

# Object Transfer Point Estimation for Fluent Human-Robot Handovers

Heramb Nemlekar, Dharini Dutia and Zhi Li

**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. This work presents an efficient method for predicting the Object Transfer Point (OTP), which synthesizes (1) an offline OTP calculated based on human preferences observed in a 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 only using 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) close 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.

## I. INTRODUCTION

The study of fluent and natural-looking human-robot handovers has been motivated by the need for physical interactions and collaborations 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(a)), and hand over medical supplies when assisting a human nurse [2]. Such handover tasks are frequently performed and therefore have a dominant effect on the overall task performance. Research on human-robot handovers has 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(b)). In this paper, 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.



(a) Robot to Human Handover (b) Human to Robot Handover

Fig. 1. Fluent and natural-looking human-robot object handover is critical to the performance of collaborative tasks.

Heramb Nemlekar, Dharini Dutia, and Zhi Li are with the Robotics Engineering Program, Worcester Polytechnic Institute, Worcester, MA 01609, USA {hsnemlekar, dkdutia, zll111}@wpi.edu

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, wherein the giver chooses a direct path to the OTP without deviating from it [3]. Similarly the giver’s arm motion in a vertical 2D plane is pre-planned, feed forward with a fixed maximum velocity [4]. The motion of the giver’s arm is also independent of the receiver [5], 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. Moreover the handovers occur half way between the giver and the receiver. Apart from interpersonal distance, factors like safety, visibility and arm comfort can be considered to postulate the point of object transfer [6]. With the 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.

Dynamic OTP estimation requires observing the human partner’s behavior in real-time for intent inference and motion prediction. Instead of inferring intent from more explicit cues (e.g., gaze or body gestures [16]), inferring the implicit intent encoded in human motion can be more efficient and less intrusive, as it minimizes end user effort for maintaining explicit or even exaggerated communication. At a high-level, the human intent inference problem can be formulated as inferring the parameters of a dynamic model [17], Bayesian network [18] or Markov decision process [19] and tackled using techniques such as inverse linear-quadratic regulation (ILQR) [20] and approximate expectation-maximization [21]. 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 [22]. But knowing that natural human reaching motions follow minimum-jerk trajectories, the timing and location of the object transfer can be predicted early, after peak velocity of the human partner’s hand has been observed [23]. To react as soon as the intent for handover is detected, the robot hand velocity can be controlled proportional to the hand velocity of the human partner [1]. The human motion can also 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 [24]. Such methods for dynamic prediction, which require significant

human motion to be observed, delay the robot’s response and do not produce a natural-looking handover.

Ideally, the robot should quickly react to the human partner and improve OTP prediction accuracy as more of human partner’s motion is observed. Dynamic Movement Primitives (DMP) can reproduce trained trajectories to new goal locations through a combination of attractor 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 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  $\pm 37$  *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 45% of the human’s hand motion. 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.

To render proactive and adaptive robot motions, we propose an OTP estimation strategy that combines a pre-computed object transfer point (*static OTP*) which addresses giver position, reachability and height, 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 Section II. 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 Section III. 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 w.r.t. 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 Section IV). Experimental results show that response time is decreased by 19.17% and estimation accuracy at the start of handover is increased by 32.5%.

## II. 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 [3]–[5]. 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.

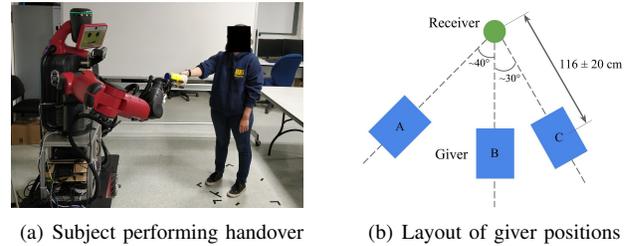


Fig. 2. Experimental setup for human to robot handover study

Shown in Fig. 2(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. 2(b), the bounding box is defined such that distances between robot and human subjects are  $116 \pm 20$  *cm* from the *receiver*, according to the social space in proxemics defined in [5]. Position *B* faces directly to robot *receiver*, while Positions *A* and *C* are chosen to be the 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.

