

# Learning Tactile Characterizations Of Object- And Pose-specific Grasps

Yasemin Bekiroglu      Renaud Detry      Danica Kragic

**Abstract**—Our aim is to predict the stability of a grasp from the perceptions available to a robot before attempting to lift up and transport an object. The percepts we consider consist of the tactile imprints and the object-gripper configuration read before and until the robot’s manipulator is fully closed around an object. Our robot is equipped with multiple tactile sensing arrays and it is able to track the pose of an object during the application of a grasp. We present a kernel-logistic-regression model of pose- and touch-conditional grasp success probability which we train on grasp data collected by letting the robot experience the effect on tactile and visual signals of grasps suggested by a teacher, and letting the robot verify which grasps can be used to rigidly control the object. We consider models defined on several subspaces of our input data – e.g., using tactile perceptions or pose information only. Our experiment demonstrates that joint tactile and pose-based perceptions carry valuable grasp-related information, as models trained on both hand poses and tactile parameters perform better than the models trained exclusively on one perceptual input.

## I. INTRODUCTION

This paper studies the exploitation of tactile, visual, and proprioceptive data for assessing stability in both planning and executing grasps.

Grasp planning relies on (1) the extraction of information from the agent’s environment (e.g., through vision), and on (2) the recovery of memories related to the current environmental configuration (e.g., previous attempts to grasp a particular object). Because of the uncertainty inherent to these two processes, designing grasp plans that are guaranteed to work in an open-loop system is difficult. Grasp execution can thus greatly benefit from a closed-loop controller which considers sensory feedback before and while issuing motor commands.

Humans make extensive use of input from several sensor modalities when executing grasps [2], [3]. Clearly, vision is one of the modalities which contribute substantially to grasp control and stability [4], [5], [6]. Touch is another one, as supported by numerous studies which show the influence of tactile feedback on different grasp sub-processes [2], [3], [7], [1], [8]. For example, Johansson and Westling [7] have shown that anesthetizing a subject’s fingers – thereby impairing his sense of touch while leaving his motor capabilities intact – directly leads to a loss in the subject’s proficiency in grasping and lifting up objects. These observations are reflected in the work of Johansson et al. [2], [7], [1] who presented an empirical formalization of the human grasping

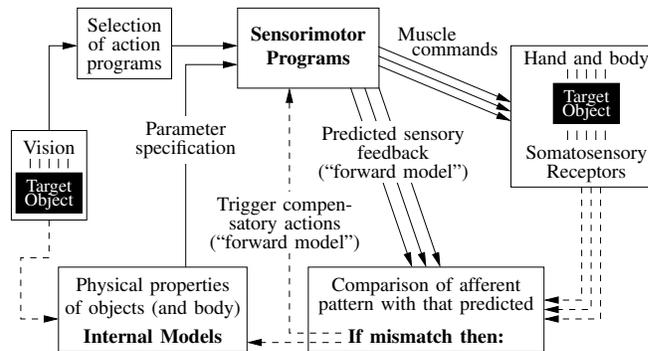


Fig. 1: Reproduction of “Predictive ‘feed-forward’ sensory control of manipulation”, from R. Johansson [1], Fig. 5, p. 56. In Johansson’s work [1], the sensorimotor planning and execution of manipulative tasks is formalized with a closed-loop system which integrates dynamic touch and visual perception with sensorimotor memories. A manipulative action is planned from visual input, which triggers an appropriate learned action program. As motor commands are being issued, sensor signals are continuously compared to values predicted by an internal forward model, which permits the detection of unexpected events. In turn, unexpected events trigger recovery procedures and the refinement of the forward model.

behavior as a closed-loop system involving visual and touch feedback and a memory-based controller, see Figure 1. A key part of this work emphasized humans’ ability to predict the repercussion of manipulative actions onto sensory channels by means of a (learned) forward model, thereby allowing us to react to unexpected situation and maintain grasp stability.

In robotics, vision-driven grasping and manipulation have been extensively studied [9], [10]. Vision has typically been used to plan grasping actions, and to update action parameters as objects move. Touch-based grasp controllers have also been studied, with emphasis on designing programs for controlling finger forces to avoid slippage and to prevent crushing objects [11], [12], [13].

In this paper, we discuss means of *learning* some of the models and sensorimotor programs which contribute to the system depicted in Figure 1. By observing the sensor signals issued during the execution of grasps demonstrated by a human, our agent learns what it feels like to grasp an object from a specific side, and learns which grasping configurations lead to a stable grasp. When planning a grasp, the agent is able to compute an initial estimate of the stability of the planned grasp. As the grasp is being executed and

Y. Bekiroglu, R. Detry and D. Kragic are with the Centre for Autonomous Systems, CSC, KTH, Stockholm, Sweden. {yaseminb,detryr,danik}@csc.kth.se.

