Journal of Intelligent & Robotic Systems
https://doi.org/10.1007/s10846-019-01125-8

Inference of Manipulation Intent in Teleoperation for Robotic Assistance

Songpo Li[1] · Michael Bowman[2] · Hamed Nobarani[2] · Xiaoli Zhang[2]

Received: 2 January 2019 / Accepted: 21 November 2019

Abstract

In teleoperation, predicting an operator's intent and providing subsequent assistance have demonstrated great advantages in reducing an operator's workload and a task's difficulty as well as enhancing the task performance. Current research aims to tackle target-approaching intent, while our work focus on inferring manipulation (task) intent after the user grasps the object. We model how an object is grasped when being utilized in different manipulation tasks (intents) and then adopt this object grasping model in teleoperation for the intent inference. Our paper focuses on determining if direct interaction models can be used for indirect interaction. As the nature of one's grasping pose may satisfy multiple tasks (intents), we explore a form of classification modeling known as multi-label classification for multiple broad categories of tasks and objects. We also comprehensively compare classification techniques to determine the most suitable method for determining manipulation intent. With knowing the manipulation intent, future robot control algorithms can provide more helpful and appropriate assistance to facilitate task accomplishment.

Keywords Object manipulation · Human intent · Teleoperation · Robotic assistant

1 Introduction

Teleoperation with a robot as a medium has brought in many advantages to augment an operator's physical capability, including increase motion precision and strength, and remote access to the work field. Due to these benefits, teleoperation has been widely used in a variety of applications, such as remote surgery, space exploration, underwater operation, and manufacturing. However, manually teleoperating a robot to accomplish a task is often difficult and complex for a human user due to problems of 1) disembodiment of indirect physical interaction with the environment and 2) the physical discrepancy between the human hands and the remotely controlled robot hands. In the conventional teleoperation, users input their control commands through a joystick [1–3] to move or rotate the robot toward a specific direction while watching visual feedback from the work field. This indirect interaction cuts off the user's sense of feeling to the physical work including senses of 3-dimensional sight , hearing, touch and especially the vestibular and proprioceptive senses [4, 5]. Moreover, the physical discrepancy between input devices and robot platforms makes the situation worse. While manually teleoperating the robot, the user must mentally and physically transform the desired robot actions to the required input at the interface. Due to those difficulties, the user easily feels lost in the virtual feedback work field.

Increasing robots' intelligence and autonomy levels to allow them to generate (semi-)autonomous behaviors and assist in the achievement of a user's intent has demonstrated great potentials. The existing research demonstrated that the shared control between the human operator and the robot could accomplish the approaching task quickly and accurately [6, 7]. However, no work has been reported to study the more complex manipulation process. Successfully

✉ Xiaoli Zhang
xlzhang@mines.edu

Songpo Li
songpo.li@duke.edu

Michael Bowman
mibowman@mines.edu

Hamed Nobarani
hnobarani@mines.edu

[1] Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA

[2] Department of Mechanical Engineering, Colorado School of Mines, Golden, CO 80401, USA

Published online: 13 February 2020

manipulating an object requires much finer motion than only approaching an object, which requires fine coordination between the arm and fingers. Moreover, an object is generally associated with different tasks, where each task could require different constraints of the approaching angle, grasping parts, and applied force. Figure 1 shows the differences in grasping a power adaptor for plug-in and handover, that different grasping positions and object covering areas could be results of the different manipulation tasks. Considering the inherent difficulties of teleoperation, successfully moving and orientating the robotic arm to satisfy the fine motions are very difficult and results in unnecessary and tedious motion adjustments. Thus, for the consideration of the task performance and operator workload, there is a great need for building an intelligent robot agent that can assist in the manipulation tasks in teleoperation.

Understanding an operator's manipulation intent is the first step toward assisting in manipulation, which is more challenging than understanding the approaching intent. To the best knowledge of the authors, no work has been reported regarding the manipulation intent inference in teleoperation, and our paper will provide the theoretical backing for the future controller and formulations developed. The object manipulation occurs after the object approaching, but we want to infer the manipulation intent before the approaching completion to achieve the proper grasping with no or fewer grasping adjustments. This suggests the manipulation intent is hidden behind the observable motions and there is a large time delay between the approaching motion and manipulation intent being fully demonstrated. Moreover, the narrow bandwidth of information provided by the conventional interface limits the information which can be utilized for intent inference, and the operator could behave differently in the indirect teleoperation. These characteristics increase the difficulty in modeling operators' manipulation behaviors and intent inference.

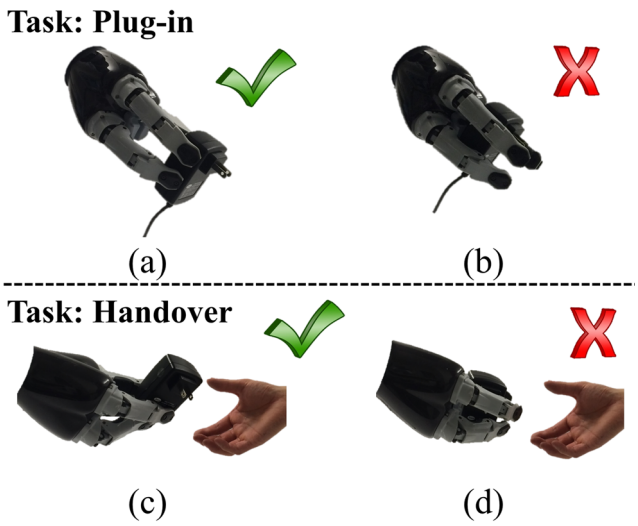


Fig. 1 **a** & **c** The robot considers the task requirement when grasping an object. **b** & **d** The robot grasps an object without task consideration

Moreover, the ambiguity between grasping poses and manipulation tasks could make the modeling more complex, which can be demonstrated in Fig. 2. An object is often associated with a variety of tasks, and there is always flexibility of handling the object for a task. For example, a human may grasp the handle of a cup when drinking from the cup or to transfer the cup to another location, and in the meantime, there is another variety of grasping poses which can be employed to accomplish the above two tasks. This unique characteristic has demonstrated the need to model the manipulation intent inference as a multi-label classification instead of treating it as a traditional single-label classification problem. Forcing it to be a single-label problem could easily confuse the classifier as the fact that one grasping instance belongs to more than one class during the classifier training(i.e. one grasp could satisfy multiple tasks at once). The wrongly inferred human intent will later direct the robot to improper assistive motions which will confuse the operator and increase the workload. Moreover, to ensure the single-label classifier achieves a reasonable performance, the operator is forced to behave in a specific way to generate behavior patterns that can be easily distinguished for minimizing the ambiguity. However, this forced behavior pattern is not intuitive for the operator and constrains the operator to carry tasks in one or several specific ways, which reduces the flexibility in handling a task and the practicability of the system in open unstructured environments. This multi-label formulation will also be beneficial to the later robotic assistance, that the robot could generate common grasping configurations to satisfy multiple tasks when none inferred task is dominant in its inference confidence.

To meet the great need in assisting an operator's manipulation intent, in this paper we focus on the inference of a user's manipulation intent in the remote or indirect teleoperation scenario. We modeled how a human operator grasped an object for primal tasks–using, transferring, and handing over the object–to achieve a fundamental understanding of correlation between grasping configurations and tasks, and the learned model was adopted to infer the potential manipulation intent(task) in the remote teleoperation. How to model the real, or direct, object interaction for teleoperation and how well this grasping model would perform in the indirect teleoperation scenario were open problems. Since this is the first attempt to model manipulation intent inferencing, a characterization of different modeling approaches is imperative in determining which overall approach is most applicable to handling this style of inferencing classification. Three modeling methods, Support Vector Machine (SVM), Neural Network (NN), and Bayesian Network (BN), were

