2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

A Novel Telerobotic Method for Human-in-the-Loop Assisted Grasping based on Intention Recognition*

Karan Khokar, Redwan Alqasemi, Sudeep Sarkar, Kyle Reed and Rajiv Dubey

***Abstract*—In this work, we present a methodology for enabling a robot to identify an object and grasp configuration of interest and assist the human teleoperating the robot, to grasp the object. The identification is carried out in real-time by detecting the motion intention of the human as they are teleoperating the remote robotic arm towards the object and the grasp configuration. Simultaneously, depending on the detected object and grasp configuration, the human user is assisted to translate and orient the remote arm gripper in order to preshape and grasp the object. The complete process occurs with the human teleoperating the arm, and without them having to interact with another interface. Motion intention recognition is carried out by using Hidden Markov Models (HMMs), trained offline by preshape trials performed by a skilled teleoperator. The environment is unstructured and comprises of a number of objects, each with multiple grasp configurations.**

Experimental tests on healthy human subjects have validated our intention recognition based assistance method. They show that the method allows objects to be grasped and placed 48% faster, and with much ease compared to unassisted teleoperation. Moreover, we have proved that the model for intention recognition, trained by a skilled teleoperator, can be used by novice users to efficiently execute a grasping task in teleoperation.

I. Introduction

Service robots have been increasingly researched recently for enabling persons with disabilities perform activities of daily living (ADL) tasks in unstructured indoor environments. They employ autonomous techniques for grasping objects. Autonomous grasping is slow and computationally intensive. The main problem with autonomous grasping is that of optimization in a high dimensional search space [1]. It was demonstrated, in [1], that human input reduces the time it takes for an autonomous planner to compute a robust grasp. The gripper was preshaped by human hand at a desired grasp configuration and the autonomous planner computed a locally optimized grasp. They demonstrated that a grasp is more likely to be successful if a human selects a grasp configuration. Human's cognitive abilities, and their knowledge gathered from experience are better than the most advanced robots. Also, in indoor tasks, involving a number of objects in cluttered unstructured environments, a human is needed in the loop for commanding the robot on the object to be grasped and its grasping configuration.

Figure 1: A subject teleoperating to grasp utensils from a dishwasher rack and lay over the table (partially visible) on the right

Teleoperation of a remote arm offers a direct way of involving human in the loop but it has been largely unused for service robot applications since teleoperation is a mentally and physically challenging activity [2, 3]. To make teleoperation easier, various computer mediated techniques such as virtual fixtures [4], potential fields [5] etc. have been developed. These techniques assist the operator by guiding the motion of the robot towards the target. In this manner, these techniques reduce the operator's workload and increase their task performance. In order to determine the appropriate trajectory and hence the direction of guidance, researchers have made use of motion intention recognition [6]. The tasks executed in [6] were simple linear trajectory following tasks.

There have been several works in which human has been involved in the robotic grasping process. [7], [8] show examples of learning by demonstration in which a human hand demonstrates a grasp and the robot creates a model for grasping objects, by learning. In [7], good grasp success rates on a large set of objects were obtained, by training on a relatively small set of objects. Human input has also enabled robotic vision systems to interpret a scene better by helping the system to segment and/or recognize the objects [9], [10]. Once the object was recognized, an autonomous planner would evaluate the grasp on the object and grasp would ensue. Better grasp success rates were obtained when a human was involved in the grasping process than when the grasp was executed autonomously. In [11] and [12], the

*Research supported by NSF, award number 0713560.

Karan Khokar is a doctoral candidate at University of South Florida, Tampa, Florida 33620 USA. (phone: 813-447-7703; e-mail: karan.khokar@gmail.com).

Redwan Alqasemi (Alqasemi@usf.edu) is the Research Professor at Center for Assistive, Rehabilitation and Robotics Technologies at the University of South Florida. Sudeep Sarkar (Sarkar@usf.edu) is a Professor in Department of Computer Science and Engineering at the University of South Florida. Kyle B. Reed (kylereed@usf.edu) is an Assistant Professor in the Department of Mechanical Engineering at the University of South Florida. Rajiv Dubey (Dubey@usf.edu) is a Professor and Chair of Department of Mechanical Engineering at the University of South Florida.

human points to the object of interest using a laser pointer and the robot autonomously completes the grasp. Only one grasp configuration and objects on flat surfaces were considered. In [13], a human was involved in the loop in scenarios when the robot failed to perform a grasp autonomously. In these cases, the human would interact with a virtual world, which was a replica of the real world, using a mouse. The human would locate and orient the virtual gripper at a desired grasp configuration over the virtual object of interest, by operating one degree of freedom (DOF) at a time. The robot in the real world would then complete the grasp autonomously. The main drawback of this system was that the user could operate only one DOF at a time, which would be time consuming.

In this work, we present a novel human-in-the-loop (HitL) grasping strategy in which the human's intention, from motion in teleoperation, is used to predict the desired object and grasp configuration. Based on the prediction, the human is assisted to preshape the remote arm gripper over the object and the grasp configuration, using scaled teleoperation [14]. The motion of the remote arm is amplified in directions that lead to the desired object and grasp configuration, and attenuated in directions that do not. Motion intention recognition is carried out by processing user motion data through an HMM based model. The model is trained by a skilled teleoperator since they have least errors and tend to follow the shortest translation and rotation path.

Our work is intended for use by wheelchair-bound individuals who are unable to reach out to surfaces in indoor environments for picking up objects, due to their disability. They need not be skilled teleoperators. The objective of our work is twofold. We would like to validate if our intention recognition algorithm is able to identify the object and grasp configuration of interest in a multi-object, multi-grasp configuration set-up. Secondly, we would like to verify whether assisted motion based on intention recognition from our model, trained by a skilled teleoperator, will improve the performance of novice users in executing a grasping task.