This work was supported by the Swedish Foundation for Strategic Research, and by the EU project COGX (FP7-IP-027657).

the manipulator’s fingers are brought around the object, the pose (3D position and 3D orientation) of the object is continuously tracked. When fingers come in contact with the object, afferent tactile signals are compared to the signals predicted by the learned feed-forward model for the current object-gripper configuration, yielding an updated estimate of the stability of the grasp. In a learning scenario, the agent can then proceed with an attempt to transport and shake the object to gather an empirical confirmation of its stability assessment, possibly updating the feed-forward model. During execution, if the stability estimate is too low, the agent may decide to move the manipulator to a better configuration before lifting up the object.

In mathematical terms, our agent learns an empirical representation of pose- and/or touch-conditional grasp success probability. This model predicts the stability of a grasp from tactile data and/or object-gripper pose parameters. We consider models defined on several subspaces of our input data – e.g., using only tactile perceptions or pose information. Models are optimized and evaluated with  $f$ -fold cross-validation. This model is presented in Section III.

To our knowledge, learning to assess grasp stability jointly from live visual and tactile feedback has not been attempted before. This experiment poses a number of technical challenges. As an object will often move while the robot is closing its hand to grasp it, the agent needs to track the pose of the object during the grasp, which is made difficult by the partial object occlusions effected by the robot hand. Section III presents an overview of the robotic system we implemented to run our experiment.

We experimentally demonstrate that considering both visual and tactile input is beneficial. While the relevance of tactile and pose data varies between objects, models trained on both pose and tactile parameters perform at least as well as the models trained exclusively on hand poses or tactile signals. This result shows that, despite the modeling difficulties associated to an increased perceptual dimensionality, our learning algorithm successfully identifies the discriminative characteristics within the joint pose and tactile channel.

## II. RELATED WORK

Our work is related to vision-based grasp planning, tactile sensing, and robot learning.

Grasp planning was often approached by approximating object shape with a number of shape primitives such as boxes, cylinders, cones, spheres [14] and superquadrics [15] in order to limit the number of possible grasps and prune the search space to find stable grasps. Borst et al. [16] reduced the number of candidate grasps by generating random grasps dependent on the object surface and filtering them with a simple heuristic. Ciocarlie et al. [17] reduced the configuration space of a robotic hand to find pre-grasp postures from which the system searched for stable grasps. Li et al. [18] utilized a user-created database of human grasps. After a shape matching algorithm found the hand shape that best matched the query object, the alignment of the hand pose

to the object shape was determined. The resulting candidate grasps were clustered and pruned depending on the task.

Learning aspects were considered in the context of grasping some of which focused on understanding human grasping strategies. Ekvall and Kragic [19] demonstrated how a robot system could learn grasping by human demonstration using a grasp experience database. The human grasp was recognized with the help of a magnetic tracking system and mapped to the kinematics of the robot hand using a predefined lookup-table. Learning was also used to infer good grasping configurations based on visual input. Saxena et al. [20] introduced a system that learned grasping points by using hand labeled training data in the form of image regions which indicated good grasping points. A probabilistic decision system was then applied to previously unseen objects to determine a good grasping point or a region. Letting a robot learn good grasping configurations based on exploration was also studied [21]. Erkan et al. [22] presented a probabilistic approach to model the success probabilities of grasp configurations obtained from visual descriptors and combined active and semisupervised learning to tackle the scarcity of labeled grasps.

In prior studies, tactile sensing was used for various purposes such as determining object pose [23] and the surface type (edge, flat, cylindrical, spherical) of the tactile contact [24], extracting object shape [25] and recognizing objects [26], [27]. Morales et al. [28] used tactile data to maximize contact surfaces to remove a book from a bookshelf. Application of force, visual and tactile feedback to open a sliding door was proposed by Prats et al. [29] where tactile control ensured that an accurate alignment between the hand and the handle was kept.

Differently from the aforementioned approaches, we use tactile feedback along with visual feedback to differentiate between successful and unsuccessful grasping configurations before further manipulating the object. The visual feedback is obtained by using a real-time tracker during grasping. Compared to our previous work on tactile learning [30] where assessing grasp stability was based on only tactile input extracted during a grasp, this work makes use of vision-based object pose to further discriminate between stable and unstable grasps.

## III. LEARNING GRASP STABILITY

Our aim is to infer grasp stability from the tactile imprints and the object-gripper configuration available *before* lifting up an object, and to provide the agent with means of learning from experience of how to make stability assessments.

