

# Tactile Perception for Teleoperated Robotic Exploration within Granular Media

SHENGXIN JIA and VERONICA J. SANTOS, Mechanical and Aerospace Engineering Department, University of California, Los Angeles

The sense of touch is essential for locating buried objects when vision-based approaches are limited. We present an approach for tactile perception when sensorized robot fingertips are used to directly interact with granular media particles in teleoperated systems. We evaluate the effects of linear and nonlinear classifier model architectures and three tactile sensor modalities (vibration, internal fluid pressure, fingerpad deformation) on the accuracy of estimates of fingertip contact state. We propose an architecture called the Sparse-Fusion Recurrent Neural Network (SF-RNN) in which sparse features are autonomously extracted prior to fusing multimodal tactile data in a fully connected RNN input layer. The multimodal SF-RNN model achieved 98.7% test accuracy and was robust to modest variations in granular media type and particle size, fingertip orientation, fingertip speed, and object location. Fingerpad deformation was the most informative modality for haptic exploration within granular media while vibration and internal fluid pressure provided additional information with appropriate signal processing. We introduce a real-time visualization of tactile percepts for remote exploration by constructing a belief map that combines probabilistic contact state estimates and fingertip location. The belief map visualizes the probability of an object being buried in the search region and could be used for planning.

CCS Concepts: • **Hardware** → **Sensor applications and deployments**; • **Computer systems organization** → **Robotic autonomy**; • **Human-centered computing** → **Displays and imagers**;

Additional Key Words and Phrases: Belief map, granular media, sensor fusion, recurrent neural networks, tactile perception, tactile sensors

## ACM Reference format:

Shengxin Jia and Veronica J. Santos. 2021. Tactile Perception for Teleoperated Robotic Exploration within Granular Media. *ACM Trans. Hum.-Robot Interact.* 10, 4, Article 34 (July 2021), 27 pages.

<https://doi.org/10.1145/3459996>

## 1 INTRODUCTION

The task of reaching into a granular material and searching for an object is an essential human skill for applications that include harvesting root vegetables, excavating archaeological artifacts, rescuing victims of landslides and avalanches, and humanitarian demining. Unique to such tasks

This work was supported in part by Office of Naval Research Award #N00014-16-1-2468 and Advanced Robotics for Manufacturing Institute Award #ARM-TEC-18-01-F-14.

Author's address: S. Jia and V. J. Santos, University of California, Los Angeles, Mechanical and Aerospace Engineering Department, 420 Westwood Plaza, Engineering IV, Room 37-140, Los Angeles, CA 90095-1597; emails: {jia0526, vjsantos}@ucla.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2021 Association for Computing Machinery.

2573-9522/2021/07-ART34 \$15.00

<https://doi.org/10.1145/3459996>

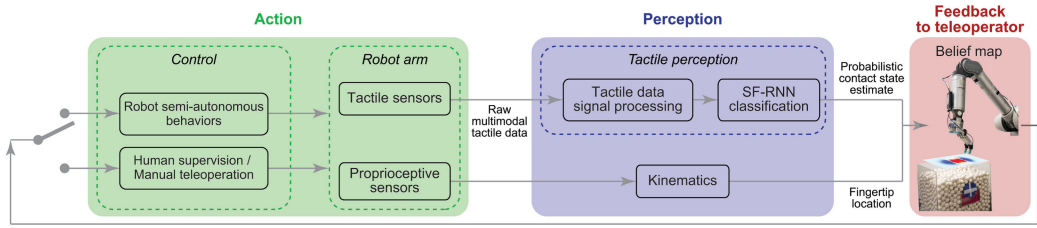


Fig. 1. A block diagram is shown for a notional real-time execution of the action, perception, and visualization processes.

is the importance of the sense of touch. This work is motivated by the need to keep humans out of harm's way by enabling robots and their teleoperators to haptically locate, excavate, retrieve, and manipulate dangerous devices that are buried within granular media, such as sand, snow, and rubble. Contact with unknown, potentially delicate and/or dangerous objects should be done with care. However, state-of-the-art robotic hands cannot match the haptic perception abilities of human hands. The challenge of remotely controlling a remote end-effector within granular media is compounded by issues that include a lack of direct, natural sensory feedback and compromised bidirectional communication between the teleoperator and robot. The cognitive burden on the teleoperators of robot manipulators could be reduced if tactile sensors were used in concert with visual feedback during interactions with granular media and buried objects. Traditionally, experiments with robot hands are performed in open air environments, where contact and no contact conditions are easily distinguished using tactile sensors. To our knowledge, the direct use of robot hands within granular media has not yet been explored. In this work, we address tactile perception challenges that arise when sensorized robot fingertips are used in granular media environments. The objective of this work is to develop the fundamental capability of tactile perception within granular media for semi-autonomous robotic systems that are supervised by human teleoperators. Specifically, the goal is to develop the tactile ability to distinguish between contact with an object embedded in granular media and ubiquitous contact with surrounding granular media.

While granular media has been manipulated by hand in virtual reality [7, 66], this work represents the first use of tactile sensors for actual robot hand movements in direct contact with granular media. The rapid development of tactile devices and skins over the last couple of years has created new opportunities for tactile information being used for a myriad of robotic applications [23, 29–31, 62] enabled by the development of novel tactile sensors [18, 22, 47]. However, unlike the rapid progress in the field of computer vision with respect to its technology and algorithm development, tactile feedback has played a smaller role in robot planning and feedback control.

While open-source tactile datasets are becoming available [70], there are still no datasets that are commonly used for evaluating tactile sensing algorithms in the way that ImageNet [17] and RoboNet [15] are used for evaluating computer vision algorithms. In contrast to image datasets, tactile datasets are expensive to collect and it is difficult to obtain ground truth for tactile perception tasks. Once tactile data are collected, a suitable computational method for deciphering the encoded percepts becomes especially important.

