Multimodal Tactile Perception of Objects in a Real Home

Tapomayukh Bhattacharjee1, Henry M. Clever2, Joshua Wade3, and Charles C. Kemp2

Abstract—When operating in human environments, such as homes, robots could use tactile sensing to better perceive objects. A challenge that has not been sufficiently addressed is the influence of an object’s surroundings on tactile perception. Prior research has focused on perception of objects in laboratory settings. Yet, a number of factors found in homes can affect multimodal tactile sensing. For example, the time-varying thermal characteristics of an object’s surroundings, such as sunlight, HVAC, and refrigeration, can affect thermal sensing. Likewise, the placement of an object with respect to other objects and surfaces will affect force sensing and alter the way an object moves when pushed. In order to investigate these and other issues, we had a mobile robot reach out and push 47 different objects found in a real home over a three day period resulting in 1340 pushing episodes. We then characterized the performance of data-driven methods (k-NNs, SVMs, HMMs, and LSTMs) for a variety of tactile perception problems using the first two seconds of force, thermal, and motion sensing data collected by the robot. We paid particular attention to the ability of these methods to generalize what they have learned to different robot velocities, times of day, and object instances. Our results demonstrate the value of multimodal tactile sensing and data-driven methods for tactile perception from short-duration contact, and also illustrate the great diversity of real-world phenomena relevant to tactile sensing.

Index Terms—Haptics and Haptic Interfaces; Force and Tactile Sensing; Perception for Grasping and Manipulation

I. INTRODUCTION

INFERRING properties of objects or distinguishing objects in the real world by touch has many interesting challenges. In laboratory settings, we have previously seen encouraging results for haptic perception with force and motion sensing [1], as well as thermal sensing [2] separately. Our objective in this work is to use these multimodal sensing modalities together for inferring properties of objects in a real home. For this, we used data-driven approaches using signals from a multimodal sensing module at the end effector of a robot which made contact with in-situ objects in a real home where three people live. However, in a real home, factors like sunlight, HVAC, and refrigeration can affect thermal sensing, whereas an object’s physical interaction with surfaces and other objects in its vicinity due to manipulation can affect force sensing. It can also change the object’s movement pattern when pushed.

One of the important aspects of haptics is that sensing depends on action. For example, changing the velocity of a robot arm when pushing an object can alter the forces generated (force sensing) or the way the object moves (motion sensing). Collecting data at a different time of the day can affect thermal sensing because the same object can be at a different temperature. Our focus in this work is on generalizing the haptic perception performance across various robot velocities, times of day and object instances using force, thermal, and motion sensing modalities.

For haptic perception, we use various widely used and state-of-the-art data-driven methods. We use k-nearest neighbors (k-NNs), support vector machines (SVMs), hidden Markov models (HMMs), as well as long short-term memory networks (LSTMs). We want robots to infer various properties of an object. Inferring whether an object is soft or hard, or if the object moved during manipulation can help devise strategies such as, avoiding a hard object but pushing through a soft object which moved, to reach a goal [1]. Detecting an object which moved may also help in deciding if the robot should move the object to access a new location or avoid it [3]. We are also interested in material based haptic labels because knowing the material of an object is informative [2].

In addition to inferring object properties, we also want robots to distinguish various objects. To distinguish objects, we identified 50 tasks relevant to ADLs (Activities of daily living) and IADLs (Instrumental activities of daily living) such as ‘putting a towel on rack’, ‘fetching a bottle from refrigerator’, ‘pushing a door handle on door surface’ to open the door, etc. Thus, conditioned on the task, we want robots to distinguish a target object (tactile foreground) from a background object in its vicinity (tactile background) that the robot may come in contact with while performing the task. We are also interested in analyzing if robots can distinguish different object parts given
an object (such as 'chair cushion' given chair, or 'mattress' given bed).

