Using Robotic Exploratory Procedures to Learn the Meaning of Haptic Adjectives

Vivian Chu[1], Ian McMahon[1], Lorenzo Riano[2], Craig G. McDonald[1], Qin He[1], Jorge Martinez Perez-Tejada[1], Michael Arrigo[1], Naomi Fitter[1], John C. Nappo[1], Trevor Darrell[2], and Katherine J. Kuchenbecker[1]

***Abstract*— Delivering on the promise of real-world robotics will require robots that can communicate with humans through natural language by learning new words and concepts through their daily experiences. Our research strives to create a robot that can learn the meaning of haptic adjectives by directly touching objects. By equipping the PR2 humanoid robot with state-of-the-art biomimetic tactile sensors that measure temperature, pressure, and fingertip deformations, we created a platform uniquely capable of feeling the physical properties of everyday objects. The robot used five exploratory procedures to touch 51 objects that were annotated by human participants with 34 binary adjective labels. We present both static and dynamic learning methods to discover the meaning of these adjectives from the labeled objects, achieving average F1 scores of 0.57 and 0.79 on a set of eight previously unfelt items.**

I. Introduction

Robots are beginning to move out of highly structured factories and laboratories into the real world, aiding humans in applications ranging from floor cleaning and flexible manufacturing to bomb disposal and surgery [1]. As robotic teammates encounter increasingly uncertain environments, they will need to communicate with the humans around them, an interaction that will most likely occur through natural language [2], [3], [4]. Robots will need to be able to learn new words and concepts through their physical experience with the world, as human children do, by seeing, hearing, and manipulating real objects and environments. To deepen our understanding of perceptually grounded language acquisition and improve robotic interaction with the physical world, this work aims to create a robot that can learn the meaning of haptic (touch-based) adjectives by physically interacting with labeled objects through sensitive fingertips, as shown in Fig. 1.

Touch is uniquely interactive among the senses, combining the ability to feel rich stimuli across the skin with the knowledge of bodily movement. Tactile sensitivity in the glabrous (non-hairy) skin of the hand is governed by four types of mechanoreceptors: the fast-adapting Meissner and Pacinian corpuscles sense vibrations and impacts, and the slow-adapting Merkel cells and Ruffini endings sense static skin deformation and stretch [5]. Glabrous skin also contains thermoreceptors and nociceptors for sensing temperature and pain [5]. Standard robotic tactile sensors do not begin to match the richness of human tactile sensitivity, though robots do usually excel at sensing their own motion. In humans, kinesthetic feedback relays the pose of the limbs and the effort being exerted at each joint by aggregating tactile mechanoreceptors with receptors that measure muscle length and muscle velocity [5]. These combined signals allow a human to know their hand position as they gently scan a wall for the light switch in a dark room. As this example illustrates, the sense of touch is inherently active – uncovering information about the environment requires exploration. Lederman and Klatzky were the first to discover that humans use stereotypical exploratory procedures (EPs), such as lateral motion for texture and pressure for hardness, to reduce their uncertainty about new objects [6].

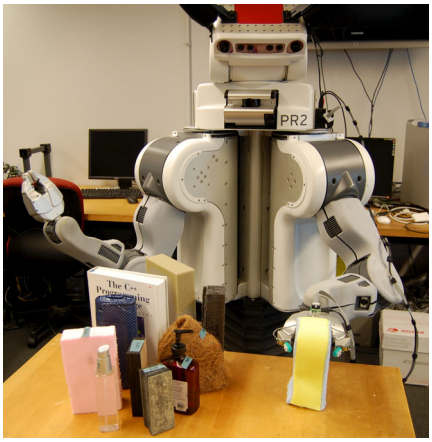


Fig. 1. A PR2 equipped with BioTac sensors explores a blue sponge. All of the objects on the table are in the Penn Haptic Adjective Corpus 1.

Other researchers have taken cues from human haptic knowledge to make robots more graceful in manipulating their environment. Romano et al. [7] imitated human grasping procedures to create a tactilely sensitive PR2 controller that holds objects with near-minimal force to avoid crushing them. Chitta et al. [8] used the PR2 to determine how much liquid remains in a beverage container by grasping the container, rolling it from side to side, and monitoring the high-frequency peaks in the PR2's tactile sensor array. Sinapov et al. [9] explored the use of different scratching motions to recognize and categorize textures; comparisons between single and multiple exploratory motions showed that joining different motions improves the results of the classification [9]. More recently, Fishel and Loeb used a SynTouch BioTac sensor and Bayesian techniques for choosing a series of movements from a repertoire of various stroking motions to classify 117 textures with 95.4% accuracy [10]. Griffith

[1]Haptics Group, GRASP Laboratory, Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. Email: {chuv, imcmahon, kuchenbe}@seas.upenn.edu

[2]Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, California 94720, USA. Email: lorenzo.riano@berkeley.edu and trevor@eecs.berkeley.edu

et al. [11] also demonstrated the benefits of having a robot perform multiple exploratory movements such as grasping, shaking, dropping, and flipping the object while analyzing the resulting audio, visual, and haptic signals.

While much previous research has focused on enabling robots to recognize particular object instances through the sense of touch, e.g., [12], [13], [14], we are interested in generalizable physical knowledge. To do so we focused on haptic adjectives – words used to describe how objects feel. Previous robotics work in this area is sparse, though human use of common haptic adjectives has been reasonably well characterized [15]. In this paper, we seek to demonstrate a set of methods that enable a tactilely sensitive robot to learn the meaning of haptic adjectives through physical interaction.