### A. Perceptual Input

Our robot platform is composed of an industrial arm, a three-finger hand and a monocular camera, see Figure 2. Tactile imprints are delivered by pressure-sensing arrays deployed on the hand. Each of the hand’s three fingers is composed of two segments, both covered by an array, yielding a total of 6 tactile arrays, see Figure 2. The tactile data is relatively high-dimensional and to some extent

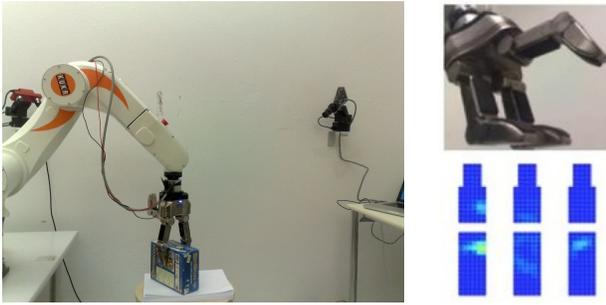


Fig. 2: Experimental robotic platform, composed of an industrial arm, a three-finger gripper equipped with tactile sensing arrays, and a camera. There are six tactile sensing arrays. The rightmost image shows an example of tactile readings obtained during a grasp.

redundant. Therefore, we start by representing the acquired data as features. Here, we borrow some ideas from image processing and consider the two-dimensional tactile patches as images. In order to reduce their dimensionality, we extract these images’ moments, as suggested in the literature [31], [32]. The general parametrization of image moments for one tactile array  $A$  is given by

$$m_{p,q} = \sum_i \sum_j i^p j^q A_{ij} \quad (1)$$

where  $p$  and  $q$  represent the order of the moment, and  $i$  and  $j$  represent the horizontal and vertical position on the tactile patch. We compute moments up to order two,  $(p + q) \in \{0, 1, 2\}$ , which yields 6 numbers that model the total pressure and the distribution of the pressure in the horizontal and vertical direction. We denote a tactile input vector by  $t$ . Such a vector contains moments from the six tactile pads and therefore holds  $6 \times 6$  numbers.

Through visual and proprioceptive feedback, our platform is able to acquire object and gripper poses in real time. Gripper poses are simply obtained from the kinematics of the robot. Obtaining object poses is more challenging. As an object will often move while the robot is closing its hand to grasp it, the agent needs to compute the pose of the object *after* having closed the hand around it. This computation is made difficult by the partial object occlusions effected by the hand. Our aim however is not to get perfectly accurate pose information, but rather a rough idea of how the object is approached. We address this issue by tracking the movement of the object for the complete duration of the grasp. We are currently using a system which tracks the pose of a textured CAD model in a monocular video stream [33]. Tracking object textures greatly helps handling partial object occlusions and distractions induced by the hand.

We aim at designing a stability predictor that is independent of the position of an object. For this reason, we do not predict stability from the manipulator and object poses directly. Instead, we base our predictions on the *relative* object-manipulator pose. Object-relative manipulator configurations allow our system to encode notions such as

“grasping a bottle from the side is better than grasping it from the top.” However, stability will often not only depend on the relative object-gripper configuration, but also on the absolute orientation of the object. When an elongated object lies on a flat surface, it is generally better to grasp it close to its center of mass. Yet, if the object is standing, grasping it near its tip is acceptable. As a result, we also base our predictions on the angle between the gripper’s approach vector and a direction aligned with gravity.

## B. Stability Classification

We predict grasp stability with object-specific classifiers trained to discriminate between percepts that lead to stable or unstable grasps for a specific object. Our agent learns an empirical representation of pose- and touch-conditional grasp stability probability. This model is learned from a set of examples denoted by

$$Z = \{(x_i, y_i)\}_{i=1, \dots, n} \quad (2)$$

where each pair  $(x_i, y_i)$  is composed of perceptual readings  $x_i \in \mathbb{R}^d$  (pose and touch) and a binary stability label  $y_i \in \{\text{stable}, \text{unstable}\}$ . Perceptual data are read during the execution of a grasping plan, shortly after the agent closed the manipulator’s fingers around the object, but before any attempt to lift or transport the object. The probability of pose- and touch-conditional grasp stability is modeled with kernel logistic regression (KLR). Logistic regression is a widely used technique to model class probabilities. Using the kernelized version of logistic regression allows us to model non-linear decision boundaries. In the next paragraph, we give an intuitive explanation of KLR applied to our problem. A short formal description follows. For further details on the theory behind logistic regression and kernel methods, we refer the reader to the work of Yamada et al. [34], Erkan et al. [22], and Schölkopf and Smola [35].

KLR models the stability probability of a grasp characterized by a perceptual vector  $x$  with the help of a weighted sum of the similarities between  $x$  and each vector in the training dataset  $Z$ . The weights associated to stable grasps will generally be positive, while those associated to unstable grasps will be negative. If  $x$  resembles percepts of  $Z$  that lead to stable grasps, its probability of stability will thus be high. In order to restrict values to the  $[0, 1]$  interval, KLR models probabilities by plugging the weighted sum described above into the logistic function  $f(z) = \frac{1}{1+e^{-z}}$ , which smoothly grows from 0 to 1 as its argument varies from minus infinity to infinity. Weights are usually chosen to maximize the probability of the training set.