In this work, we present a multimodal tactile perception model that can be used for remote robotic tactile exploration within granular media. Extending beyond a prior workshop paper [38], the model performs tactile perception by combining signal processing, sparse coding, and a recurrent neural network to integrate multiple tactile sensing modalities (vibration, pressure, fingerpad deformation). We refer to our model as a **Sparse-Fusion Recurrent Neural Network (SF-RNN)** (Figure 1), since sparse features are extracted in a pre-processing step that precedes a fusion layer

of an RNN. The tactile percepts are visualized as a spatial probabilistic distribution using a “belief map” [68]. Figure 1 shows a high-level diagram of a notional computational flow during real-time execution. This work focuses primarily on the “Tactile perception” and “Feedback to teleoperator” portions of the block diagram shown in blue and red, respectively.

This article is organized as follows. Section 2 introduces related work. Section 3 describes the multimodal tactile perception algorithm and Section 4 describes the belief map representation of tactile information. Sections 5 and 6 detail the experimental procedure for robot data collection and model evaluation, respectively. Section 7 discusses notable trends in the tactile data and real-time tactile perception and visualization for human control of remote robotic systems. Section 8 concludes the work with a summary of contributions, limitations, and future directions.

## 2 RELATED WORK

### 2.1 Complex Behaviors of Granular Media

Granular media, such as sand, grain, and powder are ubiquitous in nature and in industrial processes. Yet, the modeling of the complex behaviors of granular media remains an active area of study. It is inappropriate to classify granular material properties as strictly solid- or liquid-like [36], and there is no universal continuum model to predict the behavior of granular media in an arbitrary geometry under arbitrary loading. Affected by its packing condition, the rigidity and flow characteristics of granular media [8] exhibit complex behaviors, such as jamming [69], dilatancy [57], flow bifurcation in the vicinity of a “intruder” [26], and force fluctuations around the intruder. Models have been developed to estimate soil properties such as internal friction, interface friction, and soil density (e.g., References [33, 75]) from large interaction forces. Grain-scale numerical models such as the Discrete Element Method [54] and Resistive Force Theory [50] have been applied to simulate behaviors of granular particles.

An important feature of granular media is the “force chain,” which is a hallmark feature of physical interactions with granular media. Under external stress, the force chain carries most of the load via complex inter-particle forces [49]. When external forces are too small to cause particle rearrangements, granular media becomes an amorphous solid and the distribution of interparticle normal forces develops a peak near the jamming transition [56]. When an intruder moves through granular media, as when a robot finger is raked through sand, the granular media packing fraction will change and particle flow and force fluctuations will occur around the intruder. When the intruder is a sensorized finger, the finger-particle interactions directly affect the tactile measurements. The countless instances of making and breaking contact between the fingertips and surrounding granular media particles effectively add noise to tactile sensing modalities that might otherwise be useful for object detection, the perception of geometric features (e.g., Reference [65]), and the detection of key haptic events.

### 2.2 Robot Interaction with Granular Media

Exploiting the sense of touch is essential for robots operating according to physical interaction cues and is often associated with computational and representation challenges [11]. As with biological systems, artificial tactile perception is an active process for sensing properties of one’s environment, especially in the absence of vision [24]. The task of locating a buried rigid object has been previously studied using a leader–follower teleoperated robot system in which a custom haptic device displayed forces and torques directly to the operator’s hand [76]. In that work, a human teleoperator manually controlled a robot that was sensorized with a force/torque sensor that remained outside of the granular media. Here, we aim to develop tactile perception capabilities for a semi-autonomous robot whose contact states are visually displayed for a human telesupervisor.

We use multimodal tactile sensors that make direct contact with granular media and an object buried within the granular media.

Prior research has been conducted on robot interactions with granular media, primarily with a focus on improving robot locomotion. For example, studies have been performed on legged [67] and flipper-based locomotion [53], sidewinding [52], and sand-swimming [51]. Numerical simulation approaches have enabled high locomotion performance [4] by predicting forces and flow fields of the granular media, which are difficult to measure during experiments without sensorized feet. In addition, particle image velocimetry analyses via high speed video have enabled researchers to measure grain kinematics and predict quasistatic resistive forces to enable robots to jump [3] and walk [82] on flat dry granular surfaces. In general, however, legged robots do not incorporate local sensing on feet, because sensors lack the robustness to repeated impact and wear.

In addition to locomotion, researchers have studied robot manipulation of granular media. For household applications, the granular media are poured from containers, such as a box [10] or powdered coffee creamer bottle [84], or manipulated with a tool, such as a spoon [72]. Unlike works in simulated environments [83, 85], it is difficult to observe the dynamics of granular media in real applications. As a result, researchers have relied heavily upon computer vision techniques. To detect the flow of granular media during pouring, Yamaguchi and Atkeson employed a stereo camera to reconstruct the three-dimensional (3D) point cloud [84] of the granular media. Schenck et al. evaluated predictive models with depth images to infer the dynamics of scooping and pouring actions [72]. In these prior works, vision was the primary sensory modality and researchers did not touch the granular media directly using sensorized robot fingers.

### 2.3 Feedback to Human Teleoperator

To assist a remote human operator to haptically locate a buried object within granular media, a robotic system needs to provide accurate tactile information promptly, convey the properties of the environment, and communicate recommended future actions. Until immersive telepresence set-ups become the norm, current common methods for feedback include audio, visual, and tactile displays. Tactile feedback, which can be subdivided into kinesthetic and cutaneous feedback, relies on the design of tactile display hardware and the ability or willingness of the human teleoperator to don the hardware. Kinesthetic feedback can be achieved by coupling a human teleoperator to devices, such as exoskeletal-based robotic arms [28], that apply forces and torques to the limbs. Kinesthetic feedback devices tend to be heavy and expensive and they additionally require custom haptic devices unique to a particular task. Cutaneous feedback can be achieved by vibrating [58, 80], stretching [5], or applying pressure/force [9, 76] or changes in temperature [39] to a teleoperator's skin. The feedback device can be placed on a body part that mirrors the sensor location on a robot surrogate (as on fingertips) or on a body part that is available (as on the upper arm of a transradial amputee).