II. RELATED WORK

To the best of our knowledge, robotic tactile perception of real home environments remains an unexplored area of research. Researchers used data collected in households for analyzing grasping [4] by humans and organization behaviors in the kitchen [5]. Researchers also developed data acquisition devices to collect multimodal data from objects in real home bathrooms [6] and sample surfaces [7]. Gandhi et al. deployed unmanned aerial vehicles to crash into objects in indoor academic campus environments to learn a navigation policy [8]. Most of the related work on haptic perception has focused on using exploratory probing behaviors (squeezing, tapping, sliding, etc.) to extract information from contact with objects in laboratory settings (see Sections II-A and II-B). In this work, we focus on extracting information using a simple linear motion with a linear actuator on a mobile robot pushing objects in a real home with two seconds of contact. In addition, few related studies have looked into generalizing the haptic perception performance to new robot and environment conditions which are different from the conditions used during training. For this paper, we focus on related work with force and thermal sensing modalities.

A. Single Modality : Force or Thermal

Researchers have used force sensing to classify objects based on their material property, shape property, functional property etc. [1], [9], [10], [10]–[14]. For a more detailed survey, refer to [1], [3]. Though thermal sensing is relatively unexplored in robotics compared to force sensing, there have also been studies on hardware development [15], [16] and material recognition using only thermal sensing. For a detailed overview of such material recognition studies, please refer to [2], [6], [17], [18].

B. Multiple Modalities : Force and Thermal

1) Hardware Development: Researchers developed multimodal sensors using various techniques such as a capacitive and thermal sensor [16], conductive rubber based force sensors and temperature sensors [19], a polymer-based tactile and thermal sensor [20], capacitive tactile sensors with temperature-dependent semiconductors [21], and a tactile and thermal sensor using single pressure-conductive rubber sheet [22].

2) Object Recognition: Researchers have also used force and thermal sensing together for object recognition. Using force and thermal sensing, Takamuku et al. [23] successfully classified 5 materials and Engel et al. [24], [25] achieved 90% accuracy over 50 trials for recognizing 5 materials. Caldwell et al. [26] used force and thermal sensing to infer texture, stiffness and object profile, temperature and thermal properties of 7 materials using exploratory behaviors. Xu et al. [27] used multimodal sensor feedback to identify various materials with a BioTAC sensor [28]. Chu et al. [28] attached two BioTAC sensors to a robotic gripper and assigned 24 adjectives to 60 objects using four exploratory behaviors. They used discrete HMMs to construct a feature vector of likelihoods and used binary SVM classifiers on the feature vector for classification [29]. Schmitz et al. [30] used power grasping of objects and multiple modalities for object recognition with deep learning. Hoelscher et al. [31] used both thermal and force features for recognizing 49 objects and found thermal features to be the most informative.

III. RELEVANT HAPTIC LABELS

A. Inferring Object Properties

1) Compliance and Mobility: We want robots to infer an object’s compliance and mobility through physical interaction. Specifically, we perform three kinds of classification tasks: (a) ‘Hard’ vs. ‘Soft’, (b) ‘Moved’ vs. ‘Unmoved’, and (c) ‘Hard-Unmoved’ vs. ‘Soft-Unmoved’ vs. ‘Hard-Moved’ vs. ‘Soft-Moved’.

2) Material: We are also interested in material property based haptic labels. We sampled 67 object parts relevant to ADLs and IADLs in the bedroom, kitchen and bathroom of a household. These object parts are made of 14 different materials. Our objective is to perform binary classification between every binary combination of these 14 materials (inspired by ‘tactile foreground’ and ‘tactile background’ [6]) thus leading to 91 such comparisons. The 14 materials are ‘Foil’, ‘Aluminium’, ‘Wood’, ‘Steel’, ‘Dry Wall’, ‘Plastic’, ‘Glass’, ‘Paper’, ‘Porcelain’, ‘Granite’, ‘Cardboard’, ‘Fiberglass’, ‘Vegetable Matter’, and ‘Fabric’.