II. Multi Object and Grasp Pose Identification Problem, and HMM For Intention Recognition

In this section we explain how we use Hidden Markov Model (HMM) theory for recognizing motion in teleoperation, to identify an object and grasp configuration of interest. Consider a cluttered environment with objects of general shapes (refer Fig. 1), which can be grasped from different configurations (refer Fig. 5). The problem is to identify the object of interest and the grasp configuration, using motion data, as the user is teleoperating towards the object.

First, we classify objects into different classes based on their shape. We also preselect orthogonal grasp configurations for each object class. An HMM is associated with each object class and the states of that HMM are the various preselected grasp configurations for that shape. The parameters of each HMM are trained by having a skilled teleoperator repeatedly preshape the remote arm gripper, from random starting poses to various preselected grasp configurations, and by recording and analyzing their teleoperation data. More details about training and HMM parameter estimation will be mentioned in Sec. III. During a grasping task, as the user is teleoperating towards the grasp configuration of interest, the likelihood of occurrence of the HMM for each object is determined and the one with the highest likelihood is the object of interest. For the HMM associated with the selected object, the most likely state sequence gives the grasp configuration of interest.

HMM is a doubly embedded stochastic model [15]. Observations in an HMM are physical quantities that can be measured whereas the states are hidden. In this work, we are determining the mental state of the user, which is hidden. The mental state corresponds to the object and grasp configuration the user is interested in. Data from teleoperation are the observations. HMM provides a good structure to model a multi-object and multi-grasp configuration problem. It provides robustness by preventing fluctuations in intention recognition, which may occur due to uncertainty in human teleoperated motion. A simpler model, in which the object and grasp configuration of interest is the one on which the projections of the end-effector translation and orientation vectors are maximum, will be unstable due to errors in human motion. Moreover, we capture an expert's teleoperated motion in a model and this can be used later by novice or unskilled users to effectively perform a grasping task. This is not possible with the simpler model, just mentioned. We will see that our HMM fuses translation and orientation motion vectors into a single feature vector, which results in a quicker and more robust intention recognition.

By associating an HMM with each object class, objects can be added, removed or moved around in the environment without affecting the accuracy of intention recognition. Hence, our method is suitable for use in unstructured environments. Once the HMMs for various object classes have been trained by a skilled teleoperator, the model can be used by novice users without any need for retraining the model. A new object class needs to be trained only once.

III. HMM Feature Vectors and Parameter Estimation

In this section, we will describe the feature vectors of our HMMs, estimate its parameters and describe the training procedure. Let us say that cylindrical objects form one of our object classes and let it have N possible preshape configurations, one each at $C_1, C_2, \ldots C_N$, as shown in Fig. 2. Thus, the HMM for cylinder has N states.

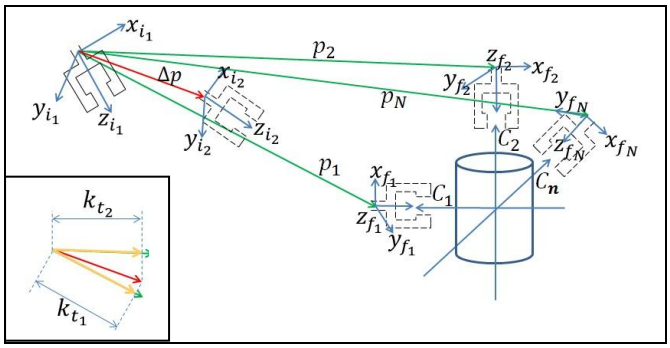

Figure 2: Preshape configurations, reference translational vectors and (inset) translational feature vector representation for a cylindrical object.

Let the remote arm gripper frame change from $(x_{i_1},y_{i_1},z_{i_1})$ to $(x_{i_2},y_{i_2},z_{i_2})$ as the user is intending to align the gripper frame with one of (x_f,y_f,z_f), at $C_1,C_2,\ldots C_N$. Let Δp be the incremental translational unit vector and Δr be the rotational vector. Let, p_1, $p_2,\ldots p_N$ be the reference unit vectors for translation, to each of the destination frames and $r_1, r_2,\ldots r_N$ be the ones for rotation. Let, $k_{t_1}, k_{t_2},\ldots k_{t_N}$ be the projection scalars of Δp on each of $p_1,p_2,\ldots p_N$, computed by taking a dot product. Similarly, let $k_{r_1},k_{r_2},\ldots k_{r_N}$ be the ones for rotation.

$$k_{t_1} = \Delta p p_1 \tag{1}$$

These projection scalars form the feature vector set, O, of the HMM i.e. $O = [k_{t_1}, k_{r_1}, k_{t_2}, k_{r_2}, \ldots, k_{t_N}, k_{r_N}]^T$. The length of O is $2N$X1 and it is generated at every time instant during teleoperation. The human, however, is interested in only one grasp configuration, out of the N possible ones, and is teleoperating the remote arm gripper towards that configuration. Now we demonstrate the estimation of HMM parameters from training data, which comprises of feature vectors.

We have observed that the distribution of the projection scalars on a desired grasp configuration, along the length of the trajectory, approximates to an exponential distribution. Thus, we have used exponential probability density function (PDF) to compute the observation probability for the HMM. Each state, j, has two exponential PDFs associated with it, one for translation and one for rotation. In order to develop the HMM by training it, we need to estimate the rate parameter, λ, of each exponential PDF for the object class. For this, we record the feature vectors that are generated as a skilled teleoperator repeatedly preshapes over a grasp configuration, starting each time from a random pose. The data from all trials for state j is concatenated.