In many cases, some feedback methods are impractical for field-specific applications. For instance, audio feedback is not an option in loud environments. In another example, a teleoperator may be wearing gear that precludes the donning of bulky kinesthetic or cutaneous feedback devices. In this work, we have elected to convey the learned contact state information from our tactile perception model in a visual manner, as a proof-of-concept. The visualization can include spatial, temporal and probabilistic information, for example. For our example application, the belief map represents a top-down view of the exploration area and uses a color map to reflect tactile sensor-driven probabilities of an object being buried in a particular region (Figure 1). Such a visualization could serve as a source of actionable information, enabling teleoperators to safely gain situational awareness and enhancing the quality and efficiency of human-robot interaction.

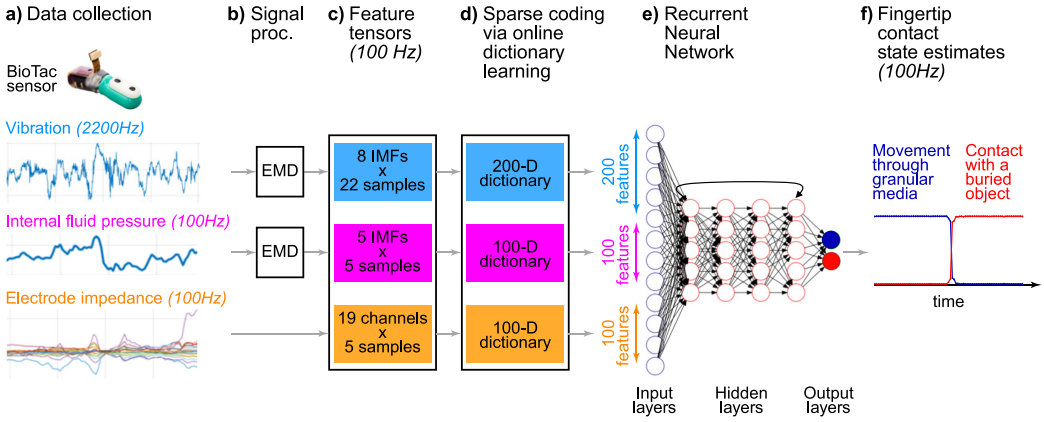


Fig. 2. SF-RNN model architecture for multimodal tactile perception in granular media. (a) Vibration, internal fluid pressure, and 19 electrodes' impedance data were collected from BioTac sensors. (b) Vibration and internal fluid pressure data were filtered by EMD. (c) Feature tensors were composed for each modality. (d) Sparse coding was used as an unsupervised pre-training step. (e) An RNN mapped a 400D input space to a 2D output space. (f) The RNN sequentially labeled tactile inputs according to the estimated fingertip contact state.

### 3 MULTIMODAL MODEL FOR TACTILE PERCEPTION

The goal of the tactile perception algorithm is to enable a robot to distinguish a buried object from its surrounding granular media through tactile exploration. The two fingertip contact states of interest are “contact with a buried object” or “free movement through granular media.” To address the tactile perception problem more generally, we propose a model architecture that we refer to as a SF-RNN (Figure 2), since sparse features are extracted via an unsupervised learning pre-training step prior to fusion through a fully connected input layer of an RNN. The model incorporates multiple sensing modalities from the BioTac tactile sensor (SynTouch LLC, Los Angeles, CA): vibration, internal fluid pressure, and electrode impedance as a proxy for fingerpad deformation (Figure 2(a)).

Raw tactile sensor signals were first pre-processed to filter out noise. Features of the vibration and internal fluid pressure data were distilled using **Empirical Mode Decomposition (EMD)** (Figure 2(b)) prior to the creation of a data structure that combined the three tactile sensing modalities (Figure 2(c)). Per well-established practices of machine learning libraries such as Tensorflow [1] and pytorch [59], we referred to the combined features from different heterogeneous sensor data as feature tensors in our SF-RNN model architecture. Sparse coding via online dictionary learning was used as an unsupervised pre-training step (Figure 2(d)), which yielded a learned feature vector for use as an input to an RNN classifier (Figure 2(e)). The use of sparse coding enabled the efficient storage of the 400D feature vector prior to the RNN mapping to a 2D output space that reflected the perceived fingertip contact state (Figure 2(f)). The accuracy of the fingertip contact state estimate is important, as a probabilistic belief map will be created to visualize the data for a human teleoperator (Section 4) under the assumption that the contact state estimate is correct.

#### 3.1 Pre-processing of Tactile Data

**3.1.1 Tactile Sensor Signal Processing.** EMD [34] (Figure 2(b)) is a powerful technique for analyzing nonlinear, non-stationary, natural signals, because EMD decomposes a time-series signal into a series of **intrinsic mode functions (IMFs)** that preserve characteristics of the original signal, such as variations in frequency (Appendix A). Physical meaning can often be associated with

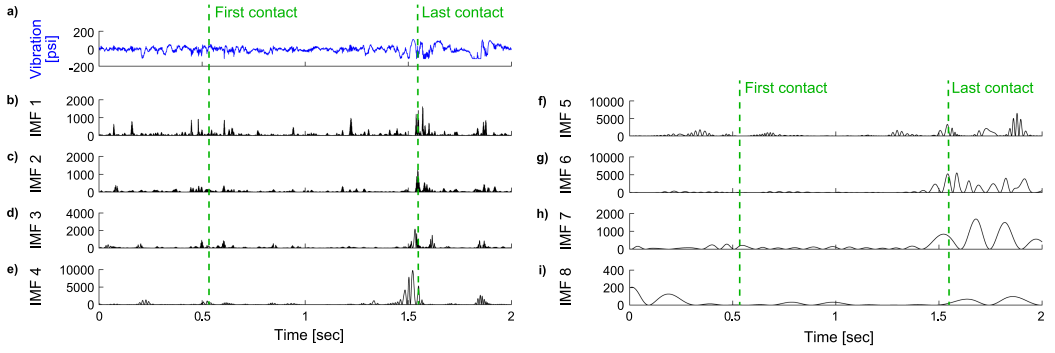


Fig. 3. (a) Raw vibration data and ((b)–(i)) squared deviations of its first eight IMFs. Each deviation was calculated as  $\text{norm}[IMF - \text{mean}(IMF)]$ .

