nemlekar
Worcester Polytechnic Institute

OBJECT TRANSFER POINT ESTIMATION FOR PROMPT HUMAN TO ROBOT HANDOVERS

by

Heramb Nemlekar

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Master of Science

in

Robotics Engineering

May 2019

APPROVED:

_____

Professor Zhi Li, Thesis Advisor

_____

Professor Jie Fu, Committee Member

_____

Professor Carlo Pinciroli, Committee Member

Abstract

Handing over objects is the foundation of many human-robot interaction and collaboration tasks. In the scenario where a human is handing over an object to a robot, the human chooses where the object needs to be transferred. The robot needs to accurately predict this point of transfer to reach out proactively, instead of waiting for the final position to be presented. We first conduct a human-to-robot handover motion study to analyze the effect of user height, arm length, position, orientation and robot gaze on the object transfer point. Our study presents new observations on the effect of robot's gaze on the point of object transfer.

Next, we present an efficient method for predicting the Object Transfer Point (OTP), which synthesizes (1) an offline OTP calculated based on human preferences observed in the human-robot motion study with (2) a dynamic OTP predicted based on the observed human motion. Our proposed OTP predictor is implemented on a humanoid nursing robot and experimentally validated in human-robot handover tasks. Compared to using only static or dynamic OTP estimators, it has better accuracy at the earlier phase of handover (up to 45% of the handover motion) and can render fluent handovers with a reach-to-grasp response time (about 3.1 secs) closer to natural human receiver's response. In addition, the OTP prediction accuracy is maintained across the robot's visible workspace by utilizing a user-adaptive reference frame.

Acknowledgements

I want to sincerely thank my advisor Prof. Jane Li for giving me the opportunity to do research with her and guiding me at every step of my master's research and studies. Without her support and unwavering faith in my abilities, probably more than I had, I may not have been able to successfully complete my graduate research. Apart from the research work, I truly appreciate all the assistance and care she has provided to ensure my well-being during and beyond the one and half years of my graduate studies.

My parents have dedicated the best part of their lives towards ensuring that I get the education I want to fulfill my dreams. Even though I have never expressed this to them, I will always be grateful for having such caring and selfless parents. I can never imagine getting through any hurdles in my life without their love and support.

Finally I would like to thank Dharini, John, Himanshu, Vishnu, Abhijeet and every other team member that I have worked with during the course of this research. All of their contributions have been instrumental in the successful completion of my thesis work.

Contents

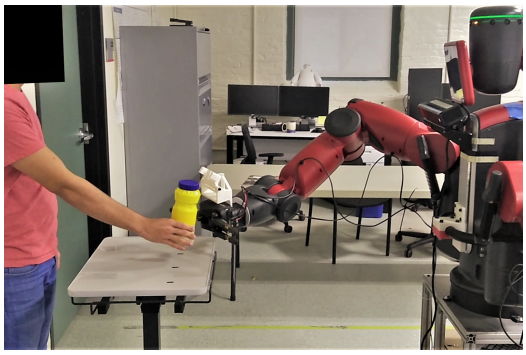
List of Figures

Chapter 1

Introduction

The study of fluent and natural-looking human-robot handovers has been motivated by the need for physical interactions and collaboration between assistive robots and their human partners [1]. For instance, a nursing robot needs to hand over food, beverages, and medicines to patients (see Fig. 1.1(a)), and hand over medical supplies when assisting a human nurse [2]. Planning these motions require consideration of human preferences like legibility, safety, comfort and reachability [3]. A nursing robot may also need to receive bottles (see Fig. 1.1(b)), food trays or clothing handed to them by patients and medical supplies handed by nurses. In this scenario, planning the robot motion not only requires similar consideration of human preferences but also involves prediction of human partner's intent and intended actions. Such handover tasks are frequently performed in many interaction and collaborative tasks, and therefore have a dominant effect on the overall task performance.

(a) Robot to Human Handover



(b) Human to Robot Handover

Figure 1.1: Fluent and natural-looking human-robot object handover is critical to the performance of collaborative tasks.

Ideally, we would like an assistive robot to perform handovers that are at least as good as a human's, if not better, in nursing tasks. Yet, even state-of-the-art robotic assistants are considerably slower than human partners causing them to be inconvenient. This is partly due to hardware limitations that prevent human-like robots from actuating their limbs with the same efficiency and dexterity as a human. Another reason for this gap in performance is the difficulty in prediction of human intent and intended motion. In a human to robot handover scenario, it is not desirable for the human to hold the object for a considerable amount of time and wait for the robot to react. Just like humans anticipate how and where an object could be handed over, we want the robot to predict and proactively respond to receive the object.

Research on human-robot handovers has predominantly focused on planning robot to human [3–11] handovers but limited work has investigated the scenario of handover from a *human giver* to a *robot receiver* [12–15] (see Fig. 1.1(b)). For a human to robot handover, the problem of planning the robot response after the final

position of the object is presented has been addressed in previous work [1, 16]. In this work, we focus on how to predict the object transfer point (OTP) in a handover process and how to render a proactive and adaptive robot reach-to-grasp response based on online OTP prediction. We do not consider the problem of inferring if and when a handover will take place and focus on the scenario where the intent for handover has been communicated, but the point of object transfer is unknown. We define our problem as prediction of the object transfer point to achieve:

- Fast and Accurate reach-to-grasp response
- Fluent, Natural and Legible robot arm motion
- Generalized OTP prediction across workspace

Contributions: We conduct a human-to-robot handover motion study to analyze how observable features like height, arm length, position, orientation and gaze of the human and robot affect the human's choice of object transfer point. Our study presents new observations on the effect of the robot's gaze on the point of object transfer. To render proactive and adaptive robot motions, we propose a unified OTP estimation strategy that combines a pre-computed object transfer point (*static OTP*) which addresses giver position, safety, reachability and height of giver, with a *dynamic OTP* estimate based on real-time handover motion phase estimation.

The parameters that determine the *static OTP* estimation are evaluated in a human-robot handover study (with 20 subjects) described in Chapter 3. To evaluate the proposed OTP estimator's performance, we also measured the receiver's

response time in natural human-human handover. The integrated OTP estimation framework is proposed in Chapter 4. We extended the Probabilistic Movement Primitive (Pro-MP) model and learned the temporal and spatial movement in a relative coordinate frame defined by the human giver's orientation with respect to the robot receiver, such that the learned model for dynamic OTP prediction generalizes across the robot's reachable workspace. Our proposed OTP-estimation strategy is implemented on a humanoid nursing platform (see Chapter 5). Experimental results show that response time is decreased by 19.17% and estimation accuracy at the start of handover is increased by 32.5%.

Chapter 2

Related Work

2.1 Motion Studies on Human-Human Handovers

The **object transfer point** (OTP) in human-robot handover tasks can be approximately predicted by the receiver before a handover motion is initiated. An analysis of handing over objects on a table showed the majority of reaching motions of the receiver to be based on experience and not on the visual feedback of the giver's arm motion [4]. Feedback is only used for grasping when the receiver's hand is close to the object. The study also showed that givers select a direct path to the OTP without deviating from it so that their motion is predictable for the receiver. Similarly, the giver's arm motion in a vertical 2D plane was observed to be pre-planned, feed forward with a fixed maximum velocity regardless of the object pose [5]. However, the velocity and OTP of the handover motion was affected by the weight and affordance of the object. Heavier objects were transferred closer to the giver and

the giver arm velocity was slower while handing over a glass filled with water as compared to an empty glass. The motion of the giver's arm is also independent of the receiver [6], with similar velocity profiles observed for handing over an object to a human and for placing the object on a table at the same distance. Here the distance maintained by the giver from the receiver was found to be independent of their height and arm length, indicating that social proxemics are potentially more important in determining the interpersonal distance than the physical limitations of the collaborative partners. Moreover, the handovers occur halfway between the giver and the receiver. Apart from interpersonal distance, factors like safety, visibility and arm comfort can be considered to postulate the preferred point of object transfer [3]. As the choice of OTP is completely up to the giver's discretion, understanding the factors that impact the giver's decision can help to estimate the OTP even before the handover starts. With a pre-computed OTP, a robot receiver can react as soon as an intent for handover has been detected. However, this OTP estimation is static and does not adapt to variations in human arm motion.

2.2 Methods for Dynamic OTP Estimation

Dynamic OTP estimation requires observing the human partner's behavior in real-time for intent inference and motion prediction. Intent in human-robot interactions is usually inferred from explicit cues like verbal commands, body gestures or gaze [17]. But maintaining explicit or even exaggerated communication requires significant effort from the end user. Instead of inferring the intent from explicit cues,

inferring the implicit intent encoded in human motion can be more efficient and less intrusive. At a high-level, the human intent inference problem can be formulated as inferring the parameters of a dynamic model [18], Bayesian network [19] or Markov decision process [20] and tackled using techniques such as inverse linear-quadratic regulation (ILQR) [21] and approximate expectation-maximization [22].

At a low-level, it hinges on whether a robot can predict its human partner's motion based on the knowledge of tempo-spatial coordination observed in interactive human motion. If the human holds the object at a fixed location, the robot's motion can be planned using random trees to the goal position or by a pseudo inverse Jacobian controller [16]. Even though such sampling-based planners can return a feasible solution, there is no guarantee that the plan will produce a natural and legible motion, unless constraints are applied to confine the random nature of inverse kinematics solutions. It is also inconvenient for the user to hold the object for a longer period and wait for the robot to react.

For early prediction of the object transfer point, the human motion can be modelled as a dynamical system and the point on the human's trajectory closest to the robot, selected as the point of object transfer [23]. But knowing that natural human reaching motions follow minimum-jerk trajectories, the timing and location of the object transfer can also be predicted early after peak velocity of the human partner's hand has been observed [24]. Such methods for dynamic prediction still require significant human motion to be observed and delay the robot's response. To react as soon as the intent for handover is detected, the robot hand velocity can be controlled proportionally to the hand velocity of the human partner [1].

But simply following the human partner's arm motion does not produce a natural-looking handover. The robot should be specifically trained to reproduce human-like handover motions so that they are legible for the human partner.

