MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Interactive Tactile Perception for Classification of Novel Object Instances

Corcodel, Radu; Jain, Siddarth; van Baar, Jeroen

TR2020-143 November 05, 2020

Abstract

In this paper, we present a novel approach for classification of unseen object instances from interactive tactile feedback. Furthermore, we demonstrate the utility of a low resolution tactile sensor array for tactile perception that can potentially close the gap between vision and physical contact for manipulation. We contrast our sensor to high-resolution camera-based tactile sensors. Our proposed approach interactively learns a one-class classification model using 3D tactile descriptors, and thus demonstrates an advantage over the existing approaches, which require pre-training on objects. We describe how we derive 3D features from the tactile sensor inputs, and exploit them for learning one-class classifiers. In addition, since our proposed method uses unsupervised learning, we do not require ground truth labels. This makes our proposed method flexible and more practical for deployment on robotic systems. We validate our proposed method on a set of household objects and results indicate good classification performance in real-world experiments

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Interactive Tactile Perception for Classification of Novel Object Instances

Radu Corcodel, Siddarth Jain and Jeroen van Baar

***Abstract*— In this paper, we present a novel approach for classification of unseen object instances from interactive tactile feedback. Furthermore, we demonstrate the utility of a low resolution tactile sensor array for tactile perception that can potentially close the gap between vision and physical contact for manipulation. We contrast our sensor to high-resolution camera-based tactile sensors. Our proposed approach interactively learns a one-class classification model using 3D tactile descriptors, and thus demonstrates an advantage over the existing approaches, which require pre-training on objects. We describe how we derive 3D features from the tactile sensor inputs, and exploit them for learning one-class classifiers. In addition, since our proposed method uses unsupervised learning, we do not require ground truth labels. This makes our proposed method flexible and more practical for deployment on robotic systems. We validate our proposed method on a set of household objects and results indicate good classification performance in real-world experiments.**

Fig. 1. Experimental setup for tactile perception. A 7-DoF (Degrees of Freedom) arm, with attached parallel jaw gripper equipped with barometric tactile sensors (zoomed-in insert).

I. INTRODUCTION

Robotic manipulation has been evolving over the years from simple pick-and-place tasks, where the robot's environment is predominantly well structured, to dexterous manipulation where neither the objects nor their poses are known to the robotic system beforehand [1], [2]. Structured pick-and-place tasks leverage the artificially reduced task complexity and thus require minimal sensing, if any, for grasping operations. Dexterous manipulation on the other hand, must rely more heavily on sensing not only to confirm successful grasp attempts, but also to localize, distinguish and track graspable objects, as well as planning grasps autonomously [3].

Typically, robotic manipulation systems rely on "classic" machine vision, *e.g.*, depth cameras, LiDAR or color cameras, which require line-of-sight with the environment. Although some of the inherent vision problems can be mitigated by using multiple points of view, in-wrist camera systems, and visual servoing, the final stage of the grasp, *i.e.*, physical contact, still remains blind and open loop. More importantly, the state of the object after grasping and during manipulation is very difficult to estimate (for example, due to material properties).

Objects that may appear similar to an advanced vision system can differ completely in terms of their material properties. Tactile feedback can close the gap between vision and physical manipulation. There have been recent advancements in tactile manipulation and state-of-the-art approaches use vision-based tactile feedback using deformable gel mounted above a camera which provides high-resolution image observations of the grasped objects [4]. Although effective, such sensors are usually bulky and may introduce computational overhead while processing high-resolution images. In this work, taking motivation from human fingertips which can account for roughly 100 taxels [1] per square centimeter [5], we propose utilization of a low resolution tactile device based on barometric MEMS devices (Micro Electro-Mechanical System) [6].

Object classification is an important task of robotic systems. Vision-based approaches require pre-training on a set of *a priori* known objects for classification. We propose instead to perform classification of novel objects based on interactive tactile perception, using unsupervised learning without any pre-training. This could make a robot system more practical and flexible. The contributions described in this paper can be summarized as:

- Using off the shelf barometric MEMS devices, we present construction and integration of a low-cost and low resolution tactile sensor array for robotic grasping and manipulation.
- We introduce a meaningful 3D representation for local geometry of objects using the proposed low-resolution tactile sensing.
- We propose an unsupervised machine learning approach for classifying novel objects, which fits a model to tactile representations acquired with interactive manipulation, without the need for pre-training and labeled ground truths across an entire training set of objects.

Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA corcodel@merl.com, sjain@merl.com, jeroen@merl.com

[1] A taxel is short for tactile element, analogous to a pixel.