Let, τ feature vectors, $O_1, O_2, \ldots O_\tau$ be recorded from all trials for one state j. Let K_t be a 1 x τ vector of translation projection scalars, such that,

$$K_t = [{k_{t_j}}^1, {k_{t_j}}^2, \ldots {k_{t_j}}^\tau] \tag{2}$$

Let I be a τx1 vector of ones, such that,

$$I = [1, 1, 1, \ldots 1]^T \tag{3}$$

The mean μ_{t_j} can be computed as,

$$\mu_{t_j} = \frac{1}{\tau} K_t I \tag{4}$$

Similarly, quantity μ_{r_j} can be computed for rotation projection scalars. The maximum likelihood estimate of the rate parameter for an exponential distribution is the inverse of the mean. Hence,

$$\lambda_{t_j} = \frac{1}{\mu_{t_j}} \text{ and } \lambda_{r_j} = \frac{1}{\mu_{r_j}} \tag{5}$$

In this way, the rate parameters associated with all N states can be computed for an HMM, and for HMMs of other object classes. In order to ensure that the distribution is probabilistic or the area under the distribution curve is unity, we evaluate co-efficient c_{t_j} as,

$$\int_{-1}^{1} c_{t_j} e^{(\lambda_{t_j}).(x)} dx = 1 \tag{6}$$

Limits of integration are the maximum and minimum values of the projection scalars on the reference vectors. Then, the observation probability for a state j, due to translation and rotation projection scalars k_{t_j} and k_{r_j}, is given by the product of p_{t_j} and p_{r_j},

$$p_{t_j} = c_{t_j} e^{\lambda_{t_j}.k_{t_j}} \text{ and } p_{r_j} = c_{r_j}.\, e^{\lambda_{r_j}.k_{r_j}} \tag{7}$$

If the user makes minor adjustments to the gripper by performing only rotation in the vicinity of a grasp configuration, the intention recognition may fluctuate between similar grasp configurations. So, we introduce a weight parameter, p_n, to consider object nearness. Its value is twice the normalized distance; normalization is with respect to all grasp points for all objects. The factor of two ensures probabilistic distribution. Another weight parameter, p_d, is introduced to take into account the direction of motion so that high probability value due to object nearness does not bias the intention recognition to the object in vicinity. This parameter is computed by adding 1 to projection scalar k_t. The total observation probability for a state at a certain time instant is then given by,

$$b_j(O) = p_{t_j} p_{r_j} p_n p_d \tag{8}$$

Thus, as the user is teleoperating towards a particular grasp configuration, $b_j(O)$ is computed for all grasp configurations. The state with the highest value of $b_j(O)$ at a time instant is the most likely state. Although this score gives a good measure of user intention, it does not take history of movements into account. For this, we use the forward-backward procedure in the form of output probability computation and Viterbi decoding [15]. We assume that once the user decides to grasp an object from a particular configuration, the user adheres to that configuration. So, our state transition matrix takes the form,

$$A_{ij} = 0.9,\ i = j$$
$$A_{ij} = 0.1/(N-1),\ i \neq j \tag{9}$$

As a result, minor deviations due to errors in teleoperation do not produce wrong intention detection. The algorithm registers an intention change only when the user makes repeated movements with the master device towards another object or grasp configuration. This makes the intention recognition robust. The accumulation of probability values in the forward algorithm computation adds to the robustness. We also assume that the user can start teleoperating towards any of the preshape configurations i.e. there is an equal likelihood. So, our initial state probability distribution has the form, $\pi = [1/N,\ 1/N\ \ldots 1/N]$. For determining state transition matrix and initial probability distribution, we have not used a standard technique such as Baum-Welch algorithm [15], due to the high cost associated with training a large number of datasets.

As the user teleoperates to a desired preshape configuration, output probability for each HMM is computed at every time instant. The one with the highest value is the most likely to occur. The object associated with that HMM is the desired object. To determine the preshape configuration

of interest, the most likely Viterbi state sequence, in a moving window of time instants, is computed for the HMM of the selected object. The mode value of this sequence gives the desired grasp configuration. For details on output probability computation and Viterbi algorithm, the reader is requested to refer to [15].

IV. Assistance using Scaled Teleoperation

Once the intention is detected, scaled teleoperation assistance in translation and orientation is provided to the user. Mathematical formulation for scaled rotation is presented next. That for scaled translation can be derived similarly.

Let R_i^O and R_f^O represent the initial and final orientations of the remote arm gripper frame with respect to the robot base frame, at each time instant as the arm is moving.

$$R_i^O = \begin{bmatrix} n_{x_i} & o_{x_i} & a_{x_i} \\ n_{y_i} & o_{y_i} & a_{y_i} \\ n_{z_i} & o_{z_i} & a_{z_i} \end{bmatrix} \text{ and } R_f^O = \begin{bmatrix} n_{x_f} & o_{x_f} & a_{x_f} \\ n_{y_f} & o_{y_f} & a_{y_f} \\ n_{z_f} & o_{z_f} & a_{z_f} \end{bmatrix} \quad (10)$$

The angular velocity of the gripper frame is given as [16]:

$$\omega_i = \frac{1}{2}\left(n_i \times n_f + o_i \times o_f + a_i \times a_f\right) \quad (11)$$

In the case of unassisted teleoperation, ω_i is determined from the master device and is used to move the remote arm. In scaled teleoperation, ω_i is modified as explained next. Let ω_d be the desired angular velocity for orienting the gripper from its current frame to the desired object frame (refer Fig. 3(a)), the latter being determined from the intention recognition algorithm. Let ω_{p1} be the unit vector in the direction of ω_d. Let, ω_{p2} and ω_{p3} be unit vectors perpendicular to ω_{p1} such that the three form a Cartesian triad. Let a_{p1}, a_{p2} and a_{p3} be the vectors generated as a result of the projection of ω_i onto ω_{p1}, ω_{p2} and ω_{p3} (refer Fig. 3(b) for two-dimensional (2D) representation).

