

# Predicting Object Transfer Position and Timing in Human-robot Handover Tasks

Zhi Li<sup>1</sup> and Kris Hauser<sup>2</sup>

Department of Electrical and Computer Engineering, Duke University, Durham, NC, 27708

Email: <sup>1</sup> zhi.li2@duke.edu; <sup>2</sup> kris.hauser@duke.edu

**Abstract**—Handing-over objects is the foundation of many human-robot interaction tasks. Toward seamless and effortless hand-overs, it is desirable for a robot to predict human motion and plan its own motion accordingly. In the scenario of human (giver)-robot (receiver), we propose to enable the robot receiver to predict when and where the object will be transferred, so that it can actively reach out and pick up the object instead of passively waiting for the object to be presented. To generalize this motion prediction problem, we collected data on human reaching motions in a 3D workspace, to test the performance of arm motion models and linear regression models we propose for end time and position prediction. Preliminary results from some of the proposed models will be reported. Future work will thoroughly compare the prediction performance of all the proposed models. These prediction algorithms will be implemented on a humanoid robot for tele-nursing, with which human-robot hand-overs are compared to human-human hand-overs.

## I. INTRODUCTION

Handing over objects is a crucial component of many human-robot interaction tasks. Seamless and effortless hand-overs are desired in intensive human-robot interactions such as in patient caring. In human-human hand-overs, both the giver and receivers play active roles by observing each other and potentially rendering expectations of when and where the object will be transferred. Particularly for a robot receiver, which is less agile than a human, it is critical to predict the end position and timing of the human giver’s hand motion and plan its responding motion ahead of time accordingly. Using motion prediction, a robot receiver can reach out to meet a human giver instead of waiting passively for the object to be presented. Such an active robot receiver may be more efficient in hand-over tasks, and behave more naturally to the human giver.

Previous research has studied online estimation of timing and object transfer position for human-robot hand-over tasks. Huber *et al* used a Kalman filter for online prediction of the duration of consecutive working steps in an assembly task. This Kalman filter estimated the time to pass the next assembly part based on an linear dependency between complexity and duration of an assembly step [3]. Furthermore, Glasauer *et al* proposed a method to determine the hand-over position, by combining an estimate based on a minimum-jerk trajectory and an naïve estimate approximately centered between the giver and receiver [2]. The procedures in a typical human (giver) - robot (receiver) task have been described in [4]. As shown in Fig. 1, the prediction of end timing and position of the human giver’s hand happens when a robot reaches out to take

the object. With an accurate prediction of timing and position, a robot receiver can bring its hand to meet the human giver’s hand at the right time and place. Compared to a scenario with a passive robot receiver, the human giver has more control on the hand-over position and timing, while the robot receiver is more adaptive and its behavior is more human-like behaviors.

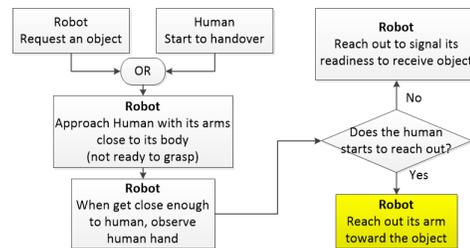


Fig. 1: Procedures of a typical human (giver)-robot (receiver) hand-over.

To generalize this hand-over prediction problem, this paper aims at predicting the end time and position of human reaching motion. We collected data of point-to-point reaching motions in a three-dimensional workspace from 10 healthy subjects, to test and compare the predictions of arm motion models (e.g., a minimum jerk model [1]) and various trained regression models (e.g., history-based linear regression, global feature-based linear regression, etc.).

## II. METHODOLOGY

This section presents our methodology for studying the end time and position of reaching motions, introducing the candidate prediction models and describing the experiment conducted to collect reaching motion data. The candidate prediction models include (1) the minimum jerk model, which renders the hand trajectory by minimizing the 3rd-order derivatives along the hand path, and (2) several linear regression models that can be trained to predict end time and position using the collected reaching motion data. These linear regression models have different input variables, but the same output variables, i.e., the predicted end time  $t_f$ , and end position  $x$  with x, y, and z components. The predictions of these models change as the reaching motion get closer to its end. For each regression model, a set of model coefficients are computed over the training data set and validated using another testing data set. With the reaching motion data we collected, the training and testing data set can be divided either by subject or the target of the reaching motion.