Formally, we model the probability of pose- and touch-conditional grasp stability as

$$p(y = \text{stable}|x; v) = \frac{1}{1 + \exp\{-\sum_{i=1}^n v_i \mathcal{K}(x, x_i)\}} \quad (3)$$

where  $p(y = \text{stable}|x)$  is the probability of success of a grasp characterized by the tactile and pose vector  $x$ ,  $\mathcal{K}$  is a kernel function that models the similarity between two perceptual

readings and  $v$  is a weight vector chosen to maximize the regularized stability probability of the data

$$-\sum_{i=1}^n \log p(y_i|x_i; v) + c \text{trace}(vKv^T) \quad (4)$$

where  $K$  is the kernel Gram matrix, with  $K_{ij} = \mathcal{K}(x_i, x_j)$ , and  $c$  is a constant. This problem can be solved, e.g., with Newton’s method. For more details, we refer the reader to the work of Yamada et al. [34]. In the experiments below, the constant  $c$  is chosen by cross-validation.

### C. Kernel Function

As explained above, we consider perceptual signals in the form of tactile readings, relative object-gripper configurations, and an angle that represents the tilt of the hand’s approach vector relative to gravity.

A vector  $x$  representing perceptual observations can be written as

$$x = (t, g, a) \quad (5)$$

where  $t$  is the tactile data,  $g$  is the object-relative gripper pose, and  $a$  is the angle between the approach vector and the vertical. The kernel  $\mathcal{K}$  is defined as

$$\mathcal{K}(x_1, x_2) = \mathcal{K}_t(t_1, t_2)\mathcal{K}_g(g_1, g_2)\mathcal{K}_a(a_1, a_2). \quad (6)$$

The kernel function  $\mathcal{K}_t$  simply corresponds to a multivariate isotropic Gaussian function

$$\mathcal{K}_t(t_1, t_2) = \mathcal{G}(t_1; t_2, \sigma_t), \quad (7)$$

where  $\sigma_t$  is a bandwidth parameter. In the next section, an optimal bandwidth is computed by cross-validation.

An object-relative gripper pose is composed of a 3D position and 3D orientation. We define the gripper pose kernel  $\mathcal{K}_g$  as the product of a position and an orientation kernel. Let us denote the decomposition of a pose  $g$  into position and orientation by  $p$  and  $o$  respectively. We define  $\mathcal{K}_g$  with

$$\mathcal{K}_g(g_1, g_2) = \mathcal{G}(p_1; p_2, \sigma_p) \frac{e^{\sigma_o o_1^T o_2} + e^{-\sigma_o o_1^T o_2}}{2} \quad (8)$$

where  $\mathcal{G}$  is a trivariate isotropic Gaussian kernel, the fraction corresponds to a pair of antipodal von-Mises Fisher distributions (Gaussian-like distribution on the rotation group [36], [37]), and the bandwidths  $\sigma_p$  and  $\sigma_o$  are fixed to allow for deviations of 20 mm and  $20^\circ$  respectively. For a more detailed mathematical description and motivation of  $SE(3)$  kernels, we refer the reader to the work of Sudderth et al. [37].

The kernel function  $\mathcal{K}_a$  corresponds to a Gaussian function

$$\mathcal{K}_a(a_1, a_2) = \mathcal{G}(a_1; a_2, \sigma_a), \quad (9)$$

where  $\sigma_a$  is a bandwidth parameter. In the next section, an optimal bandwidth is computed by cross-validation.

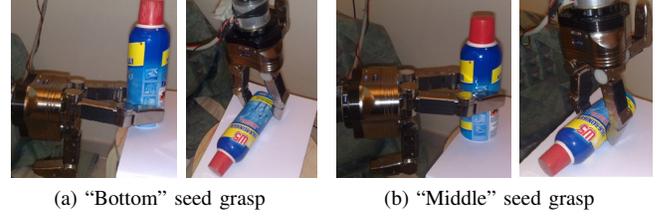


Fig. 3: Seed grasps for a detergent bottle. Each seed grasp is shown when the bottle is standing and lying.

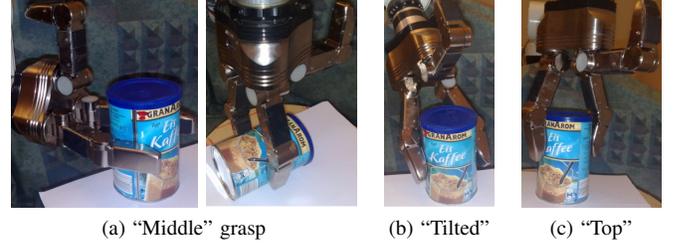


Fig. 4: Seed grasps for a coffee pot.

