Haptic Classification and Recognition of Objects Using a Tactile Sensing Forearm

Tapomayukh Bhattacharjee*, James M. Rehg, and Charles C. Kemp

Abstract—In this paper, we demonstrate data-driven inference of mechanical properties of objects using a tactile sensor array (skin) covering a robot’s forearm. We focus on the mobility (sliding vs. fixed), compliance (soft vs. hard), and identity of objects in the environment, as this information could be useful for efficient manipulation and search. By using the large surface area of the forearm, a robot could potentially search and map a cluttered volume more efficiently, and be informed by incidental contact during other manipulation tasks. Our approach tracks a contact region on the forearm over time in order to generate time series of select features, such as the maximum force, contact area, and contact motion. We then process and reduce the dimensionality of these time series to generate a feature vector to characterize the contact. Finally, we use the k-nearest neighbor algorithm (k-NN) to classify a new feature vector based on a set of previously collected feature vectors. Our results show a high cross-validation accuracy in both classification of mechanical properties and object recognition. In addition, we analyze the effect of taxel resolution, duration of observation, feature selection, and feature scaling on the classification accuracy.

I. INTRODUCTION

Autonomous manipulation in cluttered environment is a difficult problem due to the possibility of unavoidable contact with obstacles. Haptic technology can serve as a useful tool for enabling effective manipulation. A robot could utilize haptic information obtained from its interaction with objects in the environment to maneuver itself through clutter. While doing so, knowledge of the mechanical properties of an object, such as its mobility, compliance, weight and surface properties like friction could be especially useful. Such information is not only useful for efficient manipulation but can also be used to haptically search for and recognize an object. By using the large surface area of the forearm, a robot could potentially search and map a cluttered volume more efficiently than if it only uses its hand.

Note that non-contact sensing modalities, such as cameras and laser scanners, are not always effective for manipulation in clutter. Such sensors have limited ability to infer mechanical properties [1], [2]. Humans rely heavily on their sense of touch for manipulation tasks and can even manipulate objects using only tactile information [3]. Visually similar objects or environments can have very different mechanical properties. For example, compliant leaves can be pushed aside without generating large forces. At the same time, if there is something fixed and rigid occluded by the leaves, such as a branch or a concealed object, then the total system or contact behavior can become quite rigid. Likewise, when searching for an object of interest in rubble with only non-contact sensing, it may be hard to distinguish between things that are stuck and those that can be pushed or pulled aside. Hence, estimating the mobility of an object could be highly valuable for manipulation. However, mobility estimation studies using haptic sensing are lacking.

In this paper, we specifically address the mobility-based classification problem and also estimate object compliance characteristics using haptic sensing techniques during manipulation tasks. The rest of the paper is arranged as follows. In Section II, we review the related work in this domain. Section III describes the approach that we have used in tackling the problem of haptic data based classification and recognition of environmental objects. In Section IV, we present the results of applying our algorithm in real-life experimental situations and analyze the effects of various conditions on the performance of our algorithm. In Section V, we present conclusions from our work.

II. RELATED WORK

Object categorization, based on their various characteristics for specific tasks, has been dealt with extensively in previous studies. Researchers have addressed the problem of object categorization based on the objects’ various characteristics such as material, shape and functional properties using single or multiple sensor modalities as given below.

A. Material Property based Classification

Previous work on material property classification is perhaps the most closely related work to our approach. Although...
we do not explicitly model material properties, the features we extract from the interactions between the robot arm and environmental objects are a direct consequence of these material properties which affect the interaction dynamics. Drimus et al. [4] classify rigid and deformable objects based on haptic feedback from a novel tactile sensor using flexible piezoresistive rubber. They represent tactile information from a palpation procedure as a time-series of features, and use k-nearest neighbor classifier to categorize the objects [4].

This is the most similar prior work of which we are aware. However, in addition to classifying compliant and rigid objects, we also classify fixed versus movable objects. This classification could be important in cluttered environments because the mobility of an object coupled with its compliance suggests how much force needs to be applied to either change its state for effective manipulation, or give up. Moreover, our method does not employ an exploratory / probing procedure used explicitly for classification as in [4]. Instead, our method extracts the required features through general contact during a stereotyped reaching motion. This scenario is more representative of incidental contact that could occur with the forearm during a manipulation task. Also, the features extracted in our method correspond with physical quantities whereas the features in [4] are tactile array images. The tactile images do not have a clear interpretation with respect to an object’s mechanical properties. This makes it difficult to understand the underlying dynamics of the factors which might contribute in the haptic object classification.