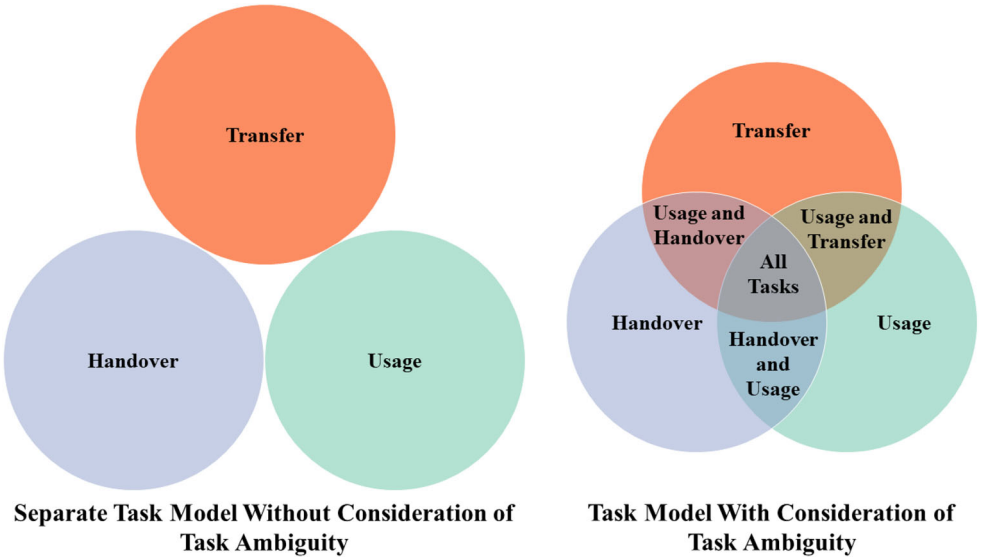


Fig. 2 Grasp modeling with and without consideration of task ambiguity. Even though there are grasping configurations which fit all tasks, they are not always preferred or applicable within the application of teleoperation with shared control. In shared teleoperation, the robot is expected to follow the human operator as much as possible, or the human operator could easily feel loss of control and result in negative attitude toward the system. This consideration constrains the potential grasping configurations and makes it impossible to always choose the one-fit-all grasping configuration as they are often greatly different from the operator's original motion inputs. This problem will be further discussed in our future work when investigating how the robot could generate action plans for assisting the manipulation intent

examined to thoroughly investigate their performance and applicability in this manipulation intent inference scenario.

The contribution of this paper can be summarized as:

1. This is the first time to consider a user's manipulation intent in teleoperation, and we formulate it as a multi-label classification problem to ensure the practicality.
2. The grasping model is adopted for the manipulation intent inference in indirect teleoperation. The effects of the uncertainties on the model caused by teleoperation is quantitatively evaluated by comparing the model's performance in real-object grasping and teleoperation.
3. The human gaze was introduced as extra information into the traditional grasp modeling to take advantage of the human eye-hand coordination for better correlating the grasping configurations with tasks and to improve the intent inference performance.
4. Three modeling methods' applicability were thoroughly examined in the tele-manipulation scenario to investigate each method's capability.
5. A comprehensive database consisting of ≈20,000 grasping samples is established and is available to share to allow more researchers to explore the manipulation intent modeling and inference.

With correctly inferred manipulation intent of a user, the robot can formulate more helpful assistance to facilitate complex object manipulation to ensure task success. Thus, teleoperation performance can be improved, and the user's workload can be reduced. The adoption of teleoperation especially in practical applications will be increased in the meantime. In the remainder of this paper, we will discuss the Related Work (Section 2) in the field of teleoperation and task inferencing; the approach of the formulation, and data collection in Methods (Section 3); the training, validation, and testing of the models as well as the posing of hypotheses in Experiments (Section 4); the specific results of the key components in the Results (Section 5); then we will discuss the hypotheses, considerations for multi-labeling and model suggestions in the Discussion (Section 6); and lastly summarize the work in the Conclusion (Section 7).

2 Related Work

In this section the review summarizes both the current robot assistance and human intent inference in teleoperation. The purposes are to point out 1) the lack of consideration of manipulation-related assistance in teleoperation and 2) the lack of corresponding technologies for the manipulation intent inference in teleoperation.

In teleoperation, great attention has been concentrated on the approaching process, which attempts to solve problems related to where is an operator's intended target/location and how to approach there precisely and quickly. To infer an operator's approaching intent, the approaching trajectory of the robot controlled by the operator is commonly used. When the motion trajectory is moving toward an object, the probability of this object being the target increases [8, 9]. With the known approaching target, robot

assistance in various formats has been investigated to assist the approaching process to be accurate and effective. In [10], virtual boundaries were generated to avoid potential damages to a beating heart in robotic surgeries while leaving a cone-shaped envelope to access the operation target. While in [11], force guidance was provided through a haptic joystick to pull the operator toward the target. Other researchers gave the robot agent more authority to allow them to directly contribute the motion of the robot platform. In [8, 9, 12], the robot agent shared the control of the robot platform with the operator, and their contributions were regulated by a leveraging weight. The idea is that robot agents take more responsibility when the task is certain and within the robot's capability. In [13, 14], the robot agent was given full autonomy to conduct the motion after the approaching target had been defined by the operator. In [12, 15], the robot assisted the operator to not only approach the target but also firmly grasp the target. However, the manipulation intent has not been considered, which means, whether the firm grasp of the target can satisfy the requirements of a specific task is uncertain.

Even though grasp modeling for autonomous robots has been studied in literature, and several have considered task-related grasp planning, how to use grasp modeling for the intent inference in the setting of teleoperation is still an open problem. Inference of manipulation intent in teleoperation is different from the task-dependent grasp planning, which makes it challenging to directly borrow the developed technologies of autonomous grasp planning. Firstly, the manipulation intent inference is an inverse process, where a human hand's grasping configuration is observable, and the manipulation intent of this grasping configuration needs to be inferred. In contrast, autonomous grasp planning for a given specific task is often to select a final grasping configuration from a candidate pool [16, 17]. The object affordance has often been used to evaluate whether the grasp can satisfy a task [18, 19], and the selection metrics varied from grasping stability [20, 21], skewness [22], and manipulability [23]. Secondly, the manipulation intent inference needs to be formulated as a multi-label classification problem when it is based on the grasping configuration and applied in practical teleoperation scenarios. How to properly formulate and how to deal with the ambiguity are open problems. Thirdly, in teleoperation, the operator uses his/her hand motions and gestures to steer a robotic arm and hand, where there is no physical contact between the human hand and the object but indirect interaction with a "virtual" object on the computer screen from the robot's perspective as shown in Fig. 3. Indirect, or virtual interaction with an object can result in different behavior patterns from the direct, or real object interaction. These differences introduce uncertainties into the intent inference, and how

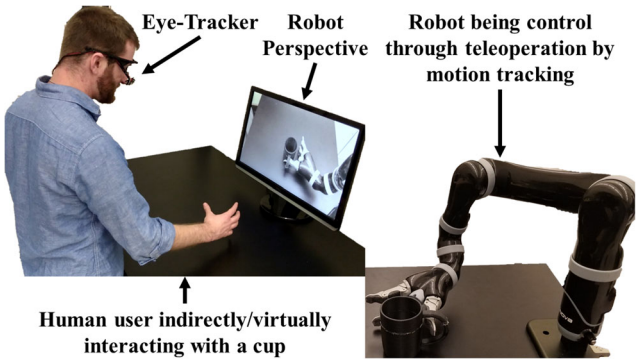


Fig. 3 The teleoperation interface where users interact with objects through the robot as a medium

well the inference performs demands further investigation. Without this investigation, robot assistance–let alone direct teleoperation–cannot provide relatively easy to use solutions as it puts too much of a burden on the operator. The burden comes with mapping their own hand to the robot end-effector for tedious fine-tuned manipulation to successfully and stably grasp the object in the robot environment (which the remote operator may not fully understand especially since they are not interacting with a real object) resulting in higher failure of attempted grasps. Thus, the robot needs to understand intent and regulate its own motion, ultimately making this system semi-autonomous.