II. Related Work

A number of prior works have studied related tactile recognition problems, particularly with supervised learning [7]. Some examples of such tactile perception problems include recognition of object instances [8], surface texture information [9], and stiffness properties [10]. Prior work has focused on recognizing object instances, when the number/types of object classes are known *a priori*. In contrast, in this work we aim to recognize novel object instances with tactile manipulation in a setting where the robot has no *a priori* information about the number of classes and the associated object labels. Our work helps address the questions whether interaction with touch can provide significant information about novel object identity and whether the global geometry and appearance properties can be approximated with such information.

A number of prior works have explored supervised learning with training datasets for classification of object categories using tactile sensing. Spiers *et al.* [8] proposed a gripper hardware comprising of a simple two-finger underactuated hand equipped with TakkTile [6] barometric pressure sensors for performing object classification. They use a random forests (RFs) classifier to learn to recognize object instances based on training data over a set of objects. Schneider *et al.* [11] identify objects with touch sensors installed in the finger tips of a manipulation robot using an approach that operates on low-resolution intensity images obtained with touch sensing. Such tactile observations are generally only partial and local views, similar as in our work. They adapt the Bag-of-Words framework to perform classification with local tactile images as features and create a feature vocabulary for the tactile observations using k-means clustering. Drimus *et al.* [12] proposed a novel tactile-array sensor based on flexible piezoresistive rubber and present an approach for classification of a number of household objects. They represent the array of tactile information as a time series of features for a k-nearest neighbors classifier with dynamic time warping to calculate the distances between different time series.

More recently, deep learning based approaches are also proposed for recognizing object instances with touch and vision. Lin *et al.* [13] proposed a convolutional neural network (CNN) for cross-modality instance recognition in which they recognize given visual and tactile observations, whether or not these observations correspond to the same object. In their work, they use two GelSight sensors [4] mounted on the fingers of a parallel jaw gripper. The GelSight tactile sensor provides high-resolution image observations, and it can detect fine surface features and material details using the deformable gel mounted above a camera in the sensor. Although their approach does not require specific class labels during training, it still needs a large dataset for training as with all deep learning based methods. Researchers have also proposed supervised techniques for inferring object properties from touch. For example, Yuan *et al.* [14] proposed estimating the hardness of objects using a convolutional neural network and the GelSight tactile sensor.

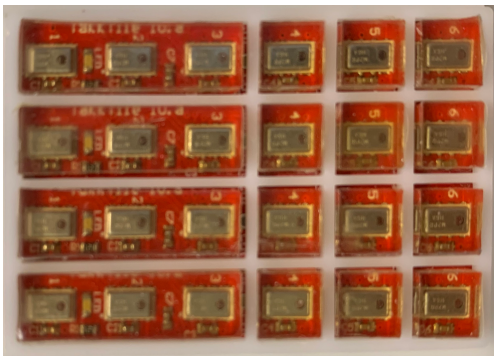

Fig. 2. Touch sensing array used to equip both the inside and outside faces of the gripper fingers. The array consists of four TakkStrip2 devices (RightHand Robotics, Inc.) connected to a main I^2C bus. Our tactile arrays consist of 48 taxels arranged in a 4×6 array, with a dot pitch of roughly 7.5mm

Other than the recognition problem, tactile sensing has also been utilized for improving robotic manipulation and grasping. Calandra *et al.* [15] proposed a multimodal sensing framework that combines vision and touch to determine the utility of touch sensing in predicting grasp outcomes. They use a deep neural network (DNN) with inputs from RGB images from the front camera and the GelSight sensors in order to predict whether the grasping will be successful or not. Hogan *et al.* [16] proposed a novel re-grasp control policy that makes use of tactile sensing for improving grasping with local grasp adjustments. In the next section, we discuss the tactile sensing hardware and the generation of tactile data.

III. Tactile Data Generation

A. Tactile sensing hardware

Our tactile sensing hardware consists of four tightly packed arrays of *Takktile* sensor strips [6], arranged as the inside and outside touch pads of a two-finger parallel jaw gripper (see Figure 2). The Takktile sensors use a series of MEMS barometric IC devices casted in a soft elastomer and packaged as strips of six taxels (tactile sensor cells). The main benefit of these devices is that they provide all the analog signal conditioning, temperature compensation and analog to digital conversion (ADC), on chip. As opposed to other tactile sensing technologies [17][18], barometric sensors read the tactile pressure *and* temperature input directly, and are thus more akin to human touch sensing. Moreover, compared with vision-based touch sensing alternatives, MEMS pressure sensors communicate over a significantly lower bandwidth while allowing for a more flexible spatial arrangement of the taxels (*i.e.* not bounded to planar touch pads).