II. Robotic Platform

As shown in Fig. 1, we augmented a humanoid robot with advanced multi-channel tactile sensors to obtain a platform capable of both controlled manipulation and rich tactile sensing.

We selected Willow Garage's PR2 (Personal Robot 2) for our robotic platform. The robot's anthropomorphic arms and head make it suited for performing tasks a human might undertake, while its low inertia and backdrivabilty make it safe to operate in a human environment. This standard platform has a widely adopted software interface, the Robot Operating System (ROS), which allowed for fast software integration with pre-written libraries such as MIT's EE Impedance Arm Controller [16] and Willow Garage's Tabletop Object Detector [17]. The PR2 robot has two 7-degree-of-freedom (7-DOF) robotic arms, each with a 1-DOF two-fingered gripper, plus a suite of cameras and LIDAR sensors. Each fingertip of each gripper houses a default tactile sensor containing 22 tactile sensing elements, 15 of which face inward to make contact with the object being grasped. These fingerpads are capable of sensing only pressure that is normal to each tactile cell, sampled at 24.4 Hz. While this combination of discrete low-frequency pressure signals is sufficient for preventing the robot from crushing most objects [7], we sought robotic fingertips capable of sensing a wider array of tactile signals at higher frequencies to more closely match the touch sensing humans have at their disposal when learning haptic adjectives.

We selected SynTouch's BioTacs (Biomimetic Tactile Sensors) for this project. Each human-fingertip-sized BioTac sensor includes a lightly ridged silicone skin filled with conductive fluid over a heated rigid core patterned with electrodes [10]. The sensor measures five types of tactile signals: low-frequency fluid pressure (P_{DC}), high-frequency fluid vibrations (P_{AC}), core temperature (T_{DC}), core temperature change (T_{AC}), and a set of nineteen electrode impedances ($E_1 \dots E_{19}$) spatially distributed across the core. P_{AC} is sampled at 2200 Hz, and the others are sampled at 100 Hz.

The metal fingers of the PR2's gripper were redesigned to mechanically accommodate a pair of BioTac sensors while maintaining compatibility with the PR2's default fingerpads. The new fingers were custom machined out of aluminum, anodized, and installed in a replacement gripper by technicians at Willow Garage. In the new design, the BioTac sensor slides into a profiled hole on the end surface of the finger and is held in place with a set screw. As with the original finger design, a default tactile sensor or a non-sensorized fingerpad attaches to the inner surface of the finger using two long machine screws. Because these screws pass through the BioTac's profiled hole, only one type of fingertip can be installed at a time. Exchanging two default fingerpads for two BioTacs takes between 5 and 10 minutes.

A 12-Volt USB hub was added to the exterior of the PR2's left shoulder for powering and communicating with the BioTacs. A Cheetah SPI-to-USB host adapter unit attaches to the USB hub for communication. The USB-power and SPI data cables were routed externally down the robot arm, attached loosely in two locations to prevent the robot from straining the wires during movement. The SynTouch BioTac hand board is mounted at the base of the gripper, and each BioTac board is mounted on the outside surface of its respective finger. This modification was successfully performed on the PR2 robots at both Penn and UC Berkeley. Full details of our mechanical and electrical integration methods are available at http://bolt-haptics.seas.upenn.edu/. To interface with ROS software, a publisher node was built to read in the data from both BioTacs and publish that data at 100 Hz over the ROS network.

III. Penn Haptic Adjective Corpus

After integrating the BioTac sensors with the PR2, we used this new system to collect a large amount of physical interaction data while the robot repeatedly touched a set of objects. We also had human participants blindly interact with the same objects to provide ground-truth ratings of the haptic adjectives that apply to each one. Together, the recorded robot interaction data and the associated adjective labels constitute the Penn Haptic Adjective Corpus 1 (PHAC-1).

A. Objects

We created an object specification list to ensure that explored items fit within the sensory and motor limits of our platform while still preserving a range of interesting tactile properties. To simplify the interactions, all objects must stand upright on a table and have two flat, parallel, vertical sides with identical surface properties. So that the PR2 gripper can surround the object and then touch it with both BioTacs, each object has a thickness between 1.5 cm and 8 cm. Each object's height is greater than 10 cm to give the PR2 gripper room to vertically slide along the object's surfaces without colliding with the table. The silicone skin of the BioTacs can be punctured or damaged by sharp, pointed, or scalding hot objects, so we excluded items with any of these dangerous properties. In addition, all objects must be clean and dry to prevent damage to any of the system's electronics.

We found 51 objects that conform to the above specification, as pictured and named in Figure 2. Many are household items, while others are constructed from raw materials. We attempted to obtain a large set of objects that represent a

Fig. 2. The 51 objects included in the Penn Haptic Adjective Corpus 1, organized by their primary material.

range of physical properties without significant redundancy. In addition to the 51 PHAC-1 objects, several extra objects that conform to the specification were collected for use in testing our system.