Ideally, the robot should quickly react to the human partner and improve OTP prediction accuracy as more of the human partner's motion is observed. Dynamic Movement Primitives (DMP) can reproduce trained trajectories to new goal locations through a combination of attractors and forcing components. [12] defines the goal of a DMP formulation to be the human's hand and uses a sigmoid weighting function to reduce the impact of the goal attractor element in the early stages of the handover. This method can lead to an initial un-natural behaviour if the human hand is farther from the training pose at the start. Triadic interaction meshes can be used to model the entire handover including the giver, receiver and the object from a single demonstration and generate the motion constraints offline [13]. This method of estimation takes 9.7 *secs* for a handover including the retraction of robot's hand but has a generalization capacity of ± 37 *cm*. Another technique [14] uses a library of human motions to obtain over 70% accuracy of time series classification after observing just one-third of the human's motion during execution. But it requires over 40% motion to be observed for any further improvement in the classification, with close to 100% accuracy requiring nearly all motion to be observed.

To address early OTP estimation for faster handovers, Maeda *et al.* proposed probabilistic models for learning and reproducing the phase matching between human and robot hands [15]. Superior to the minimal jerk model, the phase estimation

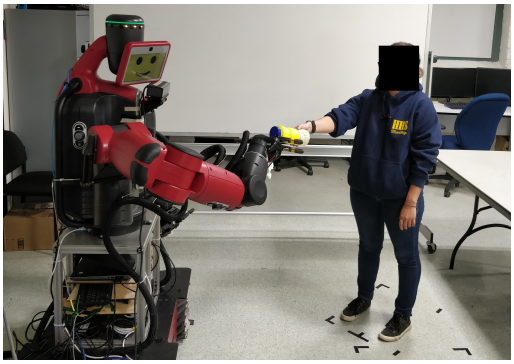
model can reliably predict the object transfer point after observing less than 45% of the human's hand motion. The phase estimation approach trains a Probabilistic Movement Primitives (Pro-MP) model to map the human's arm motion to the robot's joint action based on the phase of the handover. Legible motion can be ensured by providing expert demonstrations during the training phase. But this model predicts the handover motion phase based on the absolute hand positions of the human and robot, and therefore will not be valid in cases where the human-robot distance and relative pose are different from the learned demonstrations. We consider this approach to be the *Baseline* for implementing the proposed handover architecture based on our human-robot handover study.

Chapter 3

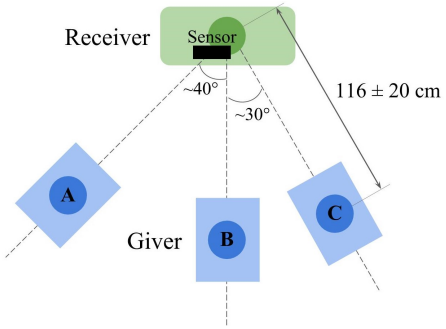
Handover Motion Studies

Motion studies for human-robot handover have analyzed human-human object handovers to determine how a robot should offer an object to a human [4–6]. For a handover from a human *giver* to robot *receiver*, it is not clear where and how the object will be handed over if the giver is allowed to hand the object from any direction in the receiver's reachable workspace. Here we conduct a human-robot handover study to analyze the effect of relative orientation, height, arm length and gaze on the point of object transfer.

Shown in Fig. 3.1(a), a robot *receiver* stands at a fixed location and orientation, while a human *giver* stands at one of the bounding boxes in the A, B, and C directions (referred as Positions A, B, and C). Shown in Fig. 3.1(b), the bounding box is defined such that distances between robot and human subjects are 116 ± 20 cm from the *receiver*, according to the social space in proxemics defined in [6]. Position B faces directly to robot *receiver*, while Positions A and C are chosen to be the



(a) Subject performing handover



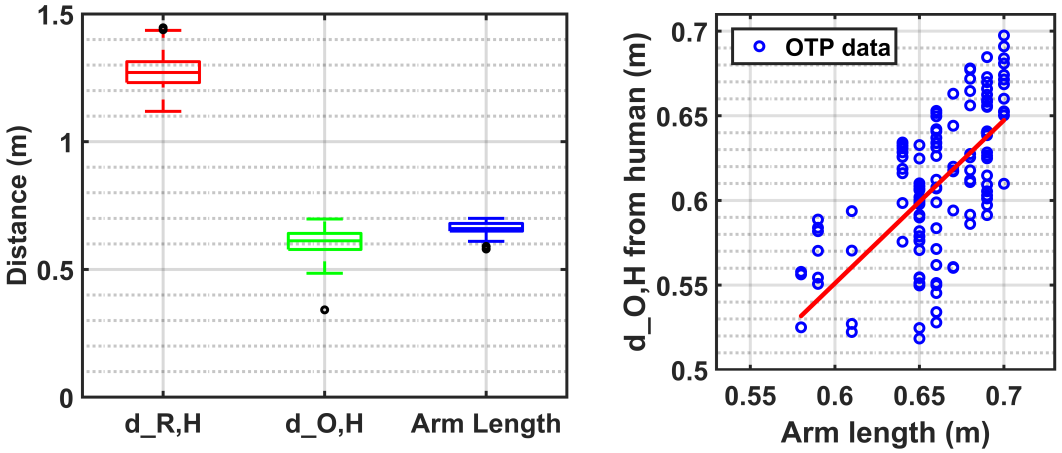
(b) Layout of giver positions

Figure 3.1: Experimental setup for human to robot handover study

boundary of the robot's motion tracking camera. Twenty subjects participated in the experiment, each performing six handovers at each position: three handovers with the robot looking at the subject, and another three with the robot looking away from the human giver (i.e. total 360 handovers). In each trial, a subject presented a bottle to the robot. As soon as the subject started to reach out, the robot responded with a pre-programmed reaching action towards the *natural reachable region* of the *giver*'s arm, which was measured in a pilot study with five subjects. In the pilot study, the experimenter kinesthetically moved the robot arm towards a human giver that reached out to hand over an object. The subjects were asked to hold the object at their preferred object transfer point until the robot's reaching motion was complete. The natural reachable regions corresponding to each position the *giver* stood at were measured as the average position that a human giver preferred to transfer the objects.

3.1 OTP in Human-Robot Handover

We evaluated the parameters that determine the OTP based on the data collected in our human-robot handover experiment. Let the distance between the human and robot be $d_{R,H}$ and between the OTP and the human giver be $d_{O,H}$.



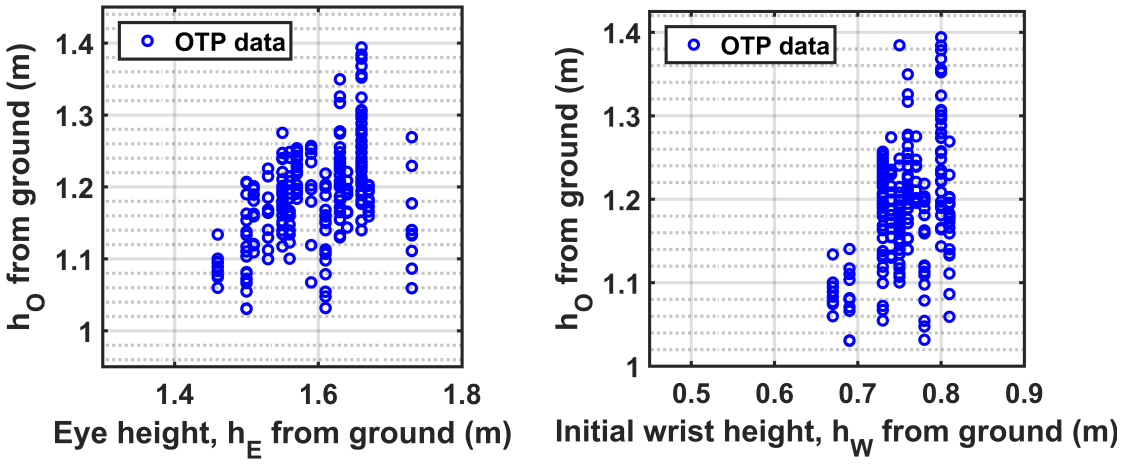
(a) Comparison of $d_{O,H}$ to $d_{R,H}$

(b) Arm length to $d_{O,H}$ relation

Figure 3.2: Effect of interpersonal distance and arm length on OTP

As seen in Fig. 3.2(a), the average $d_{O,H}$ is close to to half of average $d_{R,H}$, differing by just 2.91 cm. We also see that the average $d_{O,H}$ is only 7 mm less than the average arm length (l_{arm}). Fig. 3.2(b) shows a positive relationship between arm length and distance of OTP from *giver*. The trend line (red line in Fig. 3.2) for the increase in $d_{O,H}$ with increase in l_{arm} has a slope of 0.96. This indicates that the average behaviour of users was to present the object at the extent of their reachability and as close as possible towards the mid-point of the $d_{R,H}$.

We further evaluated how the height of human givers affects the height of the OTP, h_O (see Fig. 3.3). Let h_E and h_{W_i} be the heights of the subject's eyes and



(a) Change in h_O with h_E

(b) Change in h_O with h_{Wi}

Figure 3.3: Effect of height of giver and initial wrist position on OTP

wrists from the ground in their initial position. A multiple regression model trained with h_E and h_{W_i} as independent variables to predict h_O has an accuracy of 41.31%. The regression coefficients for h_E and h_{W_i} were 0.143 and 0.119, respectively. Thus, increase in h_E or h_{W_i} leads to small increase in h_O. These predictors had p-values of 0.003 and 0.01, respectively for predicting the h_O. Although the range of data is small, h_E and h_O have a positive relationship.