## IV. EXPERIMENTS

In this section, we present the perceptual data collected by the robot (392 grasps in total), and we discuss classification error rates for pose-based classification, tactile-based classification, and tactile-and-pose-based classification. We present in Section IV-A an experiment in which the agent explores grasping configurations around grasps demonstrated by a human. In Section IV-B, the agent tries grasps uniformly along one edge of an object.

### A. Exploration around Demonstrated Grasps

We ran the first experiment on the two objects shown in Figure 3 and Figure 4. For each object, we demonstrated to the agent sets of two and three “seed” grasps that should be interesting to explore. Each of these grasps was parametrized by the pose of the hand with respect to the object. The agent was then tasked to explore the objects around these grasps. Each grasp trial worked as follows: An object was laid in front of the robot, at an arbitrary position. The standing/lying configuration of the objects also varied. For instance, Figure 3 shows the detergent bottle grasped when standing and when lying on the table. The agent estimated the pose of the object and selected one of the seed grasps available for that object. Let us denote that grasp by  $g_s$ . In order to explore the object in the neighborhood of  $g_s$ , the agent generated a random grasp  $\hat{g}_r$  from  $P(g_r) \propto \mathcal{K}_g(g_s, g_r)$ , where  $\mathcal{K}_g$  is defined by Eq. 8. In effect, this led the agent to explore grasps distributed a few millimeters/degrees away from  $g_s$ . The grasp  $\hat{g}_r$  was executed by the robot. As the hand is rather big with respect to the objects, only two fingers were used. Grasping was run by simultaneously closing the fingers and applying a constant closing force on all joints. Once the hand had stopped, the agent recorded the pose of the object (which was usually different from the initial object pose) and



Fig. 5: Examples of grasps and associated tactile readings. Tactile images correspond to the readings obtained from the frontmost distal array. The rightmost image of each image pair shows an overlay of the object’s shape model aligned to the pose computed by the pose tracker [33].

the tactile imprints. It finally attempted to lift up the object. If lift-up could be achieved robustly, the grasp was marked as stable. If the object slipped or rotated in the hand while being lifted up, the grasp was marked as unstable.

For the detergent bottle of Figure 3, seed grasps were defined at half height, and at the bottom end. Both seed grasps were explored in standing and lying configurations. The seed grasps of the coffee pot are shown in Figure 4. For standing configurations, all three seed grasps were explored. When lying on the table, only the “middle” was tried. Figure 5 shows examples of grasps, associated tactile readings and tracking results.

In this experiment, a total of 342 grasps were collected, i.e., 232 and 110 for the detergent bottle and the coffee pot respectively, with half of these stable and the other half unstable. Each grasp  $i$  consists of the grasp’s perceptual readings  $t_i, g_i, a_i$  as defined by Eq. 5. For each object, tactile moments  $t_i$  were normalized to zero mean and unit variance. In order to evaluate the relevancy of tactile and visual feedback for stability estimation, we studied rates of correct classification for classifiers based on (1) tactile feedback alone, (2) pose feedback alone, and (3) both tactile feedback and pose feedback together. We note that as pose parameters cannot be shared across objects, each classifier is specific to one object – a classifier is learned and evaluated with the data collected for a single object. Stability classification was computed from the probabilistic stability model defined above (3). A grasp characterized by  $x$  was predicted to be stable if  $P(y = \text{stable}|x) > \frac{1}{2}$ . When classifying on tactile imprints or pose exclusively, the kernel of Eq. 6 was redefined as  $\mathcal{K}(x_1, x_2) = \mathcal{K}_t(t_1, t_2)$  or  $\mathcal{K}(x_1, x_2) = \mathcal{K}_g(g_1, g_2)\mathcal{K}_a(a_1, a_2)$  respectively. We computed success rates by ten-fold cross-validation. Cross-validation was run for several values of the tactile kernel bandwidth parameter  $\sigma_t$  (values between 0.5 and 5), and several values of the regularization constant  $c$  (see Eq. 4). Rates obtained with the best parameters are presented in Table I. For the detergent bottle, considering pose and tactile feedback jointly yields a higher classification rate than considering either pose or tactile alone. The bottle was explored around two seed points, both when standing and lying. When standing, both seed grasps led to stable and unstable grasps. However, when lying, most grasps around the bottom of the bottle were unstable, while grasps around its center were both stable

	Detergent	Coffee pot
Tactile feedback only	82%	82%
Pose feedback only	90%	73%
Pose and tactile feedback	93%	82%

TABLE I: Correct classification rates from ten-fold cross-validation of three variants of the stability classification model for the detergent bottle and the coffee pot.

and unstable. Tactile feedback alone can difficultly make a difference between a grasp applied to the bottom of the standing bottle and the same grasp while the bottle is lying. For these grasps, the pose information (in the form of the angle of the grasp approach with the vertical) allows the classifier to separate stable and unstable grasps. For grasps applied around the center of the bottle, pose information allows the model to make reasonably good predictions, but taking tactile feedback into account refines these predictions.