B. Distinguishing Objects

1) Tactile Foreground vs. Tactile Background: During a manipulation task while moving a target object (tactile foreground), a robot may come in contact with another object (tactile background) in its immediate vicinity [6]. We have identified 50 such comparisons for the objects relevant to ADLs and IADLs in a household environment, such as ‘rack vs. towel’, ‘book cover vs. book spine’, ‘TV vs. TV remote’, ‘kitchen faucet vs. kitchen backsplash’, ‘chair cushion vs. chair frame’, ‘socket vs dry wall’ etc. Our objective is to see if our algorithms can distinguish objects in each of the 50 foreground-background pairs.

2) Object Parts given Object: Finally, we analyze if robots can distinguish different object parts given an object. This could be useful in assistive manipulation scenarios such as assisting a human sitting on a chair or lying in a bed. Knowing whether the robot made contact with ‘chair cushion or chair frame given its manipulating near a chair’ could be useful in devising manipulation or control strategies used by the robot. Similarly, knowing if the robot is in ‘contact with bed frame or the mattress or pillow on the bed given its performing a task near the bed’ could provide relevant and useful information to the robot’s manipulation task. We identified 13 such comparisons involving two to three object parts given an object in the bedroom, kitchen, and bathroom of a household.

IV. GENERALIZATION TASKS

We focus on generalizing the classification results using force and thermal sensing modalities across different velocities, times of day, and instances.
which includes the objects.

A. Experimental Setup

Hardware for analog-to-digital conversion (ADC) and buck converter in this paper.

We did not perform any analysis with the audio and acceleration sensing modalities in analog-to-digital converters (ADCs). We attached a 3D printed base. Two Teensy 3.2 microcontrollers and a buck converter on a separate module provide power to the sensors module with force, audio, acceleration, active and passive thermal sensing modalities. Individual sensors are sampled the space such that at every trial, the robot collected data from the same spot over multiple trials. The waiting period of 5 seconds also ensured that the object was at the same initial temperature for the next trial.

B. Generalizing Across Time of Day

We want to see if results obtained using data-driven methods trained on data collected at one time could generalize to data collected at another time on another day. We collected the data from the same set of objects at two different times over a period of three days.

C. Generalizing Across Instance

We want to see if results obtained using data-driven methods trained on a set of objects in one specific category with a particular velocity can generalize to another object in the same category with the same velocity. We focused this analysis on the ‘Hard vs. Soft’ and ‘Moved vs. Unmoved’ classification problems. We used a ‘leave-one-object-out’ cross-validation scheme to test for generalization, meaning that data from the same object does not overlap both training and testing sets.

V. METHODS

A. Experimental Setup

Our experimental setup consists of a multimodal sensor attached at the end of a linear actuator on a mobile robot (a 2-DOF Parallax Arlo robot base), and the real-home environment which includes the objects.

1) The Multimodal Sensor: Figure 2 shows the multimodal tactile sensor module with force, audio, acceleration, active and passive thermal sensing modalities. Individual sensors are attached to a 3D printed base. Two Teensy 3.2 microcontrollers and a buck converter on a separate module provide power to and read data from the sensors using the microcontrollers’ built-in analog-to-digital converters (ADCs). We did not perform any analysis with the audio and acceleration sensing modalities in this paper.

The fabric-based force sensor is based on the design in [32]. It reads force data at 1kHz. We used a voltage dividing reference resistor, $R_{\text{ref}} = 1 \, \text{k}\Omega$. Because the fabric-based electrodes are small compared to the touched objects, we assumed the contact would cover the entire sensor’s area. In this work we used the fabric-based force sensor to detect the start of contact with a force threshold of 0.1 N.

The active thermal sensor uses a self-heated 10 kΩ B57541G1103F NTC thermistor [33], with adjustable voltage input from a digitally-controlled buck converter (see Fig. 2). We use a closed-loop temperature controller to heat the thermistor to 55°C prior to contact with the object. The sensor reads data at 100 Hz. Once the fabric-based force sensor detects contact, this closed-loop temperature controller turns off by holding the desired buck converter voltage, $V_{\text{des}}$, constant. This is necessary to ensure that the temperature controller does not cancel potentially-informative temperature changes resulting from contact.

