© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Learning-Based Modular Task-Oriented Grasp Stability Assessment

Jingyi Xu, Amit Bhardwaj, Ge Sun, Tamay Aykut, Nicolas Alt, Mojtaba Karimi, and Eckehard Steinbach

***Abstract*— Assessing grasp stability is essential to prevent the failure of robotic manipulation tasks due to sensory data and object uncertainties. Learning-based approaches are widely deployed to infer the success of a grasp. Typically, the underlying model used to estimate the grasp stability is trained for a specific task, such as lifting, hand-over, or pouring. Since every task has individual stability demands, it is important to adapt the trained model to new manipulation actions. If the same trained model is directly applied to a new task, unnecessary grasp adaptations might be triggered, or in the worst case, the manipulation might fail.**

To address this issue, we divide the manipulation task used for training into seven sub-tasks, defined as modular tasks. We deploy a learning-based approach and assess the stability for each modular task separately. We further propose analytical features to reduce the dimensionality and the redundancy of the tactile sensor readings. A main task can thereby be represented as a sequence of relevant modular tasks. The stability prediction of the main task is computed based on the inferred success labels of the modular tasks. Our experimental evaluation shows that the proposed feature set lowers the prediction error up to 5.69% compared to other sets used in state-of-the-art methods. Robotic experiments demonstrate that our modular task-oriented stability assessment avoids unnecessary grasp force adaptations and regrasps for various manipulation tasks.

I. INTRODUCTION

Intelligent robots should be able to operate in unstructured environments with partially unknown objects. The sensory data and object uncertainties often lead to failure of manipulation tasks. Hence, it is essential to predict the grasp stability prior to a manipulation attempt. The stability is affected by various properties of the object, such as weight, center of gravity, material properties etc. Analytical approaches are hence suboptimal to infer grasp stability in unstructured environments, since they typically require complete knowledge of the object and its contact with the end effector. To cope with such uncertainties, learning-based approaches are generally applied to estimate grasp stability.

The stability of a grasp highly depends on the specific manipulation task, since each task comes with different challenges. For instance, a pick-and-place task is likely to cause object sliding, while a pouring task possibly causes in-hand rotations. Moreover, manipulating deformable fragile objects or open containers is more difficult than handling rigid objects. It is further important to not damage the object and to avoid undesired effects such as content spilling of an open container caused by the deformation. Minimum grasp force should therefore be used to manipulate such objects.

Jingyi Xu, Amit Bhardwaj, Ge Sun, Tamay Aykut, Nicolas Alt, Mojtaba Karimi, and Eckehard Steinbach are with the Chair of Media Technology, Technical University of Munich, Germany. Email: {jingyi.xu, amit.bhardwaj, ge.sun, tamay.aykut, n.alt, mojtaba.karimi, eckehard.steinbach}@tum.de

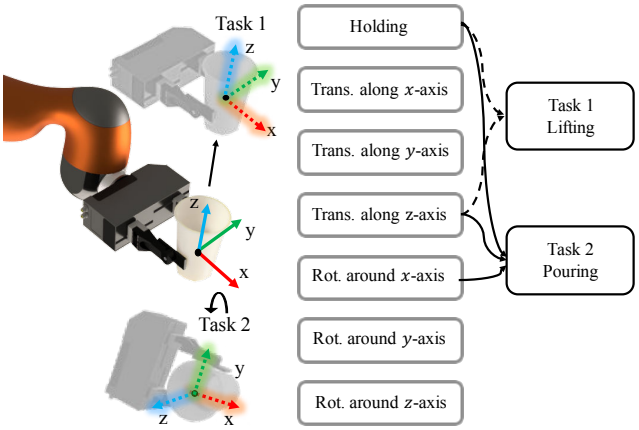


Fig. 1: We assess grasp stability for seven modular tasks: the holding task and six motion tasks. The holding task predicts whether a fragile object will be damaged or the content will spill for the planned grasp force. New manipulation tasks can be flexibly represented as sequences of modular tasks. Two example tasks are depicted. The lifting task includes the holding task and a translation along the z-axis. The pouring task includes the holding task, a translation along the z-axis and a rotation around the x-axis.

Assessing grasp stability for different manipulation tasks is studied in [1]–[6]. Typically, the success label of a grasp is collected by moving the arm according to one or several pre-defined tasks, like lifting, hand-over, or pouring. The underlying model of the grasp stability will then be trained and validated with the same tasks. The collected data and trained underlying model are only suitable for specific actions that are used in training. Adjusting the trained model to new manipulation tasks remains a challenge. Directly applying the model to a different task may cause unnecessary grasp adaptations or a failure of the action.

To address this issue, we propose a modular task-oriented grasp stability assessment. A manipulation task can be divided into several sub-tasks, defined as modular tasks in the remainder. They can be flexibly combined to represent new tasks. We define seven modular tasks: a holding task and six motion tasks, as illustrated in Fig. 1. The holding task infers whether a fragile object will be damaged or the content will spill from a plastic container with an open lid. The six motion tasks are the three-dimensional translational/rotational movements of the gripper along/around the x, y, z-axes, respectively. Each of them estimates the grasp stability for the object manipulation with the corresponding

motion. In general, a manipulation can be considered as a trajectory or a target pose of the end effector, while the object is safely grasped during the movement. Therefore, a manipulation task can be considered as a task sequence of the modular tasks followed by the grasp. Fig. 1 shows an example pouring task, which can be considered as a task sequence of lifting (a translation along the z-axis) and a $90°$ rotation around the x-axis.

For the proposed approach, grasp stability for each modular task is assessed independently. During the training phase, various objects are grasped with two approach directions for the gripper at several locations spread vertically along the object. We consider the grasping position, object deformability, and the tactile sensor readings to infer the stability. We propose a new approach to reduce the high dimensionality and redundancy of the tactile sensor readings. Analytical features are extracted from the readings, including the contact area, the pressure-weighted friction center, the friction, and the work. A decision forest is applied as a classifier and trained for each task independently.