For the grasps tried on the coffee pot, tactile feedback provides a better classification than pose, and considering both tactile and pose yields the same rate as tactile alone. The coffee pot is a rather light object compared to the detergent bottle. As a result of its low weight, the dependency of grasp stability on the standing/lying configuration of the object was less important than for the bottle. Tactile imprints however provided equally good stability assessments.

We also evaluated classification rates as a function of the amount of data available to the agent. Using fixed values for  $\sigma_t$  and  $c$ , we ran ten-fold cross-validations on increasingly large subsets of the collected data. We considered fractions of the data going from 20% to 100% of the total collected data, for each of which we ran multiple cross-validations. The mean classification rates are shown in green in Figure 6. These graphics show that for the detergent bottle, even small numbers of examples allow for robust pose-based classification. For the coffee pot, collecting more data allows for more robust pose predictions. Red-shaded areas show one standard deviation. In all cases, the variance of the predictor decreases as the number of examples increases.

### B. Exploration along the Top Edge of a Box

In the second experiment, the agent explored grasp poses distributed alongside the top side of a box (see Figure 7). Our aim with this experiment was to study how the

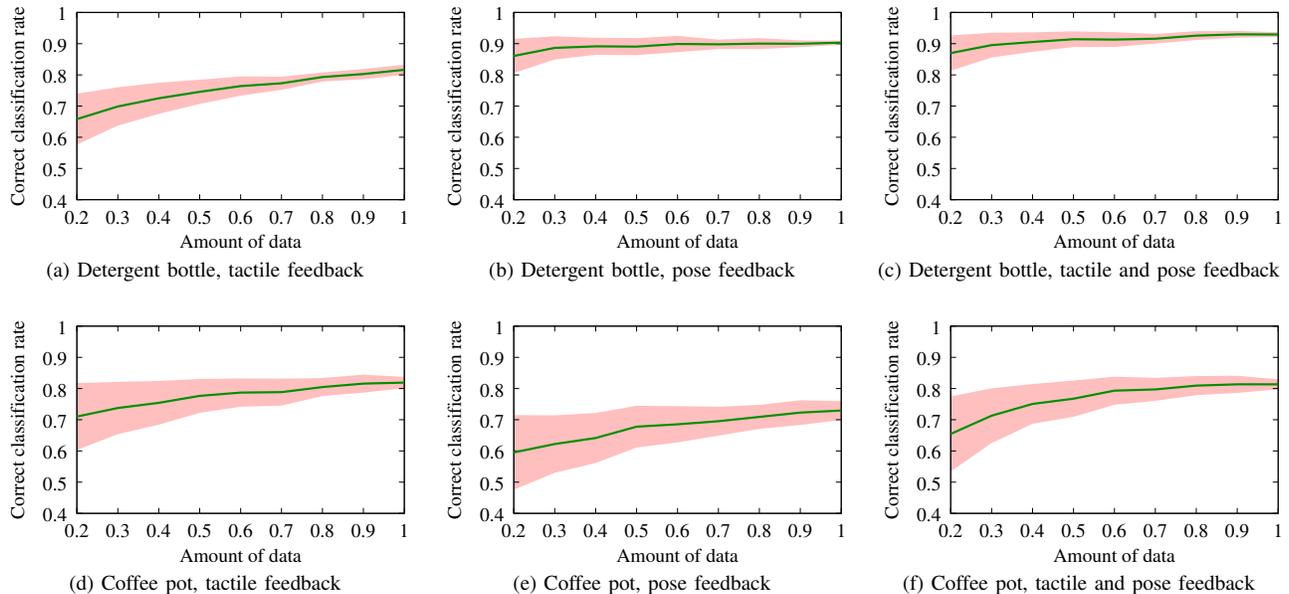


Fig. 6: Rates of correct classification as a function of the available amount of data.