Each gripper finger is fitted with eight Takktile strips, divided into two groups: one for exterior grasps and the other for interior grasps, totalling a number of 48 taxels per finger. For convenience, the touch pads are planar, although this is not a design limitation. In fact, each sensor cell can be isolated and addressed with minimal hardware changes,

while the device footprint can be further reduced by using equivalent MEMS barometric devices. The current iteration of the touch sensing array used in our experiments measures 30×45mm and contains 4×6 taxels (thus a dot pitch of 7.5mm).

All devices communicate over a single I^2C standard bus. Data collision and other transfer safeties are handled "on-strip" by a traffic controller that, when addressed by a master I^2C controller, wakes each connected device in a loop which triggers it to load the pressure data on the bus (detailed information about the Takktile's communication protocol can be found in [6]). Using a I^2C-USB device interface, the sensors are connected to a *Raspberry Pi 4* acting as a physical ROS node which publishes raw tactile data to our ROS-enabled robot controller. With this setup we achieve a 64Hz sampling rate with all 96 taxels connected.

B. Raw signal processing

The tactile raw data consists of an array of 96 pressure and temperature values, corresponding to each taxel. As a first processing step, the individual pressure values are temperature compensated as stated in the sensor manufacturer's datasheet [6]. Despite a relatively low noise (0.01 N) and good linearity (less than 1% *typ.*) the barometer cells exhibit a slow drift after a few hours of use. To compensate for the noise drift, we use a simple moving average filter with an arbitrarily chosen window of 30 samples and uniform weight across all data points. This filter doubles as a measure of the unloaded sensor state. Though not a requirement of our tactile sensing pipeline, we offset the measured steady state for each taxel (given by the moving average filter), so that our filtered readout reflects only tactile load and temperature information.

C. Generating the pressure maps

An initial goal of this paper consists of generating a meaningful 3D representation of the objects' local geometry using a low-resolution tactile device. To achieve this we represent the contact between the touch pad and the manipulated object as a continuous 3D pressure map. We generate this pressure map by uniformly sampling a Non-rational Uniform B-spline (NURBS) surface patch, where each node $\mathbf{p}_{ij}$ in the control net (represented as a quadrilateral mesh), is computed from a linear combination of taxels' location in the 3D space $\mathbf{x}_{T_{ij}}$, and their respective filtered pressure reading, expressed as a displacement on the z-axis in the spline's reference frame:

$$\mathbf{p}_{ij} = \mathbf{x}_{T_{ij}} + k \begin{bmatrix} 0 & 0 & P_{ij}^f \end{bmatrix}^T \tag{1}$$

where k is an arbitrary scaling constant which controls the surface's z-range, and T_{ij} is the taxel at grid location $\{i, j\}$ with its corresponding filtered pressure value P_{ij}^f. We uniformly evaluate the NURBS surface using the well known formulation of [19]:

$$S(u,v) = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} N_{ip}(u) N_{jq}(v) w_{ij} \mathbf{p}_{ij}}{\sum_{i=0}^{n} \sum_{j=0}^{m} N_{ip}(u) N_{jq}(v) w_{ij}} \tag{2}$$

where N_{ip} and N_{jq} are B-spline basis functions and the degree of each NURBS curve generator (n and m, respectively) is the number of control points less one, along each parametric coordinate (*i.e.* no internal knots in the two knot vectors). The weight is kept at $w_{ij} = 1$ for all control points. The surface is uniformly sampled by sweeping the normalized parametric domain $\{u, v\} = [0, 1] \times [0, 1]$ with a constant parameter increment du, and respectively dv, calculated based on a user-defined resolution and the aspect ratio of the NURBS control mesh. In our testing we used a surface sampling resolution of 2166 and an aspect ratio of $2 \div 3$.

Additionally, we also compute the surface normal for each sample of the pressure map, which is required for computing geometric descriptors (*i.e.* 3D feature vectors, see Section IV-B). To calculate the surface normal, we follow a similar approach to [20] in which the surface normal is obtained by taking the cross product of the partial derivatives with respect to the u and v parameters. To prevent poles while computing the partial derivatives, we enforce that the weights $w_{ij} > 0$.

IV. Framework: Interactive Tactile Classification of Novel Objects

Our goal is to use tactile feedback to classify objects as novel or seen before. In recent years, DNNs have achieved good performance on various classification tasks, *e.g.* [21], [16]. The networks are trained with supervisory signals, i.e. ground truth class labels, and thus fall under the umbrella of supervised learning methods. In addition, DNNs require copious amounts of training data to achieve good performance. Due to these requirements, using DNNs is not a practical solution to achieve our goal.