B. Robotic Data Collection

The PR2 was programmed to execute a fixed set of movements to tease out interesting haptic signals from each object in the corpus. To expedite training, the object was placed within a fully open gripper with a purposeful variation in position to simulate the imprecision of object detection. For the full testing and demonstration system, we integrated standard ROS tabletop object detection and arm motion planning algorithms to place the gripper around the object. Once the gripper is in position, a touch-reactive controller using the BioTac sensors begins to explore the object. This custom controller was largely based on the "pr2_gripper_sensor_controller" ROS package written by Romano et al. [7].

Lederman and Klatzky demonstrated that humans employ a stereotypical set of exploratory procedures (EPs) when haptically evaluating novel objects [6]. For this work, we identified a subset of these procedures that are feasible for the PR2 to perform: "Enclosure" (for perceiving object volume), "Pressure" (hardness), "Static Contact" (temperature), and "Lateral Motion" (texture). While these four EPs might be sufficient for a robot to learn many haptic adjectives, the "Lateral Motion" EP in particular can be performed in a variety of ways. Fishel and Loeb recently showed that different textures can be recognized with greater accuracy if a BioTac performs multiple lateral motions of varying pressure and speed [10]. Their results indicated that higher force with lower speed (1.262 N, 1.0 cm/s) was good for perceiving traction, while low force with high speed (0.20 N, 6.31 cm/s) reduced uncertainty about the surface's roughness. Combining this information with knowledge about human EPs, we refined and expanded the robot explorations into five predefined robot motions that yield streams of haptic data: *Tap*, *Squeeze*, *Static Hold*, *Slow Slide*, and *Fast Slide*. The entire sequence can be seen in the video that accompanies this paper.

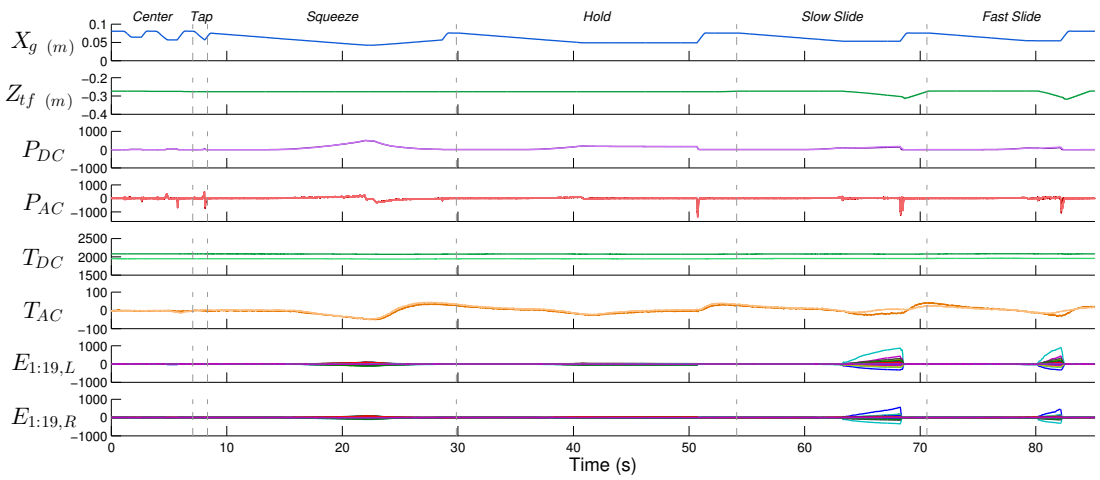
Figure 3 presents the haptic signals for one complete robotic interaction with an object, with the controller states labeled across the top. The *Center* state is not intended for data analysis, but rather to achieve approximately equal contact pressure on the two BioTacs. Without centering, the finger that first contacts the object tends to develop a high contact force, while the other finger may not touch the object at all, adding unnecessary variability between trials. Centering involves a simple bang-bang control based on the two P_{DC} signals; the gripper closes both fingers simultaneously, makes light contact with one finger first, re-centers the gripper by a small amount, and repeats this process several times until the two fingers contact the object at approximately the same instant.

During the *Tap* phase, the PR2's gripper quickly closes around the object until contact occurs on both BioTacs. When both fingers' P_{DC} readings exceed a small predefined value, the gripper opens to release the object. *Squeeze* slowly closes the gripper at constant velocity until a moderately large predefined P_{DC} value is achieved on at least one of the two fingers, then the gripper slowly opens. During *Static Hold*, the robot gently holds the object for ten seconds to let the warm fingers reach thermal equilibrium with the object, which is typically at room temperature. During this phase, the desired aperture for the robot's gripper is 50% of the distance between the position where contact is first detected and the position where the P_{DC} threshold is reached during *Squeeze*. Our initial attempts to achieve consistent light contact based only on P_{DC} were largely unsuccessful when tested on the range of objects in the PHAC-1 because the fingertips must penetrate soft objects significantly more than hard objects in order to reach the same finger pressures. For *Slow Slide* and *Fast Slide*, the PR2 contacts the object with both fingers and then moves downward for a distance of 5 cm, releasing the object in between and at the end of the motion. *Slow Slide* uses a stronger contact (20% of the total distance during *Squeeze*) and slides at 1 cm/s, compared to the lighter contact (10%) and faster 2.5 cm/s speed of *Fast Slide*.

The robot used this sequence of five EPs to touch each of the 51 objects ten times. Each of these 510 trials was recorded as a ROS bagfile that contains time histories of all PR2 transforms, left arm joint efforts, positions, and velocities, left gripper accelerometer readings, the narrow stereo left camera video feed, all readings from both BioTac sensors, and the timing of the controller states and sub-states. As shown in Figure 3, a subset of these signals was chosen for the current analysis, including gripper aperture, X_g, the vertical position of the gripper in the torso frame, Z_{tf}, and all signals from both left and right BioTacs.