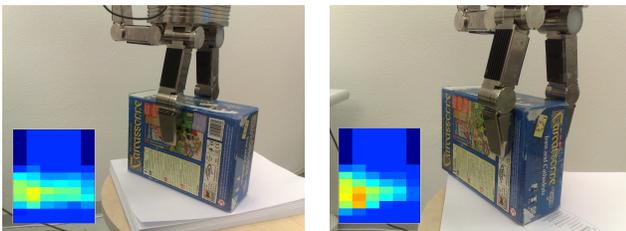


Fig. 7: Illustration of the grasps executed by the robot and the corresponding tactile images from one distal array. The left image shows a “middle” grasp, which always succeeded, while the right image shows an “extremity” grasp which always failed. The object was continuously explored between these two points. The tactile images show that during the unstable grasp, the leftmost side of the finger applies a strong pressure onto the object, while, for the stable grasp, pressure is more evenly distributed across the whole width of the finger.

transition between stable and unstable grasps occurred, and how accurately it was reflected by pre-grasp perceptual data. The wrist poses of the grasps executed by the robot were demonstrated by a human by teleoperation. The box was grasped by simultaneously closing the two fingers and continuously applying a constant closing force on all joints. A total of 50 grasps were executed, amongst which 25 were stable and 25 were unstable. Grasps applied near the middle of the top face of the box were always stable (see left image in Figure 7). As grasps were tried closer to the extremity of the box, they remained stable for a few centimeters, then abruptly became unstable. Unstable grasps were characterized by a rotation of the object when the robot tried to lift it up.

Stability classification was evaluated as explained above.

Classification rates computed from tactile data alone yielded a 94% rate. Rates computed from pose data alone, and from pose plus tactile data, led to 100% correct predictions. Several comments can be made on these results. First, pose perfectly separated stable grasps from unstable ones. We note however that in our setup, the camera is pointed directly at the objects, and the objects cover a large fraction of the camera’s field of view. If the camera were to cover a larger field, such as the whole robot workspace, pose estimation would be less accurate, and pose-based classification would be less reliable, therefore motivating the use of additional perceptual modalities. Second, in this experiment, tactile imprints can discriminate surprisingly well grasps applied on both ends of the explored region of the box. Although this discrimination may be useful in certain situations, it is likely that it is limited to the specific part of the box that was explored by the robot. As discussed below, one of our future aims is to learn models that characterize a *part* of an object, and which would thus be applicable to novel objects that share the same part. In this context, it will be interesting to study how tactile characterizations such as those learned for the box, or the objects of the previous section, generalize to novel objects.

## V. CONCLUSION

This paper studied the viability of concurrent object pose tracking and tactile sensing for assessing grasp stability on a physical robotic platform. We presented a kernel-logistic-regression model of pose- and touch-conditional grasp success probability, and a robotic platform that can track the pose of an object while it is grasping it, and that can acquire tactile imprints of the grasps it executes. We showed that the robot is able to use data collected by human demonstrations to learn grasp stability classifiers. Our results showed that stability assessments based on both tactile and pose data can

provide better rates than assessments based on tactile data alone.

Because models rely on the pose of an object, each model that the agent learns is only usable with that particular object. It is not realistic to imagine that an agent would learn different model of every object that exists. To overcome this limitation, we project to learn models that characterize only a *part* of an object, and which would thus be applicable to novel objects that share the same part.

## VI. ACKNOWLEDGEMENTS

We warmly thank T. Mörwald, M. Zillich, and M. Vincze for their support with the Blort pose tracker [33]. We thank M. Yamada for providing us with his KLR library.