Sukhoy et al. investigate the use of a vibrotactile sensor for surface texture recognition using a Support Vector Machine (SVM) classifier [5]. Ho and Jones develop a thermal display for simulating the thermal cues associated with making contact with various materials of different properties [6]. Kim and Kesavadas present a methodology for estimating the material properties of objects by an active tapping procedure [7]. Takamuku et al. use a simplified version of an artificial skin with strain gauges and PVDF films and estimate the material properties of objects with the help of exploratory procedures like tapping and squeezing [8]. Hosoda and Iwase [9] use a Bionic hand and utilize its hand compliance to grip an object to obtain haptic data. They use a recurrent neural network to classify objects based on learned haptic cues from dynamic interactions [9].

Frank et al. [10], [11] address the problem of determining the elasticity properties of deformable objects by minimizing the difference between the actual deformed surface of an object and its corresponding finite element model. They use a 3D registration technique based on point clouds obtained from a depth camera for this purpose. Ueda et al. [12] also address the issue of extracting rheological properties of deformable objects based on haptic vision. They monitor the surface deformation of an object by exerting a known contact force and then observe how the object returns to its original shape after the contact force is withdrawn. Nizar et al. [13] address the problem of the classification of material type and surface properties by developing a sensor which uses a lightweight plunger probe to detect surface properties. They also used an optical mouse sensor to obtain surface images and used a Radial Basis Function Neural Network for classification. Platt et al. [14] use proprioceptive and load-based tactile information to localize features such as a bump, a snap and a grommet embedded in flexible materials like a fabric. They claim that using both tactile and proprioceptive data results in a gain in the localization performance. Matheus and Dollar [15] estimate the static friction between different object-surface pairs while sliding a variety of objects which affects the mobility of an object on a specific surface.

B. Shape based Classification

Kikuuwe and Yoshikawa use impedance perception schemes to extract information on the local properties of object surfaces and categorize objects into two classes such as flat and convex cylindrical surfaces [16]. Schneider et al. [17] use touch sensors installed in the fingertips of a manipulation robot to get low-resolution intensity images obtained from multiple grasping interactions. They apply a bag-of-words approach and unsupervised clustering techniques to categorize objects using only the haptic feedback [17]. Allen et al. use superquadric primitives for model-based haptic object recognition and perform object recognition using the similarity between the parameters of the recovered superquadrics [18]. Caselli et al. also use volumetric models for dynamic integration of geometric information with haptic exploration data and formulate the problem as a match-to-sample scheme using the recovered model features [19]. Faldella et al. utilize an unsupervised Kohonen self-organizing feature map for performing a match-to-sample classification of 3-D objects using a volumetric model called a wrapping polyhedron [20]. Pezzementi et al. view tactile sensor readings as images and apply PCA techniques to identify the principal components of identified features, and then cluster them as well as build per-class histograms as a class characteristic [21]. Gorges et al. [22] additionally include some passive joints in the tactile sensor system of their robot hand so that the tactile sensor conforms to the object shape during interaction which could help to acquire more information for shape reconstruction. They use Self-Organizing Maps (SOMs) for identifying the haptic key features and use a Bayes Classifier to classify the objects based on their features [22].

C. Functional Property based Classification

Sinapov et al. use acoustic properties of objects during specific interaction schemes to classify the objects and the behavioral interactions performed with them such as grasping, shaking, dropping, pushing and tapping behaviors on 36 different household objects [23]. Berquist et al. monitor the changes in the joint torques of a robot while it performs five exploratory procedures such as lift, shake, crush, drop, and push on several objects and show that the robot can learn to recognize objects solely on the basis of proprioceptive information [24]. Griffith et al. use multiple exploratory behaviors and employ clustering techniques for categorizing containers and non-containers by
extracting visual and acoustic features from its interaction with objects and then employing unsupervised clustering techniques to form several categorizations [25]. Sinapov et. al. also combine proprioceptive and auditory feedback and use a behavior-grounded relational classification model to recognize categories of household objects [26].

III. METHODS

We used supervised machine learning to analyze data from a skin sensor covering the forearm of a humanoid robot named ‘Cody’. Our goal was to classify an object that the robot has not previously interacted with as being in one of four categories: Rigid-Fixed, 2) Rigid-Movable, 3) Soft-Fixed, and 4) Soft-Movable. We also used the same methods to haptically identify a specific object that the robot has previously interacted with. In Secs. IV-B and IV-C we show the effects of the spatial resolution of the taxels, and the duration of the haptic interaction, on the classification accuracy. Section IV-D shows the effect of different feature scaling schemes on the performance of the algorithm while Section IV-E highlights the importance of the individual features for both classification and recognition purposes.