3.2 Effect of Gaze and Receiver Orientation

Moreover, we studied how the robot receiver's gaze direction affects the static OTP. We instructed human subjects to stand at Positions A and C. For each position, the robot receiver gazed directly at the human giver in three handover trials. For other handover trials, the robot gazed to the opposite side (e.g., looked in the direction of Position C if the subject stood at Position A). Human-human handover studies

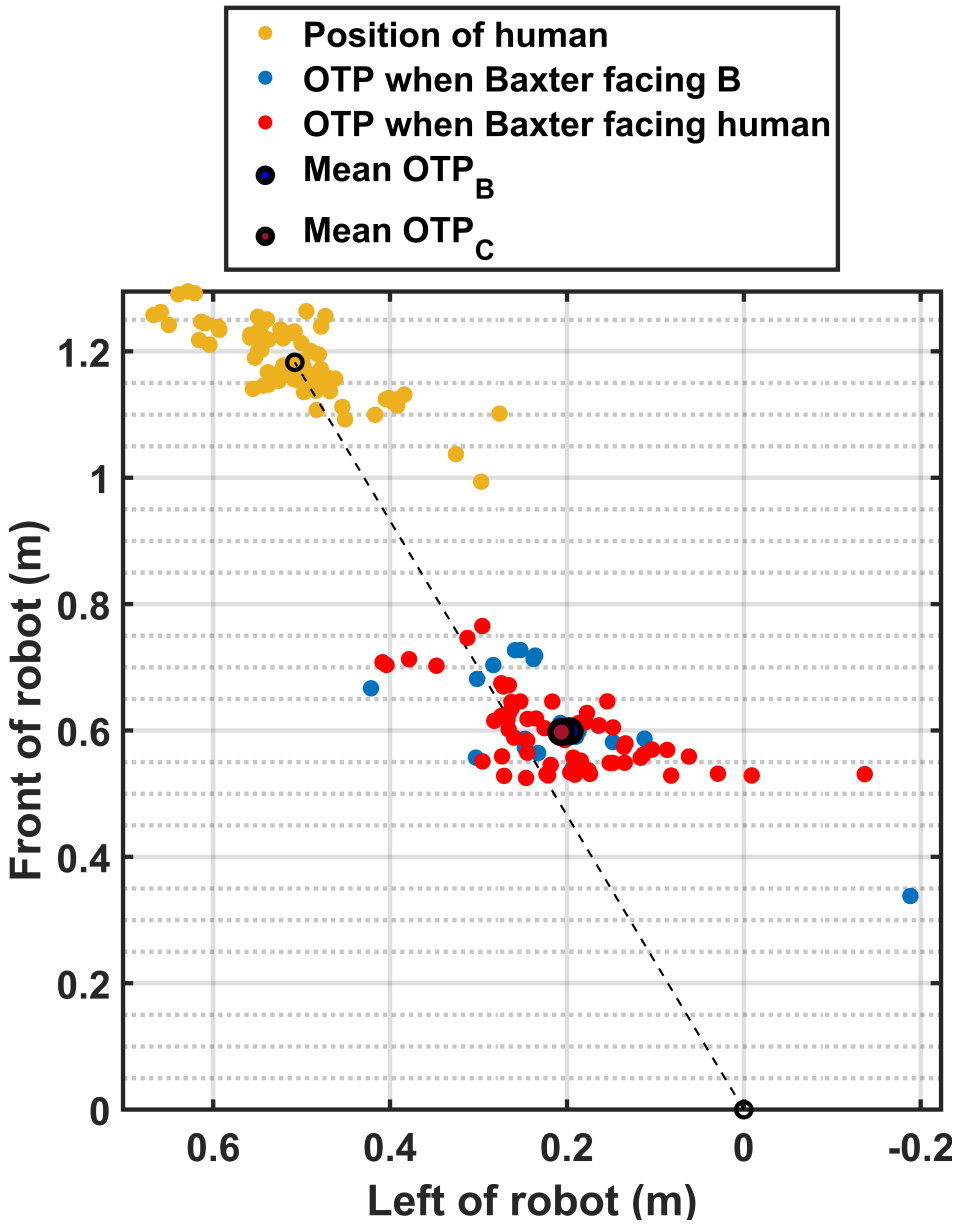


Figure 3.4: The effect of robot gaze on the OTP chosen by human givers: Human standing at position C

in [25, 26] pointed out that gazing at the partner's face and the handover location helps to communicate the handover intent. Our experiment showed that if the

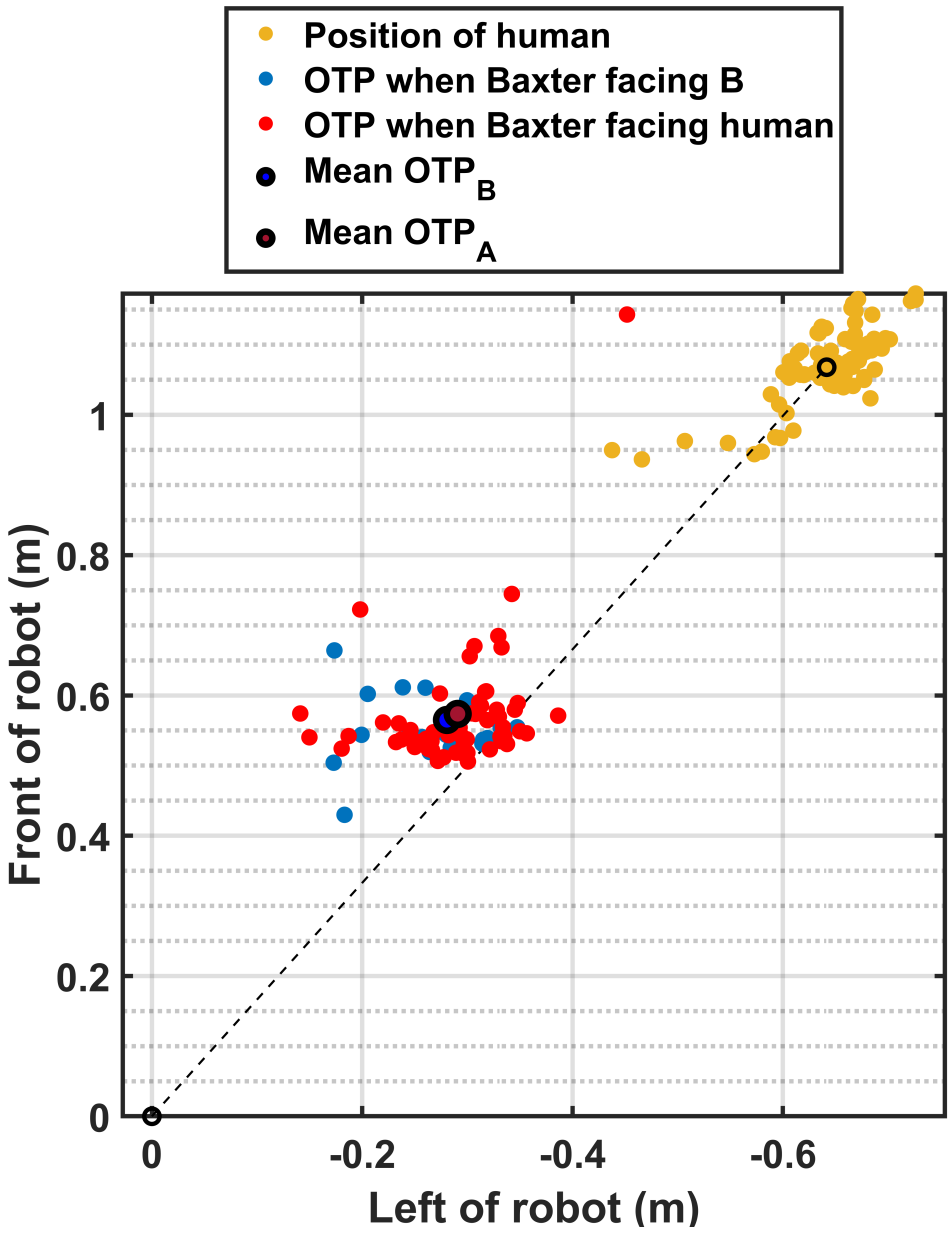


Figure 3.5: The effect of robot gaze on the OTP chosen by human givers: Human standing at position A

robot looks at the human directly, the OTP will be along the line connecting the positions of the human and the robot. However, as shown in Fig. 3.4 and Fig. 3.5,

even when the robot diverted its gaze, the average OTP position remained very close to the line connecting the human and the robot, and only shifted slightly towards the gaze direction of the robot. In previous work [26], the handovers occurred from the robot to the human. As the robot was the *giver* and needed to take the initiative of handing over an object, it was important for the robot to use its gaze to communicate its intent to the human *receiver*. In human to robot handovers, the human is the *giver* and may use his gaze to communicate the handover intent. The robot being the *receiver* does not need to communicate its intent unless explicitly required and has to simply respond to the *giver's* handover motion. Thus the *receiver's* gaze may not be an important factor that affects the *giver's* choice of OTP. This might be because the human subjects did not associate robot gaze direction closely with the direction it can sense, as long as the robot responded to the human's initial choice of OTP. Hence even when the robot's gaze was directed away from the human givers, they still chose the OTP along the plane connecting their positions in the workspace.

3.3 Natural Response Time of Human Receiver

We measured the *receiver*'s response time in human-human handover, to set up the evaluation standard for the robot receiver's response. Two human subjects performed 30 handovers (see Fig. 3.6), each taking turns to be the *giver* while the other was the *receiver*. Markers were placed on the wrists, shoulders, head, and torso of the subjects and tracked using a Vicon Motion Capture system. The **reaction**

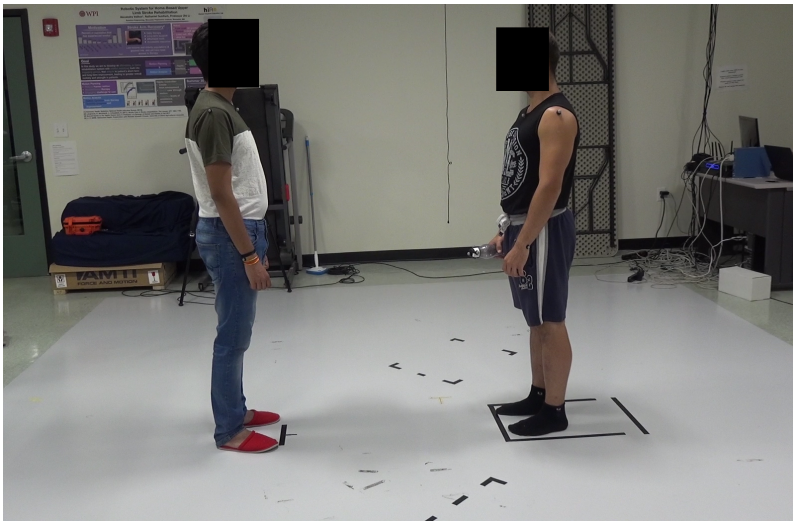


Figure 3.6: Human-human handover study to establish ideal human reaction and reach-to-grasp response time in a handover task.

time for a handover was measured from the instant the *giver* started moving their hand, to the instant the *receiver* started their reaching motion. The reaction time was observed to be $\mathbf{0.425 \pm 0.035}$ **secs**, while the observed **response time**, which was the time from the giver starting their motion to the receiver reaching to the object, was $\mathbf{1.212 \pm 0.051}$ **secs**. For an efficient handover response, a robot *receiver* must react and deliver a reach-to-grasp response as fast as a human *giver*.