## A. Prediction Methods

1) *Minimum Jerk Model*: The minimum-jerk model is proposed to predict hand trajectory in task space given the initial/final times and positions of a hand motion. As the cost function shown in Eq. (1), it minimizes the time-integral of the squared 3rd-order derivatives of the two-dimensional hand coordinates (i.e.,  $x(t)$  and  $y(t)$ ), and predicts velocity profiles that are most similar to the experimental observations by the ratio of peak velocity to average velocity along hand trajectories.

$$C_n = \int_0^{t_f} dt \left[ \left( \frac{d^3 x(t)}{dt^3} \right)^2 + \left( \frac{d^3 y(t)}{dt^3} \right)^2 \right] \quad (1)$$

The advantage of the minimum-jerk model is that it provides a simple analytical formula to predict hand position and velocity, which can be easily generalized from 2D to 3D movements. In Equations (2) and (2),  $\mathbf{x} = [x, y, z]^T$  denotes hand position in a 3D workspace. The hand position and velocity at time  $t_c$ , denoted by  $\mathbf{x}_c$  and  $\dot{\mathbf{x}}_c$  can be calculated given the initial and final hand position (denoted by  $\mathbf{x}_0$  and  $\mathbf{x}_f$ ), and the start and end times of the movement (denoted by  $t_0$  and  $t_f$ ).

$$\begin{aligned} \mathbf{x}_c &= \mathbf{x}_0 + (\mathbf{x}_f - \mathbf{x}_0) \cdot \left[ 10 \left( \frac{t}{d} \right)^3 - 15 \left( \frac{t}{d} \right)^4 + 6 \left( \frac{t}{d} \right)^5 \right] \\ \dot{\mathbf{x}}_c &= (\mathbf{x}_f - \mathbf{x}_0) \cdot \left[ 30 \left( \frac{t}{d} \right)^2 - 60 \left( \frac{t}{d} \right)^3 + 30 \left( \frac{t}{d} \right)^4 \right] \end{aligned} \quad (2)$$

where  $t = t_c - t_0$  and  $d = t_f - t_0$ . Inversely, the final time and position of hand movement  $\mathbf{x}_f$  and  $t_f$  can be found by resolving Eq. (2) analytically.

2) *History-based Linear Regression Model*: The history-based linear regression model predicts the current output of a time-varying variable based on the recent history of its dependent variables. As shown in Eq. (3), given a linear model of order  $n$ , the current predictions of final time and position, denoted by  $\mathbf{x}_f$  and  $t_f$  are linear combinations of most recent  $n$ -step hand positions.

$$\begin{aligned} \mathbf{x}_f(t_i) &= \mathbf{x}(t_{i-n})\theta_0 + \mathbf{x}(t_{i-(n-1)})\theta_1 + \cdots + \mathbf{x}(t_{i-1})\theta_{i-1} \\ t_f(t_i) &= \mathbf{x}(t_{i-n})\phi_0 + \mathbf{x}(t_{i-(n-1)})\phi_1 + \cdots + \mathbf{x}(t_{i-1})\phi_{i-1} \end{aligned} \quad (3)$$

3) *Smoothed Derivatives Linear Regression Model*: As shown in Eq. (4), the smoothed high-order-derivatives linear regression model predicts  $\mathbf{x}_f$  and  $t_f$  as a linear combination of the most recent derivatives of hand position. Tentatively, we use the first to fifth-order of derivatives, and each higher order of derivative is computed using a smoothed lower derivative, recursively. To be specific, the 3rd-order derivative (i.e. the acceleration) is computed from the smoothed velocity, which is further based on the smoothed position. The smoothing window has a size of five steps. As a result, a higher-order derivative represents more global features.

$$\begin{aligned} \mathbf{x}_f(t_i) &= \mathbf{x}(t_{i-1}) + \sum_{k=1}^5 \theta_k \frac{d^k \mathbf{x}(t_{i-1})}{dt} \\ t_f(t_i) &= \mathbf{x}(t_{i-1}) + \sum_{k=1}^5 \phi_k \frac{d^k \mathbf{x}(t_{i-1})}{dt} \end{aligned} \quad (4)$$