A. Experimental Setup

The experimental setup for our data collection is described below.

1) The Robot ‘Cody’: Cody, as shown in Fig. 1, is a statically stable mobile manipulator weighing roughly 160 kg. The components of the robot are: Meka A1 arms, a Segway omni-directional base and a Festo 1-D.O.F. linear actuator. The arms consist of two 7-D.O.F. anthropomorphic arms with series elastic actuators. When we control these arms, each joint simulates a low-stiffness visco-elastic torsional spring. We control the robot’s arms by changing the equilibrium angles of these simulated springs over time.

Cody has a force sensitive skin across its entire forearm. Meka Robotics and the Georgia-Tech Healthcare Robotics Lab developed the forearm tactile skin sensor, which is based on Stanfords capacitive sensing technology, as described in Ulmen et. al. [27]. The skin consists of a capacitive pressure-sensor array. We refer to the elements of this array as taxels (tactile pixels). There are 384 taxels on the entire skin which are distributed into a 24 X 16 array with each taxel being 9mm X 9mm in size. The array of taxels reports the estimated force applied to each taxel at 100Hz.

2) Data Collection: For our experiments, we used a set of 18 objects, shown in Fig. 2. We selected large objects that have mostly uniform material properties and vary widely in their mass, friction, and compliance. For each object, we collected haptic data by commanding the same equilibrium point trajectory for the arm and recording the sensor readings from the taxels of the forearm skin at approximately 100Hz.

We labeled each of these objects as soft or rigid. We considered pillow-like materials, foam, bubble-wrap, and vegetation to be soft, and all other objects to be rigid. For objects that could be pushed aside by the robot’s motion, we also fixed them with a clamp or a heavy weight, so that we could have both movable and fixed conditions. We collected a dataset of 5 trials for each of the 18 different objects, 10 of them in both fixed and movable conditions, 4 of them in only fixed conditions and the remaining 4 in movable conditions. Fig. 3 shows three images from one trial of the robot interacting with a plant. It also shows the data from the forearm sensor, visualized as an image.

B. Preprocessing, Feature Selection, and Dimensionality Reduction

We recorded data from the forearm taxel array at a 100Hz sampling rate. We truncated this time series data to begin at the estimated onset of contact between the robot and the object. We then represented the data at every time step as a gray-scale image, as shown in Fig. 3. We converted this image to a binary image representing the taxels in contact by applying a threshold to each taxel. Note that this hand-tuned threshold was not same for all objects. This was done to account for some of the more rigid or coarser objects for which a covering was put over the otherwise bare skin-sensor to ensure its safety. Then, we computed connected components to segment the contact regions. For the connected component with the largest area, we computed three features. Figure 4 depicts the complete experimental protocol.

Fig. 5 shows the three features. The first feature was the maximum force that the robot applied to the object at every time step. Second, we estimated the area of the contact between the arm and the object as the number of taxels in the connected component. Third, we estimated the distance the 3D position of the centroid of the connected component traveled in the world frame from its 3D position at the onset of contact. We assumed that the robot’s torso did not move throughout the trials and used the forward kinematics from the robot’s torso to the contact location center on the robot’s forearm to estimate the 3D positions and distance. We expected these three features to be informative about the object’s softness and movability. For example, with increasing force applied to a soft, fixed object, we
Fig. 3. Sequence of images that illustrate our data collection for our experiments on inferring mechanical properties of objects (foliage). Each image shows a picture of the robot Cody, and a visualization of the data from the forearm skin sensor as a 24X16 image (dark pixels correspond to larger forces). The leftmost picture shows a non-contact situation, the middle one corresponds to the situation just after the onset of contact while the rightmost picture shows the situation when the robot has pushed the foliage to the maximum extent consistent with its motion-limits.

would expect the contact area to increase. Likewise, we would expect the 3D position of the contact to travel when encountering movable and soft objects. When making contact with a rigid and fixed object, we would expect the maximum force to go up.