The passive thermal sensor uses a second 10 kΩ B57541G1103F NTC thermistor [33] similar to the active thermal sensor reading data at 100 Hz. The passive thermal sensor is powered by a constant 3.3 V input from the microcontroller rather than a higher voltage buck converter.

2) The Environment: The environment was the household of one of the co-authors of this paper where three people live. It consisted of objects in the bedroom, bathroom, and the kitchen of the household, relevant to ADLs and IADLs. There were 12 objects with 22 object parts in the bedroom, 9 objects with 15 object parts in the bathroom, and 26 objects with 30 object parts in the kitchen, as shown in Fig. 3.

B. Experimental Procedure

Our experimental procedure consisted of a teleoperation phase and an autonomous phase. First, in the teleoperation phase, we used a joystick controller to manually position the mobile robot at a location feasible for reaching a specific object before starting the trials. Then in the autonomous phase, for each experimental trial, the robot autonomously reached towards the object to push it. We define a trial as a single manipulation behavior of linear actuator extending and pushing an object and then bringing the actuator back. We set the robot-arm linear actuator to move at a specific velocity (‘slow : 3 cm/s’ or ‘fast : 6 cm/s’) until it reached a contact force threshold of 5 N. We programmed the robot-arm to be in contact with the object for 5 seconds or until the robot arm pushes the object for 2 cm, whichever is earlier. After the contact duration, the linear actuator fully retracted and waited for 20 seconds for the thermal sensors in the multimodal sensor module to reach a consistent initial condition prior to beginning the next trial. Note that for objects with large surface area, we sampled the space such that at every trial, the robot collected data from a different part of the object surface. For objects with small surface area, we collected data from the same spot over multiple trials. The waiting period of 20 seconds also ensured that the object was at the same initial temperature for the next trial.

Fig. 2: Multimodal sensor module and Teensy 3.2 microcontroller Hardware for analog-to-digital conversion (ADC) and buck converter to power active thermal sensor.

A. Generalizing Across Velocity

We want to see if results obtained using data-driven methods trained on data collected at one velocity from various objects can generalize to data collected from the same set of objects at another velocity. We varied the robot-arm velocity to two distinct velocities.

B. Generalizing Across Time of Day

We want to see if results obtained using data-driven methods trained on data collected at one time could generalize to data collected at another time on another day. We collected the data from the same set of objects at two different times over a period of three days.

C. Generalizing Across Instance

We want to see if results obtained using data-driven methods trained on a set of objects in one specific category with a particular velocity can generalize to another object in the same category with the same velocity. We focused this analysis on the ‘Hard vs. Soft’ and ‘Moved vs. Unmoved’ classification problems. We used a ‘leave-one-object-out’ cross-validation scheme to test for generalization, meaning that data from the same object does not overlap both training and testing sets.

C. Data Collection and Preprocessing

We collected data from all the objects over a period of three days. Our first day’s data collection started at 11.50 AM and continued until 6.50 PM, the second day from 9 AM until 10.45 PM, and the third day from 9.40 AM until 10.40 PM. During the data collection, we collected data from the same object at different times of the day (either ‘Morning - early Afternoon’ or ‘Afternoon - late Night’).

For each object part at one specific time, once we manually positioned the robot’s mobile base at a particular location, the robot autonomously reached to touch a given object 5 times with ‘slow’ velocity (3 cm/s) and then 5 times with ‘fast’ velocity (6 cm/s). If the object moved during ‘slow’ velocity trials, we reset the object back to the original position before the ‘fast’ velocity trials. We did this to ensure uniform initial conditions for the first trial in each of the ‘slow’ and ‘fast’ group of trials. We collected 20 trials (5 trials x 2 velocities x 2 times) for each of the 67 object parts, totaling 1340 trials. For each, we collected time-series vectors of force (f), motion (relative position of the robot arm after onset of contact) (m), active thermal (h), and passive thermal (t) modalities. We sampled force and motion signals at 1000 Hz to match the frequency of active and passive thermal signals (100 Hz). Note, when we used combinations of features, we normalized each feature by subtracting the mean and dividing by the variance of the feature across all the data.