3 Methods

The target application scenario can be demonstrated in Fig. 3 as a sample for providing assistances in daily living. A human operator cooperates with the motion planner of a robotic agent in tele-operating its physical arm (so called shared control) to perform object-manipulation tasks. A

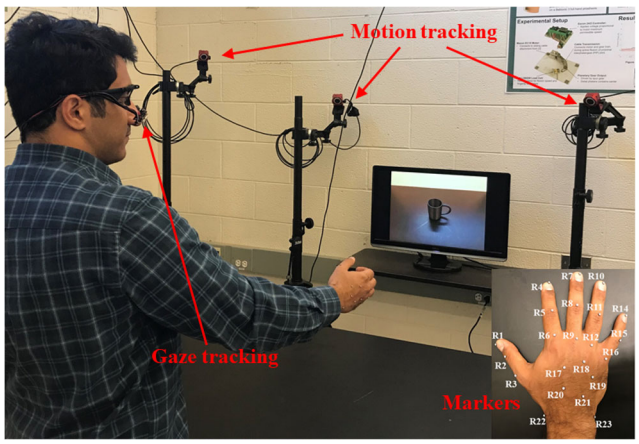


Fig. 4 The teleoperation interface with hand motion and gesture tracking and gaze tracking. Corresponding label locations on the hand are also shown

marker-based infrared motion tracking system tracks the operator's hand motions as shown in Fig. 4, and the hand motions are converted into control commends to control the robotic arm. In the meantime, the robotic agent observes the operator's eye-hand inputs to infer what manipulation the operator intents to perform with the object. Based on the inferred human manipulation intent, the robotic arm regulates its motion to approach the object from an appropriate angle and grasp it at appropriate locations, so that the robot can satisfy the constraints for accomplishing the specific task whiling avoiding the exhausting, fine position and angle adjustments. In this section we will discuss how the general teleoperation gesture control should be setup (Section 3.1 Hand Motion and Gesture for Robot Control), human grasp modeling including features definitions, processing the inputs, and building the model (Section 3.2 Grasp Modeling), and how to use multi-label classification for intent inferencing (Section 3.2.1 Intent Inference with Multi-Label Classification).

3.1 Hand Motion and Gesture for Robot Control

Instead of using a common joystick as the input device, an optical motion tracking system is used to track the operator's dexterous hand motion, as shown in Fig. 4, which is mapped to the robotic hand's motion. This new interface can improve the information bandwidth that transmits the dexterous hand motion and will be intuitive, which requires less training than the joystick. The optical tracking system consists of eight cameras, which track the position of small reflective markers that are attached to the user's hand. For a more consistent dataset across subjects, relative angles of each finger joint were calculated from the marker positions, and palm features such as palm center and palm direction were also determined. The mapping between a human hand and a robotic hand is not discussed in detail here, as the focus of this paper is on the grasp modeling and manipulation intent inference from the hand grasping configuration, however, a general representation of the teleoperation can be seen in Fig. 3. Moreover, an eye tracker is integrated to track where the user is looking on the virtual object on the computer screen. The hand configuration and the gaze information are combined to improve the performance of the modeling and inference, which takes advantage of the eye-hand coordination.

3.2 Grasp Modeling

How humans grasp an object for various tasks is modeled using data-driven methods, and the model will be employed in the teleoperation for the manipulation intent inference. How well the model performs needs to be investigated even for the common modeling methods because of the uncertainties presented in teleoperation interaction. Three common modeling methods, NN, SVM, and BN, are examined to investigate their applicability in the teleoperation scenario. These methods are selected based on their popularity and capability. BNs are helpful in understanding how the grasping pose is a result of the task and object through examining their dependencies. However, discretization of the continuous raw features into discrete features normally associates certain information loss. In contrast, SVM and NN can accept raw data without information loss but do not provide any insight into the grasping model and intent inference. These modeling methods are comprehensively analyzed to examine their applicability in this intent inference problem. Here, we only focus on the BN method to describe its data processing and modeling procedure.

The BN represents the conditional dependencies between variables through a Directed Acyclic Graph (DAG). Let I be the set of human intent of manipulating various objects, O be the set of object-related variables, and G be the set of hand kinematic variables that represents a grasping configuration. Thus, the BN grasping model represents the conditional probability distribution of I given O and G as $Prob(I \mid O, G)$. The intent inference is then the process that determines the most probable intent i^* that maximizes the marginal probability as Eq. 1.

$$i^* = \underset{i \in I}{\operatorname{argmax}}\, prob(i|o, g), o \in O, \text{and } g \in G \tag{1}$$

The Bayesian Network created by using Equation (1) is considered a discriminative Bayesian Network [24]. It is possible to adapt this to a generative network by using Bayes' rule which relates conditional probability to joint probability functions. By creating a Bayesian Network in this manner and by attempting to find the best posterior probability to determine manipulation intent, maximum-a-posteriori (MAP) estimation is used.

3.2.1 Feature Definition

Variables O and G are parameterized by a set of features. O consists of features of object identity (oi), dimensions (dim), orientation (ori), and object section (p). p indicates the part on the object the user is looking at when approaching and grasping the object for a task, and the division of an object into multiple sections is based on an object's geometry and affordance. G represents the object-centered hand kinematic features, which is collected when the operator has firmly grasped the object. G includes features of the palm orientation ($pdir$), palm center location ($pcen$), and fingers' configurations ($f_k, k \in [1, 3]$, corresponding to thumb, index, and middle fingers). Instead of using joints' positions to represent a finger's

configuration, the motion of a finger is represented by three rotation angles, the proximal phalange's rotations on and into the hand plane (fp_k) and intermediate phalange's rotation into the hand plane (pi_k). The small rotation of the distal phalange is not considered. This rotation representation is more generic to indicate a hand's absolute conditions rather than relative positions. The tracking marker was nominated as R_i,i[1,23], which also represented the 3D coordinates of the marker as shown in Fig. 4. Relative angles for each finger joint were obtained from vector calculus and inverse kinematics techniques using markers R1-R16. The palm center was obtained by taking the average across all 5 palm locations(R17-R21). The palm direction was computed by fitting a fixed frame to the hand where out of the palm is the positive z-axis and from the palm center towards the thumb(R3) was the positive y-axis. Thus, the x-axis was from the palm center towards the wrist.

3.2.2 Data Processing

It is difficult to integrate discrete and continuous variables into one BN model, especially for the high-dimension continuous variables like the grasping configuration. Thus, continuous variables are discretized with a self-organizing map (SOM) [25, 26] method. SOM is notated as a special type of an artificial neural network which is trained using an unsupervised learning method to produce a low-dimensional, discretized representation of the input space. After training, SOM maps the original input data, $X \in R^n$, onto a two-dimension map of discrete neurons, $P \in R^2$, that resembles the density of the original data. Then, the neurons are linearly indexed to create a one-dimensional representation, $Q \in R$, of the original data X, which indicates which neuron data $x \in X$ has been mapped to.

Redundant and irrelevant features may exist in the feature pool, which makes the grasping model bulky and may decrease the robustness of the modeling and inference. To reduce those unnecessary features, the HITON algorithm [27, 28] was used. This algorithm starts with the Markov Blanket of the target variable (intent I) to be classified, which is a set of variables that are most relevant to the target variable. Then, the unnecessary variables in the Markov Blanket are further removed through a greedy search process.

3.2.3 Model Building

The learning process of a BN model consists of structure learning and parameter learning. The structure of the model is critical for the effectiveness of the model, and data-driven and human-intuition strategies are employed to find the best BN structure separately. In the data-driven strategy, the maximum weight spanning tree algorithm [29] is used to find an oriented tree structure as the initial structure. With this initial structure, a greedy search algorithm [30] is then utilized to find the network structure in a neighborhood of graphs which maximizes the network score. While in the human-intuition strategy, the BN structure is defined with human intuition. In the parameter learning procedure, the joint distribution parameters will be learned using an Expectation-Maximization approach based on the BN structure.

3.3 Intent Inference with Multi-Label Classification

Task inferencing models will hold some degree of ambiguity. The ambiguity between tasks discussed is shown in Fig. 2, where it may be necessary for a single grasp configuration to satisfy multiple tasks. Further, it is natural that the same grasping configuration can be used for different manipulation tasks, the manipulation intent inference will be formulated as one multi-label classification problem to infer the most possible tasks when the operator attempts to grasp the object in a particular manner. This multi-label classification problem is tackled through problem transformation with a binary relevance method [31, 32]. It transforms the multi-label problem into a set of simpler binary classification problems, which can be handled using existing classifiers. Thus, multiple BN models, $BN^i, i \in I$, are built for each type of manipulation intent to classify whether a certain grasping configuration could be caused by this type of intent. BN^I is used to notate the set of BN models. In the testing of BN^I, each set of object-related features and grasping configuration features (o, g) will be labeled with a binary label vector $\boldsymbol{y}$, where $|\boldsymbol{y}| = |I| = L$ is the number of manipulation intent type in I. The element $\boldsymbol{y}_i, i \in I$ is one if the grasping configuration g can be used for the manipulation intent i; otherwise it is assigned as zero. Similarly, a set of binary classifiers, NN^I, and SVM^I, for each task will also be constructed for the NN and SVM methods separately. For example, one grasping configuration has been labeled as $\boldsymbol{y} = [1, 0, 1], |\boldsymbol{y}| = 3$, and it indicates that this grasping configuration can be used for the first and third task.