We instead propose to learn online, one object at a time, without any need for pre-training. Object instances that have been manipulated before by the robot should be classified as such, and novel objects should be detected, learned, classified and added to the set of previously manipulated objects. The main motivation behind our approach is data efficiency and active exploration. For a practical manipulation task, a real robotic system can only "afford" a short amount of time to determine if the object is novel, which implies too few tactile samples for deep learning. Moreover, knowing the span of object geometry and material properties, i.e., the range of tactile *feel*, beforehand, defeats the purpose of a generic tactile manipulation framework, and would simply fall into the usual robotic pick-and-place in a structured environment.

A. Problem Formulation

The problem we address in this paper can be formulated as follows. Given several palpations or sampled grasps on objects, can the robot automatically classify different unseen objects from tactile feedback alone? Initially the set of objects is empty, and each object with which the robot interacts should be classified as either known (belonging to the same category that it has seen before) or unknown (novel instance). In the case of an unknown object, it should then learn a representation of that object based on interactive

Fig. 3. Household objects used in our experiment spanning a wider range of material properties (specially hardness) and geometry.

tactile manipulation, and update the instance as a known category. We propose to solve this problem using 3D tactile features obtained by grasps interaction on the objects and unsupervised learning with one class classification.

B. Learning Local Tactile Representations for Novel Objects

Given an ensemble of objects to be sorted by object category, the robot would perform the following tasks:

1) Use a depth camera to capture a depth map of the scene containing unknown objects.
2) Record a depth map of individual object in the scene and determine candidate grasp poses.
3) Palpate an object in the scene with a selection of grasps chosen from the candidate grasps using tactile sensing.
4) Generate 3D tactile features and learn an unsupervised model using the local tactile information.
5) Record a depth map of other individual objects in the scene and determine candidate grasp poses.
6) Determine if the object is of known or unknown type, using the learned model (object specific).
7) If unknown, classify the object as novel and compute a representation based on the pressure maps recorded for each grasp and learn a new unsupervised model for the object.
8) Classify all objects in the scene, using the learned models and place them in representative bins.

For each grasp in Step 3 above we record the point cloud derived from the pressure map (see Section III-C). From the point cloud we compute a Viewpoint Feature Histogram (VFH) [22]. Each VFH is a 308-dimensional feature vector. We compute one for each finger of the gripper, and store a grasp as a combination of two VFHs, a 616-dimensional feature vector. The robot grasps (palpates) an object n_g times to acquire a set of local feature representations for an object.

As stated above, we want to avoid pre-training on objects, and handle objects as they are manipulated by the robot. Furthermore, we want to eliminate the need for known object labels which are required in supervised learning methods. We use an unsupervised learning approach based on once class classification. One class classification aims to learn a representation for the grasps, and then classify seen *vs.* unseen objects. We choose the One Class SVM (OC-SVM) classifier [23], which can be formulated as:

$$\begin{aligned} \min_{w,\xi_i,\rho} \frac{1}{2}||w||^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho \\ \text{s.t.}(w\cdot\phi(x_i)) \geq \rho - \xi_i \text{ for all } i = 1,\cdots,n \\ \xi_i \geq 0 \text{ for all } i = 1,\cdots,n \end{aligned} \tag{3}$$

where, ξ_i is the slack variable for sample i, n is the size of training samples and ν is the regularization parameter. The SVM hyperplane is represented by w and ρ. Points on one side of this hyperplane are classified as inliers, and points on the other side as outliers. For details on Eq. 3 we refer the reader to [23]. For unseen objects, we consider the VFH features for all $n = n_g$ grasps simultaneously, and fit the OC-SVM to this data. We then store this OC-SVM as a representation for the object.

C. Classifying Objects

Using the OC-SVM representation, we classify an object by evaluating the decision function, defined as:

$$\begin{aligned} f(x) &= \text{sgn}((w\cdot\phi(x_i)) - \rho) \\ &= \text{sgn}\left(\sum_{i=1}^{n}\alpha_i K(x,x_i) - \rho\right), \end{aligned} \tag{4}$$

where $\alpha_i K(x,x_i)$ is $w\cdot\phi(x_i)$ expressed with a kernel function K [23]. We use an RBF kernel function for all experiments in this paper. Each value within the sgn() represents a signed distance to the hyperplane. Positive distances represent inliers, while negative distances represent outliers.

D. Novel Object Discovery

We can compute the decision function from Eq. 4 for each of the VFH representations corresponding to the n_g grasps, and determine for each whether they are inlier or outlier. However, this ignores the signed distances $\alpha_i K(x,x_i) - \rho$ to the decision boundary. Instead, we compute the mean over the n_g signed distances corresponding to the grasps. The final classification of the object as inlier (seen) or outlier (unseen) is then based on the mean signed distance. We repeat this process with OC-SVM for each previously manipulated object type.