### A. 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  $d_{O,H}$ .

As seen in Fig. 3(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. Fig. 3(b) shows a positive relationship between

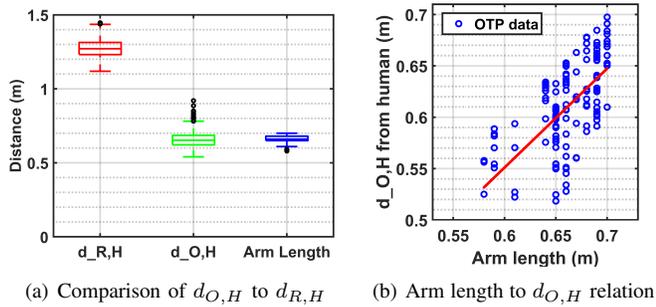


Fig. 3. Effect of interpersonal distance and arm length on OTP

arm length and distance of OTP from *giver*. The regression coefficient for arm length is 0.96, indicating that the users presented the object at the extent of their reachability.

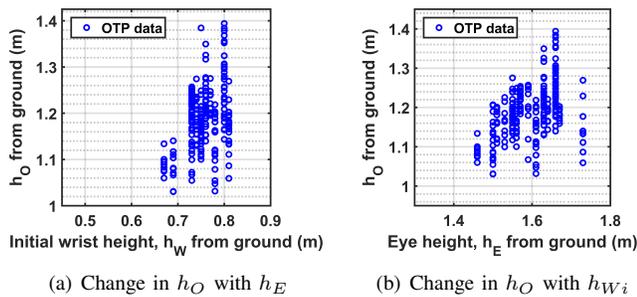
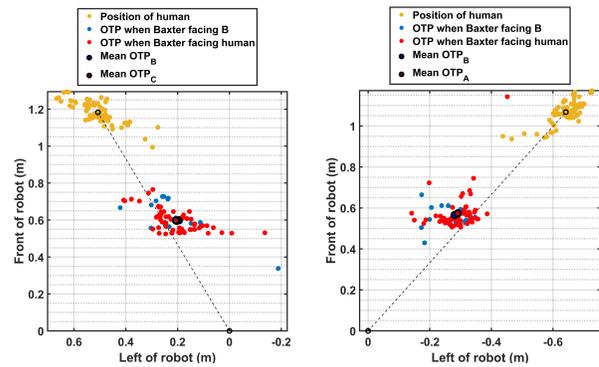


Fig. 4. Effect of height of giver and initial wrist position on OTP

We further evaluated how the height of human givers affects the height of the OTP,  $h_O$  (see Fig. 4). Let  $h_E$  and  $h_{W_i}$  be the heights of the subject's eyes and 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, which indicates that they may be significant parameters. But as the value regression coefficients and the range of data is small,  $h_E$  and  $h_{W_i}$  may not be useful in predicting the height of OTP from ground. Rather,  $h_O$  can be modelled as a Gaussian distribution with a mean of 1.182 m, which is close to the height of the robot *receiver*.

### B. 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 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 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. 5, even when the robot diverted its gaze, the average



(a) Human standing at position C (b) Human standing at position A

Fig. 5. The effect of robot gaze on the OTP chosen by human givers

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. This might be because the human subjects did not associate robot gaze direction closely with the direction it can sense. Even when the robot's gaze was directed away from the human givers, they still chose the OTP based on the orientation of the robot's body.

### C. Natural response time in human-human handover

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, 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 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  $0.425 \pm 0.035$  secs, while the observed **response time**, which was the time from the giver starting their motion to the receiver reaching to the object, was  $1.212 \pm 0.051$  secs. We aim to enable a robot *receiver* to react and deliver a reach-to-grasp response as fast as a human *giver*.