Chapter 4

Methodology

We propose an OTP estimator that integrates static OTP estimation based on our human-robot handover study, with a dynamic OTP estimator which updates the OTP prediction based on observed human motion. Shown in Fig. 4.1, the **OTP estimator module** takes input from the **sensing module** which observes the object's and human partner's motion in real-time. Within the estimation module, the offline components- *human-robot handover demonstrations* and *user study data*, are responsible for training (1) a Probabilistic Movement Primitives (Pro-MP) model to reproduce legible robot motion, as well as (2) a **static OTP estimator**, respectively before the handover starts. As soon as the human partner starts a handover, the **integrated OTP estimator** takes in the static OTP estimate and updates it with the estimate from the **dynamic OTP estimator** by determining the phase of the human partner's observed motion. The **robot controller** receives the integrated OTP and controls the robot end-effector to reach toward it.

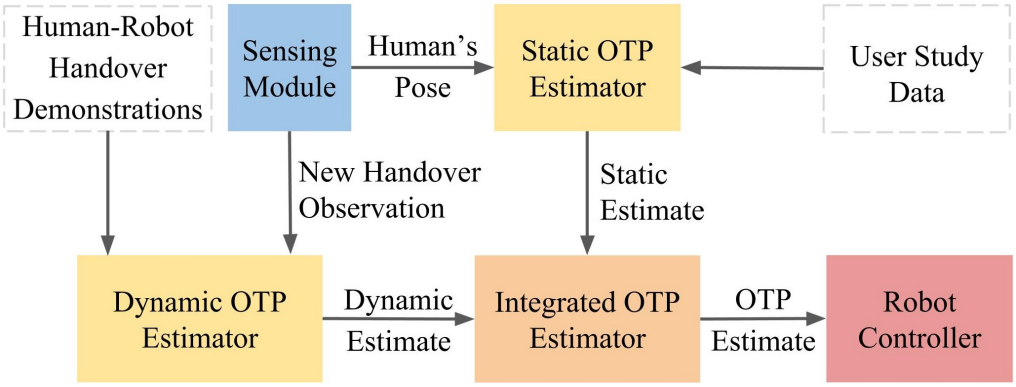


Figure 4.1: OTP Estimation Module: sensing module communicates skeleton data and grasping points to the OTP estimators generating a required trajectory which is executed by the robot controller.

4.1 Sensing Module

The sensing module tracks the human giver's motion for OTP prediction and detects the object to plan a grasping motion during object transfer. The human skeleton data is obtained as Cartesian coordinates of the wrist, elbow, and shoulder joints using the *NI Mate* [27] motion capture system. The robot joint angles were obtained from its internal functions. The object recognition is done by a Convolutional Neural Network trained on the COCO data set [28] and detection of the free region is simplified by using pure coloured objects. Grasp points are determined on the free region based on a representation of the object contour using Elliptic Fourier Descriptors [29] and Curvature Maximization [30]. This process involves calculating the Fourier coefficients of a desired order n (number of harmonics) over a closed contour. A generalized model for the Fourier approximation of a contour can be

shown as:

$$P_x(t) = A_0 + \sum_{n=1}^{k} (a_n cos\frac{2n\pi t}{T} + b_n sin\frac{2n\pi t}{T}) \tag{4.1}$$

$$P_y(t) = C_0 + \sum_{n=1}^{k} (c_n cos\frac{2n\pi t}{T} + d_n sin\frac{2n\pi t}{T}) \tag{4.2}$$

Curvature of the contour is used to select the model grasp points [30]. Taking the first and second derivatives on the model allows us to compute the relative curvature of each (x,y) location, as well as the sign of the function (concave up, concave down). The first derivative of the closed contour will yield us the tangent vectors of the contour:

$$Z(t) = \sum_{n=1}^{k} -a_n \frac{2n\pi}{T} sin\frac{2n\pi t}{T} + b_n \frac{2n\pi}{T} cos\frac{2n\pi t}{T} \tag{4.3}$$

Note that time can be converted to a parameterized range moving clockwise around the x,y contour of the object. The second derivative, i.e. the derivative of the normalized tangent vectors yields the normal vectors:

$$N = \frac{d\frac{Z}{\|Z\|}}{dt} \tag{4.4}$$

To determine the direction of the normal vector, the dot product of the tangent and normal vectors is taken. This is equivalent to finding the sign of the curvature:

$$Curvature = sign(\|Z \cdot N\|) \tag{4.5}$$

Algorithm 1 describes the process to find the grasp point pair residing in optimal curvature regions. The robot gripper is modelled as a pair of friction-less contact points. Grasp points must pass a force closure test determined by the geometry of the Fourier descriptor.

Algorithm 1 Compute Optimal Grasping Pair

Form all possible sets of x,y
For each set $\alpha = Curvature_x + Curvature_y$
Rank sets by descending α
for each set x,y with positive α **do**
    $\beta = PerformForceClosure()$
    if β above threshold **return**
Rank sets by ascending α
for each set x,y with negative α **do**
    $\beta = PerformForceClosure()$
    if β above threshold **return**

procedure PERFORM FORCE CLOSURE(x, y)
    $A = \frac{N_m 1}{\|N_m 1\|} \cdot \frac{Pm_1 - Pm_2}{\|Pm_1 - Pm_2\|}$
    $B = \frac{Nm_2}{\|Nm_2\|} \cdot \frac{Pm_1 - Pm_2}{\|Pm_1 - Pm_2\|}$
    $fc = A^2 + (\pi - B)^2$
    return

4.2 Static OTP Estimator

Before a handover is initiated, the static OTP-estimator computes the initial object transfer point (OTP_s) in the task space based on three criteria: (a) The **Initial Pose** criterion constrains the handover region to be bounded in a 3D space defined by the relative position and orientation of the *giver* and *receiver*. (b) The **Midpoint of Actors** which is the centre of the plane passing through the positions of the *giver*

and the *receiver*. And (c) the **Reachability**, which considers the accessible region based on position, height and arm length of the *giver*.

We represent the initial pose of the *giver* in terms of the relative orientation ($O_{R,H}$) to the *receiver*. The orientation is measured in terms of the angle between the *giver's* current position and the position B as shown in Fig. 3.1(b). The mid-point criteria is determined by the interpersonal distance ($d_{R,H}$) between the *giver* and the *receiver*. And the reachability is measured based on the *giver's* height (h_E) and arm length (l_{arm}) as perceived by the *receiver*. A **tree ensemble** model is trained over these predictors to estimate the 3-D position of the static OTP i.e. $[OTP_x, OTP_y, OTP_z]$. Our trained static OTP estimator has a testing mean squared error of 0.5 cm, given that the users hand over an object in their natural reachable region. In practice, the users may choose to hand over the object at any point within their total reachability. Thus, we model a dynamic OTP estimator to account for the variability in user's choice of OTP.

4.3 Dynamic OTP Estimator

The core of our dynamic estimation method is to train Multi-dimensional Interaction Probabilistic Movement Primitives (Pro-MP) with multiple human-robot handover demonstrations [31]. This method creates a combined probabilistic representation over the motions of both human and robot, thus capturing their interaction. The Pro-MP model classifies the observed human's motion by predicting the phase (timing) of the movement and generates the corresponding robot trajectory

based on the association learned in the human-robot demonstrations. The Interaction Pro-MP model can predict the entire human and robot trajectory even from partial observations in real-time. Hence, we build upon this method to model our Dynamic OTP Estimator for human to robot handovers.

Learning phase: The arm of the nursing robot is by default in an "elbow-up" configuration (see Fig. 1.1(b)). End effector control of the arm to reach the OTP results in an un-natural behavior. Therefore during the learning phase, the arm of the robot is moved by a human teacher to produce a natural reaching motion in response to the human partner's initiation of handover. At each time step t, the seven observed degrees-of-freedom (DOF) of the robot arm, six DOFs of the grasp points on the object and the three observed DOFs of the human partner's hand are concatenated into the following human-robot state vector:

$$\mathbf{y}_t = [y^H_{1,t}, \cdots, y^H_{3,t}, y^O_{1,t}, \cdots, y^O_{6,t}, y^R_{1,t}, \cdots, y^R_{7,t}]^T \tag{4.6}$$

Including grasp points as states in the learning phase makes the model sensitive to the object's grasp configuration and produces accurate reach-to-grasp trajectories. The trajectory of each DOF is further parameterized by weights ($\bar{w}$) such that:

$$p(\mathbf{y}_t|\bar{\mathbf{w}}) = \mathcal{N}(\mathbf{y_t}|\mathbf{H_t}^{\mathcal{T}}\bar{\mathbf{w}}, \mathbf{\Sigma}_\dagger) \tag{4.7}$$

where $\mathbf{H}^T_t = diag((\Psi^T_t)_1, \cdots, (\Psi^T_t)_3, (\Psi^T_t)_1, \cdots, (\Psi^T_t)_6, (\Psi^T_t)_1, \cdots, (\Psi^T_t)_7)$ is the diagonal matrix of the Gaussian basis functions. Among the M handover demonstra-

tions, the i-th demonstration correlates the observed DOFs of human and robot in the handover such that:

$$\bar{\mathbf{w}}_i = [(\mathbf{w}_1^H)^T, \cdots, (\mathbf{w}_3^H)^T, (\mathbf{w}_1^O)^T, \cdots, (\mathbf{w}_6^O)^T, (\mathbf{w}_1^R)^T, \cdots, (\mathbf{w}_7^R)^T]^T \tag{4.8}$$

Parameters of the normal distribution of the weights over all M demonstrations ($w \sim p(w;\theta)$), are used to create the joint probability distribution for each DOF.

$$\mathbf{p}(\mathbf{y_t};\theta) = \int \mathbf{p}(\mathbf{y_t}|\mathbf{w})\mathbf{p}(\mathbf{w};\theta)\mathbf{dw} \tag{4.9}$$

Reproduction phase: Given a new observation of the human's handover motion at time t' (Eqn. 4.10) during execution, the phase of the observation is determined based on correlation of the observed data with sampled trajectories from the training demonstrations.

$$\mathbf{y}_{t'} = [\mathbf{y}_{1,t'}^H, \cdots, \mathbf{y}_{3,t'}^H, \mathbf{y}_{1,t'}^O, \cdots, \mathbf{y}_{6,t'}^O, \mathbf{0}_{1,t'}^R, \cdots, \mathbf{0}_{7,t'}^R]^T \tag{4.10}$$

The joint probability ($p(y_t;\theta)$ is conditioned to get the new weight distribution $\theta_{new} = \{\mu_w^{new}, \Sigma_w^{new}\}$. The robot and human handover trajectories are generated by substituting the weights $\mathbf{w}$ conditioned on the observed human motion in the basis function model (Eqn. 4.7).