Two experimenters (co-authors) independently labeled the objects for material-based labels, ‘hard’ or ‘soft’ labels, and ‘moved’ or ‘unmoved’ labels. We convened after the experiments and found no disagreements. For material-based labels, we noted both the ‘surface material’ (i.e. paper label) and the ‘dominant material’ (i.e. glass bottle) for each object, but used the dominant material label for analysis. We also measured the stiffness of the objects using a compression spring. The compression spring was unable to measure the stiffness for very hard objects with $k > 100,000$ N/m. Both the experimenters tended to classify objects with $k < 10,000$ N/m as ‘soft’ and objects with $k > 10,000$ N/m as ‘hard’. The label ratio was skewed towards hard objects, at 83%, and towards unmoved objects, at 75%. Note, this is a property of both the home and the objects (objects relevant to ADLs and IADLs in the bedroom, bathroom, and kitchen of that specific home).

D. Data-driven Haptic Perception

We used four data-driven algorithms for classifying the objects according to haptic labels. We selected these data-driven machine-learning algorithms based on their wide usage and suitability for time series. We used various combinations of the force, motion, active and passive thermal time-series vectors as the features for performance analysis. For each algorithm, we used the same set of corresponding parameters with which we found success after exhaustive grid search in their corresponding parameter space in [2], [3] and [6].

1) Support Vector Machines: We implemented binary support vector machines (SVMs) using the scikit-learn package [34] in Python. We used a linear kernel. To produce feature vectors for training, we truncated each feature to 2.0 seconds after the onset of contact (detected using the fabric-based force sensor). For the active thermal modality, we estimated the slope of the raw temperature data using first central difference. We concatenated the features to obtain a $200 \times (n + 1)$-dimensional feature vector where ‘$n$’ is the number of sensing modalities. One feature vector corresponds to the estimated slope of the active heat-transfer data.

2) k-Nearest Neighbors: We also implemented 1-nearest neighbor (k-NN with k=1) using the scikit-learn package [34] in Python. We used the same concatenated feature vector. To reduce the effect of noise and overfitting, we computed a low-dimensional representation of the training data with principal component analysis (PCA) before classification with 1-NN [1], using three principal components for dimensionality reduction. Three principal components could account for more than 95% of the variance of the data.
We characterize thermal variability by qualitative observations. We used ‘RMSprop’ as the optimizer and with the same truncated features as in Section V-D1. In this section, we characterize real-world variability that could affect tactile perception. For example, due to various sources of real-world variability, force, motion and thermal signals from the same object can differ considerably across different trials which poses a challenge for tactile perception. We found that 12/67 object parts showed variability in the mechanics of interactions across trials. Table I provides a overview of the different types of object movement following robot-object interactions. Figure 4 shows examples of varied multimodal data from the robot touching some objects.

<table>
<thead>
<tr>
<th>Object</th>
<th>Slide</th>
<th>Slide + Contact</th>
<th>Roll / Tip</th>
<th>Does Not Move</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blender Container</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>Chair Cushion (Bedroom)</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Chair Cushion (Kitchen)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>Food Box</td>
<td>9</td>
<td>0</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Fruit (on Countertop)</td>
<td>18</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bottle (in Fridge)</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Fruit (in Fridge)</td>
<td>11</td>
<td>5</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Jug (in Fridge)</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Lamp</td>
<td>15</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Pot</td>
<td>13</td>
<td>0</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Toaster</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Water Pitcher</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

3) Hidden Markov Models: We used a multivariate continuous left-right HMM with 10 hidden states and n+1 dimensional Gaussian emissions for each category using the GHMM toolkit [35] in Python. With the same truncated features as Section V-D1, we set a uniform prior to all the states. We used a spherical covariance matrix structure for initialization. We chose these specifications based on our previous results [1], [11]. We trained these HMMs with the standard Baum-Welch algorithm. For testing, we ran the Viterbi algorithm to find the HMM with the most probable state sequence given the observations, classifying the category as being the category associated with this HMM [36].