To infer the grasp stability for a main manipulation task, the desired gripper pose for the task will first be analyzed, whether it can be reached through a sequence of the motion tasks. If all tasks within the chosen sequence and the holding task can succeed, the manipulation task is estimated to be successful. The relevance of the modular tasks can be further improved when the manipulation task is known prior to the grasp assessment (e.g. in an industry environment) or has a human in the loop (e.g. in a tele-operation scenario). For instance, when the objects to be manipulated are deformable and non-fragile, such as sponges or soft toys, the holding task can be removed from the task sequence and the prediction accuracy can be further improved.

We present an evaluation of the trained model for each modular task. We show that with our proposed feature set, a higher prediction accuracy can be achieved for six of seven modular tasks. We further compare the grasping and manipulation process with and without considering the modular tasks. Experimental evaluation shows that much less grasp adaptation steps are required with the proposed method to successfully execute the manipulation task.

Our contributions are summarized as follows:

- We propose a modular task-oriented grasp stability assessment, where the trained underlying models can be flexibly adapted to different manipulation tasks.
- We propose a combination of analytical features extracted from the tactile sensor readings to reduce the data redundancy.

II. RELATED WORK

Grasp stability estimation using tactile sensor readings can be divided into analytical and learning-based approaches. For analytical ones, Krug et al. [7] predict the grasp stability for a lifting task based on the grasp wrench space. The effect of uncertainties including object pose, weight and friction coefficient is evaluated. Conservative approximation of object properties is recommended to avoid false positive predictions. The limitations of this approach include the difficulty of formulating the disturbance wrenches for dynamic or complicated tasks.

For learning-based grasp stability assessment, different hand-crafted features are extracted from the tactile sensor readings to reduce the high dimensionality and redundancy, such as k-means clustering applied in [3], principle components analysis (PCA) in [4], [8], and image moments in [1], [2], [5], [9]–[11]. Laaksonen et al. [12] evaluate the performance of various features extracted from the tactile readings and machine learning methods in determining the grasp stability. Recently, unsupervised learning [6] or deep neural networks [13], [14] are applied to estimate the grasp stability without hand-crafted features.

Different manipulation tasks are considered for labeling the grasp stability, such as lifting [3], [6], [10], [13], [15], lifting with accelerations and decelerations [14], lift and rotate $[-120°, +120°]$ around the approach vector [2], [12], or rotate $+90°$ around the x, y-axis, where z-axis is the direction of the lift [12]. Bekiroglu et al. [1] integrate the task-dependency with grasp stability estimation. A Bayesian network is used to model the conditional relations between the sensory streams and three tasks including a hand-over task (parallel transportation), a pouring task ($90°$ rotation), and a dish washing task ($180°$ rotation).

To our knowledge, no related work has provided an extensive evaluation of the grasp stability for all 6D translational and rotational moving actions and how the stability is affected when manipulating deformable, fragile objects or open containers. Furthermore, how to adapt the trained model of the stability for specific tasks to new manipulation tasks has not been attempted yet. The consequences of applying the same model to new tasks should be analyzed. We propose grasp stability assessment for seven modular tasks and provide the flexibility to estimate the stability for new tasks by representing them as sequences of modular tasks. The proposed approach guides a grasping and manipulation process with better time and energy efficiency.

III. PROPOSED APPROACH

First, we describe the notation used in this paper:

- $\mathbb{T} = \{\mathcal{T}_{\text{holding}}, \mathcal{T}_{T_x}, \mathcal{T}_{T_y}, \mathcal{T}_{T_z}, \mathcal{T}_{R_x}, \mathcal{T}_{R_y}, \mathcal{T}_{R_z}\}$ denotes the task space for seven modular tasks: the holding task and the six motion tasks.
- $\mathbb{S} = \{\mathcal{S}_T, \mathcal{S}_R\}$ denotes the sequence of motion tasks for a main manipulation task.
- $\mathbb{R} = \{\mathcal{R}_{\mathcal{S}_T}, \mathcal{R}_{\mathcal{S}_R}\}$ denotes the relevant modular tasks for the sequence $\{\mathcal{S}_T, \mathcal{S}_R\}$.
- $\mathbb{F} = \{(H^k, D^k, \boldsymbol{X}^k)\}_{k=1\ldots N}$ denotes a feature set with N observations, where
  - H^k denotes the grasping position in the object-centered coordinate system.
  - D^k denotes the object deformation.
  - $\boldsymbol{X}^k = \{\mathcal{C}^k, \boldsymbol{p}^k, \tilde{f}^k, \tilde{\tau}^k, w^k\}$ denotes the analytical features extracted from the tactile sensor readings, including the contact area, the friction-weighted pres-

sure center, the estimated friction, and the work performed by a grasp finger.

A. Analytical features

Here we describe the analytical features extracted from the tactile sensor readings. The feature sets used for the classification are described in Section IV-C.

In analytical approaches, friction of a contact is constantly analyzed to build the grasp wrench space (GWS), where its volume is a classical quality metric of a grasp [16], [17]. The work is applied as a quality metric for grasping deformable objects [18]. Therefore, we consider them as features to analyze grasp stability. Here, we briefly introduce the computation for the friction-weighted pressure center, the maximal possible friction [19], [20], and the work performed by the grasp finger.

For a taxel t_{xy} from a tactile sensor reading with an index $[x, y]^T$, we denote μ_{xy} as the friction coefficient, ρ_{xy} as the pressure value, a_{xy} as the area of the taxel, $\boldsymbol{c} = [c_x, c_y]^T$ as the center coordinate of the taxel t_{xy}.