E. Sampling Grasps for Objects

We rely on vision *only* to determine grasp candidates for objects that the robot interacts with. The robot has an on-board RGBD camera which provides a 3D point cloud of the scene. There exists a number of approaches to autonomously generate robotic grasps on objects [24], [25], [26], [27], [28], In this work, we use grasp pose detection (GPD) [28] to propose a set of possible autonomy grasps. GPD can directly operate on point clouds and can provide a ranked set of potential grasp candidates. The grasps are filtered to avoid collisions of the robot with the environment. We select n_g grasps from the proposed set of grasps for an object, according to filtered grasp directions.

TABLE I

PERFORMANCE OF HUMAN-DIRECTED TACTILE CLASSIFICATION. TABLE REPORTS SCALED MEAN SIGNED DISTANCE TO THE DECISION BOUNDARY (0 MEANS THE OBJECT IS STRONGLY CONSIDERED AN OUTLIER, WHEREAS A VALUE OF 1 INDICATES THAT IT IS STRONGLY CONSIDERED AN INLIER). SEE TEXT UNDER PERFORMANCE MEASURE AND EXPERIMENT I IN SECTION V FOR A MORE DETAILED EXPLANATION.

	pc	sc	tb	wb	ob	ap	wg	box	kb	le	fb
paper cup (pc)	0.88	0.45	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.04
soda can (sc)	0.0	0.93	0.0	0.70	0.0	0.12	0.0	0.81	0.0	0.63	0.0
tennis ball (tb)	0.0	0.0	0.83	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
water bottle (wb)	0.0	0.16	0.25	0.86	0.0	0.0	0.0	0.21	0.0	0.46	0.0
oil bottle (ob)	0.0	0.41	0.0	0.20	0.63	0.0	0.0	0.99	0.0	0.20	0.0
apple (ap)	0.06	0.0	0.69	0.0	0.0	0.78	0.04	0.0	0.0	0.08	0.0
wine glass (wg)	0.0	0.0	0.0	0.0	0.0	0.0	0.87	0.14	0.0	0.0	0.0
box	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.99	0.0	0.0	0.0
koala bear (kb)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.82	0.0	0.0
camera lens (le)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.99	0.0
football (fb)	0.93	0.0	0.45	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.95

The tactile features on the selected grasps should essentially form some sort of basis when fitting the OC-SVM. The more we can uniformly sample an object across its surface, the more likely the model can classify it correctly. In the next section, we present the evaluation of our proposed method in two experiments.

V. EXPERIMENTS & RESULTS

Our experimental work aims to evaluate the performance of our approach for using interactive tactile perception in order to classify novel object instances. Our research platform for the experiments described in this section is the Gen3 Lightweight robot arm (Kinova Robotics, Canada), a 7-DoF manipulator with a Robotiq parallel jaw gripper. We have instrumented the fingers with the pressure sensor arrays to enable tactile sensing (See Fig. 1 and Section III).

We used a test set of $n_o = 11$ household objects in our experiments, namely: apple, box, camera lens, koala bear, plush toy, glass oil bottle, paper cup, foam football, soda can, tennis ball, plastic water bottle, and a wine glass. The set of objects are shown in Fig. 3.

Performance Measure: To evaluate our proposed approach we first define a performance measure based on the signed distances to the decision boundary for OC-SVM. In OC-SVM, a larger distance to the decision boundary denotes more confidence in the classification as inlier or outlier. As explained in Sec. IV-D, we compute the mean of the signed distances $\alpha_i K(x, x_i) - \rho$ to the decision boundary for the set of grasps for which we fit the OC-SVM, and we denote this mean signed distance as d_{fit}. Note that d_{fit} is a positive number [23]. Next, we define a range $[-d_{fit}, d_{fit}]$ around the decision boundary, and scale this range to $[0, 1]$. In this scaled range < 0.5 is classified as outlier, and ≥ 0.5 as inlier. For clarity, a value of 0 means that the object is strongly considered an outlier, whereas a value of 1 means that it's strongly considered an inlier. A value of 0.5 lies on the decision boundary, and it can be considered either inlier or outlier.

Using the same OC-SVM model for a given object, we compute the mean signed distance for the n_g grasps for each of the remaining test objects, clip them by d_{fit}, and scale to $[0, 1]$ range. Given the normalized range, we compare how strongly an object is classified as previously seen category or a novel instance.

Experiment I: Human-directed Grasps: To test our proposed method, we first generated a set of human-directed grasps in which the objects were presented in a grasp pose selected manually by the user to the touch-enabled gripper. Each object was grasped by the robot at $n_g = 25$ different user selected pose configurations and the 3D tactile features were recorded for each touch based interaction during the grasping. We learned an OC-SVM model representation for each object using the 3D features. For each learned OC-SVM model ($n_o = 11$), we presented the unseen test set of $n_o - 1$ objects to the robot twice in random order not including the particular object instance for which the OC-SVM model was learned. The grasp poses for the unseen test set were again selected manually by the user when the test objects are presented to the touch-enabled gripper. For each object, 3D tactile features are computed for each tactile interaction, which are then used by the learned OC-SVM model to make seen category vs. unseen object instance predictions.