4) Long Short-term Memory Networks: Finally, we also implemented long short-term memory (LSTM) networks [37] with the same truncated features as in Section V-D1. In this paper, we used an LSTM structure where each memory cell has an input gate, a forget gate, and an output gate. We implemented a stacked-LSTM structure of 2 layers with 50 cells each. We also added a dropout layer in between the two layers, which helps in regularization. We added a dense output layer which was fully connected. The LSTM has a total of 31,004 parameters. We initialized the parameters with uniform distribution, used ‘softsign’ activation functions for the hidden layers, and ‘softmax’ activation function [38] for the fully connected output layer. Our dropout probability was 0.2. We used ‘RMSprop’ [38] as the optimizer and ‘categorical_crossentropy’ [38] as our loss function because our task is a classification task. We used ‘MinMaxScaler’ function [38] to scale multivariate features for LSTMs. These parameters and choice of functions are similar to our implementation in [3] with which we found the best results after an exhaustive search in the parameter space.

VI. VARIABILITY IN THE REAL WORLD

In this section, we characterize real-world variability that could affect tactile perception. For example, due to various sources of real-world variability, force, motion and thermal signals from the same object can differ considerably across different trials which poses a challenge for tactile perception. We characterize thermal variability by qualitative observations in the real home environment. We characterize variability in mechanics of interactions by analyzing all experimental videos for the different types of object movement following robot-object interactions. Figure 4 shows examples of varied multimodal data from the robot touching some objects.
Fig. 5: Variability in mechanics of interactions. Dashed line shows the normal to an object’s orientation. The pot (a) slid, (b) rolled/tipped due to geometrical features, (c) tipped at the edge of burner, and (d) did not move due to the backboard. Stack of books slid on (e) wooden surface, and (f) on other book surface. A lamp (g) moved backwards and (h) tipped down due to its internal mechanics. (i) A swivel chair twisted about a center point. (j,k) A pitcher did not move when water level increased. A food box (l) slid back in a kitchen cabinet and (m) rolled/tipped due to another object. A fruit in the fridge (n) slid and (o) rolled/tipped due to contact with another object. (p) A fruit on kitchen countertop slid on a towel and changed direction due to contact with a carrot behind it.

• Does Not Move : No movement

Fig. 5 (a)-(f) detail examples of notable pushing phenomena resulting from contact interactions with the underlying surface. For example, the pot sometimes tipped on the electric stove burner and sometimes slid. For the tipping cases, four times the pot snagged on some geometric feature of the burner, and two times the pot tipped off of the burner. Other examples include motion of books on a bedside table, and pushing a blender container and toaster until they reached a wall behind the counter and no longer moved.

Examples of mechanical variability in the object itself are shown in Fig. 5 (g)-(k). The tensioning springs in the balanced-arm lamp appeared to affect how the lamp moved when pushed. Similarly, the presence of ball bearings in a swivel chair caused it to twist upon pushing the cushion. Another example is the water level in a water pitcher. On different days, the pitcher had different levels of water in it. Sometimes, the pitcher could not be moved by the robot’s motion with slow velocity when the water level was high.

In Fig. 5 (l)-(p), we consider variability from contacting other objects. During some pushes of a food box, it slid back normally. In others, the food box tipped over a short object behind it. When pushing a fruit in a fridge, the fruit usually slid back normal to gravity, but sometimes changed direction due to contact with other objects or rolled / tipped due to the underlying surface. The fruit on the kitchen countertop never tipped/rolled during a push. It rested on a high friction surface (a rough towel). The bottle and jug in the fridge also showed variability from contacting objects during pushes.