To evaluate the multi-label classification performance, the following accuracy α is defined (2), where XNOR is the logic operation whose function is the logical complement of the exclusive OR (XOR) operation. This accuracy measure is opposite to the definition of Hamming loss [33], which has been used in multi-label classification as a common performance measure. The new criterion is the fraction of the correctly inferred labels to the total number of labels. Let

$\boldsymbol{Y}$ indicate the correct labels of the testing samples and $\boldsymbol{Z}$ be the inference result, which are both a matrix of N by L. N is the number of the testing samples and is enumerated by j.

$$\alpha = \frac{1}{|I| \cdot N} \sum_{j \in N} \sum_{i \in I} xnor(\boldsymbol{Y}_{j,i}, \boldsymbol{Z}_{j,i}) \tag{2}$$

4 Experiments

In this section we will discuss the general experimental design, the direct fair comparisons across methods, and the hypotheses which will be tested.

To train the grasping model, a database collected from 20 subjects while interacting with four objects, a hammer, a cup, a power adaptor, and a dishwashing liquid bottle, for three different tasks of usage, handover to another person, and transfer to another location, was established. Although the database consisted of 20,000 samples, the data used is only a subset from two scenarios, the real-object interaction scenario–1,200 data points used where 960 are for training and 240 are for testing–and teleoperation interaction–where 480 data points were only used for testing. The subset of data shown in this paper is for one specific orientation for all objects for a more direct comparison; since across the 20,000 grasp samples, each object was placed in eight separate orientations for subjects to grasp 10 times. During the real-object interaction, each object was presented to the subjects, and they performed a given task with the object. The motion tracking system was tracking the motion of the subject's hand and the object. The part of the object the subject was looking at was recorded. Later, the collected data was converted into the object-related features and hand kinematic features, which are notated by o^r and g^r separately. Each grasping instant was associated with an intent label, i^r, which was the task the subject was performing. Ten repetitive trials were performed by each subject for each task and object to capture the grasping varieties for the same task. Sample photos while a user was interacting with the real objects are shown in Fig. 5. Moreover, sample photos that demonstrate the same grasping configuration could be associated with various tasks are shown in Fig. 6. For example, grasping the power adaptor using the grasping configuration in sample (a) is applicable of performing the task of usage and transfer, but it cannot be used for the task of handover as it does not leave sufficient open space for the receiver. In our modeling process, the grasping configurations which are not preferred for a task are labeled as inapplicable.

During the indirect or teleoperation interaction, the setup shown in Fig. 4 was used to simulate the teleoperation scenario. Virtual objects were displayed to the subjects in the robot's perspective, and they attempted to use the hand gesture to control the robot arm to grasp the object. The object-related features, o^v, and hand kinematic features, g^v, were recorded. Each grasping instant was labeled, $\boldsymbol{y}^v$, with all potential usability by surveying the tasks the grasping configuration could satisfy. Six repetitive trials were performed to avoid random uncertainties for each object.

The BN^I, NN^I, and SVM^I were trained using 80% of the data from the real-object scenario, where the data was randomly selected. The data used for training has a single task label, which can be used directly to train the binary classifiers. During the training, each model was carefully and iteratively tuned to achieve the best performance. For NN, its hidden layer number and neuron

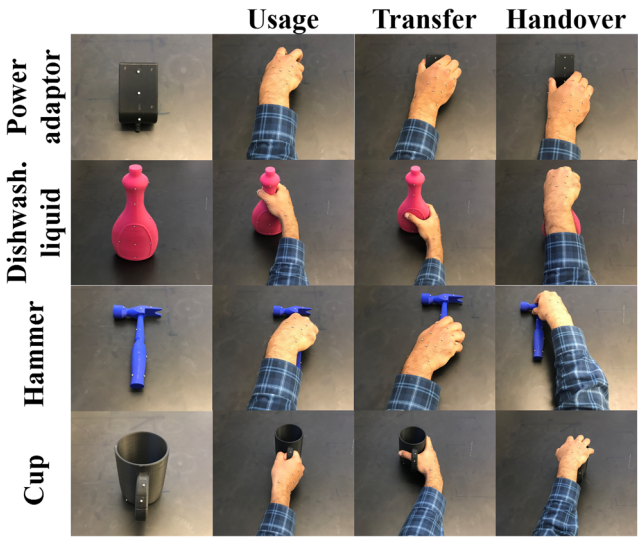


Fig. 5 Sample photos while a user is interacting with four objects for various tasks

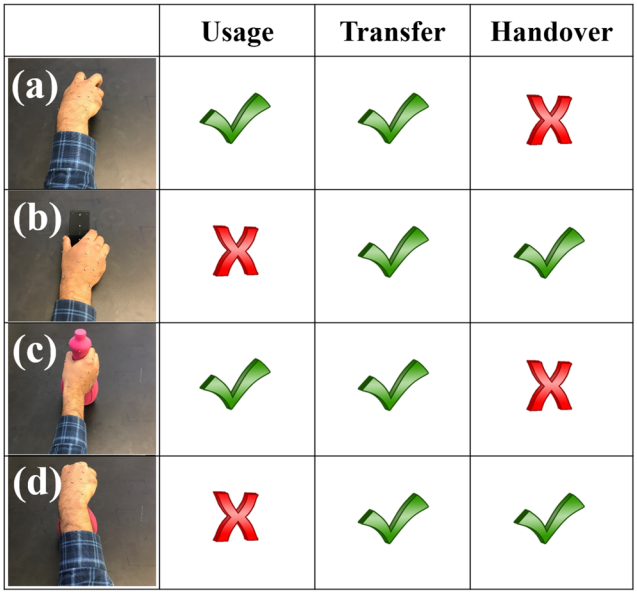


Fig. 6 Sample photos which demonstrate one grasping configuration satisfying various tasks

Table 1 Summary of models that achieve the best accuracy in various cases

		Best real-object scenario model			Best teleoperation scenario model		
		NN	SVM	BN	NN	SVM	BN
Without gaze	Real	**0.74**	**0.77**	**0.78**	0.67	0.68	0.70
	Virtual	0.67	0.56	0.73	**0.77**	**0.85**	**0.74**
With gaze	Real	**0.74**	**0.78**	**0.71**	0.67	0.78	0.63
	Virtual	0.64	0.62	0.78	**0.84**	**0.87**	**0.82**

The higher accuracy comparing the real-object case with the virtual-object case is in bold

amounts were tuned. The kernel function used for the SVM varied. For BN, its structure and SOM size were varied. After training, the grasping models were tested to perform the intent inference using the rest of the data from the real-object scenario and the data from the teleoperation scenario. The accuracy of the models in both scenarios were compared and analyzed. Additionally, the models that used and did not use gaze information were separately trained. The performances from two conditions were compared to evaluate the improvements that was brought in by gaze.

Through the experiments, the following hypotheses will be tested.

H1. It is feasible to adopt the real-object interaction model for the teleoperation interaction.
H2. The learned grasping model performs differently in the real-object and teleoperation scenarios.
H3. The model achieves the best accuracy in teleoperation scenario will be different from the model that achieves the best accuracy in real-object scenario.
H4. Gaze as an extra information resource is helpful to improve the inference accuracy in both scenarios.

5 Results

In this section we discuss the overall accuracy of the methods (Section 5.1 Accuracy), the overall tuning effort across different objects and tasks (Section 5.2 Tuning Statistics), a comparison across different BN structures (Section 5.2.1 BN Structures), and details about the available database of the data collected (Section 5.2.2 Database for Manipulation Intent Modeling).

5.1 Accuracy

Each trained grasping model was tested in both real-object and teleoperation scenarios, where the teleoperation scenario is the target application scenario. The best performance that each modeling method achieved in one scenario and the corresponding performance in the other scenario are summarized in Table 1. The top half of Table 1 are those when the gaze is not used as the extra information, and the bottom half are those when the gaze is used.

Overall, the three methods' best performances are at the similar level, and the SVM achieves the best accuracy, which was 0.85 and 0.87 without and with gaze as the extra information. The results show that when the grasp model functions well for the real-object scenario the model will not function well in the teleoperation scenario, and vice versa. Also, the gaze information is very helpful to improve the inference performance in the teleoperation scenario but does not show significant improvement to the real-object scenario.

The performance details of the models that achieved the best inference accuracy with gaze in teleoperation is displayed in Table 2. It details each model's object-specific and task-specific accuracy. The results show that the SVM and BN methods both suffer the polarized performance problem, where the methods function extremely well in some cases but very poorly in some other cases. For example, the BN method achieved a 100% accuracy in adaptor's transfer and handover inference but a poor accuracy in the adaptor's usage and cup's usage inference. In contrast, the performance of the NN method is medium, which does not suffer significant failure in any cases.