EMD-filtered components, as in the analysis of electroencephalogram [63], water wave [34], and earthquake time-series data [35].

The original tactile signal  $x(t)$  is thus decomposed in the time domain with a small-enough residue  $r_n$ ,

$$x(t) = \sum_{j=1}^n IMF_j + r_n, \quad (1)$$

As shown in Figure 2(b), EMD was applied to the BioTac’s vibration and internal fluid pressure data. Per Reference [34], we set the residue threshold at 10% of the norm of  $IMF_{n-1}$  for each the vibration data and internal fluid pressure data. We selected the first  $n = 8$  IMFs for the vibration data and the first  $n = 5$  IMFs for the internal fluid pressure data, leaving only monotonic final residues. Figure 3 shows the original vibration data for 2 s of a representative trial along with deviations of its first 8 IMFs from the mean IMF.

**3.1.2 Reconciliation of Heterogeneous Datastreams.** Our model incorporated features from vibration, internal fluid pressure and electrode impedance data from a multimodal tactile sensor. These tactile sensor data were heterogeneous due to the different transduction mechanisms, dimensionality of sensing channels, and sampling frequencies unique to each modality. One obvious way to fuse heterogeneous sensor data is by concatenating the datastreams. However, we found that models that utilized this simple concatenation approach performed poorly, as will be described in Section 6.1 and shown in columns 1 and 4 of Tables 2 and 3.

Ultimately, we used a sensor fusion architecture to leverage well-established computer vision techniques while reconciling the differences in the heterogeneous datastreams. We selected the update frequency of the feature tensors outputted from step “(c)” in Figure 2 to be 100 Hz. For the vibration data sampled at 2,200 Hz and then filtered using EMD, the most recent 22 samples were extracted for each of 8 IMFs. For the internal fluid pressure data sampled at 100 Hz and then filtered using EMD, the most recent 5 samples were extracted for each of 5 IMFs. Finally, for the electrode impedance data sampled at 100 Hz, the most recent 5 samples were extracted. In this way, all three modalities would be represented by a spatiotemporal series of data. With an update frequency of 100 Hz, the following feature tensors were updated every 10 ms: 8 IMFs  $\times$  22 samples for vibration data, 5 IMFs  $\times$  5 samples for internal fluid pressure data, and 19 channels  $\times$  5 samples for electrode impedance data (Figure 2(c)).

**3.1.3 Sparse Representation and Online Dictionary Learning.** Upon creating feature tensors from three tactile sensing modalities, we used sparse coding (Figure 2(d)) to represent the multimodal



tactile data as linear combinations of basis functions that capture higher-level features (Appendix B). These features were identified autonomously and adaptively using an unsupervised online optimization algorithm for dictionary learning [48]. This unsupervised, pre-training step was introduced specifically to improve the generalization of the multimodal tactile perception model [20].

Using sparse coding, we compactly represented the overcomplete model of the tactile sensor data. By solving a matrix factorization problem with a sparsity penalty (Equation (4) in Appendix B), the algorithm learns a dictionary  $D = [d_1, \dots, d_k] \in R^{m \times k}$  and a sparse matrix of coefficients  $\alpha = [\alpha_1, \dots, \alpha_t] \in R^{k \times t}$  that can be used to approximate a signal  $X = [x_1, \dots, x_t] \in R^{m \times t}$  in a linearized manner as  $\hat{X}$ ,

$$\hat{X} = D\alpha. \quad (2)$$

Each dictionary size  $k$ , also referred to as feature vector length, was chosen heuristically as  $k = 200, 100$ , and  $100$  for vibration, internal fluid pressure, and electrode impedance modalities, respectively. According to Reference [48], the sparsity penalty was set with a scaling parameter related to the dimension of the original signal. The feature tensors from step “(c)” in Figure 2 were flattened and passed through the sparse coding algorithm in parallel.

### 3.2 RNN with GRUs for Binary Classification

The final stage of our multimodal tactile perception pipeline was comprised of a binary RNN classifier. Through the use of an RNN architecture, we aim to develop a model capable of learning rich representations of tactile signals within granular media and capturing temporal dependencies. The inputs to the RNN were the sparse, overcomplete 200D, 100D, and 100D feature vectors for the vibration, internal fluid pressure, and electrode impedance data, respectively. These feature vectors were fused within the RNN’s fully connected input layer. The output of the RNN was one of two fingertip contact states: “contact with a buried object” or “free movement through granular media”

The RNN was comprised of four hidden layers with 50 neurons per layer with parameters selected as described below. Model training and sequence labeling were performed with a **Long Short Term Memory (LSTM)** variation called **Gated Recurrent Units (GRUs)** [12] (Appendix C), which perform well for sequence learning and translation applications [74]. By combining RNNs with GRUs, an RNN can learn a nonlinear mapping between multimodal tactile feature inputs and contact state outputs. The long-term memory duration was designed as 500 ms with a short-term memory duration of 10 ms, per the 100-Hz update rate selected for step “(c)” of Figure 2. RNN weights were updated using Adaptive Momentum Estimation Optimization, which computes adaptive learning rates for each parameter [40]. The cost function was based on cross-entropy loss and also included a penalty for heavy weights via L2-regulation. Dropout was applied with a rate of 0.4 to prevent overfitting and to enhance robustness of the model to the loss of input features, which will be evaluated using an ablation study in Section 6.1.