For a contact area $\mathcal{C}$, the friction-weighted pressure center $\boldsymbol{p} = [p_x, p_y]$ is computed by:

$$p_x = \frac{\int_{\mathcal{C}} x\mu_{xy}\rho_{xy}d\mathcal{C}}{\int_{\mathcal{C}} \mu_{xy}\rho_{xy}d\mathcal{C}},\ p_y = \frac{\int_{\mathcal{C}} y\mu_{xy}\rho_{xy}d\mathcal{C}}{\int_{\mathcal{C}} \mu_{xy}\rho_{xy}d\mathcal{C}}. \tag{1}$$

The maximal possible frictional force f_{xy} and torque τ_{xy} of a taxel t_{xy} is computed by:

$$\begin{aligned} &f_{xy} = \mu_{xy}\rho_{xy}a_{xy},\ \tau_{xy} = f_{xy}l_{xy}, \\ &\text{with } l_{xy} = \sqrt{(p_x - c_x)^2 + (p_y - c_y)^2}. \end{aligned} \tag{2}$$

In practical experiments, it is difficult to accurately estimate the friction coefficient. Inspired by [7], we assume a Coulomb friction model and use a conservatively approximated friction coefficient $\tilde{\mu}_{xy} = 0.45$ based on the table of frictional coefficients [21]. Detailed friction sensitivity analysis for the grasp quality can be found in [22].

The approximated maximal frictional force $\tilde{f}$ and torque $\tilde{\tau}$ of the tactile sensor reading with size $S_x \times S_y$ are computed by:

$$\begin{aligned} \tilde{f} &= \sum_{x=1}^{S_x}\sum_{y=1}^{S_y} \tilde{\mu}_{xy}\rho_{xy}a_{xy}, \\ \tilde{\tau} &= \sum_{x=1}^{S_x}\sum_{y=1}^{S_y} \tilde{\mu}_{xy}\rho_{xy}a_{xy}l_{xy}. \end{aligned} \tag{3}$$

The contact area $\mathcal{C}$ is computed by:

$$\mathcal{C} = \sum_{x=1}^{S_x}\sum_{y=1}^{S_y} \mathcal{C}_{xy}, \text{ with } \begin{cases} \mathcal{C}_{xy} = a_{xy}, \text{ if } \rho_{xy} > 0 \\ \mathcal{C}_{xy} = 0, \quad \text{else.} \end{cases} \tag{4}$$

The amount of work w performed by a grasp finger on the object is computed by:

$$w = \sum_{x=1}^{S_x}\sum_{y=1}^{S_y} \rho_{xy}a_{xy}d, \tag{5}$$

where d is the deformation of the object caused by the grasp force.

B. Grasp stability assessment for modular tasks

We divide the task for training into modular tasks, such that new manipulation tasks can be assessed without recollecting the data or retraining the underlying model. We define seven modular tasks, which consist of the holding task and six motion tasks, i.e. $\mathcal{T}_{T_x}, \mathcal{T}_{T_y}, \mathcal{T}_{T_z}, \mathcal{T}_{R_x}, \mathcal{T}_{R_y}, \mathcal{T}_{R_z}$. The task $\mathcal{T}_{\text{holding}}$ infers whether holding a deformable and fragile object will succeed with a planned grasp force.

The six motion tasks represent the six degrees of freedom of the object movement. The defined x, y, z-axes for the motion tasks coincide with the gripper's coordinate system, as shown in Fig. 1. The y-axis is perpendicular to the tactile sensor, which represents the direction of the grasp force. In this work, only grasps from the side are considered. This means that the initial z-axis of the gripper's coordinate system is parallel to the direction of gravity.

The coordinate system for the modular tasks is selected based on their different stability demands. For instance, the frictional force is required to compensate the gravity for all the translational movements. However, the acceleration along the y-axis for the task $\mathcal{T}_{T_y}$ can be provided by the grasp force, while the acceleration along the x-axis needs to be provided by the frictional force. Therefore, the task $\mathcal{T}_{T_x}$ is more demanding than $\mathcal{T}_{T_y}$, when the magnitude of the acceleration along the two axes is the same. For a rotation around the y-axis, the gravitational torque of the object should be compensated with the frictional torque, while it can be balanced by the torque of the grasp force for a rotation around the x-axis. The task $\mathcal{T}_{R_y}$ is hence expected to be more difficult than $\mathcal{T}_{R_x}$.

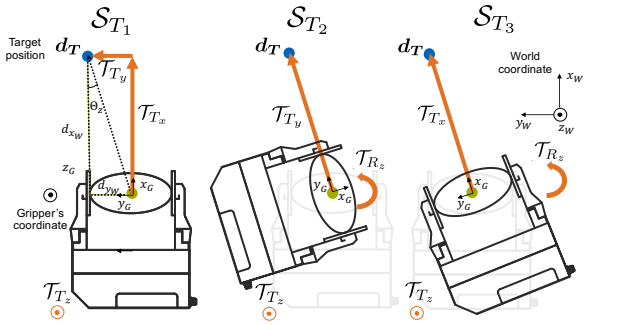
To estimate the grasp stability, we use the random forest classifier [23], which is trained separately for each modular task. The out-of-bag (OOB) data is utilized to test the performance of the classifier and to estimate the feature importance.

C. Grasp stability assessment for main tasks

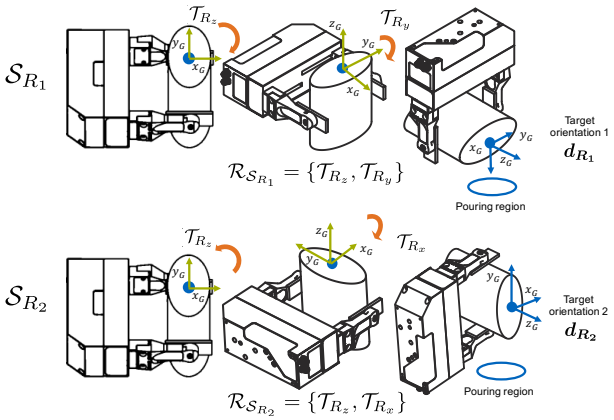
A manipulation task can be considered in general as a trajectory or a target pose of the end effector. In this work, we define a main task as a pose (position and orientation) of the gripper. The object will be first transported to a target position without possibly spilling the content. A rotation is then followed to reach a target orientation.