C. Haptic Adjective Labels

Ground-truth labels are needed to enable the robot to associate the quantitative data it collects with subjective ratings of how the objects are perceived to feel. We started



3. The haptic data collected as the PR2 touched the blue sponge shown in Fig. 1. The BioTac signals are analog-to-digital readings shown in bits; all are initially zeroed except T_{DC}. The signals from the left and right fingers are shown together, except for the electrode arrays. The majority of the human raters labeled this object as compressible, fuzzy, nice, soft, squishy, and thick.

Fig. 4. The 34 adjective labels that the human participants used to describe the objects shown in Fig. 2. The adjectives are shown here in fonts that suggest their meaning, but participants saw them all in a standard font. The three adjectives marked by a * were selected by the majority of participants for only one object in the corpus, giving little basis for learning their meaning.

this labeling process by having pilot participants touch the objects under conditions matched to those used by the PR2 – using only two fingers, wearing noise-canceling headphones playing pink noise, and with the object occluded from view by a visual barrier. We first tested these pilot participants in free response to see what adjectives they used. We also surveyed the literature to discover what adjectives have previously been used to describe the feel of objects [15], [18]. This process allowed us to assemble the list of 34 adjectives shown in Figure 4. We treat each adjective as a binary label that can apply to any object, while each object can have any number of positive labels from this set. We purposefully avoided assuming any antonym or synonym relationships because we want to develop methods that can uncover such correlations from the data. Four volunteers then touched each PHAC-1 object using the above procedure and assigned each object a binary (yes/no) rating for each of the 34 selected adjectives. The objects were presented in random order.

For each object-adjective pair, we counted the number of participants who used that adjective to describe that object, a value that ranged from 0 to 4. These ratings were subjected to a majority voting scheme to determine the final binary labels: for our set of four participants, an agreement of three or more individuals was required to yield a positive label, whereas two or fewer individual positive labels produced an overall negative label for that object-adjective pair. As noted in Fig. 4, three adjectives (elastic, grainy, and porous) were found to apply to only one object in the corpus. Learning the

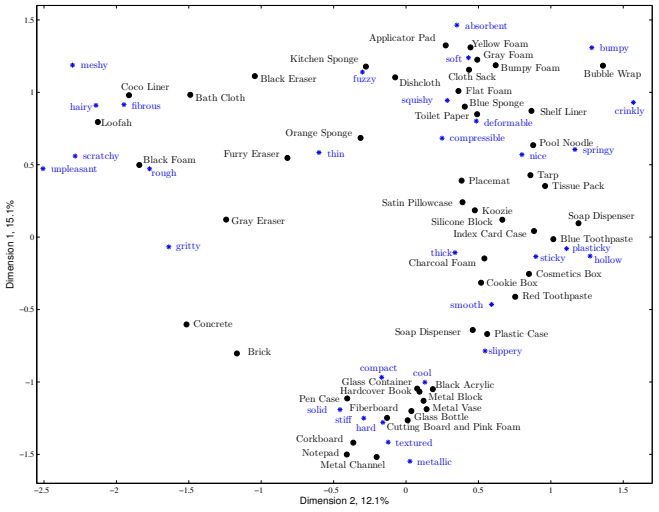


Fig. 5. The first two dimensions of the haptic object-adjective space created using correspondence analysis on the majority voting ratings.

meaning of an adjective from a single example is a difficult problem, and the lack of additional examples precludes testing any developed learning methods, so we eliminated these three adjectives from further consideration, leaving 31 in the set.

Since this dataset of adjective ratings is categorical, non-negative, and uniformly scaled, we used correspondence analysis (CA) to investigate the validity of these labels. As Picard et al. explains [18], "CA has been traditionally used to posit both stimuli and descriptors in an n-dimensional space, where the distance between the stimuli reflects their perceived dissimilarity and the location of the descriptors reflects their degree of association with the stimuli." In our study, the stimuli are the objects, and the descriptors are the adjectives. We found that 10 dimensions contribute significantly to the space described by the 51 objects and the 31 binary adjectives. Figure 5 shows where each adjective and each object sit in the first two dimensions. We observe that similar objects, such as the concrete and the brick, appear close to one another, as do similar adjectives, such as compressible and deformable. This investigation demonstrates the validity of the labels.

IV. Machine Learning Techniques

We tested two approaches for learning the associations between what the robot felt when exploring the objects and the ground-truth haptic adjective labels. Our first method follows existing robotic texture recognition systems such as those by Fishel and Loeb [10] and Sinapov et al. [9], which use carefully crafted static features in conjunction with machine learning algorithms. Our second method is inspired by the observation that many of the haptic data channels resemble audio signals, a research area that typically uses dynamic learning techniques. All of our machine learning algorithms were implemented using scikit-learn [19], an open-source package for Python.