Generalization across workspace In [15], the demonstrations for training the Pro-MPs are recorded in the robot's body frame (F_R) or the world frame (F_W)

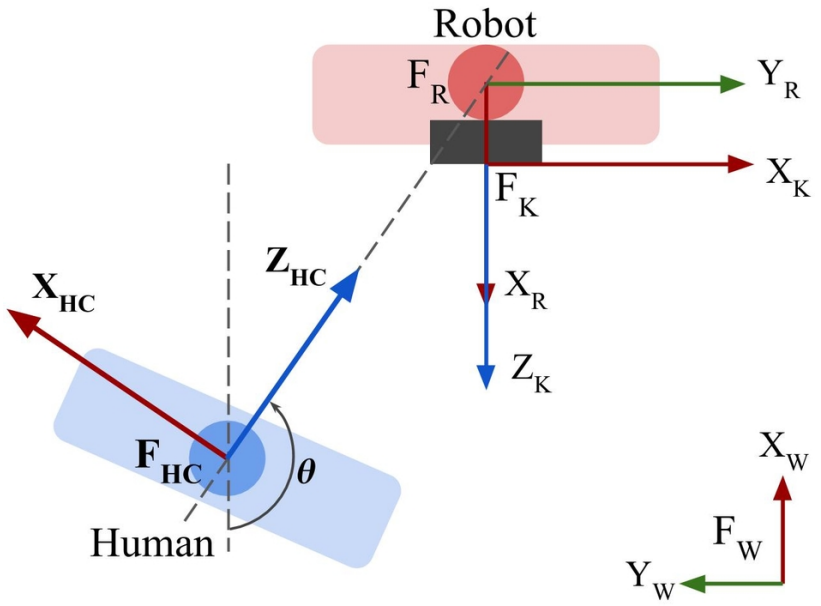


Figure 4.2: Representation of the user-adaptive reference frame

depending on the sensor placement. As a result, the motion of the human arm differs from the training demonstrations if the human stands in a new position. This causes the Pro-MP estimation of the OTP to be inaccurate.

It is highly inefficient to train the Pro-MP with many demonstrations of all possible handover configurations. Therefore, we learned a *dynamic Pro-MP model* from demonstration data collected in a **user-adaptive frame**. From the user study (in Section 3.2), we observed that the *giver's* handover motion is correlated to the plane connecting the positions of human and robot, provided the human is within the robot's field of view. The user-adaptive frame is thus defined based on the robot's frame and human's position with respect to the Kinect camera's frame. The robot to Kinect frame transformation matrix ${}^{K}_{R}T$ is found from the position of the sensor on the robot as shown in Fig. 4.2. The human-centric frame F_{HC}

can be defined with the Z-axis pointing towards the robot's position and the Y-axis perpendicular to the ground. The shoulder positions tracked by the Kinect are used to calculate the origin $(P_{HCx}, P_{HCy}, P_{HCz})$ which is chosen as the midpoint of the shoulder positions and the orientation θ of the frame by trigonometric evaluations. A point in this human-centric frame is found by:

$$p_{HC} = {}^{HC}_{K}T\,{}^{K}_{R}T\,p_R \tag{4.11}$$

In this reference frame, the robot's end effector and human wrist positions are recorded and saved from the perspective of the human partner. Since the object transfer points can be calculated with respect to this user-adaptive frame, the accuracy of the predicted points is not affected by the changes in position and orientation of the human partner with respect to the robot. Overall, using a **user-adaptive frame** improves the generalization capability of the Pro-MP model.

4.4 Integrated OTP Estimator

The dynamic Pro-MP model needs to observe at least 45% of the human's motion from start of handover to accurately estimate the OTP without further feedback [31]. Considering that the robot's arm movement is not as quick as a human's, waiting to observe the human partner's motion further increases the handover time. Also, a slow response by the robot increases the discomfort felt by the human as per the Robot Social Attributes Scale (RoSAS) [32].

The reaction time can be reduced by starting the reaching motion of the robot's

arm as soon as intent for handover has been detected. Here, we don't consider the intent communication problem and define the start of a handover as:

1. The human is nearby ($d_H < 1.5m$) and oriented towards the robot ($\pi/2 < \theta < 3\pi/2$).

2. The object is in hand.

$$\|p_{object} - p_{hand}\| < 0.1m \tag{4.12}$$

3. The hand is moving towards the OTP_s estimate.

$$|d(p_h, OTP_s)_t - d(p_h, OTP_s)_{t-1}| > 0.001m \tag{4.13}$$

As the dynamic OTP estimate is inaccurate during the initial phase of handover, moving robot end effector to the dynamic OTP estimate as soon as the handover intent is detected will result in an irregular motion of the robot's arm. To ensure fluent robot response, accuracy of the OTP estimate in the initial phase of handover needs to be improved such that the difference between the OTP estimate at consecutive time steps is small. We propose an Integrated OTP Estimator that starts with the static OTP estimate, which has better accuracy at the initial phase of handover, and smoothly transitions toward the dynamic OTP estimate as more handover motion is observed.

When the human partner initiates the handover, the integrated OTP (OTP_I) is calculated as the weighted sum of OTP_s and the dynamic OTP estimate (OTP_d)

and updated until the giver's motion is complete. This deformation from static estimation to dynamic estimation is done by tracking the following homotopy function:

$$OTP_I = (1 - \lambda) \cdot OTP_s + \lambda \cdot OTP_d \tag{4.14}$$

As more of the human partner's motion is observed, the homotopy parameter λ is updated as a cubic function based on the estimated phase ϕ of the human's motion.

$$\lambda = (\phi - 1)^3 + 1 \tag{4.15}$$

The phase goes from 0 at the start of the giver's handover motion, to 1 at the end of receiver's reach-to-grasp response. Thus the Dynamic OTP estimate is assigned 0 weight at the start of handover and 1 by the end of the handover motion. We use a cubic function for updating λ based on our observations of the prediction error of the Pro-MP model and is specific to our dynamic estimation method. Still, the strategy for the Integrated OTP estimation remains universal. The Static OTP prediction has to be given more weight in the initial phase of handover and the Dynamic OTP prediction should be used in the later stages of a handover motion. The trained Pro-MP model is then used to generate a natural human-like trajectory to the current estimate of OTP_I. Direct feedback can be used once the final position of the object has been observed.

Chapter 5

Implementation

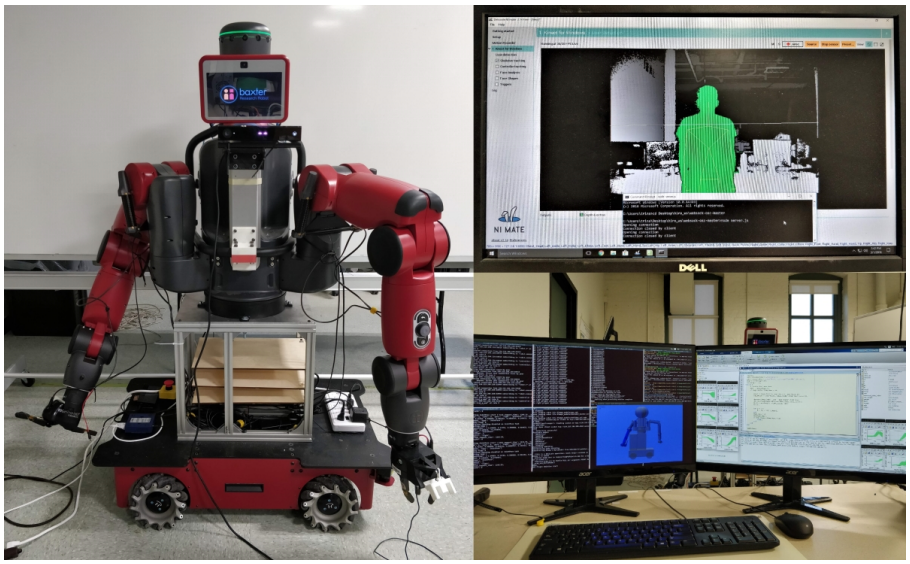


Figure 5.1: (Left) The Tele-robotic Intelligent Nursing Assistant (TRINA) system. (Right top) The sensing server computer that runs skeleton tracking system, and (Right Bottom) the operator console displayed on the robot control computer.

We implemented the proposed OTP estimation method on the Tele-robotic Intelligent Nursing Assistant (TRINA) system shown in Fig. 5.1, which was developed

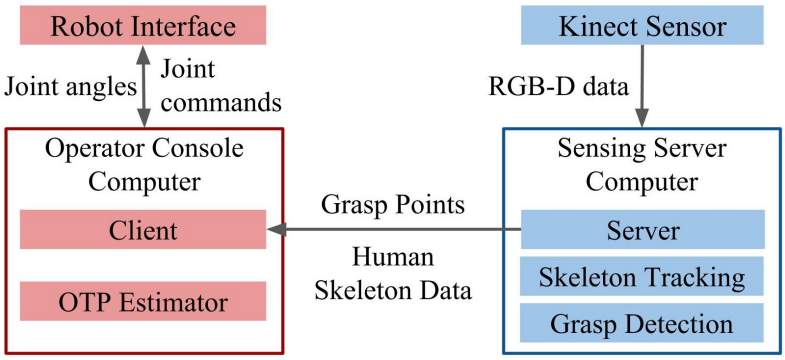


Figure 5.2: Software Architecture: Human skeleton data is streamed over local network by the sensing server and received by the operator console client that runs the OTP estimator.

for nursing tasks [2]. This robotic platform consists of a dual-armed humanoid torso (Rethink Robotics Baxter), an omni-directional mobile base (HStar AMP-I), and two three-fingered grippers (Righthand Robotics ReFlex grippers). A variety of sensors are placed on the robot to provide visual feedback, including the ultrasonic range-finders that come with Baxter for detecting people in its vicinity, a Microsoft Kinect 2 attached to the robot's chest, two Intel RealSense F200 3D cameras attached to the robot's wrists, and two Huokuyo LIDAR sensors attached to the mobile base.