## 4 BELIEF MAP REPRESENTATION OF TACTILE INFORMATION

To convey the ever-changing fingertip state from our multimodal tactile perception pipeline to a human teleoperator, we lay the foundation for the visualization of an unknown object buried in granular media using a “belief map” [68]. The belief map combines spatial information related to fingertip location (derived from known robot kinematics) with probabilistic fingertip contact state estimates provided by our SF-RNN model architecture (Figure 2). A belief map can be used by a robot for semi-autonomous trajectory planning to further reduce uncertainty in the map or by a human teleoperator for high-level decision-making.

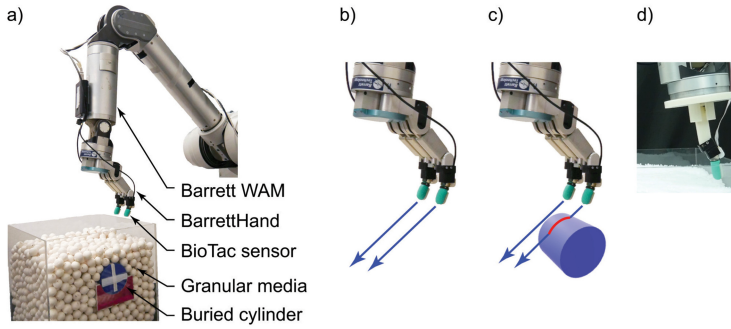


Fig. 4. (a) Experimental set-up used to collect multimodal tactile sensor training data by sweeping sensorized fingertips through granular media (cottonwood beads) and across a buried cylinder. (b) Blue arrows indicate the linear trajectory of two fingertips through granular media (not shown for clarity). (c) In the presence of a buried cylinder, one fingertip was swept briefly across the cylinder (finger-object contact is indicated in red). (d) For a robustness study, different granular media types were used (plastic pellets are shown) and a 3D-printed adapter was used to change the orientation of the fingertip to simulate  $20^\circ$  of finger extension relative to the pose in (a)–(c).

As a robot explores a region of interest, it samples from an unknown distribution and obtains new information. Specific to our application, a robotic hand with tactile sensors was used to explore and estimate the location of a buried object. As the robotic hand moved through space, it sampled the underlying spatial distribution by estimating the fingertip contact state and simultaneously associating the contact state with the fingertip location in space. The probability values for the entire spatial distribution were updated at each timestep with each new observation through a probabilistic classification method, specifically **Support Vector Classification (SVC)** with a **radial basis function (RBF)** kernel (Appendix D). Spatial information was stored in an RBF kernel that maps the influence of contact state at its spatial location into an infinite dimensional feature space. The spatial reach, or sphere of influence, of the radial basis function was customized for best performance for the size of our testbed, tactile sensor, and buried object. Since the SVC score function sums the learned mapping from all previous timesteps, the belief map is capable of tracking the states it has previously sampled. Finally, the SVC scores were calibrated using Platt Scaling [64] to produce a probability of whether an object is buried at a specific location. This probability integrates information about estimated fingertip contact state and known fingertip location. In this work, the Python Scikit-Learn package [60] was used for optimization and calibration. The distribution that represents the probability of fingertip contact with a buried object is updated for every 10 iterations of the RNN, a belief map update rate that was limited by our current computational power and is discussed further in Section 7.2.

## 5 EXPERIMENTAL PROCEDURE: ROBOT DATA COLLECTION

To equip robots with the ability to locate objects buried within granular media, we focus on perceiving fingertip contact states for a preliminary set of fingertip-object-media interactions. In this section, we describe our robot testbed and datasets for model training and model testing.

### 5.1 Robot Testbed

Our robot testbed consisted of a 7 **degree-of-freedom (DOF)** Barrett Whole Arm Manipulator with a 4 DOF BarrettHand (Barrett Technology, Cambridge, MA) (Figure 4). Two BioTac sensors (SynTouch LLC, Los Angeles, CA) were mounted on the BarrettHand fingertips. The multimodal



BioTac sensor measures vibration at 2,200 Hz, internal fluid pressure at 100 Hz, and fingerpad deformation data at 100 Hz via an array of 19 spatially distributed electrodes that sense impedance [22].

The sensorized robot fingertips were commanded, using position control, to perform linear exploratory movements within a rigid, 30 cm cube containing granular media. A 9.1-cm-diameter, rigid plastic cylinder was buried within the granular media. The cylinder was fixed in place such that its long axis was perpendicular to the linear exploratory movements. Interactions with buried objects that can move with or within granular media are beyond the scope of this work. The top, curved surface of the cylinder was set at a depth of 4 cm below the top surface of the granular media. The robot fingertips were swept at a constant commanded velocity through the granular media at a depth of 5 cm to ensure finger-object contact during some portion of the exploratory movements.

## 5.2 Dataset for Model Training

As a precursor to future experiments in soil, we used 20-mm-diameter wooden beads made of lightweight cottonwood as granular media for model training. The robot digits were swept through the granular media at a constant commanded speed of 4 cm/s. For one block of 250 trials, both digits were moved through the granular media without contacting the cylinder (Figure 4(b)). Pooling trials across both digits for this case resulted in a total of 500 trials in which the fingertips did not contact a buried object. As a first step toward multi-finger tactile perception, the commanded trajectory was such that only one digit contacted the cylinder while the other digit moved through the granular media (Figure 4(c)) for another block of 250 trials. This scenario generated a total of 250 trials in which a single fingertip made contact with a buried object. In summary, a total of 750 trials were collected (250 with contact and 500 without contact with a buried object).