A. Static Feature Learning

Before extracting features from a robot-object interaction trial, we zero each of the 46 channels of BioTac data $(2 \times (4 + 19))$ by subtracting the mean of the channel's first 100 measurements. We then split the run into the five EPs: *Tap*, *Squeeze*, *Static Hold*, *Slow Slide*, and *Fast Slide*. Each EP is passed through our feature extraction pipeline, with the same set of features being calculated for each one. This systematic feature extraction process was designed to enable the machine learning algorithms to determine which EP and which features are best for learning a given adjective. We selected a total of 47 features: 22 for each BioTac and 3 for the robot transforms.

1) Features: The first set of features focuses on P_{DC}, the internal fluid pressure experienced by the BioTac. We calculate the maximum and the mean of the signal over the duration of the EP as the first two features. For the third, we smooth P_{DC} using an 11-point Hanning window before computing the greatest change in the signal's slope over time. These features were designed to capture the compliance of the object as P_{DC} varies significantly with object stiffness.

The next set of features is based on the high-frequency pressure variations recorded in the BioTac's P_{AC} channel. Past research has demonstrated good robotic texture recognition using simple frequency-domain features constructed from this same signal [10] and from measurements by a high-bandwidth accelerometer [9]. We first convert zeroed P_{AC} into a non-normalized energy spectral density, $ESD(\omega)$, where ω is the vector of frequencies. To capture the shape of this curve using only a small number of features, we then compute the total energy, the area under the *ESD* curve; the spectral centroid, the weighted average of the *ESD*; the spectral variance, the statistical variance of *ESD*; the spectral skewness; and the spectral kurtosis.

The third set of features is constructed from the two BioTac temperature signals. As shown by Lin et al. [20], the thermal conductivity of a material can be measured using T_{AC} and T_{DC}. Under the same contact conditions, thermally conductive materials (e.g., metal) pull heat from the BioTacs more quickly than thermally insulating materials. The first feature we selected in this domain is the area under T_{AC} over the duration of the EP, calculated via trapezoidal integration, to capture the amount of heat transferred out of the sensor. The second selected feature is the time constant of an exponential fit of T_{DC} over time, which directly reflects how quickly the BioTac core temperature changes after coming into contact with an object at room temperature.

The fourth group of tactile features is based on the readings from the BioTac's 19 electrodes. These signals are highly coupled due to their spatial proximity within the fluid of the finger. Consequently, we used Principal Component Analysis (PCA) to find the synergies that arise between the electrodes for each of the five EPs (analyzed separately to obtain principal components specific to each motion). The first two principal components were found to capture the majority of the electrode signal variance in the training data for all five EPs. The features selected were the coefficients of fifth-order polynomial fits of the first two principal components over time, which yields twelve features.

The last set of features stems from the movement of the PR2 robot, particularly the gripper aperture distance, X_g, and the vertical position of the gripper, Z_{tf}. We take the mean and minimum of the aperture distance and the range of the gripper's vertical position. These features were designed to reveal the size of the object and the extent of the gripper's vertical movement during the trial. Vertical movement was selected because the friction of the object affects how the PR2's hand moves during the sliding EPs.

2) Learning Algorithms: The first stage of the static-feature-based pipeline divides the data into train (two thirds) and test (one third) sets for each adjective, with an equal proportion of positive and negative examples in both sets. For example, an adjective that has 9 positive examples (objects) and 42 negative examples produces a training set with 6 positive and 28 negative examples, and a testing set with 3 positive and 14 negative examples; all 10 trials recorded for each object are grouped together. The features in the training sets are normalized and used to create classifiers specific to each adjective-EP combination.

We compared initial results from a one-element feature vector against the results of classifying on the entire 47-element feature vector; the one-element feature vector generally outperformed the full feature vector. Given the strength of the individual features, we decided to train many weak classifiers from these individual features and boost them using Gradient Tree Boosting (GTB), which is advantageous for heterogeneous data sources and also tolerant of outliers [21]. The specific implementation from scikit-learn was Gradient Boosting Classification and LOOCV performed on learning rate, boosting stages, and maximum tree depth. With 31 adjectives and 5 motions, we ended the first training stage with 155 adjective-EP-specific classifiers. We then fed the test set to these classifiers and computed an average F_1 score for each classifier. These testing results enable us to select the best EP-specific classifier for each adjective, which we use for final classification.

B. Dynamic Feature Learning

As seen in Fig. 3, time plays a major role in shaping the signals from the robot's tactile and kinesthetic sensors. This

consideration led us to test an alternative machine learning approach that analyzes the temporal fluctuation of the haptic signals using Hidden Markov Models (HMMs) [22].

Using HMMs with our recorded haptic signals poses two significant problems. First, HMMs require discrete signals rather than continuously varying signals. Second, signals that have many dimensions or many samples lead to numerical instability and an overall poor performance when using HMMs. The first problem can be overcome by discretizing the data, for example via k-means. The second problem required dimensionality reduction via PCA and resampling using interpolation. Given an input signal $x \in \mathbb{R}^{n \times d}$, the processing pipeline to train an HMM is:

1) Reduce the signal dimensionality to $p \leq d$ via PCA.
2) Resample the signal to m elements using interpolation.
3) Cluster the resulting signal $y \in \mathbb{R}^{m \times p}$ into k symbols using k-means.
4) Train an HMM using a discrete signal with length m and an output alphabet of k symbols.