The target pose can be reached with a sequence of modular tasks. An object can be transported to a target position $\boldsymbol{d_T} = [d_{x_W}, d_{y_W}, d_{z_W}]^T$ in the world coordinate system through three sequences $\mathcal{S}_{T_{1,2,3}}$, as depicted in Fig. 2(a).

$$\begin{aligned} &\mathcal{S}_{T_1} = \{d_{z_W}\mathcal{T}_{T_z}, d_{x_W}\mathcal{T}_{T_x}, d_{y_W}\mathcal{T}_{T_y}\}, \\ &\mathcal{S}_{T_2} = \{d_{z_W}\mathcal{T}_{T_z}, (90^\circ - \theta_z)\mathcal{T}_{R_z}, \sqrt{d_{x_W}^2 + d_{y_W}^2}\mathcal{T}_{T_y}\}, \\ &\mathcal{S}_{T_3} = \{d_{z_W}\mathcal{T}_{T_z}, \theta_z\mathcal{T}_{R_z}, \sqrt{d_{x_W}^2 + d_{y_W}^2}\mathcal{T}_{T_x}\}, \\ &\text{with } \theta_z = atan(\frac{d_{y_W}}{d_{x_W}}), \end{aligned} \tag{6}$$



(a) Three sequences of modular tasks to reach a target position $\boldsymbol{d_T}$.



(b) Sequences of modular tasks to reach two target orientations $\boldsymbol{d_{R_1}}, \boldsymbol{d_{R_2}}$ for a pouring task.

Fig. 2: Example sequences of modular tasks to (a): reach a target position $\boldsymbol{d_T}$, (b): pour liquid within a target region. The relevant modular tasks to reach the target are summarized below each sequence.

where the multiplication is defined as the translation/rotation in the direction of the corresponding modular task.

If $d_{x_W}, d_{y_W}, d_{z_W} \neq 0$, the relevant modular tasks $\mathcal{R}_{\mathcal{S}_{T_i}}$ for the corresponding sequence $\mathcal{S}_{T_i}, i = 1, 2, 3$ are:

$$\begin{aligned} \mathcal{R}_{\mathcal{S}_{T_1}} &= \{\mathcal{T}_{T_z}, \mathcal{T}_{T_x}, \mathcal{T}_{T_y}\}, \\ \mathcal{R}_{\mathcal{S}_{T_2}} &= \{\mathcal{T}_{T_z}, \mathcal{T}_{R_z}, \mathcal{T}_{T_y}\}, \\ \mathcal{R}_{\mathcal{S}_{T_3}} &= \{\mathcal{T}_{T_z}, \mathcal{T}_{R_z}, \mathcal{T}_{T_x}\}. \end{aligned} \tag{7}$$

A rotation task is followed when the transportation task is executed. A desired orientation $\boldsymbol{d_R} = [\theta_x, \theta_y, \theta_z]^T$ can be reached through three movements, where each of them is a rotation around one of the x, y, z-axes. Therefore, $\boldsymbol{d_R}$ can be reached through the task sequence $\mathcal{S}_R$:

$$\mathcal{S}_R = \{\theta_z \mathcal{T}_{R_z}, \theta_y \mathcal{T}_{R_y}, \theta_x \mathcal{T}_{R_x}\}. \tag{8}$$

Partial terms of $\boldsymbol{d_R}$ can be zero for different rotation tasks, e.g. to pour liquid from a container for a target pouring region. Two desired orientations $\boldsymbol{d_{R_1}} = [0, \theta_{y_1}, \theta_{z_1}]^T$ and $\boldsymbol{d_{R_2}} = [\theta_{x_2}, 0, \theta_{z_2}]^T$ are typically used for such tasks, as illustrated in Fig. 2(b). The sequences $\mathcal{S}_{R_{1,2}}$ and the relevant modular tasks $\mathcal{R}_{\mathcal{S}_{R_{1,2}}}$ to reach the orientations are:

$$\begin{aligned} \mathcal{S}_{R_1} &= \{\theta_{z_1} \mathcal{T}_{R_z}, \theta_{y_1} \mathcal{T}_{R_y}\}, \mathcal{R}_{\mathcal{S}_{R_1}} = \{\mathcal{T}_{R_z}, \mathcal{T}_{R_y}\}, \\ \mathcal{S}_{R_2} &= \{\theta_{z_2} \mathcal{T}_{R_z}, \theta_{x_2} \mathcal{T}_{R_x}\}, \mathcal{R}_{\mathcal{S}_{R_2}} = \{\mathcal{T}_{R_z}, \mathcal{T}_{R_x}\}. \end{aligned} \tag{9}$$

The main task will be estimated to be successful, if the following conditions hold:

1. The desired position $\boldsymbol{d_T}$ and orientation $\boldsymbol{d_R}$ can be reached through a sequence of the six motion tasks $\{\mathcal{S}_T, \mathcal{S}_R\}$.
2. T_{holding} and all the relevant motion tasks $\{\mathcal{R}_{\mathcal{S}_T}, \mathcal{R}_{\mathcal{S}_R}\}$ of the chosen sequence are inferred to succeed.

Moreover, the relevance of the modular tasks may be further improved with human hints. For instance, in an industry environment or a tele-operation scenario, the human operator can remove the holding task when the objects are deformable and non-fragile, such as sponges or soft toys. The false negatives can thereby be further reduced and unnecessary grasp adaptations can be avoided.

IV. DATA ACQUISITION

A. Hardware setup

The hardware setup used for the data acquisition is shown in Fig. 3. We use a Schunk parallel gripper, which is mounted on a KUKA lightweight robot arm. We attach an Intel®RealSense™ SR300 RGBD camera on top of the gripper to localize the object and to estimate the object size directly from the captured point cloud. We apply the statistical outlier removal from the Point Cloud Library (PCL) [24] to remove noise. A Weiss Robotics WTS tactile sensor [25] with 14×6 taxels is mounted on one gripper finger, where the size of each taxel is 3.4 mm $\times$ 3.4 mm. In addition, we cover both fingers with a thin layer of rubber sheet to smooth the surface of the tactile sensor and to increase the grasp stability.

B. Data collection

We describe the grasp strategy to collect data, including the approach vectors, the grasping locations, and the grasp forces. To reduce the space of possible grasping postures, the object shape can be represented with shape primitives [26], [27]. In this work, we estimate the three-dimensional object size from the point cloud and approximate it with a cuboid. When collecting data for each object, we use two approach directions $\vec{a}_1$ and $\vec{a}_2$, which are perpendicular to the planes of both sides of the cuboid, respectively. For each approach direction, N_G locations are equally spread vertically along the object.