Table I shows the results for human-directed grasps using the performance measure score explained above. We emphasize that despite the appearance, this table should not be confused with a typical confusion matrix. Almost all objects will be strongly classified as seen, when presented again to the tactile sensors (see the numbers along the diagonal). Some of the off-diagonal entries are non-zero, but below 0.5, and thus will be correctly classified as unseen. For example, a novel instance of apple will not be mistaken for a previously seen soda can (second row in Table I). However, some other off-diagonal entries indicate that an unseen object may be wrongly classified as seen. In the case of the glass oil bottle, the box object is more strongly classified as oil bottle (fifth row in Table I), and a novel instance of water bottle may be classified as a seen soda can (second row in Table I). These

TABLE II

PERFORMANCE OF REAL ROBOT TACTILE CLASSIFICATION WITH AUTONOMY GRASPS. TABLE REPORTS SCALED MEAN SIGNED DISTANCE TO THE DECISION BOUNDARY (0 MEANS THE OBJECT IS STRONGLY CONSIDERED AN OUTLIER, WHEREAS A VALUE OF 1 MEANS IT IS STRONGLY CONSIDERED AN INLIER). SEE TEXT UNDER PERFORMANCE MEASURE AND EXPERIMENT II IN SECTION V FOR A MORE DETAILED EXPLANATION.

	pc	sc	tb	wb	ob	ap	wg	box	kb	le	fb
paper cup (pc)	0.65	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
soda can (sc)	0.0	0.43	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
tennis ball (tb)	0.0	0.0	0.38	0.0	0.0	0.64	0.0	0.0	0.0	0.0	0.0
water bottle (wb)	0.0	0.06	0.70	0.70	0.0	0.79	0.0	0.0	0.0	0.17	0.0
oil bottle (ob)	0.0	0.22	0.0	0.0	0.42	0.0	0.0	0.0	0.0	0.0	0.0
apple (ap)	0.0	0.0	0.0	0.0	0.0	0.20	0.0	0.0	0.0	0.0	0.0
wine glass (wg)	0.16	0.07	0.65	0.51	0.0	0.95	0.67	0.0	0.0	0.42	0.0
box	0.0	0.78	0.32	0.53	0.43	0.63	0.0	0.89	0.0	0.0	0.0
koala bear (kb)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.72	0.0	0.0
camera lens (le)	0.0	0.0	0.74	0.64	0.0	0.98	0.09	0.0	0.0	0.57	0.0
football (fb)	0.0	0.0	0.05	0.0	0.0	0.12	0.0	0.0	0.0	0.0	0.87

possible misclassifications are for objects with very similar tactile characteristics. Please see Section VI for a further discussion.

Experiment II: Autonomy Grasps: In this experiment we used a learned grasp pose detector (GPD)[28] to calculate grasp candidates and have the robot arm autonomously execute a tactile-enabled palpation on the object for each of the autonomy grasp candidates. The depth information needed by the GPD is provided by the wrist mounted RGB-D camera available on the Kinova Gen3 arm. The list of grasp candidates as computed by GPD are ranked based on a grasp quality index (see [28] for more details), and for each element in the list we enforce restrictions on approach angles relative to the robot base, inverse kinematics, and grasp location relative to the table. It is important to note that the particular choice of a grasp pose detector has no impact to our system. In fact any grasp detection approach capable of producing uniformly distributed grasp candidates relative to an object pose will suffice.

In this exploratory experiment, the objects presented to the robot were separated (not cluttered) so that we can group the grasps candidates generated for each individual object. Fig. 4 shows examples of grasp candidates generated by the GPD for some of the household objects used in our experiments. Each experiments consists of at least $n_g = 15$ palpations per object, in descending order of grasp quality. Ideally, we would have $n_g = 25$ as seen in the the human-directed grasps. However, the GPD only provide a limited number of good quality grasp proposals. If after grasp selection pruning we have $n_g < 15$, we abort the experiment and retry with a different camera point of view. Similar to Experiment I, we learned an OC-SVM model representation for each object using the 3D features which are computed on the autonomy generated grasp palpations. For each learned model ($n_o = 11$), we then presented the unseen test set of $n_o - 1$ remaining objects to the robot twice in random order not including the particular object instance for which the OC-SVM model is learned. In this case, the robot autonomously generated grasp poses on the test objects and then computed the 3D tactile