Sensing Module: The Microsoft Kinect 2 sensor is interfaced with a Windows 10 sensing server computer Fig. 5.2. The sensing server computer uses *NI Mate* [27] to track the human partner's motions and streams the human skeleton data for training the dynamic OTP estimation model as well as for predicting the OTP in real-time. The human motion data is recorded as the Cartesian coordinates of all the arm joints in the Kinect's frame, which is fixed with respect to the robot's

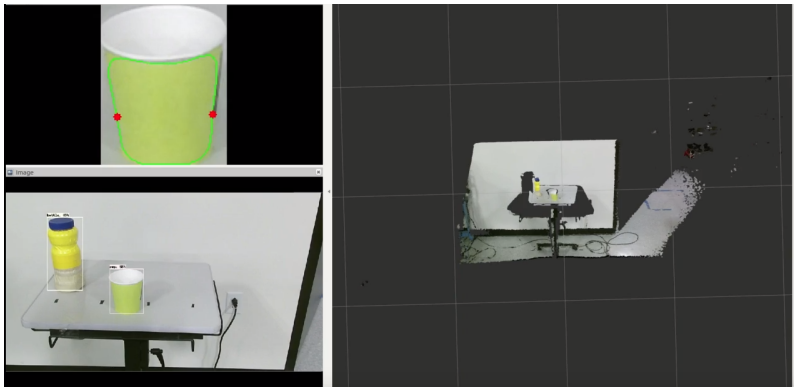


Figure 5.3: An example of grasp point detection for a cup. The cup is identified and cropped from the background. Contour (green line) and grasp points (red) are calculated on the cropped image.

torso. This data is published in the form of OSC (Open Sound Control) messages. A JavaScript server transmits this data to any client in the local network. On the operator console computer, a Python Websocket client converts it into ROS messages and publishes it to a topic which is subscribed by the OTP Estimator.

Grasp Points Detection: The Kinect 2 sensor data is also used to detect the object held by the user (Fig. 5.3). Once an object has been recognized, a bounding box is created, which is used to crop the image out of the scene. The object is segmented from the residue of the environment that was passed in the cropped image by converting to HSV format and applying a threshold. The contour of the segmented object is extracted using built in Open CV functions based on [33]. If multiple contours are found, the one with the largest area is selected. Once the contour has been extracted, the grasp planner calculates the grasp points using Elliptic Fourier Descriptors and Curvature Maximization (see Section 4.1).

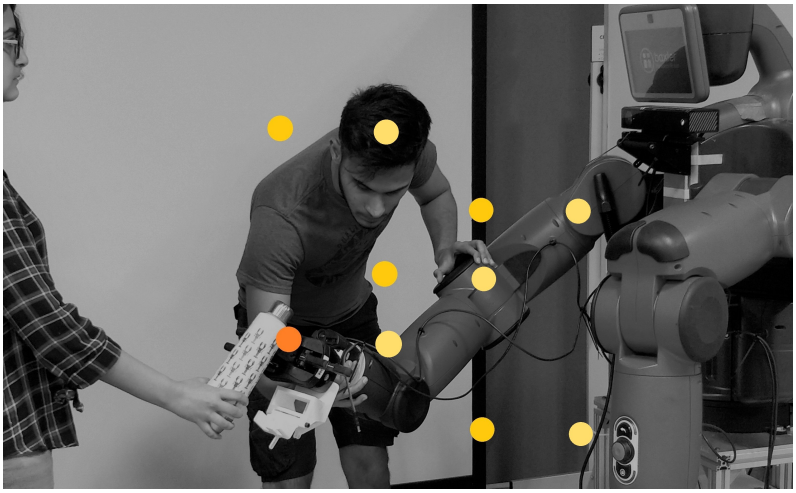


Figure 5.4: Kinesthetic learning for 10 different goal locations

Dynamic OTP Estimation: The estimation model is trained using thirty human-robot handover demonstrations, in which an experimenter moves the robot arm in a natural "elbow down" configuration corresponding to a human giver's handover motion. In these demonstrations, the human giver stands at Position B (Fig. 3.1(b)) and reaches to hand over the object at ten different OTPs in the natural reachable region identified in our user study. A total of 30 demonstrations, three at each OTP, are used to train the dynamic OTP estimation model.

During execution, the human skeleton and grasp points data are constantly updated to the dynamic OTP estimator. For every observation, the dynamic OTP estimator predicts the robot joint angles, and the object position and grasp points for rest of the handover motion. Each joint angle prediction is given to the robot's built-in controller to command the robot arm to the predicted configuration.

Chapter 6

Experimental Evaluation

We compare the proposed Integrated OTP estimation approach with the *Baseline* i.e. the standard Pro-MP estimation method [31] over three criteria:

- **Accuracy** of OTP estimation in the initial phase of handover
- **Time** taken to generate a reach-to-grasp response
- **Generalization** of OTP estimation for new positions of human giver

6.1 Improved Accuracy at Initial Phase

In Experiment 1, we compare OTP estimation accuracy using the Pro-MP model proposed in [31] (i.e. *Baseline*) and the proposed OTP Estimation method (i.e. *Proposed*). The subject stands at the same position as in the training demonstration (position B) and initiates handovers towards different positions within their natural reachable region. The *Estimation Error* is defined as the Euclidean distance

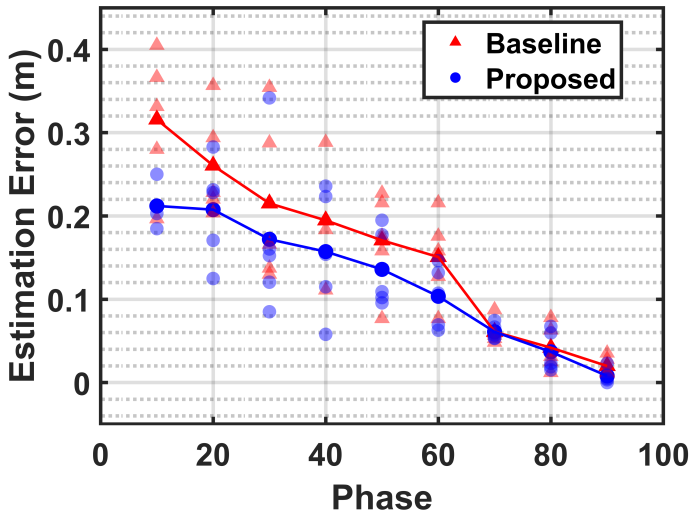


Figure 6.1: Comparison of the prediction error between the baseline (red) and the proposed method (blue).

between the estimated and observed final position of the object. The estimation error was measured at different phases of the handover, when $10\%, 20\%, \cdots, 90\%$ of the human giver's handover motion had been observed. Shown in Fig. 6.1, the estimation errors of the *Baseline* and *Proposed* method decrease as more of handover motion is observed. The *Baseline* method has higher estimation errors at earlier phases of the handover which cause irregular motion at the start of reaching phase. The *Proposed* method assigns a smaller weight to the dynamic OTP estimator (using the Pro-MP model) before its estimation accuracy is better than the Static OTP, and thus can achieve 32.5% more accurate estimation at the start of handover. The smooth weight shifting from static to dynamic OTP estimation leads to a fluent robot reaching motion.

6.2 Faster Handover Response

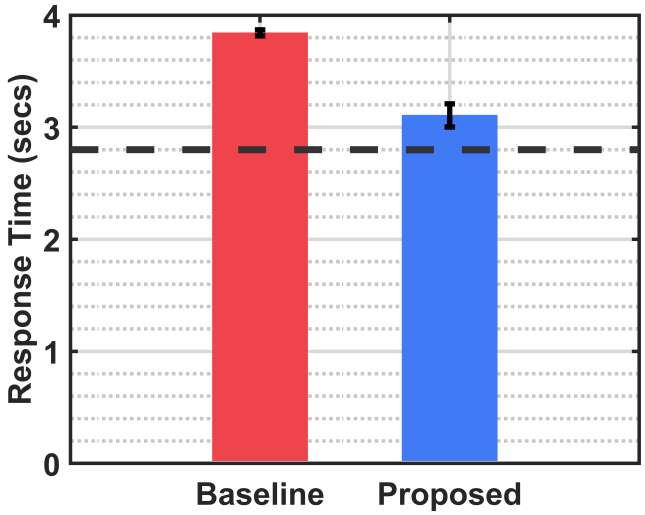


Figure 6.2: Comparison of the response time between the baseline (red) and the proposed method (blue) while ensuring fluent motion.

We further compare the *response time* of the *Proposed* and *Baseline* methods. The Proposed method can start immediately, because at early handover phase it primarily relies on the prediction of static OTP estimator, which has reasonable prediction performance. But the *Baseline* method primarily depends on the dynamic OTP estimation at early handover phase. For safety concern, the robot was set to move only when the estimation error is below 0.2 m, according to Fig. 6.1. The *response time* measures the time from when the robot starts the OTP estimation (as soon as it observes the human givers initiates an handover), to when the robot hand has arrived at the estimated OTP. Shown in Fig. 6.2, the average response times of the *Baseline* and *Proposed* methods are 3.842 *secs* and 3.105 *secs*, respectively. The average time the robot takes to plan and execute the reaching motion

is 2.816 *secs* (as the dotted line in Fig. 6.2 indicates), given an accurate enough OTP is specified. Thus the *Proposed* method reacts in 0.29 *secs* and reduces robot response time by 19.17%.

6.3 Improved Generalization Capability

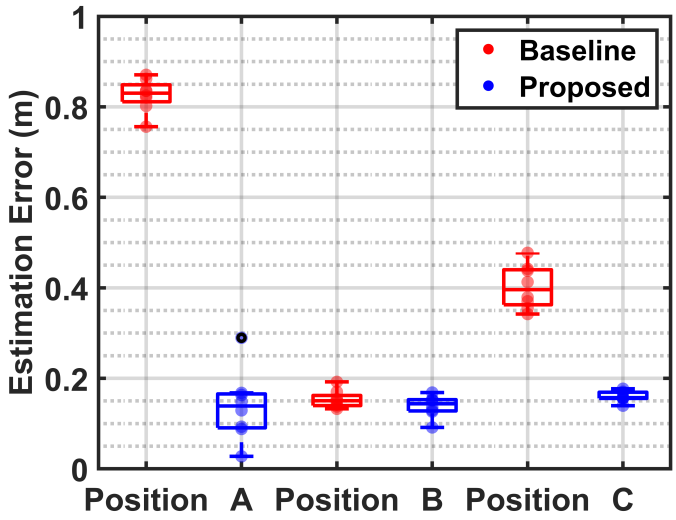


Figure 6.3: Comparison of the generalization capability between the baseline (red) and proposed method (blue) across the workspace.