Table 2 Performance details of the best models in teleoperation with gaze

		Adaptor	Bottle	Hammer	Cup
NN (0.84)	U	0.85	0.88	0.92	0.69
	T	0.84	0.67	0.95	0.96
	H	0.95	0.62	0.73	1.00
SVM (0.87)	U	0.97	**0.48**	0.90	1.00
	T	1.00	0.92	0.98	1.00
	H	0.97	0.78	**0.46**	1.00
BN (0.82)	U	**0.29**	1.00	1.00	**0.12**
	T	1.00	1.00	0.95	0.85
	H	1.00	0.65	1.00	1.00

U: Usage, T: Transfer, H: Handover

The lowest accuracy of an object's intent inference is in bold

In the results, the accuracies of the three methods are not sufficiently high, and we believe two factors could have contributed to this, which all lead to great uncertainties in our collected data. 1) The differences between grasping configurations for various tasks are naturally difficult to capture. It is the nature that a person could grasp an object in the same way when performing various tasks with that object. On the other hand, one person could grasp the same object differently when performing the same task. Great ambiguities are associated with the grasping across various objects plus various tasks. While collecting the grasping data where subjects were freely performing the tasks the data consisted of the great ambiguity naturally. In the meantime, subject differences were also increasing the embedded ambiguity. 2) The data amount was not sufficient to capture the subtle differences between grasping configurations for various tasks. Even though a database consists of ≈20,000 samples were collected, given the numerous object conditions(i.e. orientations, objects, and tasks) it is still relatively small since each object condition set model may only be a few hundred data points. We collected grasping poses when the object was facing various directions, and, in each direction, the data amount is very limited–where ≈80 training samples for each object-task pair for each object orientation. This limited data for each condition with great uncertainties makes it very challenging to capture the commonness of various grasping configurations for a task and the differences of various grasping configurations for different tasks.

Given these two possible reasons, in the future, we could first try to develop an individual grasping model then generalize it to capture the general grasping behaviors among humans. In the meantime, more data could be collected to ensure the data can sufficiently represent the commonness and differences.

5.2 Tuning Statistics

Even though the SVM achieved the best accuracy shown in Table 1, it is arbitrary to say the SVM method is the best approach that should be selected. During the iterative fine-tuning of each modeling method, each iteration's performance was recorded and statistically summarized to provide a comprehensive understanding of the tuning effort of each modeling method. Each method has different tuning space as its parameter's range, and each tuning iteration results in a different model. For the NN method, two-layer and three-layer NN structures were examined, and the neuron number in each layer ranged from 16 to 96. In total, there were 811 tuning iterations for the NN method. There were 6 tuning iterations for the SVM resulting from different kernel functions. The size of the SOM had been considered as a tuning factor that affected the modeling performance, and we attempted combinations of length and width of 3, 4,5,7,10. For example, the SOM sizes of 3x3,7x5,10x10 were attempted in the training process. There are 75 tuning iterations for the BN method when varying the SOM size and BN structures. For the target teleoperation scenario, size of 3x5 is recommended without using gaze information, and size of 7x5 is recommended when using gaze information.

5.2.1 Average Accuracy Statistics over Iterative Training

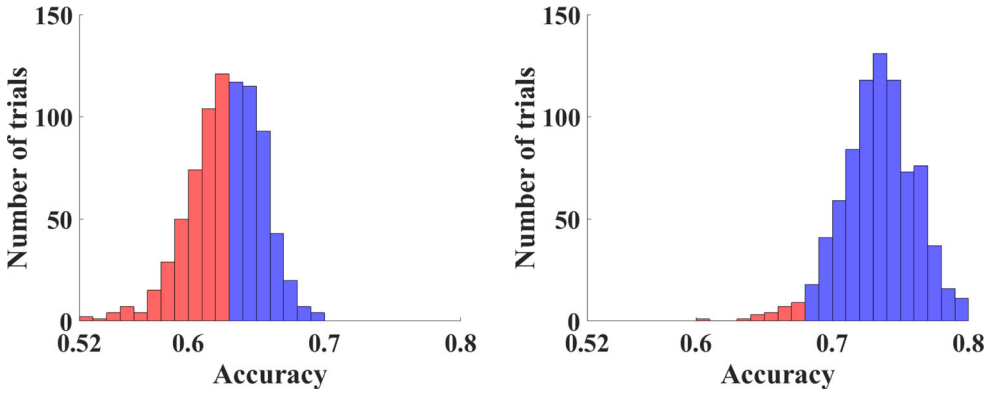
The models are separated into two categories based on in which scenario the model achieves a higher accuracy. The statistical average accuracy of each model over all the training iterations are summarized in Table 3. The models which achieved a higher accuracy in the real-object scenario are summarized in Table 3. While, the models that achieved a higher accuracy in the teleoperation scenario

Table 3 Summary of the statistical results of models with mean(standard deviation) and e denotes the tuning effort

			NN	SVM	BN
Best model in the real-object scenario	Without gaze	Real	**0.66(0.034)**	**0.56(X)**	**0.72(0.024)**
		Virtual	0.62(0.027)	0.45(X)	0.70(0.016)
	With Gaze	Real	**0.72(0.024)**	**0.59(0.035)**	**X(X)**
		Virtual	0.70(0.031)	0.51(0.004)	X(X)
Best model in the tele-operation scenario	Without gaze	e	0.452	0.004	0
		Real	0.58(0.41)	0.55(0.105)	0.66(0.032)
		Virtual	**0.64(0.024)**	**0.61(0.069)**	**0.71(0.021)**
	With Gaze	e	0.020	0.006	0
		Real	0.63(0.041)	0.59(0.085)	0.65(0.027)
		Virtual	**0.73(0.027)**	**0.68(0.023)**	**0.79(0.012)**

Note, there is no iteration that the BN functions better in real object scenario with gaze

The higher accuracy comparing the real-object case with the virtual-object case is in bold



(a) Without gaze. (b) With gaze.

Fig. 7 Histograms of the NN models based on their accuracy. The blue region is the models that achieve an accuracy that is within the 20% range of the best performance, and the red region is those that fall off this 20% range

are summarized in Table 3. The average accuracy and the standard deviation in brackets are listed.

In addition, we define the tuning effort to achieve a good accuracy in teleoperation for each modeling method as $e_m = (N_m - n_m)/N$, where $N_m, m \in \{NN, SVM, BN\}$ is the tuning iteration number for a modeling method and N is the total tuning number for all three methods. n_m is the number of iterations that a method achieved a good accuracy. Here, we define that a model achieves a good accuracy when its accuracy is within the 20% range of the average of the three methods' best accuracy. Thus, the smaller the tuning effort is the less failure a modeling method has and the more robust this method is.[1] Figure 7 is the histograms of the NN models' accuracy from the iterative tuning process. The blue region represents the models that achieve an accuracy within the 20% range. They show that when the gaze information is used there are more NN models which can achieve an accuracy within the 20% range.

From Table 1, results show that the BN functions the best, which has the highest average accuracy. Even though the SVM method has the highest accuracy (Table 1), this method has the poorest performance in overall tuning. Also, the standard deviations of SVM methods are the highest in most of the cases. Moreover, the tables and histograms show that the gaze information is very helpful to increase the inference accuracy in both real-object and teleoperation scenarios, and this improvement is relatively larger for the teleoperation scenario.

In both conditions, with and without gaze information, the BN method has a tuning effort of zero. This means all BN models in the tuning process achieved an accuracy that was within the 20% range. The NN method has the largest tuning effort regarding its large parameter space. In addition, the gaze information can greatly reduce the tuning effort (from 0.452 to 0.020 for the NN method). This demonstrates the helpfulness of the gaze information. The SVM method's tuning effort is small due to the large total tuning number. However, its relative tuning effort is large. There is only one iteration the SVM got an accuracy that is within the 20% range when using gaze and two iterations when not using gaze.