## III. PROPOSED FRAMEWORK FOR OTP ESTIMATION

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. 6, the **OTP estimator module** takes input from the **sensing module** which observes the robot state, the object position, and human partner's motion in real-time. Within the estimation module, the *offline training* components are responsible for (1) training a Probabilistic Movement Primitives (Pro-MP) model to reproduce legible robot motion using demonstrations of human-robot handovers, as well as (2) generating a static OTP estimation before the handover starts. As soon as the human partner starts a handover, the integrated OTP generator takes in the static OTP estimation and updates it with the estimate from the dynamic OTP estimator by determining the phase of the human partner's observed

motion. The motion controller receives the integrated OTP and controls the robot end-effector to reach toward it.

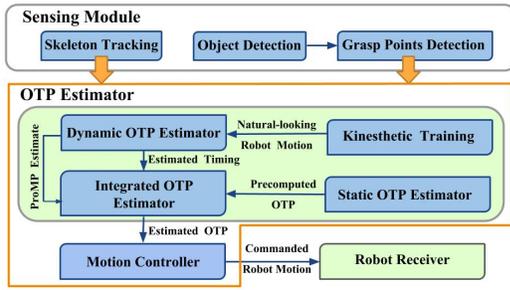


Fig. 6. Handover System Architecture: sensing module communicates skeleton data and grasping points to the OTP estimator generating a required trajectory which is implemented by the robot.

### A. 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 correspondence grouping [28] and detection is simplified by using pure coloured objects. Grasp points are determined based on a representation of the object contour using Elliptic Fourier Descriptors [29]. Curvature is used in selecting model grasp points, and can be determined as the sign of the dot product between normal and tangent vectors [30]:

$$Curvature = sign(\|Z \cdot N\|) \quad (1)$$

**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

---

- 1: Rank all possible pair sets by descending  $x + y$  curvature
  - 2: **for** each set  $x, y$  with positive  $\alpha$  **do**
  - 3:      $\beta = PerformForceClosure()$
  - 4:     **if**  $\beta$  above threshold **return**
  - 5: Rank sets by ascending and repeat for negative  $\alpha$
  - 6:
  - 7: **procedure** PERFORM FORCE CLOSURE( $x, y$ )
  - 8:      $A = \frac{N_{m1}}{\|N_{m1}\|} \cdot \frac{P_{m1} - P_{m2}}{\|P_{m1} - P_{m2}\|}$
  - 9:      $B = \frac{N_{m2}}{\|N_{m2}\|} \cdot \frac{P_{m1} - P_{m2}}{\|P_{m1} - P_{m2}\|}$
  - 10: **return**  $A^2 + (\pi - B)^2$
- 

### B. Static OTP Estimator

Before 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 *giver’s* position and *receiver’s* orientation. (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*.

### C. Dynamic OTP Estimator

The core of our method is to train Multi-dimensional Interaction Probabilistic Movement Primitives (Pro-MP) with multiple human-robot handover demonstrations [31].

*Learning phase:* The arm of the nursing robot is by default in an “elbow-up” configuration. 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:

$$\mathbf{y}_t = [y_{1,t}^H, \dots, y_{3,t}^H, y_{1,t}^O, \dots, y_{6,t}^O, y_{1,t}^R, \dots, y_{7,t}^R]^T \quad (2)$$

The trajectory of each DOF is further parameterized by weights such that:

$$p(\mathbf{y}_t | \bar{\mathbf{w}}) = \mathcal{N}(\mathbf{y}_t | \mathbf{H}_t^T \bar{\mathbf{w}}, \Sigma_t) \quad (3)$$

where  $\mathbf{H}_t^T = \text{diag}((\Psi_t^T)_1, \dots, (\Psi_t^T)_3, (\Psi_t^T)_1, \dots, (\Psi_t^T)_6, (\Psi_t^T)_1, \dots, (\Psi_t^T)_7)$  is the diagonal matrix of the Gaussian basis functions. Among the  $M$  handover demonstrations, the  $i$ -th demonstration correlates the observed DOFs of human and robot in the handover such that:

$$\bar{\mathbf{w}}_i = [(\mathbf{w}_1^H)^T, \dots, (\mathbf{w}_3^H)^T, (\mathbf{w}_1^O)^T, \dots, (\mathbf{w}_3^O)^T, (\mathbf{w}_1^R)^T, \dots, (\mathbf{w}_7^R)^T]^T \quad (4)$$

*Reproduction phase:* Using the learned Pro-MP model, the robot end-effector trajectory can be inferred by computing the posterior probability distribution of the weights  $\mathbf{w}$  conditioned on the observed human motion. The phase of the observation is determined based on correlation of the observed data with sampled trajectories from the training demonstrations. 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.