To train and test the model, a 2 sec window of data was extracted from each trial in which the sensorized fingertip moved freely through granular media, made contact and broke contact with a buried object, and again moved freely through granular media (Figure 5). The making and breaking of contact with the buried cylinder will be referred to as “first contact” and “last contact” events, respectively. Trials without the buried cylinder featured only free fingertip movement through granular media. To mark the timing of the “first contact” and “last contact” events, trials were conducted with the cylinder exposed (in open air) for use as a “ground truth” (Figure 5).

## 5.3 Dataset for Model Testing

To directly test the robustness of our models to variation in the granular media, we collected additional data for model testing that included a wide range of granular media types. We selected granular media that were representative of aggregates ranging from coarse sand to coarse gravel according to the Wentworth scale [81], which is the most frequently used geological grade scale for classifying the size of sediment particles. As shown in Table 1, we used 1-mm-diameter poppyseed, 3-mm-diameter plastic pellets for injection molding, 6-mm-diameter plastic BBs for pellet guns, 16-mm-diameter wooden beads made of pine, and 20-mm-diameter wooden beads made of cottonwood.

We also tested the robustness of our models to variations in finger-object contact conditions. First, we used a 3D-printed adapter to change the orientation of the fingertip and simulate 20° of extension (Figure 4(d)) relative to the original pose of the BioTac when collecting data for model training (Figure 4(a)–(c)). Second, we swept the robot digit at a slower speed of 2 cm/s as compared to the 4 cm/s used for the training dataset. Finally, we moved the buried object 3.5 cm downstream from the original location such that the robot digit would not make contact until later in the

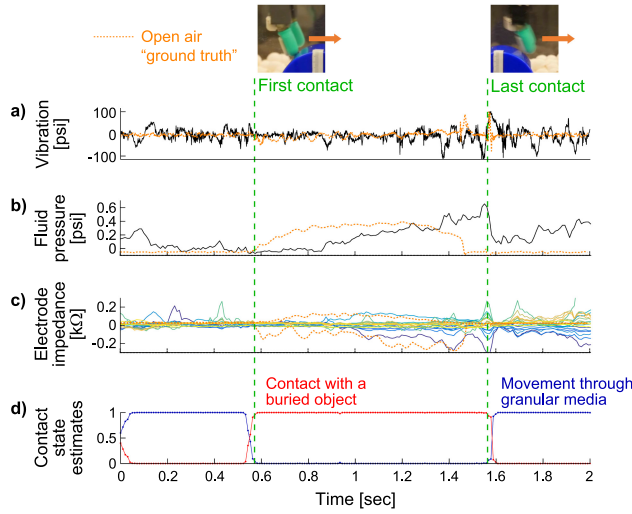


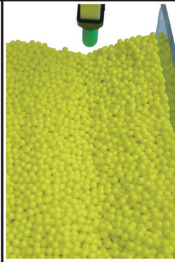




Fig. 5. (a) Vibration, (b) internal fluid pressure, (c) electrode impedance, and (d) fingertip contact state estimates from an SF-RNN model are shown for a representative 2-s trial. “First contact” and “last contact” events were identified from open air “ground truth” trials (screenshots, orange dotted lines). For clarity, impedance values from an open air trial are shown for two electrodes only (min, max traces).

Table 1. Granular Media Types: To Test Model Robustness, We Selected Five Distinct Granular Media Types to Represent the Five Aggregate Types Ranging from Coarse Sand to Coarse Gravel According to the Wentworth Scale [81]

Aggregate type	Coarse sand	Very fine gravel	Fine gravel	Medium gravel	Coarse gravel
Particle diam.	1-2 mm	2-4 mm	4-8 mm	8-16 mm	16-32 mm
Selected media	Poppysseed	Plastic pellets	Plastic BBs	Wooden beads (pine)	Wooden beads (cottonwood)
Particle diam.	1 mm	3 mm	6 mm	16 mm	20 mm
Testbed					

exploratory movement. In doing so, we could directly test whether the spatiotemporal model trained with cottonwood beads was overfit.

For each of the five granular media types, we collected 50 trials in which a single fingertip contacted the buried object and 25 trials in which neither digit contacted the buried object. Since fingertip speed was halved for the test dataset, the window of data that was extracted from each test trial was doubled (i.e., 4 s of data). Data labeling proceeded as described in Section 5.2.

## 6 EXPERIMENTAL PROCEDURE: MODEL EVALUATION

We extensively evaluated the SF-RNN on a tactile dataset with 750 trials of exploration within granular media. The dataset was randomized, then separated into 60% for training, 20% for validation, and 20% for testing. The validation dataset was used for tuning of model hyperparameters while the test dataset was used to evaluate the performance of the trained and validated model. We trained a number of classifier models with varying complexity following steps in the tactile perception pipeline (Figure 2). Models were trained for a sequence labeling task to distinguish between two fingertip states: “contact with a buried object” or “free movement through granular media.” The “contact with a buried object” class includes cases in which the fingertip touches the object directly and when the fingertip encounters the object indirectly through a force chain of granular particles.

We evaluated tactile perception models with both linear **support vector machine (SVM)** [21] and RNN classifiers on the following input architectures:

- “(b)”: IMFs extracted via EMD from vibration and internal fluid pressure data plus raw electrode impedance data (Figure 2(b)).
- “(c)”: Feature tensors constructed from EMD-processed vibration and internal fluid pressure data and raw electrode impedance data (Figure 2(c)).
- “(d)”: Sparse features identified autonomously and adaptively using an unsupervised online optimization algorithm for dictionary learning (Figure 2(d)).

For the linear SVM classifier models, feature tensors were flattened and concatenated. The linear classifiers were implemented in MATLAB (MathWorks, Natick, MA) while RNN architectures were implemented in Python using TensorFlow [1] with two NVIDIA GeForce GTX 970 graphics cards.