The main issue with this approach is that several parameters need to be tuned to avoid over-fitting. The most important ones are the new number of dimensions p, the resampling size m, the resampling approach (linear, cubic, spline or nearest neighbors), the number of clusters k in k-means (which corresponds to the number of output symbols in the HMM), and the number of hidden states in the HMM. We coupled grid search with cross-validation to find the set of parameters that yielded the best generalization. Similar to the static-feature-based approach, we used two thirds of the positive examples in the corpus for training and the remaining third for testing. The score of each HMM is calculated as the log-probability of the observed sequence found by using the forward algorithm.

The dynamic-feature-based approach did not consider the *Tap* EP because of its short duration. Furthermore, this analysis included only the main four tactile channels, omitting X_g, Z_{tf}, and T_{DC}. This choice yielded 16 signals (4 EPs $\times$ 4 tactile channels) for each trial. We then trained 16 HMMs for each adjective using the pipeline described above. Although each HMM had its own set of parameters, cross-validation highlighted some common patterns in the corpus:

- An aggressive resampling with as few as 50 elements proved to be sufficient for a high HMM score. However, given the probabilistic nature of HMMs, usually the shorter the sequence the higher the probability of observing it. We believe that this effect, while mitigated by cross-validation and the need for generalization, may have biased the effective resampling size. Further analysis will be required to investigate how much the aggressive resampling affected the results.
- The same applied to the number of dimensions, where 7 to 8 dimensions were found to be sufficient to explain 97% of the variance of an otherwise 38-dimensional space (in the case of the electrodes for two fingers).
- The cardinality of the HMM alphabet varied between 12 and 20 symbols. This number suggests that the training data is probably ambiguous, and fine details are required to distinguish between otherwise very similar signals.
- The number of hidden states is of the same order as the alphabet. Though expected, this result negatively affected the required training time for the HMMs.

Once the HMMs were trained, for each adjective we have a set of 16 log-probabilities that describe the confidence each model has for having seen a particular pattern. The classifier that decides if an adjective applies to a certain trial is a linear SVM trained using the 16 outputs of the HMMs. The penalty factor C was chosen using cross-validation on a different split of the training data. Given the nature of the training set, the number of negative examples is greater than the number of positive ones.

V. Results

We tested both of our classification approaches on data reserved from the corpus and also on the 8 previously unfelt objects shown in Figure 6, labeled by participants in the same way as those in the corpus.

The following sections report the performance of our machine learning techniques in terms of precision, recall, and F_1 score, as defined in Table I. Here, tp is the true positive count (the number of correct positive results returned by the classifier), fp is false positives (the number of results incorrectly labeled as positive by the classifier), and fn is false negatives (the number of positive results missed by the classifier). To satisfy space constraints, we report only brief results for our second method.

A. Static Feature Learning

For the static feature learning approach (Sec. IV-A), we first analyzed the feature vectors to confirm that they capture meaningful differences in the feel of the objects. All 47 features were calculated for every *Squeeze* trial in the corpus, and each object's 10 trials were averaged to yield 51 object-specific average feature vectors. Performing PCA on these average feature vectors revealed the top two principal components and the distribution of objects shown

Fig. 6. The final test set. In order from left to right: Cardboard Box, Dark Foam, Light Foam, Hardcover Book, Cushioned Envelope, Layered Cork, and Rough/Smooth Acrylic (different sides of the same object).

TABLE I

Metric Equations

Precision	Recall	F_1
$\frac{\text{tp}}{\text{tp}+\text{fp}}$	$\frac{\text{tp}}{\text{tp}+\text{fn}}$	$2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

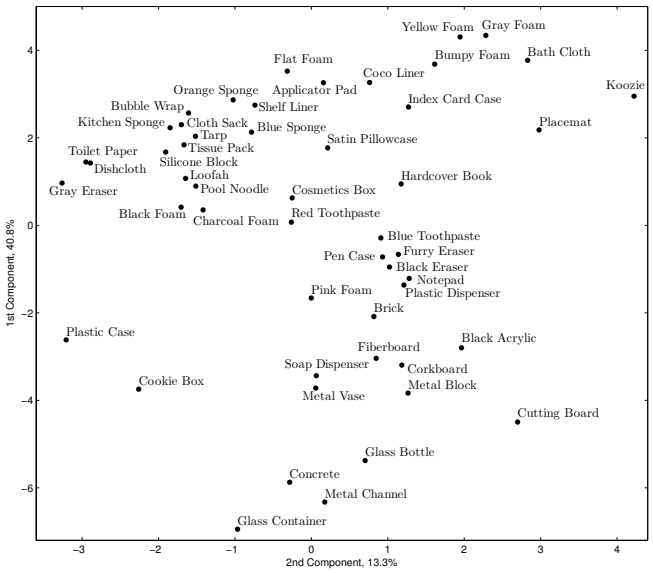


Fig. 7. The first two principal components of the feature space created from the average *Squeeze* feature vector of each object. Haptically similar objects generally appear close to one another, much like the adjective-based space (Fig. 5).

in Figure 7. By comparing this plot with the adjective space seen in Figure 5, we observe that both of the first principal components seem to indicate compliance, a highly salient object property, with a greater proportion of the feature variance attributable to this dimension. The second components do not seem to correlate cleanly, though objects close together in feature space do generally seem to feel similar, substantiating the value of these features.

Having intermediate classifiers trained on the features from just one exploratory procedure allows us to see how each EP performs for each of the 31 adjectives. The F_1 scores for each adjective-EP combination are shown in Table II; these results were generated by testing on the third of the corpus not seen during the intermediate classifier training.