*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 ( $F_W$ ) 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

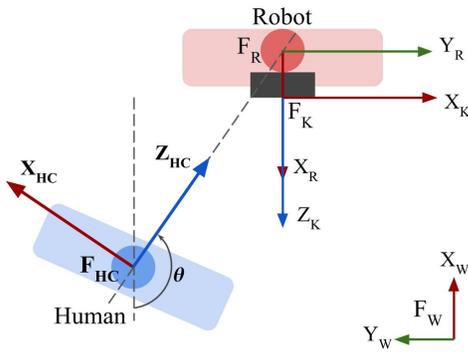


Fig. 7. User-adaptive Frame Representations

demonstration data collected in a **user-adaptive frame**. From the user study (in Section II-B), 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. 7. 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  $\theta$  of the frame by trigonometric evaluations. A point in this human-centric frame is found by:

$$p_{HC} = {}^{HC}_K T {}^K_R T p_R \quad (5)$$

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.

#### D. 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. 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.

$$\|p_{object} - p_{hand}\| < 0.1m \quad (6)$$

- 3) The hand is moving towards the  $OTP_s$  estimate.

$$|d(p_h, OTP_s)_t - d(p_h, OTP_s)_{t-1}| > 0.001m \quad (7)$$

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:

$$OTP_I = (1 - \lambda) \cdot OTP_s + \lambda \cdot OTP_d \quad (8)$$

As more of the human partner's motion is observed, the homotopy parameter  $\lambda$  is updated as a cubic function based on the estimated phase  $\phi$  of the human's motion and the prediction error of the Pro-MP model.

$$\lambda = (\phi - 1)^3 + 1 \quad (9)$$

The trained 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.

## IV. EXPERIMENTS & RESULTS

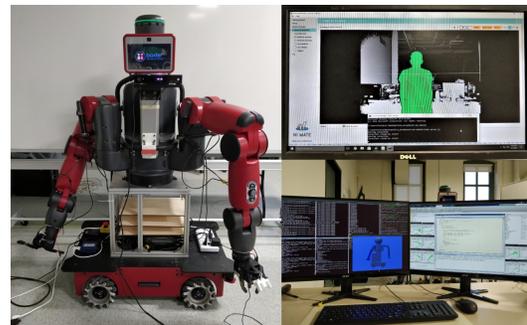


Fig. 8. (Left) The Tele-robotic Intelligent Nursing Assistant (TRINA) system. (Right top) The sensing server computer that runs skeleton tracking system, and (Right Bottom) the operator console displayed on the robot control computer.

We implemented the proposed OTP estimation method on the Tele-robotic Intelligent Nursing Assistant (TRINA) system shown in Fig. 8, which was developed for nursing tasks [2]. A Microsoft Kinect 2 sensor is attached to the robot's chest and interfaced with a sensing server computer that streams the human partner's skeleton data on the local network. The estimation model is trained using twenty-five human-robot handover demonstrations, in which an experimenter holds the robot arm to reach to a human giver. In these demonstrations, the human givers stand at Position  $B$  (Fig. 2(b)) and reach to hand over the object at five different OTPs in their natural reachable region.