4) *Global Feature-based Linear Regression Model*: The global feature-based linear regression model predicts  $\mathbf{x}_f$  and  $t_f$  based on global features such as the average hand position, the average velocity, etc.. These global features are computed using all existing hand position history. As shown in Eq. (5), we tentatively renders the predictions as linear combinations of the average hand position  $\bar{\mathbf{x}}$  as well as average first- to fifth-order derivatives, denoted by  $\frac{d^k \mathbf{x}(t_{i-1})}{dt}$  with  $k = 1 \cdots 5$ .

$$\begin{aligned} \mathbf{x}_f(t_i) &= \bar{\mathbf{x}}(t_{i-1}) + \sum_{k=1}^5 \theta_k \frac{d^k \mathbf{x}(t_{i-1})}{dt} \\ t_f(t_i) &= \bar{\mathbf{x}}(t_{i-1}) + \sum_{k=1}^5 \phi_k \frac{d^k \mathbf{x}(t_{i-1})}{dt} \end{aligned} \quad (5)$$

5) *Naïve KNN Method*: Using Naïve K-nearest-neighbor (KNN) method, we compare the global features of a reaching motion at/until the current time step (the current ‘‘motion status’’) to the motion statuses of our recorded reaching motions. Note that we can only compare the motion statuses at which the hand has traveled for the same duration. Using such comparisons, we can find K reaching motion trials that are most similar to the motion to be predicted, and predict its end time and position as the mean of those of the known trials. Same global features in Section II-A4 can also be used here. This naïve KNN method is expected provide better predictions as its reaching motion database get bigger.

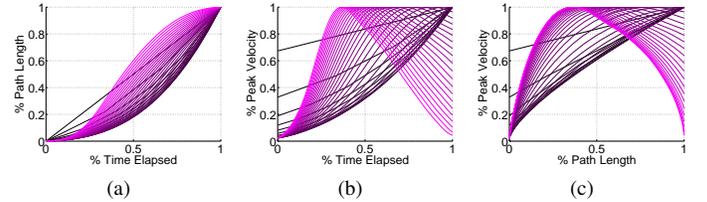


Fig. 2: As the hand moves closer to its end, a reaching motion can be normalized with respect to its traveled time, path length and peak velocity.

6) *KNN Method with Normalized Motion Features*: Instead of directly comparing motion status as in Section II-A5, for this method we first normalize the hand position and velocity profiles of a reaching motion with respect to the percentages of path length traveled and time elapsed. As shown in Fig. 2a, at each time step, the percentage of path length traveled and the percentage of time elapsed are computed for an existing history and are plotted against each other. The darker lines are the plots at the earlier stage of the motion, while the bright magenta lines are the plots when the motion is close to its end. Similarly, Fig. 2b and 2c normalize the velocity profiles with respect to the peak velocity of the motion and

then plot it against the percentage-wise path length and time, respectively. Such normalized plots represent the characteristics of a reaching motion at its different percentages regardless of speed and magnitude of the motion. These plots can be approximated by high-order polynomial fitting, which results in characteristic coefficients that represent the motion status at different motion stages. To predict the final position and total duration of a motion, we first compute this characteristic coefficients for the motion to be predict with its existing history. These coefficients will be compared to the coefficients of known motion, to find out with respect to the total distance and time to be traveled, how many percentages the hand has traveled. Given these percentages, and the distance traveled and time elapsed, we can computer how much further and longer the hand will keep traveling.

7) *Position Tracking - A Trick to Meet at the Same Time and Position:* Instead of predicting the end position and time of a reaching motion, the robot receiver's hand can continuously track and move toward the human giver's hand, which also guarantees that they will meet at the same time and position. In [4], the velocity of the robot hand is controlled to be proportional to the distance from the human's giving hand. In our study, this position tracking method can be considered as a baseline to evaluate our controller with motion prediction algorithms, since one important goal we want to achieve is handover efficiency.