To collect data for each object, we select a set of N_F grasping forces $\boldsymbol{F} = \{F_1, \ldots, F_{N_F}\}$ to squeeze each grasp location. Each object has an individual set of N_F grasp forces. However, the ratio ξ between each grasp force and the mass of the object is a constant over all N_O objects, where the mass is estimated by the volume of the cuboid multiplied by the water density. This implies that for the i-th object with the mass m^i, the set of grasp forces F_j^i with $j = 1, \ldots, N_F$ is computed by:

$$\frac{F_j^i}{m^i} = \xi_j, \forall i = 1, \ldots, N_O, \text{ where } \xi_j \text{ are constants.} \tag{10}$$

The grasp forces are determined such that the classifier for grasp stability is invariant to the object weight. The equally

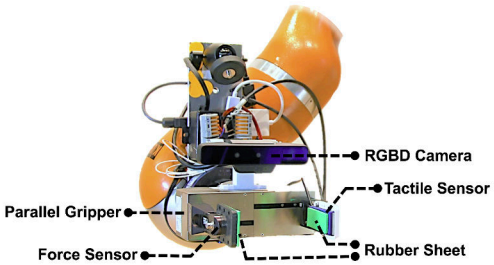


Fig. 3: Our hardware setup includes a parallel gripper, a RGBD camera, and a tactile sensor, which provides a 14×6 2D pressure array.

scaled forces are used for force adaptation as well. When a grasp is predicted to be unstable, the force with the next larger scale will be applied.

Finally, when the gripper has reached a grasp location, the fingers are closed until the planned grasp force is applied. The data including the grasp location, object deformation and the tactile sensor readings are recorded prior to the execution of each modular task.

To label the grasp data, the six motion tasks are evaluated separately by moving the arm accordantly with a maximum speed after reaching the planned grasp. However, the object is first lifted along the z-axis prior to each movement, such that no support-force acts on the object during the manipulation. Each motion task has an individual moving range due to the limited workspace, where the translation along the x, y-axes are within $[-0.5\,\mathrm{m}, 0.5\,\mathrm{m}]$, the range for lifting the object is $[0.0\,\mathrm{m}, 0.7\,\mathrm{m}]$. The rotation around the x, z-axes is within $[-90°, 90°]$, while the range of the rotation around the y-axis is $[0, 90°]$.

The task $\mathcal{T}_{\text{holding}}$ uses the data with a grasp force to predict the success for the force with the next larger scale in the set $\{F_1, \ldots, F_{N_F}\}$. Therefore, the set of grasp forces for the $\mathcal{T}_{\text{holding}}$ is $\{F_0, F_1, \ldots, F_{N_F-1}\}$, where $F_0 = 0.5 * F_1$.

C. Feature representation

The features used in this work are the grasping location H, the object deformation D, and analytical features extracted from the tactile sensor readings $\boldsymbol{X}$.

The grasping location H^k of the observation k is computed by the center location of two fingertips in the object-centered coordinate system and normalized by dividing it by the object height, such that it becomes invariant to the object scale.

The object deformation D^k of the k-th observation is computed by:

$$D^k = \frac{L_S^k - L_E^k}{L_S^k}, \tag{11}$$

where L_S^k is the start grasp length of the gripper and is recorded when both fingers and the object are just in contact. L_E^k is the end grasp length when the contact force reached the planned force. The deformation is normalized by dividing it by L_S^k, since it should be relative to the original length of the object at this location. The proposed analytical features

Fig. 4: Selected objects for the experiments, including rigid objects, plastic cups, non-fragile objects, and plastic bottles. The bottles with open and close states are marked with a star. The dashed box marks the objects for testing.

extracted from the tactile sensor readings are described in Sec. III-A.

Similar to [2], we normalize the features to zero-mean and unit standard deviation, with the exception of the grasping location and the friction-weighted pressure center, since they have a fixed range.

Other methods for dimensionality reduction of the tactile sensor readings exist, such as principal component analysis (PCA) applied in [4] and the image moments applied in [1], [2], [5], [9]–[11]. We compare the proposed analytical features to these two methods. Eight principal components are used in the experiments, which explain ca. 90% of the data. The image moments $m_{p,q}$ for the tactile sensor readings are computed by:

$$m_{p,q} = \sum_{x=1}^{S_x} \sum_{y=1}^{S_y} x^p y^q \rho_{xy}. \tag{12}$$

We use the image moments up to order two, which means $(p+q) \in \{0, 1, 2\}$. Therefore, there are six moments in total.

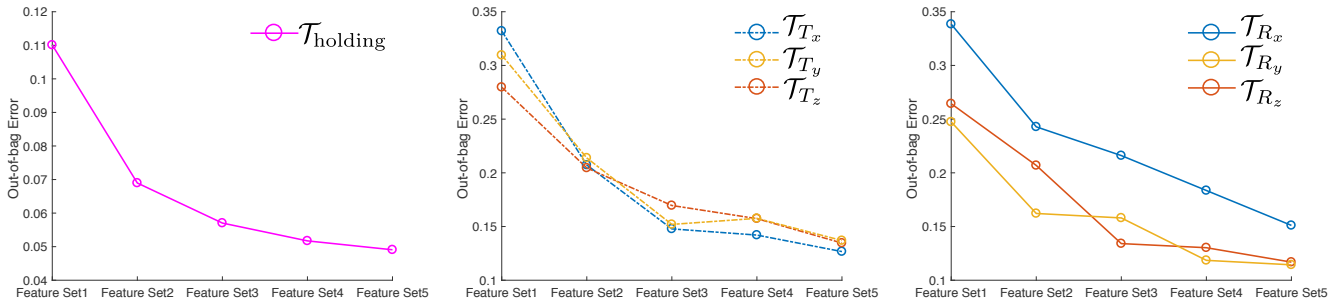
V. EXPERIMENTS AND RESULTS

A. Experimental setup

We evaluate two aspects of the proposed approach: the classification performance for each modular task with the proposed feature set and the prediction result for the main tasks with and without dividing the training task into modular ones.