We created 40 element vectors for each of the feature time-series by uniformly sub-sampling the 100Hz measurements. We then concatenated the resulting vectors of maximum force, number of taxels in the contact region, and motion of the centroid of the contact region to form a feature vector of length 120 for each trial considering the first 1.2s time-window after the onset of contact. To reduce the influence of noise and overfitting, we computed a low dimensional representation of the data with principal component analysis (PCA) before classification with a k-nearest neighbor classifier (k-NN). In our classification experiments, we used a maximum of 20 principal components for dimensionality reduction.

IV. RESULTS AND DISCUSSION

A. Classification Results

We used a k-NN classifier to test the classification accuracy for two different classification problems. In each case, we picked the number of principal components and the value
for \( k \) by performing a grid search over these two parameters and picking the values associated with the highest leave-one-out cross-validation accuracy.

Fig. 6 shows the confusion matrix for the classification into four categories: 1) Rigid-Fixed, 2) Rigid-Movable, 3) Soft-Fixed, and 4) Soft-Movable. The classification accuracy was 80% with \( k = 2 \) and dimensionality 20. Many of the classification errors were between the Rigid-Movable and Soft-Movable classes.

Fig. 7 shows the confusion matrix for a two category classification problem where we used the data to classify an object as either fixed or movable. The classification accuracy was 91.43% with \( k = 4 \) and dimensionality 3.

Fig. 8 shows the confusion matrix for recognizing the specific object that the robot interacted with. The classification accuracy was 72.14% with \( k = 1 \) and dimensionality 7.

The next subsections analyze the effect of various conditions, parameters and features on the classification and recognition accuracy.

B. Effect of Taxel Resolution

We performed the four category classification experiment and the object recognition experiment for different spatial resolutions of the taxels.

Table I shows the leave-one-out cross-validation accuracy, the values for the number of neighbors and the dimensionality of the subspace that resulted in the highest accuracy for each taxel resolution. Fig. 9 shows the best classification accuracy that we obtained for the different resolutions. Compared to 1 taxel/meter, 112 taxel/meter resolution improved the classification accuracy by 24.44% for the four category classification problem and by 25.01% for object recognition.

C. Effect of Time Window

We also investigated the effect of varying the time-window of the sensor data on the classification accuracy. To do this, we used the same methods, except that we uniformly sampled 40 measurements of each feature type over a shorter time-window. Fig. 10 shows that a shorter time-window of 0.8 seconds resulted in substantially lower classification accuracy. This is unfortunate, since faster estimation could improve the robot’s efficiency. Estimation over shorter periods of time might be possible with faster motions. Note that for our experiments the robot forearm joint velocity was around 0.35 rad/s prior to contact. Other measurements, such as higher frequency tactile information and different modalities, such as shear force and temperature, might enable more rapid estimation. For example, the surface texture of an object could potentially be sensed soon after initial contact. On the other hand, determining whether or not an object will slide depends on the applied force. So, we would expect that there would be some delay as the force applied by the robot ramps up and potentially overcomes static friction. For an object to be considered movable in our experiments, it needed to be moved by the robot’s stereotyped motion.

D. Effect of Feature Scaling

Researchers often argue that proper scaling of different feature vectors might be necessary for high performance [28]. To analyze this aspect, we employed several scaling schemes to our original data, which were in units of taxels (contact region area), Newtons (maximum force), and Meters (displacement of the contact region center), to see how the performance was affected by the choice of scaling function. We used four different scaling methods as described by Eqs. 1-5 denoted as Methods I to V respectively. Methods I to IV
given by Eqs. 1-4 scale all three features within a uniform range. However, Method V, given by Eq. 5 scales up the contact motion feature to the range of the other two features such as contact area and maximum force. We tested this since both the contact area and maximum force features are in a comparable range of values while the values for the contact motion feature were much smaller as seen from Fig. 5.

\[
S_f = \frac{(f - \text{mean}(f))}{\text{std}(f)}
\]

∀ \( f \in \{\text{Max. Force, Contact Area, Contact Motion}\} \)

The results of the 4-category, 2-category classification accuracy and object recognition performance, with feature scaling, are given in Table II.

Results from Table II show that Method V has the highest 4-category classification accuracy while Method I
and II have the highest 2-category classification accuracy. The highest object recognition performance was obtained using the original units without additional scaling. Also, the accuracy enhancement for classification algorithm was negligible compared to the unscaled feature based results. None of the scaling schemes showed a consistent increase in accuracy for all the object classification and recognition cases when compared to the unscaled data. Overall, scaling the original units did not have clear benefits.