The selected best EP-specific classifiers achieved an average precision, recall, and F_1 score of 0.91, 0.88, 0.88 respectively, across all adjectives on the training set. However, selecting the maximum EP based on the test set, rather than merging the results of all of the EPs (as the HMM approach does), introduces biases that can be seen in the final results. The final results of the classifiers were calculated by object rather than by adjective. For example, the human participants had labeled the dark foam as squishy, thick, compressible, fuzzy, absorbent, and soft, whereas the classifiers predicted it to be squishy, compressible, and soft. The static feature learning's average F_1 score on the previously unfelt objects shown in Figure 6 was 0.57. This relatively low score supports our belief that multiple motions should be combined to increase the recognition of haptic object properties.

B. Dynamic Feature Learning

The adjective-specific SVMs trained on the HMM outputs (Sec. IV-B) achieved an impressive average precision score of 0.98 across all adjectives on the reserved corpus data, due to a low rate of false positives. However, the classifiers had a somewhat higher number of false negatives, resulting in an average F_1 score of 0.81 on the reserved training corpus. To investigate the generalization capabilities of this approach, we also tested these classifiers on the previously unfelt objects shown in Figure 6. The average F_1 score over this new set was 0.79, which supports the hypothesis that our methods can produce a meaningful set of adjectives for completely new objects when using all EPs.

TABLE II

F_1 SCORE ACROSS ADJECTIVES AND EXPLORATORY PROCEDURES

	Tap	*Squeeze*	*Hold*	*Slide*	*Fast Slide*	PE*
absorbent	0.108	0.370	0.000	0.000	0.333	3
bumpy	0.526	0.143	0.000	0.182	0.333	2
compact	0.644	0.795	0.675	0.487	0.574	11
compressible	0.821	0.847	0.806	0.833	0.829	26
cool	0.348	0.324	0.431	0.419	0.417	6
crinkly	0.000	0.143	0.250	0.000	0.000	1
deformable	0.000	0.000	0.190	0.000	0.000	3
fibrous	0.000	0.000	0.000	0.000	0.000	3
fuzzy	0.388	0.097	0.267	0.448	0.327	8
gritty	0.000	0.000	0.000	0.000	0.000	3
hairy	0.000	0.000	0.000	0.000	0.000	2
hard	0.814	0.835	0.913	0.850	0.764	13
hollow	0.000	0.154	0.000	0.667	0.143	4
meshy	0.714	0.000	0.133	0.000	0.000	1
metallic	0.000	0.000	0.000	0.471	0.125	2
nice	0.333	0.000	0.000	0.500	0.488	5
plasticky	0.000	0.286	0.000	0.000	0.000	4
rough	0.386	0.051	0.391	0.491	0.667	5
scratchy	0.526	0.385	0.000	0.000	0.390	4
slippery	0.000	0.000	0.000	0.061	0.378	7
smooth	0.494	0.667	0.510	0.582	0.593	18
soft	0.333	0.769	0.277	0.108	0.150	9
solid	0.787	0.779	0.808	0.787	0.788	12
springy	0.000	0.000	0.000	0.000	0.000	3
squishy	0.686	0.815	0.860	0.772	0.693	18
sticky	0.233	0.143	0.386	0.225	0.170	12
stiff	0.807	0.868	0.855	0.885	0.916	13
textured	0.000	0.000	0.000	0.000	0.000	2
thick	0.500	0.000	0.000	0.606	0.000	4
thin	0.800	0.500	0.857	0.579	0.711	5
unpleasant	0.000	0.000	0.000	0.000	0.000	2

*PE indicates the number of positive examples in the training set.

VI. DISCUSSION AND CONCLUSION

The presented results prove that a robot equipped with rich multi-channel tactile sensors can discover the haptic properties of objects through physical interaction and then generalize this understanding across previously unfelt objects. Furthermore, we have shown that these object properties can be related to subjective human labels in the form of haptic adjectives, a task that has rarely been explored in the literature, though it stands to benefit a wide range of future applications in robotics.

Both the static and dynamic learning methods showed promise in the reported experiments, including the strict test involving everyday objects the robot had never before encountered. The static-feature approach yielded an average F_1 score of 0.57 for the 8 objects in this test set, and the dynamic-feature approach achieved an average F_1 score of 0.79. While these raw performance scores are lower than those achieved by existing research in areas of texture and object classification, generalizing adjectives to new objects is a harder problem than instance recognition. These results establish a baseline against which future efforts in this area can be compared.

As shown in Table II, the individual scores from the intermediate classifiers in the static feature learning approach differ by adjective and by exploratory procedure. Adjectives with a large number of positive examples achieve good results across all five EPs. However, when there are only a handful of positive examples, the results differ more between the EPs. For example, rough's F_1 score ranges from 0.051 for *Squeeze* up to 0.667 for *Fast Slide*, showing that certain object properties are most easily felt through certain actions, and confirming the fundamental motivation for using different exploratory procedures. Adjectives that have only one or two positive examples in the training set tend to achieve low F_1 scores regardless of the EP, implying that our current static-feature-based approach cannot learn the meanings of these adjectives with a single best EP from so few examples. This limitation will become a key focus of our future research, as we shift toward online learning and no longer have the luxury of many positive examples. Also, the low overall F_1 score on the test set by a single best EP in the static-feature-based approach suggests that multiple EPs are necessary to determine properties of objects. The more novel dynamic-feature-based learning approach that combined all EPs has already demonstrated strong results for adjectives with only a small number of examples, so we will look for ways to combine the two approaches as well as to use information from multiple EPs. Future work will also involve Bayesian techniques for selecting among several available EPs, as done by Fishel and Loeb for texture recognition [10].