$$a_{p1} = \omega_i \omega_{p1},\ a_{p2} = \omega_i \omega_{p2},\ a_{p3} = \omega_i \omega_{p3} \quad (12)$$

In order to implement scaled teleoperation in orientation, we need to scale up those components of gripper angular velocity which are along the direction of the desired angular velocity, and scale down those which are in the directions perpendicular to the desired. Let, a'_{p1}, a'_{p2} and a'_{p3} be vectors such that,

$$\begin{aligned} a'_{p1} &= s_{up}\, a_{p1} \\ a'_{p2} &= s_{down}\, a_{p2} \\ a'_{p3} &= s_{down}\, a_{p3} \end{aligned} \quad (13)$$

where s_{up} is a scalar of a relatively higher value than s_{down}. The difference depends on the amount of assistance to be provided to the user. The higher the difference the faster the user is able to align the gripper with the desired configuration. Finally, the angular velocity that needs to be sent to the arm control program to move the arm with scaled orientation is given by,

$$\omega_{new} = a'_{p1} + a'_{p2} + a'_{p3} \quad (14)$$

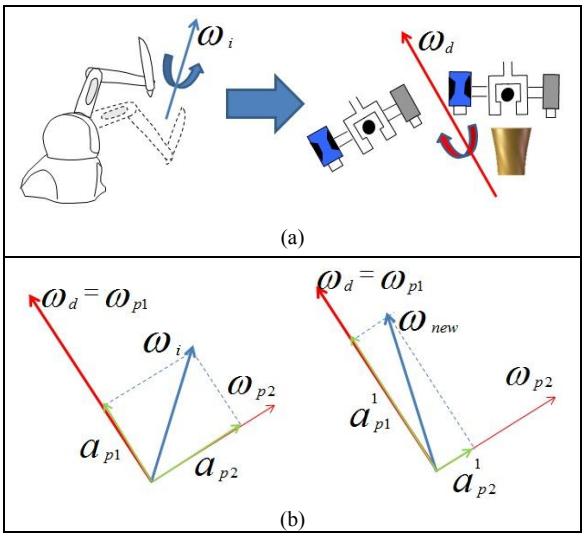

Figure 3: (a) Concept diagram (b) 2D Vector representation of scaled teleoperation in orientation. Scaled translation representation is similar.

V. Implementation

In this section we will describe the hardware and software implementation of our methodology.

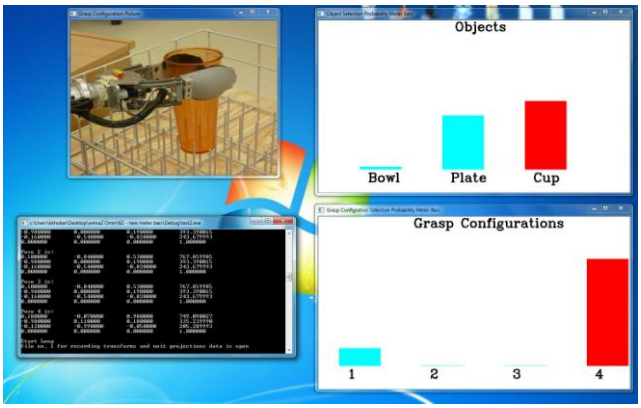

Figure 4: Image of predicted grasp configuration and probability meter bars

The teleoperation test-bed (refer Fig. 1) consists of an in-house developed 7 DOF wheelchair mounted robotic arm (WMRA) [17], teleoperated using a 6 DOF Phantom Omni device [18]. A parallel jaw gripper was mounted on the WMRA. The wheelchair was stationary throughout the experiments. The base frame for WMRA kinematics and for the poses of the objects was where the arm is mounted to the wheelchair. The workspace for grasping the objects is directly visible to the user. The amount of separation between the gripper paddles to grasp an object is controlled by the user. Keyboard keys are used to open and close the gripper. Joint velocity vectors from Omni space are mapped to the WMRA space and then singularity-robust inverse of the Jacobian is used to compute WMRA joint velocities. Optimization criteria based on weighted least norm of joint velocities, and joint limit avoidance was used for redundancy resolution of the WMRA. The Omni and the WMRA control loop ran at 500 HZ and 70~80 Hz respectively but this delay does not affect the task performance.