## REFERENCES

- [1] R. Johansson, "Sensory input and control of grip," in *Novartis Foundation Symposium*, 1998, pp. 45–59.
- [2] —, "Sensory control of dexterous manipulation in humans," *Hand and brain: The neurophysiology and psychology of hand movements*, pp. 381–414, 1996.
- [3] P. Jenmalm, S. Dahlstedt, and R. Johansson, "Visual and tactile information about object-curvature control fingertip forces and grasp kinematics in human dexterous manipulation," *Journal of Neurophysiology*, vol. 84, no. 6, p. 2984, 2000.
- [4] R. Woodworth, "The accuracy of voluntary movement," *The Journal of Nervous and Mental Disease*, vol. 26, no. 12, p. 743, 1899.
- [5] A. Milner and M. Goodale, *The visual brain in action*. Oxford University Press, USA, 2006.
- [6] C. Hesse and V. Franz, "Grasping remembered objects: Exponential decay of the visual memory," *Vision Research*, 2010.
- [7] R. Johansson and G. Westling, "Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects," *Experimental Brain Research*, vol. 56, no. 3, pp. 550–564, 1984.
- [8] A. Kritikos and C. Brasch, "Visual and tactile integration in action comprehension and execution," *Brain Research*, vol. 1242, pp. 73–86, 2008.
- [9] B. Yoshimi and P. Allen, "Closed-loop visual grasping and manipulation," in *IEEE International Conference on Robotics and Automation*, 1996.
- [10] D. Kragic, A. T. Miller, and P. K. Allen, "Real-time tracking meets online grasp planning," in *IEEE International Conference on Robotics and Automation*, 2001, pp. 2460–2465.
- [11] A. Bicchi, J. Salisbury, and P. Dario, "Augmentation of grasp robustness using intrinsic tactile sensing," in *IEEE International Conference on Robotics and Automation*, 1989.
- [12] R. Howe, N. Popp, P. Akella, I. Kao, and M. Cutkosky, "Grasping, manipulation, and control with tactile sensing," in *IEEE International Conference on Robotics and Automation*, 1990.
- [13] R. Howe, "Tactile sensing and control of robotic manipulation," *Advanced Robotics*, vol. 8, no. 3, pp. 245–261, 1993.
- [14] 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*, 2003, pp. 1824–1829.
- [15] C. Goldfeder, P. K. Allen, C. Lackner, and R. Pelossof, "Grasp Planning Via Decomposition Trees," in *IEEE International Conference on Robotics and Automation*, 2007, pp. 4679–4684.
- [16] C. Borst, M. Fischer, and G. Hirzinger, "Grasping the dice by dicing the grasp," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2003, pp. 3692–3697.
- [17] M. Ciocarlie, C. Goldfeder, and P. Allen, "Dexterous grasping via eigengrasps: A low-dimensional approach to a high-complexity problem," in *In Robotics: Science and Systems Manipulation Workshop*, 2007.
- [18] Y. Li, J. L. Fu, and N. S. Pollard, "Data-driven grasp synthesis using shape matching and task-based pruning," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, pp. 732–747, 2007.
- [19] S. Ekvall and D. Kragic, "Learning and Evaluation of the Approach Vector for Automatic Grasp Generation and Planning," in *IEEE International Conference on Robotics and Automation*, 2007, pp. 4715–4720.
- [20] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 157–173, 2008.
- [21] R. Detry, E. Baseski, M. Popovic, Y. Touati, N. Krueger, O. Kroemer, J. Peters, and J. Piater, "Learning continuous grasp affordances by sensorimotor exploration," in *From Motor Learning To Interaction Learning in Robots*. Springer-Verlag, 2010, pp. 451–465.
- [22] A. Erkan, O. Kroemer, R. Detry, Y. Altun, J. Piater, and J. Peters, "Learning probabilistic discriminative models of grasp affordances under limited supervision," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 1586–1591.
- [23] A. Petrovskaya, O. Khatib, S. Thrun, and A. Y. Ng, "Bayesian estimation for autonomous object manipulation based on tactile sensors," in *IEEE International Conference on Robotics and Automation*, 2006, pp. 707–714.
- [24] A. Jiménez, A. Soembagijo, D. Reynaerts, H. V. Brussel, R. Ceres, and J. Pons., "Featureless classification of tactile contacts in a gripper using neural networks," *Sensors and Actuators A: Physical*, vol. 62, no. 1-3, pp. 488–491, 1997.
- [25] A. Bierbaum, M. Rambow, T. Asfour, and R. Dillmann, "A potential field approach to dexterous tactile exploration," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, 2008.
- [26] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard, "Object identification with tactile sensors using bag-of-features," in *IEEE/RSJ international Conference on Intelligent Robots and Systems*, 2009, pp. 243–248.
- [27] N. Gorges, S. E. Navarro, D. Göger, and H. Wörn, "Haptic object recognition using passive joints and haptic key features," in *IEEE International Conference on Robotics and Automation*, 2010.
- [28] A. Morales, M. Prats, P. Sanz, and A. P. Pobil, "An experiment in the use of manipulation primitives and tactile perception for reactive grasping," in *Robotics: Science and Systems, Workshop on Robot Manipulation: Sensing and Adapting to the Real World*, Atlanta, USA, 2007.
- [29] M. Prats, P. Sanz, and A. del Pobil, "Vision-tactile-force integration and robot physical interaction," in *IEEE International Conference on Robotics and Automation*, Kobe, Japan, 2009, pp. 3975–3980.
- [30] Y. Bekiroglu, J. Laaksonen, J. 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.
- [31] J. Laaksonen, V. Kyrki, and D. Kragic, "Evaluation of feature representation and machine learning methods in grasp stability learning," in *IEEE-RAS International Conference on Humanoid Robots*, 2010.
- [32] H. Zhang and N. Chen, "Control of contact via tactile sensing," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 5, pp. 482–495, 2000.
- [33] T. Mörwald, J. Prankl, A. Richtsfeld, M. Zillich, and M. Vincze, "BLORT—the blocks world robotic vision toolbox," *Best Practice in 3D Perception and Modeling for Mobile Manipulation (Workshop at ICRA 2010)*, 2010.
- [34] M. Yamada, M. Sugiyama, and T. Matsui, "Semi-supervised speaker identification under covariate shift," *Signal Processing*, vol. 90, no. 8, pp. 2353–2361, 2010.
- [35] B. Schölkopf and A. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. the MIT Press, 2002.
- [36] R. A. Fisher, "Dispersion on a sphere," in *Proc. Roy. Soc. London Ser. A.*, 1953.
- [37] E. B. Sudderth, "Graphical models for visual object recognition and tracking," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006.