E. Effect of Different Features

Lastly, we analyzed the effect of individual features and their combinations on the performance of the classification and object recognition tasks. We implemented the algorithm with different combinations of features for both 4-category and 2-category classification schemes as well as object recognition scheme. Table III shows the cross-validation accuracy.

Table III shows that using both maximum force and contact area features gave better overall performance. The addition of contact motion feature did not improve the performance considerably. If only one feature was to be used, the probable choice would be to use the maximum force over time feature. Also, the choice of a particular feature had little effect on the performance of the 2-category classification scheme. The lack of influence due to the motion feature may be due to the robot's stereotyped motion. Although the robot’s compliance resulted in different contact motion over time, the motions resulted from the same controller commands over time, and thus had a form of temporal consistency.

V. CONCLUSION

In this paper, we developed an object classification and recognition algorithm using haptic information obtained from interactions of a tactile sensing forearm with environment objects. Our algorithm classified objects into one of the four categories: Rigid-Fixed, 2) Rigid-Movable, 3) Soft-Fixed, and 4) Soft-Movable. We extracted features such as time-trends of maximum force, contact area and contact motion from the haptic interactions and preprocessed those to show the information from the onset of contact. We computed a low-dimensional representation of the data using Principal component analysis and used a Nearest Neighbor classifier for classification and recognition purposes. Results showed that the classification and recognition algorithms worked well. We studied the effect of the skin-sensor resolution on the performance of the algorithm. It showed that the skin-sensor with higher resolution (384 taxels) enhanced the performance of the algorithm compared to 1 taxel resolution. We also analyzed the effects of time-window of haptic interaction, feature-scaling, and selection of specific features on the overall performance. These studies could provide useful intuitions on the various aspects of this task at hand and might serve as valuable guidelines for our future work in this domain.

VI. ACKNOWLEDGMENTS

We thank Advait Jain for his technical input. We gratefully acknowledge support from DARPA Maximum Mobility and Manipulation (M3) Contract W911NF-11-1-603. We thank Mark Cutkosky and the Stanford Biomimetics and Dexterous Manipulation Lab for their contributions to the forearm tactile skin sensor.

REFERENCES

TABLE II

<table>
<thead>
<tr>
<th>Scheme</th>
<th>No Scaling</th>
<th>Method-I</th>
<th>Method-II</th>
<th>Method-III</th>
<th>Method-IV</th>
<th>Method-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification into 4 categories</td>
<td>80%</td>
<td>66.43%</td>
<td>67.86%</td>
<td>75%</td>
<td>61.43%</td>
<td>82.14%</td>
</tr>
<tr>
<td>Classification into 2 categories</td>
<td>91.43%</td>
<td>92.14%</td>
<td>92.14%</td>
<td>87.14%</td>
<td>92.86%</td>
<td>90.71%</td>
</tr>
<tr>
<td>Object Recognition</td>
<td>72.14%</td>
<td>65%</td>
<td>67.86%</td>
<td>62.86%</td>
<td>62.14%</td>
<td>70.71%</td>
</tr>
</tbody>
</table>

TABLE III

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Features</th>
<th>Maximum Force</th>
<th>Contact Area</th>
<th>Contact Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-Category Classification Accuracy</td>
<td>Maximum Force</td>
<td>81.43%</td>
<td>80%</td>
<td>75.71%</td>
</tr>
<tr>
<td></td>
<td>Contact Area</td>
<td>80%</td>
<td>73.37%</td>
<td>72.86%</td>
</tr>
<tr>
<td></td>
<td>Contact Motion</td>
<td>75.71%</td>
<td>73.37%</td>
<td>72.86%</td>
</tr>
<tr>
<td>2-Category Classification Accuracy</td>
<td>Maximum Force</td>
<td>90%</td>
<td>91.43%</td>
<td>84.29%</td>
</tr>
<tr>
<td></td>
<td>Contact Area</td>
<td>91.43%</td>
<td>90%</td>
<td>89.29%</td>
</tr>
<tr>
<td></td>
<td>Contact Motion</td>
<td>84.29%</td>
<td>89.29%</td>
<td>82.14%</td>
</tr>
<tr>
<td>Object Recognition</td>
<td>Maximum Force</td>
<td>64.29%</td>
<td>72.14%</td>
<td>66.43%</td>
</tr>
<tr>
<td></td>
<td>Contact Area</td>
<td>72.14%</td>
<td>56.43%</td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td>Contact Motion</td>
<td>66.43%</td>
<td>60%</td>
<td>49.29%</td>
</tr>
</tbody>
</table>