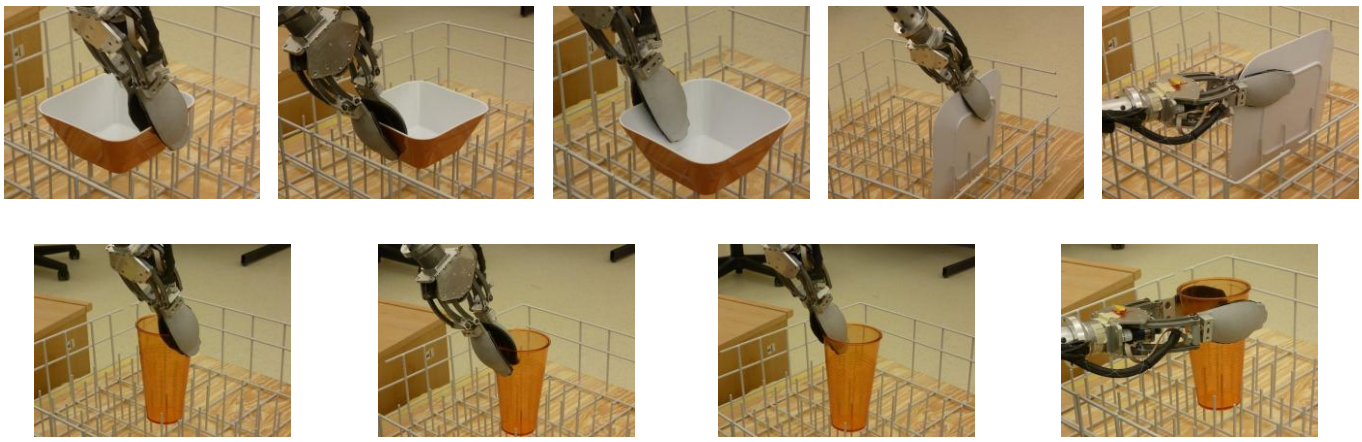

Figure 5: Objects and grasp configurations in our experiments. First row has 3 configurations for bowl (B1, B2 and B3) and 2 for plate (P1 and P2). Second row has 4 configurations for cup (C1, C2, C3 and C4)

An image of the predicted object and grasp configuration, and probability meter bars (refer Fig. 4) are shown on the computer screen in front of the user, as the intention is detected. These were used in the validation of our algorithm and as a feedback to the user. The top window on the right half of the screen shows the normalized output probability of each object's HMM. The tallest column represents the predicted object. The bottom window shows the normalized count of the states from the Viterbi sequence of the selected object. The tallest column in this window represents the predicted grasp configuration. All the objects and their grasp configurations, that we have considered, are shown in Fig. 5. Note the acronyms used to describe grasp configurations for each object in the Fig. 5 caption. The same images are the ones produced on the screen just described. In this proof-of-concept, we recorded the pose of each preshape or grasp configuration, before a task began, by teleoperating to each one of them and running forward kinematics. These were then fed to the intention recognition algorithm. In future, the robot will automatically determine the poses by using a depth based vision system, which we are in the process of developing and integrating. Our method, however, will still be needed to select the pose the user desires.

Since, we have three distinctly shaped objects, viz. bowl, plate and cup, we have three object classes and hence, three HMMs. The HMM for the bowl has three states since it has three grasp configurations. Similarly, the HMMs for the plate and the cup have two and four states respectively. We know that two exponential distributions, one for translation and one for rotation, are associated with each state and they are used for observation probability computation. For computing the rate parameters and coefficients for each exponential distribution, as explained in Sec. III, a skilled teleoperator repeatedly preshaped over each grasp configuration of each object ten times, starting from random remote arm end-effector poses each time. Ten trials were chosen since convergence of rate parameter values was obtained after 10 trials. In all, the skilled teleoperator performed 90 trials. Table I gives the values of rate parameters computed for the plate. Fig. 6 shows the histogram and the exponential distribution for rotation, obtained from training trials on the second grasp configuration of the plate. Values and distributions for the bowl and the cup were similarly obtained. It does not matter if a starting pose was not covered during training because the training data, consisting of projections of the incremental vectors on the reference vectors, is relative in nature. Incremental and reference vectors are measured with respect to the WMRA base frame.

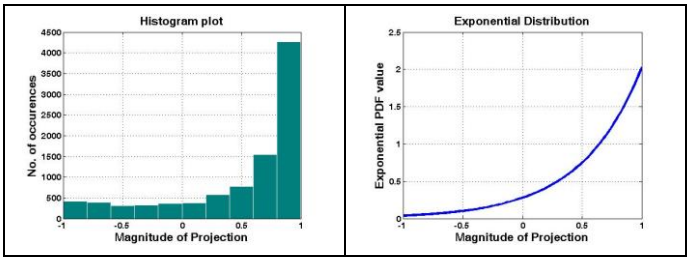

Figure 6: Histogram and distribution for rotation projection scalars from training for second grasp configuration of plate

Table I: Rate Parameters for Plate

Grasp Configuration	λ_{t_j}	λ_{r_j}
1	1.2459	2.5344
2	1.3488	1.9924

For implementing HMM, we have used logarithm of probabilities in order to avoid overflow of the double data type of C++, the programming language we used. The difference in the output probabilities of the detected object and all other objects had an upper bound of 10^{200}, without which it would take a very long time for the algorithm to register an intention change. The backtracking window for Viterbi algorithm was 50 Viterbi steps; higher values would slow down the intention change detection whereas lower values would cause more fluctuations in intention recognition. Mode value of the most likely state sequence was used as the detected grasp configuration. For assistance, the motion of the WMRA was amplified by two times in the desired direction and attenuated by five times in the undesired direction, for every motion input from the Omni. In the case of no assistance, no scaling was provided.

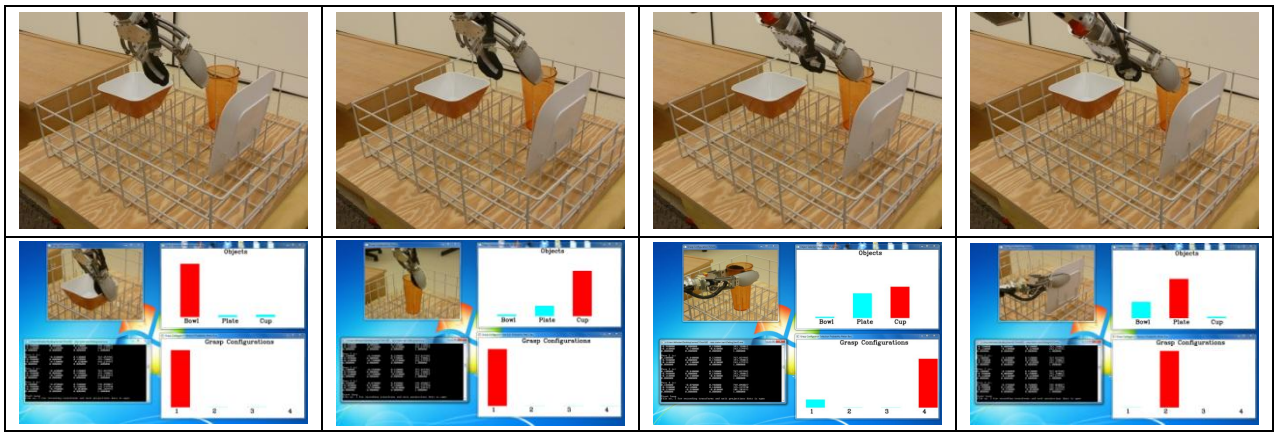

Figure 7: Results from validation of our motion intention recognition algorithm – Set 1