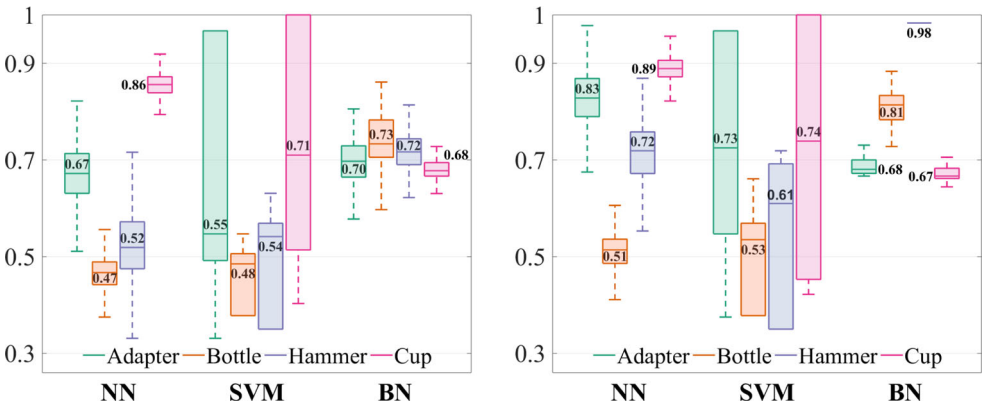
5.2.2 Object-Specific Statistics

During manipulation, objects demonstrate various manipulation patterns due to the object's size, shape, and affordance. In addition, the modeling method's capability of capturing those manipulation patterns varies, which results in object-specific performance distributions. This performance distribution in the teleoperation scenario is shown in Fig. 8. It shows each modeling method performs variously for different objects, and the gaze information demonstrates various improvements for each object. The NN and SVM methods demonstrate a similar object-specific accuracy distribution pattern. They both function poorly for the bottle and hammer with and without gaze information compared to the adaptor and cup. Moreover, the gaze information is particularly helpful for the adaptor and hammer as there is a large accuracy improvement. However, the BN has a different distribution pattern, where BN functions well relatively for the bottle and hammer, and the gaze information is greatly helpful for these objects too. The BN method has the best stability compared to the NN and SVM methods. In contrast, the SVM's stability is the poorest.

5.2.3 Task-Specific Statistics

Figure 9 shows the task-specific manipulation patterns of various modeling methods. Three methods all function well for the transfer task relative to usage and handover. The reason may be the transfer task yields the least constraints. The NN method demonstrates great stability across all

[1] For example, the average best accuracy is $0.843 = (0.84 + 0.87 + 0.82)/3$ when the gaze information was used, and the 20% range bar is 0.674. For the NN method, there are 793 iterations that the models achieved an accuracy higher than 0.674 with gaze as the extra information. Thus, the tuning effort is $(811 - 793)/(811 + 6 + 75) = 0.020$.



(a) Without gaze. (b) With gaze.

Fig. 8 Statistical object-specific accuracy in the teleoperation scenario when using different modeling methods. The distribution median of the accuracy of each task is shown

tasks, and the stability of the SVM is still the poorest. The SVM and BN are relatively aggressive as they provide chances to achieve a high accuracy but also face chances of performing poorly. The performance improvement of using the gaze information is apparent for the transfer task with the SVM method and transfer and handover tasks with the BN method.

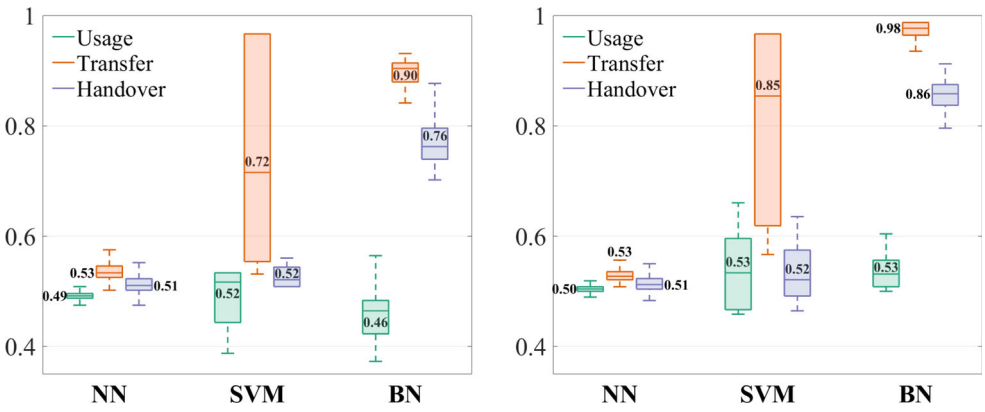
5.3 BN Structures

The BN structures that were built in the experiment are shown in Fig. 10. Figure 10a is obtained using the data-driven strategy. Figure 10b and c were built through the human-intuition strategy based on different beliefs about how the hand motion is generated. In the finger-driven model, it is believed that the fingers are primarily driving the motion of the hand and arm. While in the palm-driven model, the palm is believed to be primarily driving the arm motion, while the fingers are passively opening and closing. The human-intuition models are more concise than the data-driven model, as there are less dependency links between variables. Here, only the BN models with the gaze feature are displayed. The BN structures, when the gaze feature was not used, are similar, which have all the dependency links related to gaze removed.

The data-driven model reflects the correlations of variables in the collected data and worked the best in the teleoperation scenario. The best manipulation inference accuracy in Table 1 was achieved when using the data-driven model. While the finger-driven model was used to achieve the best accuracy in the real-object scenario in Table 1.

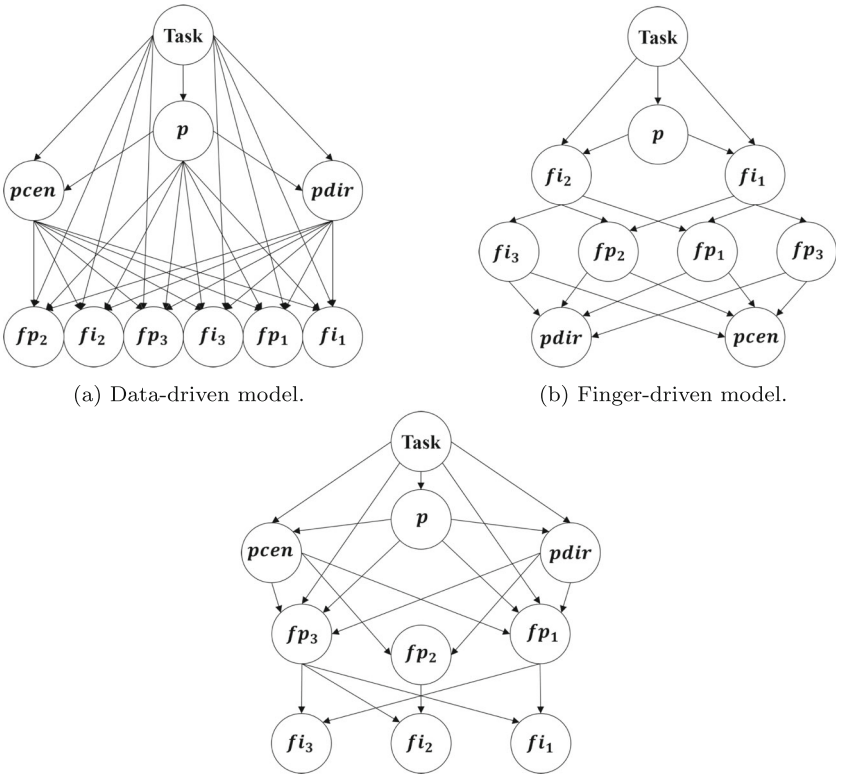
5.4 Database for Manipulation Intent Modeling

Through the previous experiment, 19600 grasping samples of real-object interaction and 1440 grasping samples through the teleoperation interface were collected. Data has been pre-processed to remove the tracking failures. All data is available on the authors' lab website to support other researchers' exploration and employment of human manipulation intent in teleoperation. In addition, the 3D models of the studied objects are provided too. In the work presented in this paper,



(a) Without gaze. (b) With gaze.

Fig. 9 Statistical task-specific accuracy in the teleoperation scenario when using different modeling methods. The distribution median of the accuracy of each task is shown



(a) Data-driven model. (b) Finger-driven model.

(c) Palm-driven model.

Fig. 10 Statistical task-specific accuracy in the teleoperation scenario when using different modeling methods. The medians of each task are shown

1/8 of the real-object interaction data and 1/3 of the teleoperation interaction data were used, in which the object was facing one direction. When considering grasping the object from any arbitrary directions, there is higher complications for the grasping model. Further investigation of the NN, SVM, and BN modeling methods and developing new methods will be conducted in the future, and their computation complexities in training and usage will be evaluated following the methods in [34–36].