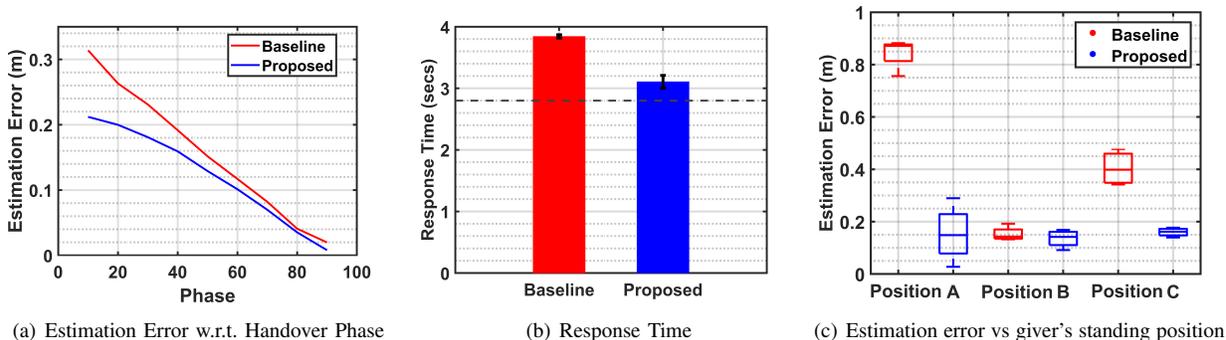


Fig. 9. Performance comparison between the baseline Pro-MP (red) and the proposed estimator (blue)

### A. Faster Handover Response

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 between the estimated and observed final position of the object. The estimation error was measured at different phases of the handover, when 10%, 20%, ..., 90% of the human giver's handover motion had been observed. Shown in Fig. 9(a), 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.

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 using the *Baseline* method is set to move when the estimation error is below 0.2 m, according to Fig. 9(a). 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. 9(b), 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. 9(b) 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%. This reaction

and response time is closest to a human *receiver's* time and is only limited by the maximum speed of the robot hardware.

### B. Improved OTP Estimation Accuracy

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 2(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. 9(c) 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.

## V. CONCLUSION

This paper looks at human to robot handovers and develops a method that enables a robot to accurately and promptly predict the object transfer point chosen by the human giver. The response time measured for human-human handovers indicates that a robot *receiver* should respond in 0.425 secs to the handover initiated by the *giver*. Our human-robot handover study showed how the position, height and arm length of the *giver*, and the orientation and height of *receiver* determine the location of the OTP. We also found that the gaze of a robotic *receiver* may not affect the choice of OTP by a human *giver* as the *giver* does not associate the robot's ability to perceive a handover to its gaze, unless gaze is highlighted through exaggerated motion. Our improved OTP estimator reacts as soon as intent for handover has been established and generates human-like handovers that are 19.17% faster than the *Baseline*. The robot response also corresponds to the object grasp configuration. Lastly, we highlight the importance of learning motions in a dynamic human-centric reference frame. Our user adaptive frame approach helps to generalize the robot motions to any new configurations of its human partner.

## REFERENCES

- [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] 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.
- [4] 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.
- [5] 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.
- [6] E. A. Sisbot and R. Alami, "A Human-Aware Manipulation Planner," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1045–1057, Oct 2012.
- [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," in *IARP/IEEE-RAS/EURON workshop on technical challenges for dependable robots in human environments*, Jan 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 handovers," 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, and H. 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] C.-M. Huang and B. Mutlu, "Anticipatory Robot Control for Efficient Human-Robot Collaboration," Tech. Rep., 2016.
- [17] 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.
- [18] R. Liu and X. Zhang, "Fuzzy context-specific intention inference for robotic caregiving," *International Journal of Advanced Robotic Systems*, vol. 13, no. 5, 2016.
- [19] 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.
- [20] M. Monfort, A. Liu, and B. Ziebart, "Intent prediction and trajectory forecasting via predictive inverse linear-quadratic regulation," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [21] 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.
- [22] V. Micelli, K. Strabala, and S. Srinivasa, "Perception and control challenges for effective human-robot handoffs," 2011.
- [23] Z. Li and K. Hauser, "Predicting object transfer position and timing in human-robot handover tasks," in *Science and Systems: 2015 Robotics (RSS 2015), Human-Robot Handover Workshop, Rome, Italy*, 2015.
- [24] J. R. Medina, F. Duvallat, 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.
- [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] F. Tombari and L. D. Stefano, "Object recognition in 3d scenes with occlusions and clutter by hough voting," in *2010 Fourth Pacific-Rim Symposium on Image and Video Technology*, Nov 2010, pp. 349–355.
- [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.