VI. Experiments

We carried out experiments to validate our intention recognition algorithm and also to validate our hypothesis, that intention based assisted method makes grasping easier and faster compared to the unassisted teleoperation method.

The experiments were conducted on five healthy human subjects who gave informed consent through an IRB approved protocol. The subjects were all males, aged 22 to 29 years. One of them was a skilled telerobot operator, having more than 5 years of experience. The other four subjects had never used a telerobotic system. Only the skilled teleoperator trained the HMMs. These trained models were used by all subjects, including the skilled teleoperator, while testing using our method.

To validate our hypothesis, the subjects executed pick-and-place tasks involving the three objects shown in Fig. 5, in the intention based assisted mode ('I' mode) and the unassisted teleoperation mode ('T' mode). The set-up used for the experiments is shown in Fig. 1. The objects are to be picked up from the dishwasher rack and placed on a table next to it. In the 'I' mode, the intention recognition algorithm detects the object and the grasp configuration of interest and assists the subject to preshape and grasp it. Assistance is also provided in the place phase but no intention is detected in this phase, since each object can be placed in only one way. In the 'T' mode, there is no intention recognition and no assistance. Each subject executed a pick-and-place task four times in each mode. This way, all distinct grasp configurations would be accounted for twice. A 'T' mode would be followed by an 'I' mode. Each T-I pair is one set and so each subject executed four sets. The order of grasp configurations and pose of objects was changed after each set in order to (i) eliminate bias in results from any learning effects and (ii) to confirm if the accuracy of intention recognition remains the same when the pose of the objects is changed. In all, each subject performed 24 pick-and-place trials.

As the subjects executed the pick-and-place tasks, we recorded the time for grasping each object, the total time for completing each pick-and-place task and the number of Omni stylus button clicks. The number of clicks gave us an idea of the amount of master device movements and hence the amount of effort expended. Since the workspace of the Omni is smaller than WMRA, users would click and drag Omni stylus to teleoperate until the workspace limit was reached, and then reposition the stylus to continue teleoperating. One Omni click was equivalent to one unit movement.

After completing the tests, the subjects were asked to rate on a scale of 1 (easy) to 10 (difficult) as to how easy it was to execute the task in the two modes. They also were asked to provide a self-rating of various factors, from NASA-TLX assessment [19], that contribute to operator workload. We replaced the factor 'effort' with 'fatigue' because fatigue is more relevant to this study and it is not a part of the NASA-TLX assessment. An Analysis of Variance (ANOVA) was carried out to statistically compare the two modes.

VII. Results and Discussion

First, we present the results of our motion intention recognition algorithm. The images in the top row of Fig. 7 are snapshots from a single run of teleoperation, performed by a subject. Those in the second row are the corresponding screen output described in Sec. V. Initially, the gripper was heading to B1 and the algorithm predicted the grasp configuration correctly. A slight movement forward predicted C1, which has grasp configuration similar to B1. This shows how probability due to translation p_{t_j} in (8) helps in determining the intention. The subject then changed his intention and desired to grasp the plate from P2. As the subject started teleoperating by pitching the gripper, the algorithm predicted C4, which has a configuration similar to P2. This is due to probability due to rotation p_{r_j} and the weight factor due to nearness p_n in (8), since the gripper is closer to the cup. A slight translation and rotation towards P2 then made the algorithm predict the right intention. Although one may argue that C4 was a wrong prediction, as mentioned previously in Sec. III, p_n makes the intention prediction more robust by safeguarding against fluctuations.

The images in the two rows of Fig. 8 are snapshots from another run of teleoperation performed by the same subject, as that in Fig. 7. The subject desired to grasp the bowl from B2 and began from a point close to the cup. The algorithm predicted C2 from the initial rotational movement.