6 Discussion

In this section we discuss the hypotheses the success of the earlier presented hypotheses (Section 6.1 Hypothesis Testing), dealing with ambiguity of multi-label classification (Section 6.2 Ambiguity of Multi-Label Classification), our recommendations for modeling manipulation intent (Section 6.3 Model Suggestion), and limitations of current assistance to incorporate manipulation intent (Section 6.4 Assistance to Manipulation Intent).

6.1 Hypothesis Testing

Table 4 shows the testing results of the hypotheses and the proofs used to derive the conclusion. All hypotheses are supported by the experimental results. When using the grasping model built in real-object interaction for the manipulation intent inference in teleoperation, a reasonably good inference accuracy is achieved, especially when the gaze information is used. Thus, it is feasible to adopt real-object interaction models for teleoperation interaction in teleoperation (H1 is supported). Even though the adoption is feasible, the same model's performance varies and polarizes in two scenarios. This conclusion is drawn from Tables 1 and 3, which show that when the model functions well in one scenario, it will very likely function worse in the other one (H2 is supported). This difference suggests that there are differences in the essential features in two scenarios for the same grasping process. This can be demonstrated by the BN results. The BN model that functions the best in the teleoperation scenario has the data-driven model structure, while the finger-driven model structure functions the best in

Table 4 Hypothesis summary

Hypothesis	Result	Proof
H1	Supported	Table II, Table IV, Table VII, and Table IX
H2	Supported	Tables I-IV, Tables VI-IX, and Fig. 5
H3	Supported	Table I *vs.* II and Table III *vs.* IV
H4	Supported	Table V *vs.* VII and Table VI *vs.* VIII

the real-object scenario. This structure difference is also a vivid evidence to support H3.

The H4 is supported with data in Table 3, which adding the gaze information to the grasp modeling takes advantage of the eye-hand coordination and improves the inference accuracy in both scenarios. This accuracy improvement is apparent in both teleoperation and real-object scenarios. Moreover, the improvement is larger in the teleoperation scenario. The possible reason could be that gaze has lower essentiality in the real-object scenario. In the real-object scenario, the subjects perform actual grasping on the object. The hand grasping configuration is explicit and can clearly demonstrate the grasping correlation. However, in the teleoperation, the subjects try to grasp a virtual object, and the hand grasping configuration becomes fuzzy. In this case, the gaze information still strongly and clearly indicates which part on the object the subject is looking at. The essentiality of the gaze information increases in the teleoperation scenario.

6.2 Ambiguity of Multi-Label Classification

The formulation of the multi-label classification makes the robot aware of the existence of other possible manipulation intent. However, there is ambiguity which task the user intends. One way to further clarify the ambiguity is to consider the context information and task-related sequential knowledge when building the intent model. However, this will require specific knowledge to model the application scenario, which loses the generality of the current model that only consists of primitive motions.

Moreover, this intent ambiguity can be handled later when the robot generates an assistance plan. If the robot is aware of the ambiguity, it can select an action plan that can satisfy all the possible types of intent. For example, if the inferred intent is handover and usage, the robot can select a grasping plan that can satisfy both tasks. In this way, the robot can have more flexible reasoning capability and behaviors. It will also improve the robot's practicality as it is able to handle the ambiguity, which is the nature of the practical scenarios.

6.3 Model Suggestion

We explicitly and comprehensively examined NN, SVM, and BN methods for the manipulation intent inference in teleoperation. Each method is scored with 1-3 from five perspectives, which is summarized in Table 5. The higher the score is, the better the model performs. Even though the best accuracy achieved using the BN method is lower than NN and SVM methods, it outperforms the other two from other four evaluation perspectives. In all the tuning iterations, the BN method's performance is the highest and very stable.

Table 5 Evaluation scores of the NN, SVM, and BN methods

	B-A	A-A	R	T-E	H-U	Overall
NN	2	2	3	1	1	9
SVM	3	1	1	1	1	7
BN	1	3	3	3	3	12

B-A: Best Accuracy, A-A: Average Accuracy, R: Robustness, T-E: Tuning Effort, and H-U: Human Understandability

This shows the robustness of the BN method for our application. Also, the BN method has the lowest tuning effort. The visibility of the BN structure makes the BN method understandable with human intuition. Thus, the BN method has the highest evaluation score and is suggested for the manipulation intent inference in teleoperation. Moreover, the palm-driven model is preferred as it is more compact, and it achieved a compared accuracy as the data-driven model. This palm-driven model is from human intuition, which is easier to understand. In terms of computational complexity of the three methods implemented, SVM is proportional to both the number of classes and the number of input features, while the NN can be determined for the basic connection layers used. Although the layers for our attempted models did change, the general complexity is a summation of all the products between each pair of layers. The BN, however, has more difficult complexity as it is structure dependent. For instance, a Naive Bayes model, where each feature is independent of one another, acts as a single layer NN where it can be calculated as matrix multiplication. The structure, number of input features, and classifications all influences the complexity where there is not a guaranteed manner of determining complexity given arbitrary inputs. Where once again, further investigation of the NN, SVM, and BN modeling methods and developing new methods will be conducted in the future, and their computation complexities in training and usage will be evaluated following the methods in [34–36].

6.4 Assistance to Manipulation Intent

To assist the achievement of an operator's motion intent, assistance formats of boundaries, guiding force, and shared motion are commonly used, as reviewed in the related work. It may be challenging to use the same formats of assistance to assist an operator's manipulation intent. For example, if guiding force format is selected, the six-dimensional haptic force will be needed to push the operator to reach a specific position with a specific orientation. How to formulate a robot's assistance to achieve effective teleoperation of object manipulation demands further investigation.

7 Conclusion

In this paper, an operator's manipulation intent in teleoperation is inferred by modeling the object grasping for various tasks. This is the first time that an operator's manipulation intent in teleoperation has been considered. The differences between the inference accuracy in real-object interaction and in teleoperation have been discussed. The manipulation intent inference was formulated as a multi-label classification problem. This formulation makes the robot aware of the existence of the ambiguity. We validated that the grasping model that is built through real-object interaction is feasible to be used for interaction in teleoperation. However, one model that well behaves in the real-object interaction does not function well at the same degree in the teleoperation, vice versa. Thus, training with the consideration of the application scenario is essential. The experimental results also show that adding the gaze information is helpful to improve the intent inference in teleoperation. With the inferred manipulation intent, the robot can generate action plans to assist the sophisticated manipulation-related tasks in teleoperation rather than just approaching the target location or object. This can bring the teleoperation into more practical applications and increase robot adoption. In the future, we may consider adding the context information to reduce the ambiguity, which, however, will reduce the generality of the current pure grasping model. In the meantime, we are looking forward to formulate a robust decision-making engine that considers the intent ambiguity when generating assistance plans, where the teleoperation performance with the robotic assistance in achieving the manipulation intent will be tested by comparing it to traditional manual control. We will also investigate how manipulation intent could benefit other open problems in teleoperation, like the problem of time delay and data loss [37, 38]. With the capability of understanding and predicting the operator's intent, the robot could generate supportive or assistive execution plans to accomplish the task even loss continuous input from the teleoperator due to the time delay or data loss.

Acknowledgements This material is based on work supported by the US NSF under grant 1652454. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation.