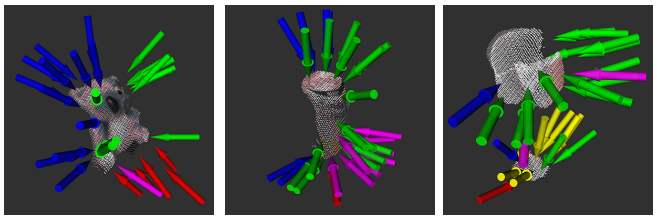

Fig. 4. Examples of grasp poses proposed by the GPD for koala bear, soda can and wine glass. The arrows indicate whether the proposed grasp is acceptable (green), or rejected (blue and magenta: unfeasible approach angle, red: grasp too close to table, yellow: invalid inverse kinematics

features for each tactile interaction on the computed grasps, which were then used by the learned OC-SVM model to make predictions.

Table II shows the results for autonomy grasps using the performance measure score. Again, this table does not represent a typical confusion matrix. For autonomous grasps, the classification performance is less compared to the human-directed grasps of Table I. A novel instance of unseen apple or tennis ball, may be misclassified as a seen water bottle category (fourth row in Table II). A known instance of an apple indicated low 0.2 performance value, yet it's distance is the only one which is even within the d_{fit} range, as for all other novel instances of objects the values along the row are 0 (sixth row in Table II). For most objects the values along the diagonal are highest, or among the higher values, in the corresponding rows, indicating the capability to correctly recognize object instances. We provide further discussion on the results in the next section.

VI. DISCUSSION AND FUTURE WORK

Our envisioned system would take as input a cluttered ensemble of unknown objects and perform classification of objects into bins based on interactive tactile sensing. The robot would grasp objects one at a time, interactively manipulate them in a designated grasping area, and then classify the object accordingly using the proposed approach.

In this paper we focus on the tactile representation learning and classification in exploratory experiments, rather than present a full manipulation system.

Our results are encouraging and it is important to note that with local tactile sensing, objects of similar shape (*e.g.* cylindrical) may be difficult to distinguish. For example, the water bottle and lens, or the apple and tennis ball have very similar local geometry and thus tactile sensing will produce similar features for the particular grasps proposed by the autonomous GPD. In addition, soft objects which may deform drastically during grasping (for example, the football or a paper cup) can also result in features that are not sufficiently discriminative based on the sampled grasps. It is important to get a good representation of the sampled grasps for the local tactile feature to capture the global geometry of the object. For example, as seen in the results a box may be misclassified as a soda can or an oil bottle in scenarios when the sampled grasps were on the flat sides of the box and oil bottle. This fails to produce good VFH features and subsequent OC-SVM representations. Moreover, for large enough cylindrical objects, such a soda can, can be misclassified (for example, as a box) primarily because its curvature is large compared to the surface area of the tactile array. Ideally, our system should identify novel objects through a mix of geometric information and material properties (especially hardness and texture). While a low resolution sensing device may limit the minimum feature size we are manipulating, our preliminary results show that a taxel pitch of 7.5mm is sufficient for a variety of household objects.

In addition to OC-SVM for one class classification, we also evaluated a one-class version of Gaussian Mixture Model (OC-GMM) [29]. In order for OC-GMM to perform well, pre-training to determine the decision boundaries or a confidence threshold for classifying the objects as inliers or outliers is required. However, any pre-training on all object classes is precisely what we want to avoid, as we focus on interactive and online learning about objects in an unsupervised manner.

Results with our current hardware prototype are encouraging. The average experiment time per object was typically 60 seconds. Only a small fraction of the time was spent learning the OC-SVM model, while most of the time was spent in performing interactive manipulation by the robot. We believe that the performance can be improved further, since the geometric features which can be resolved with the current tactile sensors configuration are rather coarse. In future work, we focus on prototyping a new version of the sensor array which has a greater taxel density and use non-planar tactile sensitive surfaces. Our choice of barometric pressure taxels was due mostly to anthropomorphism to animal tactile sensing where touch is primarily sensed by pressure and temperature. The advantage of using a barometric device becomes apparent while grasping hyper-elastic objects; while the object deforms beyond its unloaded state and complies to the gripper finger, the pressure applied to the array is proportional to its undeformed geometry. Even though we severely deform the object during manipulation, the tactile sensor can determine, to some degree, the undeformed geometry from the sensed pressure distribution.

We empirically chose a fixed number of sample grasps to fit the OC-SVM in the case of autonomy grasps experiment. Our future work will focus on a more thorough evaluation of how many grasps, and particularly which grasp poses should be selected autonomously to further optimize the performance. Fewer, but more informative grasps are preferred from a practical standpoint. The point cloud can further be analyzed in terms of geometric properties (*i.e.* a ridge or extrusion) and grasps proposed by the GPD near those geometric properties could be favored for sampling the object. Having a better reconstruction of the object under consideration could also help with sampling. Furthermore, an externally mounted depth sensor can overcome the limited point cloud reconstructions or alternatively, the object may also be scanned from different directions with the robot mounted RGBD sensor.