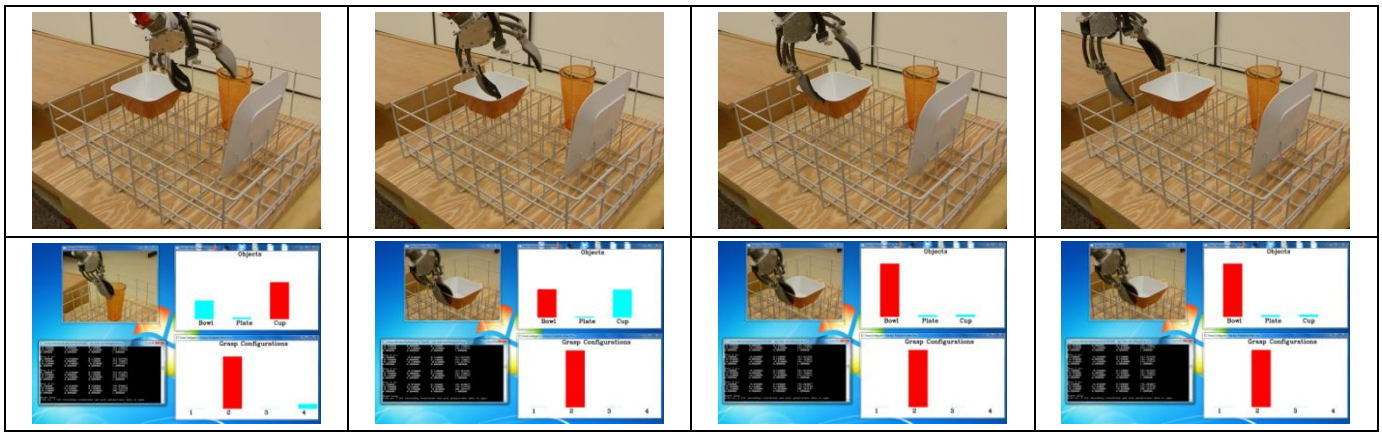

Figure 8: Results from validation of our motion intention recognition algorithm – Set 2

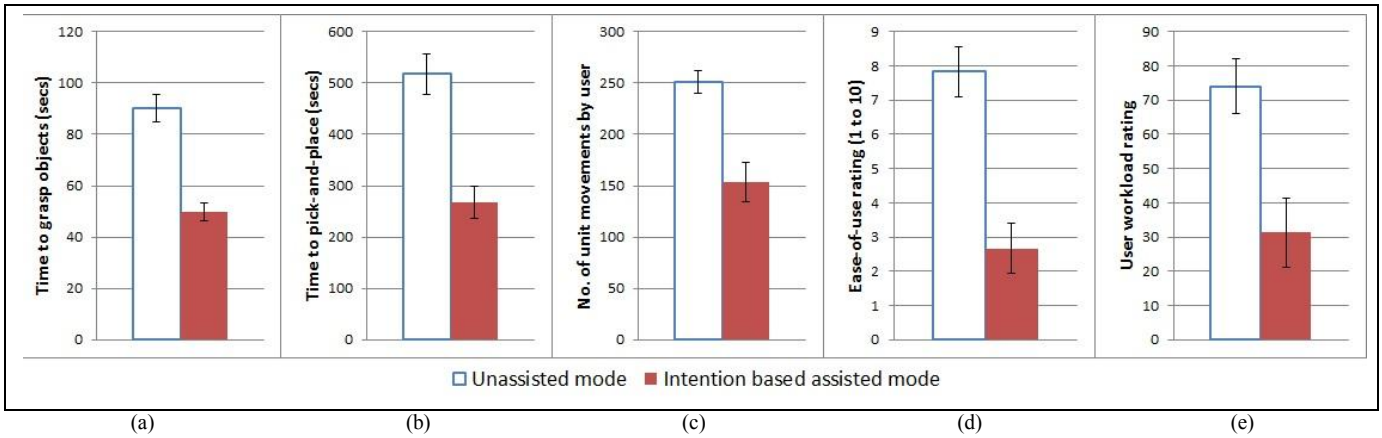

Figure 9: Quantitative and qualitative results of the user performing grasping and pick-and-place task in the unassisted and intention based assisted mode

C2 has a configuration similar to B2 and the gripper was closer to the cup. Then, as the subject rotated and translated the gripper, the correct intention was detected. As the gripper was continuously teleoperated towards B2, correct intention was detected throughout the course of the trajectory. This demonstrates the stability of our intention detection. For the last portion of this trajectory, there was no rotational and only translational input from the subject. In this portion, the subject often deviated from the trajectory and was unable to keep orientation constant, all the time. A simpler algorithm, that detects the object and grasp configuration based on the maximum projections of incremental translational and rotational vectors on the reference, gives B1, B2 or B3 as the detected intention. It fluctuates between the three whereas our algorithm is robust to these fluctuations. The attached video demonstrates these results. In the video, the user teleoperates from B2 to C4, and the order of the objects and grasp configurations in the meter bar windows is the same as that shown in Fig. 4.

We now present quantitative and qualitative results to validate our hypothesis. The plots in Fig. 9 give a comparison of executing the grasping and pick-and-place task in the two modes. The error bars in Fig. 9 are the standard errors. Fig. 9(a) and 9(b) compare the average time, over all trials and all subjects, to complete the grasp and to complete the pick-and-place task respectively, in the two modes. Fig. 9(c) gives the average number of unit movements of the master device.

From the plots of Fig. 9(a), 9(b) and 9(c), we observe that it took the subjects less time and fewer movements to grasp the objects and complete the pick-and-place task in the 'I' mode than in the 'T' mode. On average, the subjects were faster with intention based assistance: 44.62% for the grasping task and 48.43% for the complete pick-and-place task. The percentage saving in the number of unit hand movements of the user was 39%. Based on ANOVA, savings in time and human efforts were found to be statistically significant at 95% confidence level ($p < 0.001$).

Fig. 9(d) presents the average ease-of-use ratings, as provided by each user at the end of their tests. This difference in the rating value in the two modes was found to be statistically significant at 95% confidence level ($p < 0.001$), based on ANOVA. Fig. 9(e) compares the overall weighted workload score between the two modes for all subjects. We can infer that the subjects found it easier to grasp and complete the pick-and-place task, and experienced less workload using our intention based assisted method. The feedback from one of the subjects, when executing the task using 'I' mode, was that he did not expect the gripper to preshape quickly because he thought his input movements were not perfect. He was expecting to make several more moves to preshape. According to him, the unassisted method was tougher and he had to put more thought into movements using the unassisted method. Another subject commented that performing the task using the unassisted method involved a lot more thinking and there was a lot of stress in figuring out

angles of movement to pick up objects. He further added that with intention based assisted method, it seemed that the gripper performed most of the orientations on its own and it was very convenient to grasp the object.