We select 21 objects for the experiment, as summarized in Fig. 4, including rigid objects, plastic cups, plastic bottles, and deformable non-fragile objects. The bottles that have two states, open and closed, are marked with a star in Fig. 4. The plastic cups and bottles are filled with liquid, which are covered with a plastic material on the top during the experiment, such that the liquid is able to overflow, but will not destroy the electrical devices. The holding task is considered successful, when the bottles are closed. It is labeled as a failure, when the objects are damaged or the liquid overflows with a planned grasp force.

Two grasp forces are selected for the experiment. The ratios ξ_1, ξ_2 between each force and the estimated mass of the object are 10 and 60, which are determined based on



(a) The holding task. (b) The translational motion tasks. (c) The rotational motion tasks.

Fig. 5: The out-of-bag (OOB) error for five sets of features for seven modular tasks. Feature set 5 (proposed) achieves the lowest OOB error for all tasks. Large improvements can be found for translational modular tasks and the task $\mathcal{T}_{R_x}$.

TABLE I: Comparison of the prediction error for seven modular tasks and the average error for all tasks.

	$\mathcal{T}_{\text{holding}}$		$\mathcal{T}_{T_x}$		$\mathcal{T}_{T_y}$		$\mathcal{T}_{T_z}$		$\mathcal{T}_{R_x}$		$\mathcal{T}_{R_y}$		$\mathcal{T}_{R_z}$		All 7 Tasks
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean
Set 3	0.4	0.370	0.188	0.132	0.189	0.136	0.219	0.160	0.225	0.100	0.274	0.285	0.181	0.144	0.2394
Set 4	**0.370**	0.299	0.200	0.146	0.202	0.129	0.224	0.148	0.280	0.110	0.320	0.267	0.177	0.122	0.2533
Set 5	0.410	0.393	**0.178**	0.151	**0.135**	0.129	**0.166**	0.150	**0.160**	0.119	**0.175**	0.101	**0.151**	0.132	**0.1964**

an overestimated ($\approx$ 1) and an underestimated ($\approx$ 0.16) friction coefficient. Each object is grasped with two approach vectors. For each approach vector, 6-12 locations are grasped depending on the height of the object. Each location is grasped with two forces for each of the motion tasks and three forces for the holding task, where the ratio ξ_0 for the smallest force F_0 is 5. This force is used to predict whether the holding task will succeed when the grasp force F_1 with the ratio $\xi_1 = 10$ is applied. The grasp data for the plastic containers, that are marked with a star in Fig. 4, are collected twice for its open and closed state.

B. Prediction results for modular tasks

To infer the success of modular tasks, we train 1000 decision trees for each task. We compare the proposed analytical features extracted from the tactile sensor readings with features based on principle components and image moments as introduced in Section IV-C. The forests are trained with the data of all 21 objects and we find the prediction error over the out-of-bag (OOB) data, which is defined as OOB error in this work.

Table II summarized the five combinations of features that are evaluated in our experiments.

TABLE II: Considered feature sets for the OOB error comparison. Feature set 5 is proposed.

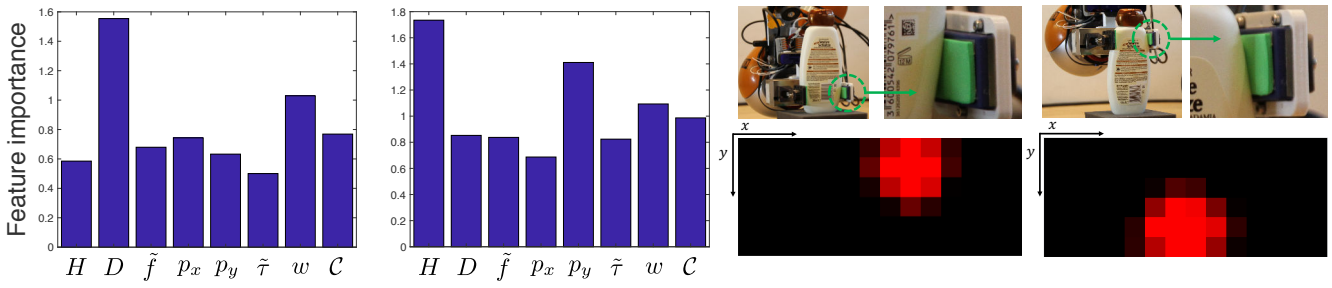
Feature set	Basic features	Tactile sensor readings
1	H	/
2	H,D	/
3	H,D	6 image moments
4	H,D	8 principle components
5	H,D	$\mathcal{C}, \boldsymbol{p}, \tilde{f}, \tilde{\tau}, w$

The OOB error for the seven modular tasks is shown in Fig. 5. The proposed feature set yields a lower OOB error for all the tasks. Large improvements can be found for translational modular tasks $\mathcal{T}_{T_x}, \mathcal{T}_{T_y}, \mathcal{T}_{T_z}$ and the task $\mathcal{T}_{R_x}$.

Next, we select 13 objects for training and 8 for testing to compare the feature sets 3-5 in Table II. The objects for testing are marked with a dashed box in Fig. 4. Table I summarizes the mean and the standard deviation (std) of the prediction error for each modular task and the average error for all tasks. We are able to achieve a prediction accuracy of 80.36% with the proposed features extracted from the tactile sensor readings. The prediction error is lowered by up to 5.69% compared to other feature sets. The error for the holding task is high compared to other tasks. In particular, the false negatives of the pink cup (a soft toy) in Fig. 4 is 100%, due to the excessive deformation caused by the grasp. When manipulating such objects, a prior-knowledge about whether the object is fragile, can avoid the false negatives. The experiment is described in Section V-C.