### 6.1 Study of Model Architecture

For an ablation study, all model architectures were evaluated as unimodal models (vibration, internal fluid pressure, or electrode impedance inputs only) and as multimodal models (with pairs of modalities, or all three modalities simultaneously) using training, validation, and test datasets based on cottonwood beads. When the linear SVM classifier was trained as a multimodal model, input streams from each modality were simply concatenated after data alignment to account for different sampling frequencies. Since this sort of simple concatenation performs poorly for an RNN architecture, data from different tactile sensor modalities were processed by separate RNNs and then passed through a fully connected (fusion) layer, which was referred to as **Fusion RNN (F-RNN)** by Jain et al. [37]. For our proposed multimodal tactile perception SF-RNN model architecture, tactile sensor modalities were passed through a sparse coding process as an unsupervised pre-training step, which preceded a fully connected fusion layer of RNN (Figure 2(e)).

The data augmentation technique of oversampling was used to address our unbalanced dataset, increase the diversity of the training data, and minimize overfitting of the trained model. For linear classifier models, observations from the minority class (“contact with a buried object” class) were randomly selected and added to the training set during each iteration during training. For RNN models, since the long-term memory duration was designed as 500 ms, the data were oversampled using a sliding window paradigm. The 500 ms tactile data streams were extracted from the selected 2-s window with different increments. Trials featuring free movement through granular media were oversampled by shifting a 500-ms sliding window in 10-frame (100-ms) increments. Since there were half as many trials featuring contact with a buried object, a 5-frame increment was used to shift the sliding window in those cases. Tables 2 and 3 summarize the results of the ablation studies on the unimodal and multimodal models, respectively.

Table 2. Ablation Study with Unimodal Models: Model Accuracy Is Reported as Mean  $\pm$  Standard Deviation in %

Modality	Category	Method					
		(b)+Linear SVM	(c)+Linear SVM	(d)+Linear SVM	(b)+RNN	(c)+RNN	(d)+RNN (SF-RNN)
Vibration	Validation (%)	51.1 $\pm$ 0.7	62.2 $\pm$ 0.6	73.5 $\pm$ 0.5	65.91 $\pm$ 0.5	66.0 $\pm$ 0.5	95.9 $\pm$ 0.2
	test (%)	48.7 $\pm$ 0.7	60.4 $\pm$ 0.6	72.0 $\pm$ 0.5	65.3 $\pm$ 0.5	65.6 $\pm$ 0.5	<b>95.7</b> $\pm$ 0.2
	process Time (ms)	1.68	2.28	2.47	1.5	0.35	2.58
	Input	8 IMFs	8 IMFs $\times$ 22 samples	200 sparse features	8 IMFs	8 IMFs $\times$ 22 samples	200 sparse features
Fluid pressure	Validation (%)	52.7 $\pm$ 0.6	73.3 $\pm$ 0.5	92.5 $\pm$ 0.3	67.5 $\pm$ 0.5	67.1 $\pm$ 0.5	85.4 $\pm$ 0.4
	test (%)	49.0 $\pm$ 0.7	69.0 $\pm$ 0.6	<b>95.8</b> $\pm$ 0.2	67.4 $\pm$ 0.5	67.2 $\pm$ 0.5	85.8 $\pm$ 0.4
	process Time (ms)	0.06	0.05	0.08	0.19	0.09	0.26
	Input	5 IMFs	5 IMFs $\times$ 5 samples	100 sparse features	5 IMFs	5 IMFs $\times$ 5 samples	100 sparse features
Electrode impedance	Validation (%)	82.2 $\pm$ 0.4	84.7 $\pm$ 0.4	93.4 $\pm$ 0.3	90.3 $\pm$ 0.3	92.1 $\pm$ 0.3	97.8 $\pm$ 0.2
	test (%)	78.6 $\pm$ 0.5	78.7 $\pm$ 0.5	85.4 $\pm$ 0.4	88.3 $\pm$ 0.4	90.1 $\pm$ 0.3	<b>97.6</b> $\pm$ 0.2
	process Time (ms)	0.22	0.17	0.24	0.23	0.17	0.4
	Input	19 channels	19 channels $\times$ 5 samples	100 sparse features	19 channels	19 channels $\times$ 5 samples	100 sparse features
Avg test accuracy (%)		58.76	69.38	84.42	73.68	79.06	<b>93.06</b>

Best performing models are shown in bold. The models were trained and evaluated with cottonwood beads. The (b)–(d) annotations refer to the perception pipeline steps shown in Figure 2.

**6.1.1 Effect of Model Inputs and Sparse Coding.** We built multiple models to assess the usefulness of different model input combinations for both model training and testing: one tactile sensing modality, two modalities, and three modalities. As expected, models trained with two or three modalities (Table 3) achieved higher test accuracy than those trained with only one modality (Table 2).

In Table 2, we present the results of the ablation study for all six model architectures for models that used unimodal tactile inputs only. When the (d)+Linear SVM architecture was used, fluid pressure was the most informative tactile sensing modality, yielding a test accuracy of 95.8%. However, when the SF-RNN architecture was used, vibration and electrode impedance became much more informative, yielding test accuracies of 95.7% and 97.6%, respectively. Use of electrode impedance data resulted in the highest test accuracy of 97.6% for all unimodal models, suggesting that electrode impedance data, generated by fingerpad deformation, was the single most informative tactile sensing modality for distinguishing between the two fingertip contact states (Table 2).

Of the unimodal models (Table 2), classifiers trained on overcomplete sparse representations of tactile data (“(d)+” cases) had obvious advantages as compared to other models. Notably, only the proposed SF-RNN (“(d)+RNN”) model architecture was capable of achieving at least an 85% test accuracy for each of the three unimodal input cases. While the “(c)+RNN” method achieved 90.1% test accuracy for unimodal models trained on electrode impedance, test accuracy decreased to 65.6% and 67.2% for models trained only on vibration data or fluid pressure data, respectively.