It was observed that most users preferred either translating or rotating and not both at the same time. In that, they would translate along one global axis viz. x, y or z, or they would roll, pitch or yaw. On the other hand, a skilled teleoperator translates and rotates simultaneously and moves along more than a single DOF. A skilled operator tries to execute a trajectory along the shortest path to preshape over the desired configuration. Our method recognizes the intention of novice users correctly and quickly, in spite of the disparity in their movement pattern and style, compared to the skilled user. This is one of the major benefits of our method.

VIII. Conclusion and Future Work

Our methodology of identifying the object and the grasp configuration of interest, from motion intention recognition, makes it much easier for a person to execute a grasping task in teleoperation. Our algorithm is able to determine the intention of the user early in the grasping task and, this combined with assistance, enables users to complete the task quickly. They are effortlessly able to align with their desired grasp configuration. The users preferred using our method as they experienced less mental load and less overall workload in executing the task. The pick-and-place task was almost twice as quick.

As a first step, we plan to integrate a depth based vision system, which will help in automatic object identification and estimation of the preshape configurations. This will obviate us from making manual measurements for preshape configurations and it will make our method widely usable. One may argue as to why our method is needed when an autonomous planner can grasp the object, after the preshape configurations are known. A human, however, will be needed to indicate the object and the grasp configuration. Our method is a means to achieve this end. We, however, plan to integrate semi-autonomous supervisory control in which the robot autonomously completes the grasp, once the intention is detected and is confirmed by the user. We plan to compare this method of grasping with our method, which involves complete teleoperation. We hypothesize that semi-autonomous method will make it easier for the user to complete the task. We would like to carry out tests on wheelchair-bound persons to determine if they will be interested in using such a methodology for grasping unreachable objects. We would like to test our model with more skilled and novice users. We plan to extend our system by using a mobile platform and providing visual feedback, so that the user can perform tasks remotely. We also plan to use haptic feedback to enable the user to sense when (s)he is teleoperating beyond the desired grasp pose, by providing a resisting force. This will enhance the user's experience.

Acknowledgment

The authors would like to thank all the members of the Center for Assistive, Rehabilitation and Robotics Technologies at the University of South Florida, USA.

References

[1] M. T. Ciocarlie and P. K. Allen, "Hand posture subspaces for dexterous robotic grasping," *The International Journal of Robotics Research,* vol. 28, pp. 851-867, 2009.

[2] B. P. DeJong, E. L. Faulring, J. E. Colgate, M. A. Peshkin, H. Kang, Y. S. Park and T. F. Ewing, "Lessons learned from a novel teleoperation testbed," *Industrial Robot: An International Journal,* vol. 33, pp. 187-193, 2006.

[3] B. DeJong, J. Colgate and M. Peshkin, "Mental transformations in human-robot interaction," in *Mixed Reality and Human-Robot Interaction*Anonymous Springer, 2011, pp. 35-51.

[4] L. Joly and C. Andriot. Motion constraints to a force reflecting telerobot through real-time simulation of a virtual mechanism. Presented at IEEE International Conference on Robotics and Automation. 1995, .

[5] P. Aigner and B. McCarragher. Human integration into robot control utilizing potential fields. Presented at IEEE International Conference on Robotics and Automation. 1997, .

[6] D. Kragic, P. Marayong, M. Li, A. Okamura and G. Hager, "Human-Machine Collaborative Systems for Microsurgical Applications," *The International Journal of Robotics Research,* vol. 24, pp. 731-741, 2005.

[7] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, T. Asfour and S. Schaal, "Template-based learning of grasp selection," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on,* 2012, pp. 2379-2384.

[8] S. Ekvall and D. Kragic, "Grasp recognition for programming by demonstration," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on,* 2005, pp. 748-753.

[9] A. Sorokin, D. Berenson, S. S. Srinivasa and M. Hebert, "People helping robots helping people: Crowdsourcing for grasping novel objects," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on,* 2010, pp. 2117-2122.

[10] B. Pitzer, M. Styer, C. Bersch, C. DuHadway and J. Becker, "Towards perceptual shared autonomy for robotic mobile manipulation," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on,* 2011, pp. 6245-6251.

[11] A. Jain and C. C. Kemp, "EL-E: an assistive mobile manipulator that autonomously fetches objects from flat surfaces," *Autonomous Robots,* vol. 28, pp. 45-64, 2010.

[12] K. Khokar, K. B. Reed, R. Alqasemi and R. Dubey, "Laser-assisted telerobotic control for enhancing manipulation capabilities of persons with disabilities," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on,* 2010, pp. 5139-5144.

[13] A. E. Leeper, K. Hsiao, M. Ciocarlie, L. Takayama and D. Gossow, "Strategies for human-in-the-loop robotic grasping," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction,* 2012, pp. 1-8.

[14] N. Pernalete, W. Yu, R. Dubey and W. Moreno. Development of a robotic haptic interface to assist the performance of vocational tasks by people with disability. Presented at IEEE International Conference on Robotics and Automation. 2002, .

[15] L. Rabiner, "A Turorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE,* vol. 77, pp. 257-286, 1989.

[16] J. Luh, M. Walker and R. Paul, "Resolved-acceleration control of mechanical manipulators," *Automatic Control, IEEE Transactions on,* vol. 25, pp. 468-474, 1980.

[17] R. Alqasemi and R. Dubey, "Combined mobility and manipulation control of a newly developed 9-DOF wheelchair-mounted robotic arm system," in *Robotics and Automation, 2007 IEEE International Conference on,* 2007, pp. 4524-4529.

[18] Sensable Technologies. (). http://geomagic.com/en/products/phantom-omni/overview.

[19] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 2006, pp. 904-908.