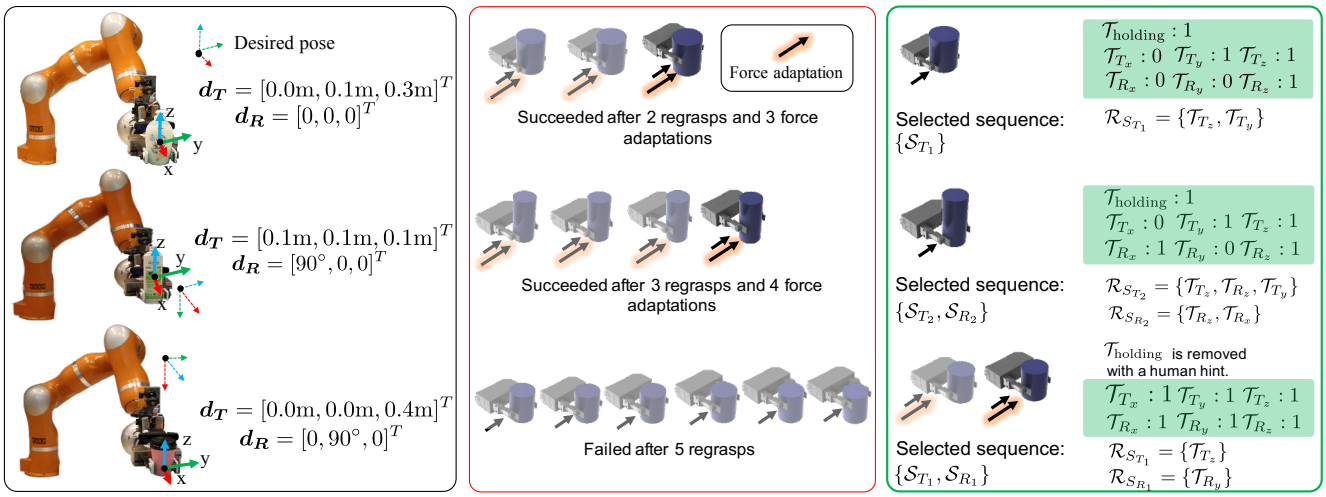
Finally, we analyze the importance of each proposed feature for two selected modular tasks in Fig. 6(a) and (b). The importance is measured by the increment of the prediction error, when permuting the values of each feature.

The object deformation carries a high importance for the holding task, as shown in Fig. 6(a). The work w is the second most important feature, since it indirectly measures the object deformation. Fig. 6(b) shows that the grasping location is essential for the lifting task $\mathcal{T}_{T_z}$. The friction-weighted pressure center $\boldsymbol{p}$ on the y-axis p_y is important for the task as well, since it encodes the object local geometry. Two grasp examples are illustrated in Fig. 6(c) and (d) with an enlarged view of each contact and the corresponding tactile sensor reading. Fig. 6(c) is a successful grasp for a lifting task while Fig. 6(d) fails. The pressure center in Fig. 6(c) is on the upper part of the reading, which indicates a lower value of p_y. It implies that the contact area of the



(a) $\mathcal{T}_{\text{holding}}$ (b) $\mathcal{T}_{T_z}$ (c) Successful grasp for the $\mathcal{T}_{T_z}$. (d) Failed grasp for the $\mathcal{T}_{T_z}$.

Fig. 6: Importance plot of features for (a) the modular task $\mathcal{T}_{\text{holding}}$ and (b) the lifting task $\mathcal{T}_{T_z}$. The vertical location p_y of the friction-weighted pressure center encodes object local geometry and therefore carries a high importance for the $\mathcal{T}_{T_z}$. Fig. (c) and (d) show a successful and a failed grasp for the $\mathcal{T}_{T_z}$ with an enlarged view of the contact and the tactile sensor readings. If the center is on the upper part of the tactile image, the lift-up task is likely to succeed.



(a) Manipulation tasks. (b) Grasp processes without modular tasks. (c) Grasp processes with modular tasks (proposed).

Fig. 7: Grasp processes for three example manipulation tasks. The considered sequences of modular tasks $\mathcal{S}_R, \mathcal{S}_T$ can be found in Fig. 2. With the proposed method, much less grasp adaptations are required to successfully execute the manipulation tasks, as depicted in (c). Note that for the last main task, $\mathcal{T}_{\text{holding}}$ is predicted to fail for all grasp locations and forces due to the excessive deformation. With a prior-knowledge that the object is non-fragile, it can be successfully manipulated with the proposed method, while the manipulation fails in (b) after all grasp candidates are attempted.

object is wider on the upper part than on the lower part. Therefore, it is likely that there is a support structure above the object part that is in contact with the gripper when the object is lifted.

C. Assessing grasp stability for main tasks

We infer the success of the main manipulation tasks and the paths for the task execution based on the prediction results of the modular tasks. We compare the manipulation process of the proposed approach with the case, where the model for stability assessment is trained for a fixed task. In the latter case, the training task is considered as a sequence of all seven tasks. The success label will only be true if the whole task sequence can be successfully executed. Such conservative training process can be expected to avoid false positives and a failure of the action, when the new manipulation task is unknown. A preliminary planner is applied to demonstrate the difference. The object will be grasped starting the lowest location among the grasp candidates. If the grasp for the main task is predicted to be unsuccessful, a grasp force adaptation will be triggered with a larger force ($\xi = 60$) used in training. If the task is further inferred to fail, a higher location will be selected for a regrasp. The process will be repeated until a feasible grasp location is detected or all grasp locations are predicted to fail. The results for three example manipulation tasks are illustrated in Fig. 7, where the target pose of each task is manually defined. Without considering modular tasks, the prediction of the success will only be true for the main task, if the whole task sequence can succeed, i.e. all modular tasks are inferred to be successful. Robotic experiments show that

numerous unnecessary grasp force adaptations and regrasps are avoided with the proposed approach.

VI. CONCLUSION

In this work, we present a learning-based approach for grasp stability assessment for seven modular tasks including a holding task and six motions tasks to manipulate deformable fragile objects. Our experiments demonstrate that the average prediction accuracy achieves 80.36% for all modular tasks. The proposed analytical features extracted from the tactile sensor readings lower the prediction error up to 5.69% compared to the classical methods for dimensionality reduction, i.e. image moments and principle component analysis. Furthermore, we show that a manipulation task can be represented as a sequence of modular tasks and the stability can be inferred thereby based on the prediction of relevant modular tasks. We demonstrate that unnecessary grasp force adaptations and regrasps can be avoided with the proposed stability assessment for modular tasks.