In experiment 2, we compare the OTP estimation accuracy using the *Baseline* and the *Proposed* methods, when the human givers stand at different positions (A, B and C as in 3.1(b)) in the visible workspace of robot motion tracking camera. Both methods have accurate OTP estimation when human givers stand at Position B, which was the position of giver during training. However, the average estimation errors of the *Baseline* method increase to 0.8322 m and 0.4075 m when the human giver stands at Positions A and C, respectively. On the other hand, the average

estimation errors of the *Proposed* method, which adopted a user-adaptive frame, are 0.174 m and 0.167 m for Position A and C, respectively. Fig. 6.3 compares the estimation errors of the *Baseline* and *Proposed* methods. Note that Position A is further away from Position B compared to Position C, and therefore has a larger increase of estimation error. Thus, using the user-adaptive frame for modelling the Pro-MPs generalizes the prediction to new user positions and is applicable to any dynamic estimation approach for human-robot handovers.

Chapter 7

Conclusion

This work studies human to robot handovers and shows how the distance between human and robot, the height, initial position of wrist and arm length of the human *giver* affect the object transfer point. We present a new inferences on the effect of gaze in human-robot handovers. Based on our observations, a robot *receiver's* gaze has no impact on the human *giver's* choice of OTP. We postulate that a human *giver* chooses the OTP simply based on a robot *receiver's* position and the human's estimate of the robot's reachability. Gaze is important in communicating intent in human-robot interaction [25, 26], but may not affect the OTP in human-robot handovers unless communicated through explicit and exaggerated motion.

We also develop a method that enables a robot to accurately and promptly predict the object transfer point chosen by the human giver. We improve upon the Probabilistic Movement Primitive model by training the model in a user-adaptive reference frame and including grasp points as predictors. Using a user-adaptive

frame helps to generalize the model predictions to new positions of the human *giver*. Our proposed user-adaptive frame can be used with any dynamic OTP estimation method that creates an interactive model of human and robot motions. Creating a model with grasp points helps to generate a reach-to-grasp response that aligns the robot end-effector with the objects grasp configuration. Natural and legible motion of the robot arm is ensured by kinesthetically training the robot arm to follow human-like trajectories. Using the Static OTP estimation improves the accuracy at the start of handover by 0.1 m and allows the robot to initiate its response as soon as the intent for handover has been established. Our Integrated OTP estimator smoothly transitions from the Static OTP to Dynamic OTP estimate to generate fluent handovers that are 19.17% faster than our *Baseline* Pro-MP model.

Chapter 8

Future Work

Apart from physical factors like height, distance, orientation and gaze; **mental factors** may also play an important role in determining the object transfer point in human-robot handover tasks. The work in this thesis assumes that both partners in a handover task equally share the work load of the interaction. But in reality, one partner may tend to do less work than the other. Therefore, the nature of a partner to be **more or less collaborative** can have an effect on their performance in an object handover tasks. This thesis work also considers an isolated handover scenario, where the choice of OTP by the *giver* is dependent on only one *receiver*. In a complex collaborative setting, a human or robot may have to perform handover with multiple agents and therefore the choice of OTP may be affected by the physical and mental condition of other receivers.

As future work, we look at one-to-one and one-to-many human-human handover scenarios to determine the effect of collaboration on the object transfer point.

8.1 Inferring the level of collaboration in handover tasks: From one-to-one to one-to-many

The study of one-to-many handover is motivated by the scenario of (1) one autonomous robot serving many humans, and (2) human supervising multiple low-autonomy robots to serve their end users. The level of collaboration is about (1) whether the end user will perceive the autonomous robot or the entire human-robot teaming system to be collaborative or not, and (2) how to design such system to behave as a collaborative partner with all (or most) of the remote users.

Research in improving robot performance in handover tasks focuses on inferring human intent and planning robot motion such that it is efficient, intuitive, safe and comfortable for the human partner. Robot efficiency in handover tasks depends on the reaction time and accuracy of the robot response. Often, observations from human-human handover studies [3, 5, 6] are used to model expected human behaviour. Human posture, arm length and gaze can be used to predict a prior static estimate of the object transfer point (Section 4.2). This static estimate can then be updated based on the observed human motion to promptly and accurately plan the robot reach-to-grasp motion (Section 4.4).

Although predictive control leads to efficient and functional handovers, planning legible motions that clearly indicate the robot's intent lead to a more fluent collaboration [34]. Characteristics of collaborative fluency, such as the subjective and objective fluency metrics, observer and participant fluency perception, etc, help to evaluate the fluency of human-robot handovers [35]. Apart from fluency, factors

like adaptability [36], compliance [37] and trust [38] also indicate the level of collaboration of the human or robot partner. For sequential tasks, adaptability can be measured based on the probability with which one partner adapts to the other partner's reward function [36]. Inferring the robot's reward function in a task also helps to build a human partner's trust in the robot's capabilities [39].

Although handovers have been studied for face-to-face, dynamic, repetitive and sequential task scenarios, the majority of the research deals with one-to-one handover tasks. A non-sequential one-to-many handover task would involve the additional problem of scheduling the robot's actions to cater to multiple users. In the case of mixed human-robot teams where a human leader has to allocate tasks to a human assistant and a robotic co-leader [40], task scheduling can be done by minimizing the maximum amount of work assigned to an agent. Constraints for this problem consider lower bounds on time, number of tasks assigned to each agent and other temporal and spatial constraints of the task. However, the study only focuses on how human satisfaction was affected by the level of robot autonomy and not the level of collaboration. In our proposed study, we aim to evaluate the aspects of a robot's performance that affect a human partner's perception of the robot's level of collaboration or vice versa.

8.1.1 One-to-One Object Handover

To determine the factors that indicate the level of collaboration of a partner in an object handover task, we conducted a one-to-one human-human handover experiment.

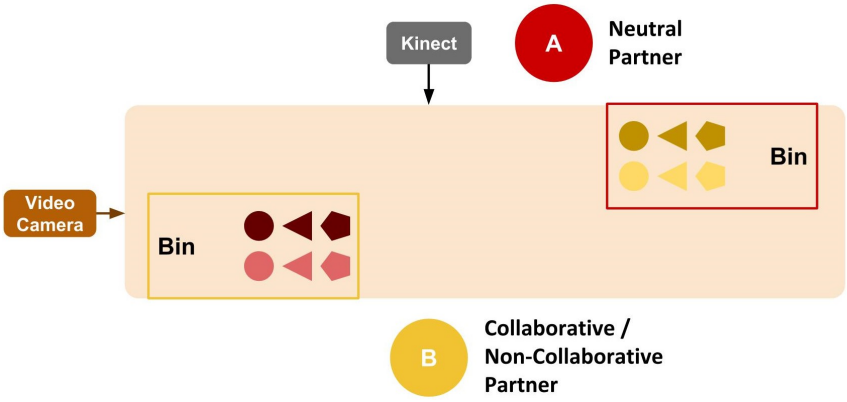


Figure 8.1: Experiment setup for Pilot Study

As shown in Fig. 8.1, the subjects A and B were asked to stand on opposite sides of a table. 6 objects with different affordances were placed in each of the bins on either side of the table. The subjects were asked to collaborate in moving all the red objects to the red bin and yellow objects to the yellow bin. They were only allowed to handle one object at a time. A trial was considered complete when all the objects were in their respective bins.

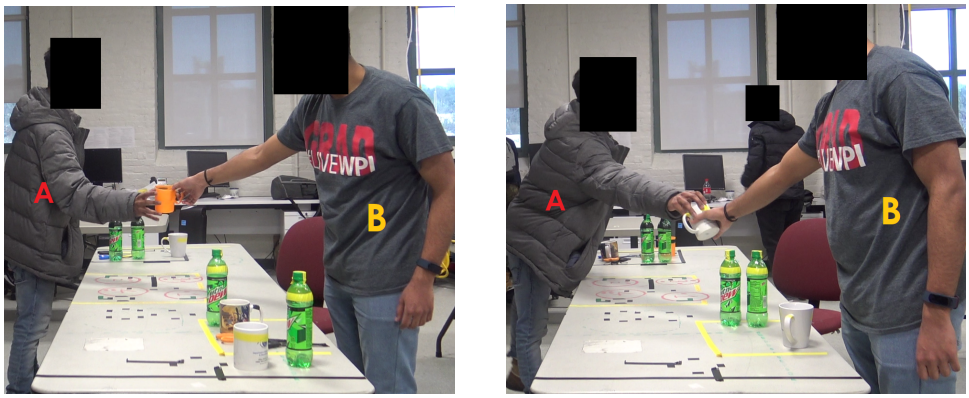
The study comprised of 2 trials. In one trial, Subject B was asked to be **Collaborative** i.e. *be helpful to their partner*. In the other trial, Subject B was asked to be **Non-Collaborative** i.e. *offer minimum help to their partner*. Subject A was provided with no specific instruction and was unaware of Subject B's instruction. Subject A's behaviour was assumed to be neutral or collaborative. The order of collaborative and non-collaborative trials for all subjects was decided based on balanced latin square.

Subject B's movements were tracked through a Kinect sensor using the NI Mate motion capture system. The skeleton data was used to calculate the **object trans-**

fer point and the **orientation** of the subject's body and head. A video camera on the side of the table captured the task scene. The video data was used used to record **verbal communication**, **object affordance**, the **timing of actions** and **total time**. At the end of the study both the subjects answered a questionnaire:

- Do you think your partner was collaborative? Explain. (Only Subject A)
- What did you do to act collaborative/non-collaborative? (Only Subject B)
- Who took the charge? Explain.
- Were there any conflicts? If yes, how were they solved?

8.1.2 Preliminary Work



(a) Collaborative trial (b) Non-collaborative trial

Figure 8.2: Object transfer point changes based on level of collaboration

A pilot study of the one-to-one handover experiment was performed with 6 pairs of subjects. Based on our observation, **affordance** of the objects had no impact on collaboration intent. As the objects did not have a function in the task, affordance

was not considered by Subject B. But the **object transfer point** during the non-collaborative trial was lower and closer to the yellow bin than in the collaborative trial. **Attention** of Subject B was modelled using the body and head orientation. During the collaborative trial Subject B paid attention to all actions initiated by Subject A. While in the non-collaborative trial Subject B gave and received objects without acknowledging Subject A's intended actions. The average **reaction time** of Subject B was consistent for all actions in the collaborative trial. While the average reaction time during the non-collaborative trial was slower or inconsistent. Conflicts occurred when both subjects tried to handover an object at the same time. **Resolution of conflicts** was much faster in the collaborative trial than the non-collaborative trial.