### B. Data Collection

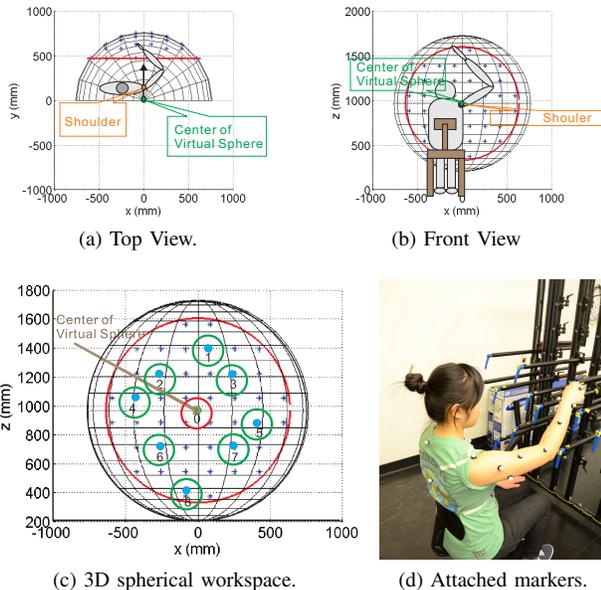


Fig. 3: (a) and (b) show the top and front views of the spherical workspace, respectively. (c) Eight targets are selected among all the available targets (denoted by blue dots in circles). (d) A subject is performing the instructed reaching movements, with markers attached to her right arm and torso.

To study the end position and timing of human reaching motions, we conducts experiments to collect data on point-to-point reaching motion. In this experiment, ten healthy subjects

(six males and four females) were instructed to conduct reaching movements with their right arms to each of the eight targets specified in the spherical workspace (Fig. 3). Each subject performs eight reaching motion sessions. Motions of the same session has the same end target, starting from one of the rest seven targets. Thus, a complete session consists of five repetitions of seven different motions. The total number of trials for each subject was  $8 \times 7 \times 5 = 280$ . During the experiment, a subject sit in a chair with a straight back support. The chair was placed such that the subject can point at the targets with comfort and with his/her elbow naturally flexed. The height of the workspace center was adjustable and was always aligned with the right shoulder of the subject. The subject's right arm was free for reaching movements, but the body of the subject was set against the chair back to minimize shoulder displacement. During the reaching movements, subjects kept the pointing fingers in line with the forearm to minimize wrist flexion.

Subjects were asked to point with the index finger tip at their comfortable paces. At the beginning of each trial, the subject was informed of the start and end targets of the trial. After receiving a "start" command, the subject moved his/her index finger from the start target to the end target. A motion capture system recorded a single file for each trial at sampling rate of 100 Hz. As shown in Fig. 3d, passive reflective markers were attached to the subject's torso and right arm. Each recording started from the time when the subject points the index finger to the start target and ended after the index finger tip becomes steady at the end target. To minimize the effect of fatigue, subjects took a rest after completing each session. Further analysis are based on the recorded shoulder, elbow and wrist positions.

### III. PRELIMINARY RESULTS

This section presents the preliminary results of the end position and time prediction performance of some of the models we proposed. Here we compare the prediction from models based on (1) jerk minimization, and (2) recent history. Due to limited space, we only present the prediction for the x-component of a 3D reaching motion, since the prediction algorithms can be applied similarly to the y- and z-components. For each prediction model, we present the representative prediction results for a single reaching trial. Among the ten subjects, we randomly select one for model training and another one for model testing, each providing 280 reaching motion trials. For each reaching motion trial, the end position prediction performance is measured by the ratio of prediction error to the motion range of that trial. With all the trials from a testing subject, we plot distribution of the prediction error ratio every 10% of the motion to show how the prediction error ratio changes as the motion develops.

Fig. 4 shows the prediction performance using minimum jerk model. The analytical formula for  $x_f$  and  $t_f$  can be derived such that there is no need for prediction model training. We randomly select a subject to test the model prediction performance against all the subject's reaching motion trials.

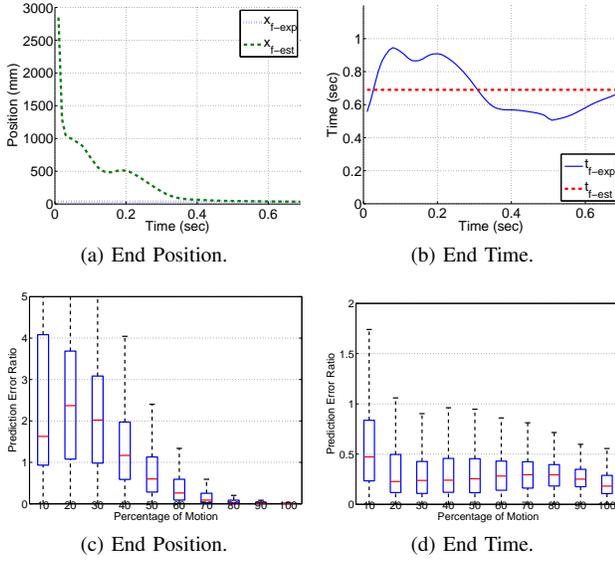


Fig. 4: End time and position (x-component) using minimum jerk model. (a) and (b) are predictions for a single reaching motion. (c) and (d) are the distributions of prediction error ratio.