Limitation analysis: The OOB error and the importance plots of features suggests that partial modular tasks might be correlated, such as $\mathcal{T}_{T_z}, \mathcal{T}_{T_y}$ and $\mathcal{T}_{R_z}$. When collecting data for these tasks, we observed that the success rates are similar among them. This may be caused by the limited acceleration of the three movements. Therefore, the division of the modular tasks needs to be further investigated. In addition, the labeling of the modular tasks are under the restriction that the approach vectors of the grasps are from the side. The prediction results might differ when the grasps are from the top. Moreover, further analysis is required to evaluate whether a complicated manipulation task can be guaranteed to be successful, when all relevant modular tasks succeeded. To investigate this, analytical approaches can be used to determine the required force and torque components for the main and modular tasks. Finally, the path to reach the target pose is restricted by the prediction results of the six motion tasks. A probabilistic framework can be investigated to combine the stability assessment with path planning. The success rate of each path, which is provided by, e.g., inverse kinematics can be estimated. The optimal path can be thereby selected for safe manipulation.

Future work: In addition to addressing the aforementioned limitations, we will consider more modular tasks, such as accelerations and decelerations, and more grasps to further investigate their dependencies. We will combine the grasp stability assessment with a more sophisticated grasp planner and grasp adaptation strategies for a more complete manipulation process.

REFERENCES

[1] Y. Bekiroglu, D. Song, L. Wang, and D. Kragic, "A probabilistic framework for task-oriented grasp stability assessment," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[2] Y. Bekiroglu, J. Laaksonen, J. A. Jorgensen, V. Kyrki, and D. Kragic, "Assessing grasp stability based on learning and haptic data," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 616–629, 2011.

[3] H. Dang and P. K. Allen, "Learning grasp stability," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012.

[4] M. Li, Y. Bekiroglu, D. Kragic, and A. Billard, "Learning of grasp adaptation through experience and tactile sensing," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014.

[5] Y. Bekiroglu, R. Detry, and D. Kragic, "Learning tactile characterizations of object-and pose-specific grasps," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011.

[6] D. Cockburn, J. Roberge, T. Le, A. Maslyczyk, and V. Duchaine, "Grasp stability assessment through unsupervised feature learning of tactile images," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[7] R. Krug, A. J. Lilienthal, D. Kragic, and Y. Bekiroglu, "Analytic grasp success prediction with tactile feedback," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[8] K. Hang, M. Li, J. A. Stork, Y. Bekiroglu, F. T. Pokorny, A. Billard, and D. Kragic, "Hierarchical fingertip space: A unified framework for grasp planning and in-hand grasp adaptation," *IEEE Transactions on robotics*, vol. 32, no. 4, pp. 960–972, 2016.

[9] Y. Bekiroglu, D. Kragic, and V. Kyrki, "Learning grasp stability based on tactile data and hmms," in *RO-MAN, IEEE*, 2010.

[10] J. Schill, J. Laaksonen, M. Przybylski, V. Kyrki, T. Asfour, and R. Dillmann, "Learning continuous grasp stability for a humanoid robot hand based on tactile sensing," in *IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012.

[11] E. Hyttinen, D. Kragic, and R. Detry, "Learning the tactile signatures of prototypical object parts for robust part-based grasping of novel objects," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[12] J. Laaksonen, V. Kyrki, and D. Kragic, "Evaluation of feature representation and machine learning methods in grasp stability learning," in *10th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2010.

[13] R. Calandra, A. Owens, M. Upadhyaya, W. Yuan, J. Lin, E. H. Adelson, and S. Levine, "The feeling of success: Does touch sensing help predict grasp outcomes?" *arXiv preprint arXiv:1710.05512*, 2017.

[14] J. Kwiatkowski, D. Cockburn, and V. Duchaine, "Grasp stability assessment through the fusion of proprioception and tactile signals using convolutional neural networks," 2017.

[15] Y. Bekiroglu, A. Damianou, R. Detry, J. A. Stork, D. Kragic, and C. H. Ek, "Probabilistic consolidation of grasp experience," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[16] C. Ferrari and J. Canny, "Planning optimal grasps," in *IEEE International Conference on Robotics and Automation (ICRA)*, 1992.

[17] J. Xu, N. Alt, Z. Zhang, and E. Steinbach, "Grasping posture estimation for a two-finger parallel gripper with soft material jaws using a curved contact area friction model," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[18] Y.-B. Jia, F. Guo, and H. Lin, "Grasping deformable planar objects: Squeeze, stick/slip analysis, and energy-based optimalities," *The International Journal of Robotics Research*, vol. 33, no. 6, pp. 866–897, 2014.

[19] R. D. Howe and M. R. Cutkosky, "Practical force-motion models for sliding manipulation," *The International Journal of Robotics Research*, vol. 15, no. 6, pp. 557–572, 1996.

[20] S. Goyal, A. Ruina, and J. Papadopoulos, "Planar sliding with dry friction part 1. limit surface and moment function," *Wear*, vol. 143, no. 2, pp. 307–330, 1991.

[21] Coefficient of friction equation and table chart. http://www.engineersedge.com/coeffients_of_friction.htm. Last visited: July 2018.

[22] K. Hang, F. T. Pokorny, and D. Kragic, "Friction coefficients and grasp synthesis," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013.

[23] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[24] Point cloud library. http://pointclouds.org. Last visited: July 2018.

[25] Weiss robotics tactile sensor. [Online]. Available: https://www.weiss-robotics.com/de/produkte/taktile-sensorik. Last visited: Sept. 2017.

[26] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen, "Automatic grasp planning using shape primitives," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2003.

[27] C. Goldfeder, P. K. Allen, C. Lackner, and R. Pelossof, "Grasp planning via decomposition trees," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2007.