VII. CONCLUSION

We have proposed a method to classify novel objects based on tactile feedback, without the need of pre-training and ground truth labels for supervision. Our proposed method uses One-Class SVM to fit a set of features derived from grasp pressure maps acquired from interactive tactile manipulation on objects, and subsequently classify instances by interacting with the objects. In real robot experiments, we have shown that the results of our proposed approach are encouraging for classification of novel objects based on interactive tactile sensing.

REFERENCES

[1] M. T. Mason, "Toward robotic manipulation," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 1–28, 2018.

[2] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, 2019.

[3] J. Tegin and J. Wikander, "Tactile sensing in intelligent robotic manipulation–a review," *Industrial Robot: An International Journal*, 2005.

[4] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.

[5] J. R.S. and J. R. Flanagan, "Tactile sensory control of object manipulation in human," in *In Handbook of the Senses. Vol.: Somatosensation.*, K. J. and G. E., Eds. Elsevier, 2007.

[6] Y. Tenzer, L. P. Jentoft, and R. D. Howe, "The feel of mems barometers: Inexpensive and easily customized tactile array sensors," *IEEE Robotics & Automation Magazine*, vol. 21, no. 3, pp. 89–95, 2014.

[7] S. Luo, J. Bimbo, R. Dahiya, and H. Liu, "Robotic tactile perception of object properties: A review," *Mechatronics*, vol. 48, pp. 54–67, 2017.

[8] A. J. Spiers, M. V. Liarokapis, B. Calli, and A. M. Dollar, "Single-grasp object classification and feature extraction with simple robot hands and tactile sensors," *IEEE transactions on haptics*, vol. 9, no. 2, pp. 207–220, 2016.

[9] J. A. Fishel and G. E. Loeb, "Bayesian exploration for intelligent identification of textures," *Frontiers in neurorobotics*, vol. 6, p. 4, 2012.

[10] Z. Su, J. A. Fishel, T. Yamamoto, and G. E. Loeb, "Use of tactile feedback to control exploratory movements to characterize object compliance," *Frontiers in neurorobotics*, vol. 6, p. 7, 2012.

[11] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard, "Object identification with tactile sensors using bag-of-features," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009.

[12] A. Drimus, G. Kootstra, A. Bilberg, and D. Kragic, "Classification of rigid and deformable objects using a novel tactile sensor," in *2011 15th International Conference on Advanced Robotics (ICAR)*, 2011.

[13] J. Lin, R. Calandra, and S. Levine, "Learning to identify object instances by touch: Tactile recognition via multimodal matching," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019.

[14] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, and E. H. Adelson, "Shape-independent hardness estimation using deep learning and a gelsight tactile sensor," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[15] 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.

[16] F. R. Hogan, M. Bauza, O. Canal, E. Donlon, and A. Rodriguez, "Tactile regrasp: Grasp adjustments via simulated tactile transformations," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[17] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez, "Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1927–1934.

[18] T. P. Tomo, S. Somlor, A. Schmitz, L. Jamone, W. Huang, H. Kristanto, and S. Sugano, "Design and characterization of a three-axis hall effect-based soft skin sensor," *Sensors*, vol. 16, no. 4, p. 491, 2016.

[19] L. Piegl and W. Tiller, *The NURBS book*. Springer Science & Business Media, 2012.

[20] A. Krishnamurthy, R. Khardekar, S. McMains, K. Haller, and G. Elber, "Performing efficient nurbs modeling operations on the gpu," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 4, pp. 530–543, 2009.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[22] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3d recognition and pose using the viewpoint feature histogram," in *Proceedings of the 23rd IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, October 2010.

[23] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support vector method for novelty detection," in *Advances in neural information processing systems*, 2000, pp. 582–588.

[24] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis—a survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2013.

[25] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," *arXiv preprint arXiv:1703.09312*, 2017.

[26] S. Jain and B. Argall, "Grasp detection for assistive robotic manipulation," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 2015–2021.

[27] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015.

[28] A. ten Pas, M. Gualtieri, K. Saenko, and R. Platt, "Grasp pose detection in point clouds," *The International Journal of Robotics Research*, vol. 36, no. 13-14, pp. 1455–1473, 2017. [Online]. Available: https://doi.org/10.1177/0278364917735594

[29] J. Ilonen, P. Paalanen, J.-K. Kamarainen, and H. Kalviainen, "Gaussian mixture pdf in one-class classification: computing and utilizing confidence values," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006.