8.1.3 One-to-Many Object Handover

The significant variables inferred from the one-to-one handover experiment will be used to model and contrast how the level of collaboration is estimated in a one-to-many handover scenario. We utilize the results of the pilot study to design a similar one-to-many handover experiment (Fig. 8.3). Here the **affordances** of objects will be enforced by defining how the objects should be placed in the bin. Along with the factors mentioned in the one-to-one scenario, **task scheduling** will now affect how the level of collaboration of subject B is perceived by Subjects A1, A2, and A3.

The human subjects can initiate a *Give* action where they would offer an object to the robotic agent or a *Demand* action where they would raise their arm to

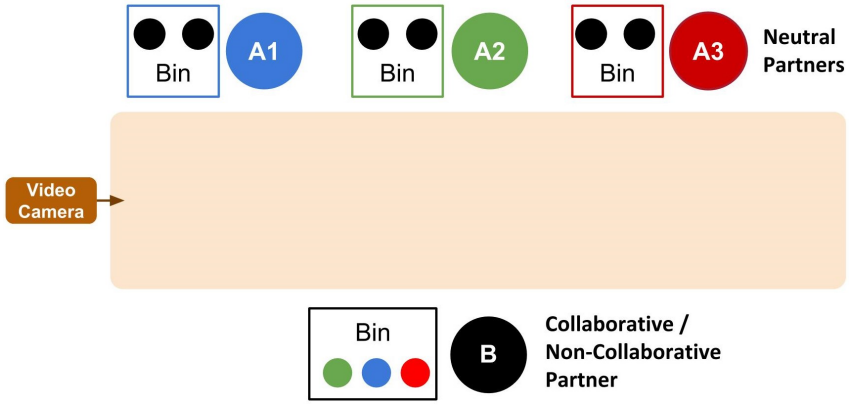


Figure 8.3: Experiment setup for one-to-many handover study

demand an object from the robotic agent. The robot can respond with a *Take* or *Give* action. The robot can also initiate a *Demand* action. The robot requires t_t time to execute the *Take* action and t_g time to execute the *Give* action. The task scheduling problem will select actions based on the following cost function:

$$min \quad \sum_{i=1}^{3}(C_{G_i} * t_{w_i} + C_{D_i} * t_{w_i}) + t_{total}$$

Where, t_{w_i} is the waiting period for subject i, $C_{G_i} * tw_i$ is the cost associated with the *Give* action and $C_{D_i} * tw_i$ is the cost associated with the *Demand* action. t_{total} is the time required to complete the total task. The problem can be formulated with additional temporal and spatial constraints.

We propose a one-to-many human-human user study to learn the cost factors C_G and C_D and the weights for the one-to-one factors: **object transfer point**, **attention**, **verbal communication** and **conflict resolution** that lead to a collaborative behaviour. This study will help to analyze the low-level and high-level factors that affect how a service robot will be perceived by its users.

Bibliography

[1] K. W. Strabala, M. K. Lee, A. D. Dragan, J. L. Forlizzi, S. Srinivasa, M. Cakmak, and V. Micelli, "Towards Seamless Human-Robot Handovers," *Journal of Human-Robot Interaction*, vol. 2, no. 1, pp. 112–132, mar 2013.

[2] Z. Li, P. Moran, Q. Dong, R. J. Shaw, and K. Hauser, "Development of a tele-nursing mobile manipulator for remote care-giving in quarantine areas," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, may 2017, pp. 3581–3586.

[3] E. A. Sisbot and R. Alami, "A Human-Aware Manipulation Planner," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1045–1057, oct 2012.

[4] S. Shibata, K. Tanaka, and A. Shimizu, "Experimental analysis of handing over," in *Proceedings 4th IEEE International Workshop on Robot and Human Communication*. IEEE, pp. 53–58.

[5] S. Parastegari, B. Abbasi, E. Noohi, and M. Zefran, "Modeling human reaching phase in human-human object handover with application in robot-human

handover," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, sep 2017, pp. 3597–3602.

[6] P. Basili, M. Huber, T. Brandt, S. Hirche, and S. Glasauer, "Investigating Human-Human Approach and Hand-Over." Springer, Berlin, Heidelberg, 2009, pp. 151–160.

[7] W. P. Chan, M. K. X. J. Pan, E. A. Croft, and M. Inaba, "Characterization of handover orientations used by humans for efficient robot to human handovers," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2015, pp. 1–6.

[8] J. Aleotti, V. Micelli, and S. Caselli, "Comfortable robot to human object hand-over," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, Sept 2012, pp. 771–776.

[9] J. Mainprice, E. Sisbot, T. Siméon, and R. Alami, "Planning safe and legible hand-over motions for human-robot interaction," 01 2010.

[10] A. H. Quispe, H. B. Amor, and M. Stilman, "Handover planning for every occasion," in *2014 IEEE-RAS International Conference on Humanoid Robots*, Nov 2014, pp. 431–436.

[11] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi, "Using spatial and temporal contrast for fluent robot-human hand-overs," in *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2011, pp. 489–496.

[12] M. Prada, A. Remazeilles, A. Koene, and S. Endo, "Implementation and experimental validation of Dynamic Movement Primitives for object handover," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, sep 2014, pp. 2146–2153.

[13] D. Vogt, S. Stepputtis, ·. B. Jung, ·. Heni, and B. Amor, "One-shot learning of human-robot handovers with triadic interaction meshes," *Autonomous Robots*, vol. 42, pp. 1053–1065, 2018.

[14] C. Perez-D'Arpino and J. A. Shah, "Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, may 2015, pp. 6175–6182.

[15] G. Maeda, M. Ewerton, G. Neumann, R. Lioutikov, and J. Peters, "Phase estimation for fast action recognition and trajectory generation in human-robot collaboration," *The International Journal of Robotics Research*, vol. 36, pp. 13–14, 2017.

[16] R. Showcase, . Cmu, V. Micelli, K. Strabala, S. Srinivasa, and S. S. Srinivasa, "Perception and Control Challenges for Effective Human-Robot Handoffs," Tech. Rep., 2011.

[17] C.-M. Huang and B. Mutlu, "Anticipatory Robot Control for Efficient Human-Robot Collaboration," Tech. Rep., 2016.

[18] Z. Wang, K. Mülling, M. P. Deisenroth, H. Ben Amor, D. Vogt, B. Schölkopf, and J. Peters, "Probabilistic movement modeling for intention inference in human–robot interaction," *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 841–858, jun 2013.

[19] R. Liu and X. Zhang, "Fuzzy context-specific intention inference for robotic caregiving."

[20] C. L. R. McGhan, A. Nasir, and E. M. Atkins, "Human Intent Prediction Using Markov Decision Processes," *Journal of Aerospace Information Systems*, vol. 12, no. 5, pp. 393–397, may 2015.

[21] M. Monfort, A. Liu, and B. D. Ziebart, "Intent Prediction and Trajectory Forecasting via Predictive Inverse Linear-Quadratic Regulation," Tech. Rep.

[22] H. C. Ravichandar and A. P. Dani, "Human Intention Inference Using Expectation-Maximization Algorithm With Online Model Learning," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 2, pp. 855–868, apr 2017.

[23] J. R. Medina, F. Duvallet, M. Karnam, and A. Billard, "A human-inspired controller for fluid human-robot handovers," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*. IEEE, nov 2016, pp. 324–331.

[24] Z. Li and K. Hauser, "Predicting Object Transfer Position and Timing in Human-robot Handover Tasks," Tech. Rep.

[25] A. Moon, D. M. Troniak, B. Gleeson, M. K. Pan, M. Zheng, B. A. Blumer, K. MacLean, and E. A. Croft, "Meet me where i'm gazing: How shared attention gaze affects human-robot handover timing," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '14. New York, NY, USA: ACM, 2014, pp. 334–341.

[26] M. Zheng, A. Moon, E. A. Croft, and M. Q.-H. Meng, "Impacts of robot head gaze on robot-to-human handovers," *International Journal of Social Robotics*, vol. 7, no. 5, pp. 783–798, Nov 2015.

[27] [Online]. Available: https://ni-mate.com/

[28] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[29] F. P Kuhl and C. R Giardina, "Elliptic Fourier Features of a Closed Contour," *Computer Graphics and Image Processing*, vol. 18, pp. 236–258, 1982.

[30] B. Calli, M. Wisse, and P. Jonker, "Grasping of unknown objects via curvature maximization using active vision," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2011, pp. 995–1001.

[31] G. J. Maeda, G. Neumann, M. Ewerton, R. Lioutikov, O. Kroemer, and J. Peters, "Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks," *Autonomous Robots*, vol. 41, no. 3, pp. 593–612, mar 2017.

[32] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas): Development and validation," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 254–262.

[33] S. Suzuki *et al.*, "Topological structural analysis of digitized binary images by border following," *Computer vision, graphics, and image processing*, vol. 30, no. 1, pp. 32–46, 1985.

[34] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 51–58.

[35] G. Hoffman, "Evaluating fluency in human-robot collaboration," in *International conference on human-robot interaction (HRI), workshop on human robot collaboration*, vol. 381, 2013, pp. 1–8.

[36] S. Nikolaidis, D. Hsu, and S. Srinivasa, "Human-robot mutual adaptation in collaborative tasks: Models and experiments," *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 618–634, 2017.

[37] S. Nikolaidis, M. Kwon, J. Forlizzi, and S. Srinivasa, "Planning with verbal communication for human-robot collaboration," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 7, no. 3, 2018.

[38] A. Freedy, E. DeVisser, G. Weltman, and N. Coeyman, "Measurement of trust in human-robot collaboration," in *2007 International Symposium on Collaborative Technologies and Systems*. IEEE, 2007.

[39] S. Nikolaidis, S. Nath, A. D. Procaccia, and S. Srinivasa, "Game-theoretic modeling of human adaptation in human-robot collaboration," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE, 2017, pp. 323–331.

[40] M. C. Gombolay, R. A. Gutierrez, S. G. Clarke, G. F. Sturla, and J. A. Shah, "Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams," *Autonomous Robots*, vol. 39, no. 3, pp. 293–312, 2015.