VI. RESULTS

We present results of haptic classification with various combinations of force (f), motion (m), active thermal (h), and passive thermal (t) features. For generalization across velocity, we trained on one velocity and tested on another (across both the velocities) and vice-versa and reported the average percentage accuracy. For generalization across instance, we used the ‘leave-one-object-out’ cross-validation scheme. We did not implement LSTMs for binary material recognition tasks and for distinguishing objects as the amount of training data was less. Table II shows the best average results.

A. Compliance and Mobility

In this section, we present the classification results for ‘Hard vs. Soft’ (HS), ‘Moved vs. Unmoved’ (MU) and 4-category ‘Hard Unmoved vs. Hard Moved vs. Soft Unmoved vs. Soft Moved’ (4C) labels.

1) Generalizing Across Velocity: Irrespective of the type of categorization, multivariate HMMs usually showed the best results (Table II) whereas 1-NN with single features showed the worst results (61% for HS, 23% for MU). This is similar to our results with laboratory objects [1], [3], showing that the results from these algorithms can apply to situations with different mechanical interactions. Univariate HMMs with temperature feature also performed the worst (50% for HS). LSTMs again showed promise (84% for HS, 92% for MU, and 81% for 4C).

2) Generalizing Across Time of Day: Irrespective of the type of categorization, SVMs with both force and motion features gave the best results (Table II). Similar to the generalization results for velocity, the worst results were for 1-NN with single features (54% for MU, 27% for 4C). Univariate HMMs with heat-transfer feature also performed the worst (50% for HS). This is intuitive because we do not expect heat-transfer feature to be informative about the compliance or mobility of an object. LSTMs showed promise (84% for HS, 92% for MU, and 81% for 4C).

3) Generalizing Across Instance: Again, irrespective of whether its the ‘Hard vs. Soft’ or ‘Moved vs. Unmoved’ categorization, multivariate HMMs performed the best (Table II) and 1-NNs performed the worst (67% for HS, 27% for 4C). Interestingly for ‘Moved vs. Unmoved’, LSTMs did the worst
TABLE II: Best Algorithm Performance with 2.0s of Contact

<table>
<thead>
<tr>
<th>Categories</th>
<th>Generalize Across</th>
<th>Algorithm</th>
<th>Features</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard vs. Soft</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velocity</td>
<td>HMM</td>
<td>f+m+h+t</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>SVM</td>
<td>f+m</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>Instance</td>
<td>HMM</td>
<td>f+m+h+t</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>Maj. Classifier</td>
<td></td>
<td></td>
<td></td>
<td>85</td>
</tr>
<tr>
<td>Rand. Guess</td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>Moved vs. Unmoved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velocity</td>
<td>HMM</td>
<td>f+m+h+t</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>SVM</td>
<td>f+m</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>Instance</td>
<td>HMM</td>
<td>f+m+h+t</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Maj. Classifier</td>
<td></td>
<td></td>
<td></td>
<td>76</td>
</tr>
<tr>
<td>Rand. Guess</td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>Hard Unmoved vs. Hard Moved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velocity</td>
<td>HMM</td>
<td>f+m+h+t</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>SVM</td>
<td>f+m</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>Instance</td>
<td>HMM</td>
<td>f+m+h+t</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Maj. Classifier</td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>Rand. Guess</td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>Material Recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velocity</td>
<td>SVM</td>
<td>f+m+h+t</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>SVM</td>
<td>f+m+h+t</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>Maj. Classifier</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Rand. Guess</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Tactile Foreground vs. Tactile Background</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velocity</td>
<td>SVM</td>
<td>f+m+h+t</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>SVM</td>
<td>f+m+h+t</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Maj. Classifier</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Rand. Guess</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Object Parts given Object</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velocity</td>
<td>SVM</td>
<td>f+m+h+t</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>SVM</td>
<td>f+m+h+t</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Maj. Classifier</td>
<td></td>
<td></td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Rand. Guess</td>
<td></td>
<td></td>
<td>42</td>
<td></td>
</tr>
</tbody>
</table>

Note: ‘f’ = force, ‘m’ = motion, ‘h’ = active thermal (heat-transfer), and ‘t’ = passive thermal (temperature) feature. ‘Maj. Classifier’ is Majority Classifier and ‘Rand. Guess’ is Random Guess Classifier.

with just the motion feature (53%), probably because of the lack of sufficient data.