References

1. Lum, M.J., Friedman, D.C., Sankaranarayanan, G., King, H., Fodero, K., Leuschke, R., Hannaford, B., Rosen, J., Sinanan, M.N.: The raven: Design and validation of a telesurgery system. Int. J. Robot. Res. **28**(9), 1183–1197 (2009)
2. Hochberg, L.R., Bacher, D., Jarosiewicz, B., Masse, N.Y., Simeral, J.D., Vogel, J., Haddadin, S., Liu, J., Cash, S.S., van der Smagt, P., et al.: Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature **485**(7398), 372 (2012)
3. Bodner, J., Wykypiel, H., Wetscher, G., Schmid, T.: First experiences with the da vinci™operating robot in thoracic surgery. Europ. J. Cardio-thoracic Surg. **25**(5), 844–851 (2004)
4. Rybarczyk, Y., Colle, E., Hoppenot, P.: Contribution of neuroscience to the teleoperation of rehabilitation robot. In: 2002 IEEE International Conference on Systems, Man and Cybernetics, vol. 4, pp. 6–pp. IEEE (2002)
5. Healey, A.N.: Speculation on the neuropsychology of teleoperation: Implications for presence research and minimally invasive surgery. Presence **17**(2), 199–211 (2008)
6. Li, Y., Tee, K.P., Chan, W.L., Yan, R., Chua, Y., Limbu, D.K.: Continuous role adaptation for human–robot shared control. IEEE Trans. Robot. **31**(3), 672–681 (2015)
7. Webb, J.D., Li, S., Zhang, X.: Using visuomotor tendencies to increase control performance in teleoperation. In: American Control Conference (ACC), 2016, pp. 7110–7116. IEEE (2016)
8. Dragan, A.D., Srinivasa, S.S.: A policy-blending formalism for shared control. Int. J. Robot. Res. **32**(7), 790–805 (2013)
9. Javdani, S., Srinivasa, S.S., Bagnell, J.A.: Shared autonomy via hindsight optimization. arXiv:1503.07619 (2015)
10. Mylonas, G.P., Kwok, K.-W., James, D.R., Leff, D., Orihuela-Espina, F., Darzi, A., Yang, G.-Z.: Gaze-contingent motor channelling, haptic constraints and associated cognitive demand for robotic mis. Medi. Image Anal. **16**(3), 612–631 (2012)
11. Ren, J., Patel, R.V., McIsaac, K.A., Guiraudon, G., Peters, T.M.: Dynamic 3-d virtual fixtures for minimally invasive beating heart procedures. IEEE Trans. Med. Imag. **27**(8), 1061–1070 (2008)
12. Muelling, K., Venkatraman, A., Valois, J.-S., Downey, J., Weiss, J., Javdani, S., Hebert, M., Schwartz, A.B., Collinger, J.L., Bagnell, J.A.: Autonomy infused teleoperation with application to bci manipulation. arXiv:1503.05451 (2015)
13. Kim, H.K., Biggs, J., Schloerb, W., Carmena, M., Lebedev, M.A., Nicolelis, M.A., Srinivasan, M.A.: Continuous shared control for stabilizing reaching and grasping with brain-machine interfaces. IEEE Trans. Biomed. Eng. **53**(6), 1164–1173 (2006)
14. Li, S., Zhang, X., Kim, F.J., da Silva, R.D., Gustafson, D., Molina, W.R.: Attention-aware robotic laparoscope based on fuzzy interpretation of eye-gaze patterns. J. Med. Dev. **9**(4), 041007 (2015)
15. Nikolaidis, S., Zhu, Y.X., Hsu, D., Srinivasa, S.: Human-robot mutual adaptation in shared autonomy. arXiv:1701.07851 (2017)
16. Romano, J.M., Hsiao, K., Niemeyer, G., Chitta, S., Kuchenbecker, K.J.: Human-inspired robotic grasp control with tactile sensing. IEEE Trans. Robot. **27**(6), 1067–1079 (2011)
17. Lenz, I., Lee, H., Saxena, A.: Deep learning for detecting robotic grasps. Int. J. Robot. Res. **34**(4-5), 705–724 (2015)
18. Montesano, L., Lopes, M., Bernardino, A., Santos-Victor, J.: Learning object affordances: From sensory–motor coordination to imitation. IEEE Trans. Robot. **24**(1), 15–26 (2008)
19. Fischinger, D., Vincze, M.: Empty the basket-a shape based learning approach for grasping piles of unknown objects. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2051–2057. IEEE (2012)
20. Trinkle, J.C.: On the stability and instantaneous velocity of grasped frictionless objects. IEEE Trans. Robot. Autom. **8**(5), 560–572 (1992)
21. Song, D., Huebner, K., Kyrki, V., Kragic, D.: Learning task constraints for robot grasping using graphical models. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1579–1585. IEEE (2010)
22. Balasubramanian, R., Xu, L., Brook, P.D., Smith, J.R., Matsuoka, Y.: Physical human interactive guidance: Identifying grasping principles from human-planned grasps. IEEE Trans. Robot. **28**(4), 899–910 (2012)

23. Huaman Quispe, A., Ben Amor, H., Christensen, H., Stilman, M.: Grasping for a purpose: Using task goals for efficient manipulation planning, arXiv:1603.04338 (2016)
24. Ng, A.Y., Jordan, M.I.: On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In: Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, ser. NIPS'01, pp. 841–848. MIT Press, Cambridge (2001). [Online]. Available: http://dl.acm.org/citation.cfm?id=2980539.2980648
25. Kohonen, T.: The self-organizing map. Neurocomputing **21**(1-3), 1–6 (1998)
26. Wehrens, R., Buydens, L.M., et al.: Self-and super-organizing maps in r: The kohonen package. J Stat Softw **21**(5), 1–19 (2007)
27. Aliferis, C.F., Tsamardinos, I., Statnikov, A.: Hiton: A novel Markov blanket algorithm for optimal variable selection. In: AMIA Annual Symposium Proceedings, vol. 2003, p. 21. American Medical Informatics Association (2003)
28. Scutari, M.: Learning bayesian networks with the bnlearn R package. J. Stat. Softw. **35**(3), 1–22 (2010)
29. Kruskal, J.B.: On the shortest spanning subtree of a graph and the traveling salesman problem. Proc. Am. Math. Soc. **7**(1), 48–50 (1956)
30. Cooper, G.F., Herskovits, E.: A bayesian method for the induction of probabilistic networks from data. Mach. Learn. **9**(4), 309–347 (1992)
31. Tsoumakas, G., Katakis, I.: Multi-label classification: An overview. Int. J. Data Warehousing Mining **3**, 3 (2006)
32. Alvares-Cherman, E., Metz, J., Monard, M.C.: Incorporating label dependency into the binary relevance framework for multi-label classification. Expert Syst. Appl. **39**(2), 1647–1655 (2012)
33. Elisseeff, A., Weston, J.: A kernel method for multi-labelled classification. In: Advances in Neural Information Processing Systems (2002)
34. Cooper, G.F.: The computational complexity of probabilistic inference using Bayesian belief networks. Artif. Intell. **42**(2–3), 393–405 (1990)
35. Abdiansah, A., Wardoyo, R.: Time complexity analysis of support vector machines (svm) in libsvm. International Journal Computer and Application (2015)
36. Rojas, R.: Neural Networks: A Systematic Introduction. Springer Science and Business Media (2013)
37. Baranitha, R., Mohajerpoor, R., Rakkiyappan, R.: Bilateral teleoperation of single-master multislave systems with semi-Markovian jump stochastic interval time-varying delayed communication channels. IEEE Trans. Cybern., 1–11 (2019)
38. Mohajerpoor, R., Sharifi, I., Talebi, H.A., Rezaei, S.M.: Adaptive bilateral teleoperation of an unknown object handled by multiple robots under unknown communication delay. In: 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 1158–1163 (2013)

Songpo Li received the B.S. degree in Computer Science and Technology from Hebei University of Economics and Business, Shijiazhuang, China, in 2010, the M.S. degree in Mechanical Engineering from Wilkes University, Wilkes-Barre, PA, USA, in 2013, and Ph. D. degree in Mechanical Engineering from Colorado School of Mines, Golden, CO, USA, in 2017. He is now a postdoctoral associate at Duke University. His research interests include human-robot interaction, human and robot understanding and behavioral modeling, robotic intelligence, and shared autonomy.

Michael Bowman received B.S. degrees in both Mechanical Engineering and Electrical Engineering from Colorado School of Mines in 2017. He is currently a PhD Candidate within the Department of Mechanical Engineering, Colorado School of Mines, Golden, CO, USA. His research interests include human-robot interaction and cooperation, human modeling, shared control, robot transparency, and decision making.

Hamed Nobarani received the B.S. degree in Mechanical Engineering from Azad University (Science and Research Branch), Tehran, Iran in 2015. He got his M.S. degree in Mechanical Engineering with the focus on Biomechanics from University of Colorado at Denver, Denver, CO, USA, in 2017. He is currently a Ph.D. student in the Department of Mechanical Engineering, Colorado School of Mines, Golden, CO, USA. His research area include intelligent human?robot interaction and cooperation and atomistic simulations of shape memory functional materials.

Xiaoli Zhang received the B.S. degree in Mechanical and Automation Engineering and the M.S. degree in Mechatronics Engineering from Xi'an Jiaotong University, Xi'an, China, in 2003 and 2006, respectively, and the Ph.D. degree in Biomedical Engineering from the University of Nebraska-Lincoln, Lincoln, NE, USA, in 2009. She is currently an Associate Professor with the Department of Mechanical Engineering, Colorado School of Mines, Golden, CO, USA. Her research interests include intelligent human?robot interaction and cooperation, human intention awareness, data-driven modeling, prediction, and control, and their applications in healthcare and industrial fields.