Looking again at the data in Table II, we notice two other interesting trends. First, even though an increase of training examples generally leads to an increase in performance, some adjectives do not follow this rule. For example, the system seems to have struggled to learn the meanings of nice and soft, though they have 5 and 9 positive examples in the training set, respectively. We suspect that ambiguities and multiple physical interpretations of the meaning of these adjectives may make them harder to learn from a handful of examples compared to more straightforward adjectives, as found in the tactile adjective study by Picard et al. [18]. Second, the adjectives related to textural properties (fibrous, gritty, hairy, rough, scratchy, slippery, smooth, and textured) seem to have lower overall performance than adjectives pertaining to other object properties. This difference may stem from how the robot executed the EPs, from shortcomings in our feature set, or from a lack of textural sensitivity in our BioTac sensors, possibly due to a gradual wearing away of its synthetic fingerprints. Our future work will investigate these trends along with questions pertaining to the connection between visual and haptic perception for adjective learning.

ACKNOWLEDGMENTS

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) in the United States as part of Activity E within the Broad Operational Language Translation (BOLT) program.

REFERENCES

[1] "A roadmap for US robotics: From Internet to robotics," http://www.us-robotics.us, May 2009.

[2] D. J. Brooks, C. Lignos, C. Finucane, M. S. Medvedev, I. Perera, V. Raman, H. Kress-Gazit, M. Marcus, and H. A. Yanco, "Make it so: Continuous, flexible natural language interaction with an autonomous robot," *AAAI*, pp. 2–8, 2012.

[3] J. F. Gorostiza and M. A. Salichs, "End-user programming of a social robot by dialog," *Robotics and Autonomous Systems*, vol. 59, no. 12, pp. 1102–1114, Dec. 2011.

[4] X. Chen, J. Ji, J. Jiang, G. Jin, F. Wang, and J. Xie, "Developing high-level cognitive functions for service robots," in *Proc. International Conference on Autonomous Agents and Multiagent Systems*, Richland, SC, 2010, pp. 989–996.

[5] L. A. Jones and S. J. Lederman, *Human Hand Function*. Oxford University Press, 2006.

[6] S. J. Lederman and R. L. Klatzky, "Extracting object properties through haptic exploration," *Acta Psychologica*, vol. 84, pp. 29–40, 1993.

[7] J. M. Romano, K. Hsiao, G. Niemeyer, S. Chitta, and K. J. Kuchenbecker, "Human-inspired robotic grasp control with tactile sensing," *IEEE Transactions on Robotics*, vol. 27, no. 6, pp. 1067–1079, December 2011.

[8] S. Chitta, J. Sturm, M. Piccoli, and W. Burgard, "Tactile sensing for mobile manipulation," *IEEE Transactions on Robotics*, pp. 558–568, 2011.

[9] J. Sinapov, V. Sukhoy, R. Sahai, and A. Stoytchev, "Vibrotactile recognition and categorization of surfaces by a humanoid robot," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 488–497, 2011.

[10] J. A. Fishel and G. E. Loeb, "Bayesian exploration for intelligent identification of textures," *Frontiers in Neurorobotics*, vol. 6, no. 4, pp. 1–20, Jun. 2012.

[11] S. Griffith, J. Sinapov, V. Sukhoy, and A. Stoytchev, "A behavior-grounded approach to forming object categories: Separating containers from noncontainers," *IEEE Transactions on Autonomous Mental Development*, pp. 54–69, 2012.

[12] S. Caselli, C. Magnanini, and F. Zanichelli, "Haptic object recognition with a dextrous hand based on volumetric shape representations," in *Proc. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, October 1994, pp. 280–287.

[13] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard, "Object identification with tactile sensors using bag-of-features," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2009, pp. 243 –248.

[14] N. Gorges, S. E. Navarro, D. Goger, and H. Wörn, "Haptic object recognition using passive joints and haptic key features," in *Proc. IEEE Int. Conference on Robotics and Automation*, 2010, pp. 2349–2355.

[15] M. Hollins, S. Bensmaïa, K. Karlof, and F. Young, "Individual differences in perceptual space for tactile textures: Evidence from multidimensional scaling," *Perception & Psychophysics*, vol. 62, no. 8, pp. 1534–44, Nov. 2000.

[16] J. Barry, M. Bollini, and H. Liu, http://ros.org/wiki/ee_cart_imped, 2012.

[17] M. Muja and M. Ciocarlie, http://ros.org/wiki/tabletop_object_detector, 2011.

[18] D. Picard, C. Dacremont, D. Valentin, and A. Giboreau, "Perceptual dimensions of tactile textures," *Acta Psychologica*, vol. 114, no. 2, pp. 165–184, Oct. 2003.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[20] C. H. Lin, T. W. Erickson, J. A. Fishel, N. Wettels, and G. E. Loeb, "Signal processing and fabrication of a biomimetic tactile sensor array with thermal, force and microvibration modalities," in *Proc. IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec. 2009, pp. 129–134.

[21] J. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, 2002.

[22] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.