B. Material

1) Generalizing Across Velocity: SVMs with all the features of force, motion, heat-transfer, and temperature showed the best results (Table II). This showed that all these modalities with complementary information were informative for getting material specific information. The worst results were with force or temperature features as these features alone could not extract relevant information for material recognition.

2) Generalizing Across Time of Day: Again, SVM with all the features gave the best results (Table II) which shows the relevance of all the features for material recognition problems. And again, univariate HMMs (60%) and 1-NNs (56%) gave the worst performance with just force and temperature features, probably because of similar reasons outlined above.

C. Tactile Foreground vs. Tactile Background

1) Generalizing Across Velocity: SVMs with all features showed the best results (Table II) for this recognition task. Similar to the binary material recognition results in Section VII-B, these showed the importance of all the features in recognizing two objects. The worst results were for 1-NN with temperature feature alone (55%). This is because passive temperature sensing is a function of the ambient temperature and not any object characteristics. This showed that heat-transfer modality (active thermal sensing) was an informative feature for object categorization task, which agrees with our previous result in [6].

2) Generalizing Across Time of Day: Here, again SVMs with all the features gave the best results (Table II) whereas 1-NNs with temperature feature (55%) and HMMs with force and motion features (55%) performed the worst.

D. Object Parts given Object

1) Generalizing Across Velocity: SVMs with all the features showed the best results (Table II) showing the relevance of all the features to recognize an object. 1-NN with heat-transfer and temperature feature (47%) showed the worst results. This is probably because without using complementary modalities of force and thermal sensing, it was difficult to extract useful information.

2) Generalizing Across Time of Day: SVMs with all the features or using features that captured the complementarity of force and thermal sensing performed the best (Table II), whereas 1-NNs with just the force and motion features (44%) gave the worst results.

E. Overall Results

Figure 6 shows the overall results. Irrespective of algorithms used, on average, force and motion features were especially informative for classifying compliance and mobility based haptic labels. All four features (force, motion, active heat, and passive temperature) features were informative for binary material recognition, distinguishing tactile foreground vs tactile background, and distinguishing object parts given object. This shows that the complementary information from force, motion, and thermal sensing modalities are informative for inferring material properties as well as distinguishing objects. However, passive temperature modality did not seem to be very informative for these tasks.

VIII. CONCLUSIONS

We demonstrated the usefulness of the complementary capabilities of force and thermal sensing modalities for haptic perception tasks in a real home. Specifically, we were interested in inferring compliance and mobility based haptic labels, material based haptic labels as well as distinguishing objects in relevant tasks. We identified various sources of variability in haptic perception with objects in the real world due to
its surroundings and the environment. Our objective was to analyze if the haptic perception results from these objects in the real world can generalize to different speeds (which may affect force sensing) or different times of day (which may affect thermal sensing) or across different object instances. We collected data using a mobile robot with a multimodal sensor module attached at the end of a linear actuator from 67 object parts in the bathroom, bedroom, and kitchen of a house over a period of three days. We implemented widely used and state-of-the-art data-driven algorithms such as 1-NNs, SVMs, HMMs, and LSTMs for generalization tasks. Our results showed the importance of using features from multiple sensing modalities in inferring haptic properties of objects or distinguishing objects. SVMs and HMMs showed the best results for different haptic label based classification tasks whereas 1-NNs failed to generalize. LSTMs showed promise for our problem but may need large amounts of data to give meaningful results.

REFERENCES