As shown in Fig. 4c the prediction error ratio of end position is reduced to less than 10% after two thirds of the motion has been observed, while the prediction error ratio for end time is mostly reduced to less 50% of the total motion time after 20% of the motion.

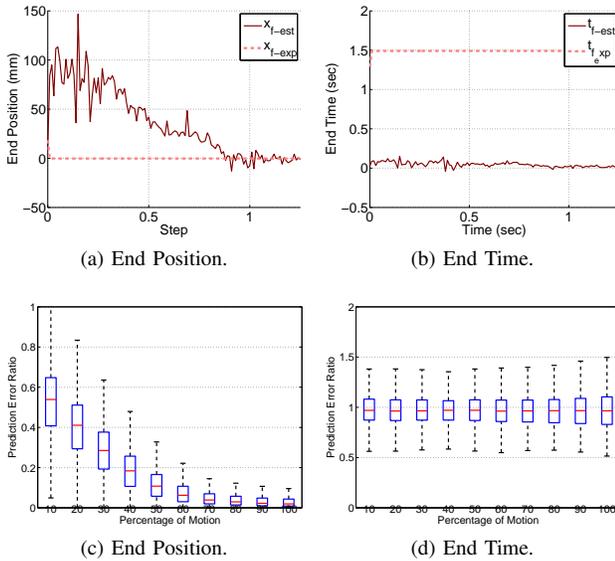


Fig. 5: End time and position (x-component) using history-based linear regression model. (a) and (b) are predictions for a single reaching motion. (c) and (d) are the distributions of prediction error ratio.

Fig. 4 shows the prediction performance using history-based linear regression model. The model coefficients are trained with one subject and tested against all reaching motion trials of another. Note that both the training and testing subjects are

randomly selected. Comparing Fig. 5 and 4d with Fig. 5c and 5d, the prediction error ratio of end position is reduced at a much earlier stage of the motion than when using the minimum jerk model, while the prediction error ratio for end time is much worse than minimum jerk model performance, all the way to the end of the motion.

#### IV. FUTURE WORK

In future work, we will thoroughly compare our proposed timing and position prediction methods to find out which one can provide the most accurate prediction at the earliest stage of a motion. The winning prediction algorithm will be implemented on a humanoid robot with a torso and two arms, which is the major component of tele-nursing robotic system. This tele-nursing robotic system is developed for attending patients with highly infectious diseases (e.g. Ebola). Fig. 6 shows the sketch of the tele-nursing robotic system, and its in-lab setup at its current system integration stage. The nursing robot, nicked-named Ebolabot, is remotely controlled via an operator console with a variety of input devices. It also serves as a research platform for developing operator assistance modules that fully or partially automate tedious and error-prone tasks, and reduce user training time. The proposed hand-over prediction algorithm will help the Ebolabot to receive objects more efficiently from patients and human coworkers.



Fig. 6: (a) Sketch of Ebolabot system - a tele-nursing robot system for caring patients with highly infectious diseases. The robot (left) consists of a humanoid torso, teleconference screen, and mobile base. The operator console (right) consists of displays for, from left to right, the 3D map, camera feedback, and GUI; two 6-DOF haptic input devices; and standard mouse and keyboard. (b) In its current state, the Ebolabot system has its major hardware components integrated, and is capable of basic teleoperation and motion planning.

#### REFERENCES

- [1] T. Flash and N. Hogan. The coordination of arm movements: an experimentally confirmed mathematical model. *J. of Neurophysiology*, 5:1688–1703, 1985.
- [2] S. Glasauer et al. Interacting in time and space: Investigating human-human and human-robot joint action. In *RO-MAN 2010*, pages 252–257, Viareggio, Italy, September 2010.
- [3] M. Huber et al. When to assist? - modelling human behaviour for hybrid assembly systems. In *ROBOTIK*, pages 1–6, Munich, Germany, June 2010.
- [4] K. W. Strabala et al. Towards seamless human-robot handovers. *Journal of Human Robot Interaction*, 2(1):1–23, 2013.