In Table 3, we present the results of the ablation study for all six model architectures for models that used multimodal tactile inputs (with pairs of modalities, or all three modalities simultaneously). Both the linear SVM and RNN multimodal models trained on sparse features (“(d)+” cases) produced high test accuracy results. The average test accuracy was 96.7% and 97.3% for the (d)+Linear SVM and SF-RNN model architectures, respectively. Although the (d)+Linear SVM model performed well in Table 3, the (d)+Linear SVM model did not perform well during the robustness tests, which will be discussed in Section 6.2.

Table 3. Ablation Study with Multimodal Models: Model Accuracy Is Reported as Mean  $\pm$  Standard Deviation in %

Modality	Category	Method					
		(b)+Linear SVM	(c)+Linear SVM	(d)+Linear SVM	(b)+ F-RNN	(c)+ F-RNN	(d)+RNN (SF-RNN)
Vibration, fluid pressure	Validation (%)	51.8 $\pm$ 0.7	78.1 $\pm$ 0.5	99.7 $\pm$ 0.1	67.1 $\pm$ 0.5	73.0 $\pm$ 0.5	96.5 $\pm$ 0.2
	test (%)	46.6 $\pm$ 0.7	72.2 $\pm$ .5	<b>99.5</b> $\pm$ 0.1	64.4 $\pm$ 0.6	67.0 $\pm$ .5	96.5 $\pm$ 0.2
	process Time (ms)	2.62	1.87	2.52	7.34	0.75	2.72
	Input	8 IMFs, 5 IMFs	8 $\times$ 22, 5 $\times$ 5 feature tensors	200D, 100D sparse features	8 IMFs, 5 IMFs	8 $\times$ 22, 5 $\times$ 5 feature tensors	200D, 100D sparse features
Vibration, electrode impedance	Validation (%)	84.7 $\pm$ 0.4	86.4 $\pm$ 0.4	93.8 $\pm$ 0.3	85.1 $\pm$ 0.4	90.4 $\pm$ 0.3	98.3 $\pm$ 0.2
	test (%)	78.3 $\pm$ 0.5	84.9 $\pm$ 0.4	91.7 $\pm$ 0.3	83.9 $\pm$ 0.4	86.4 $\pm$ 0.4	<b>97.8</b> $\pm$ 0.2
	process Time (ms)	1.68	2.24	2.73	7.41	0.77	2.83
	Input	8 IMFs, 19 channels	8 $\times$ 22, 19 $\times$ 5 feature tensors	200D, 100D sparse features	8 IMFs, 19 channels	8 $\times$ 22, 19 $\times$ 5 feature tensors	200D, 100D sparse features
Fluid pressure, electrode impedance	Validation (%)	83.4 $\pm$ 0.4	88.3 $\pm$ 0.3	98.7 $\pm$ 0.1	89.4 $\pm$ 0.4	92.1 $\pm$ 0.3	98.1 $\pm$ 0.2
	test (%)	81.7 $\pm$ 0.4	85.8 $\pm$ 0.4	<b>98.9</b> $\pm$ 0.1	89.0 $\pm$ .4	91.0 $\pm$ 0.3	97.5 $\pm$ 0.2
	process Time (ms)	0.27	0.19	0.302	0.43	0.54	0.51
	Input	5 IMFs, 19 channels	5 $\times$ 5, 19 $\times$ 5 feature tensors	100D, 100D sparse features	5 IMFs, 19 channels	5 $\times$ 5, 19 $\times$ 5 feature tensors	100D, 100D sparse features
All modalities	Validation (%)	85.0 $\pm$ 0.4	89.6 $\pm$ 0.3	99.9 $\pm$ 0.1	84.7 $\pm$ 0.4	92.0 $\pm$ 0.3	98.4 $\pm$ 0.1
	test (%)	80.8 $\pm$ 0.4	89.2 $\pm$ 0.3	<b>99.7</b> $\pm$ 0.1	84.1 $\pm$ 0.4	84.3 $\pm$ 0.4	98.7 $\pm$ 0.1
	process Time (ms)	8.08	2.39	2.82	8.06	0.85	2.96
	Input	8 IMFs, 5 IMFs, 19 channels	8 $\times$ 22, 5 $\times$ 5, 19 $\times$ 5 feature tensors	200D, 100D, 100D sparse features	8 IMFs, 5 IMFs, 19 channels	8 $\times$ 22, 5 $\times$ 5, 19 $\times$ 5 feature tensors	200D, 100D, 100D sparse features
Avg test accuracy (%)		68.8	81	96.7	79.1	81.5	<b>97.3</b>

Best performing models are shown in bold. The models were trained and evaluated with cottonwood beads. The (b)–(d) annotations refer to the perception pipeline steps shown in Figure 2.

Training on three modalities did not provide much of an advantage over training on two modalities. For example, the SF-RNN model trained and tested on three modalities had a test accuracy of 98.7%, which was only slightly better than the levels of performance for the two-modality SF-RNN models, which ranged from 96.5% to 97.8% accuracy. One advantage of training on three tactile sensing modalities appears to be a robustness to incomplete input data. Without having to retrain the model, one could provide data from any of the tactile sensing modalities, or combinations thereof, and achieve comparable performance to models trained more specifically on fewer modalities (results of the SF-RNN model tested with incomplete inputs are shown in Table 5 in Appendix E). This could be useful if one or more modalities becomes temporarily unavailable during tactile exploration. Considering the complexity and harsh working environment of interacting with and maneuvering within granular media, we recommend a multimodal SF-RNN structure instead of relying on a single modality for haptically perceiving contact state within granular media.

**6.1.2 Effect of Sensor Data Representation.** To check the sensor data representation and model training, we visualized the data and feature spaces throughout the multimodal modeling pipeline (Figure 2) using a dimensionality reduction algorithm called **t-Distributed Stochastic Neighbor Embedding (tSNE)** [78]. Feature spaces were visualized in Figure 6, in which each dot represented a labeled multimodal tactile sensor feature tensor that was updated at 100 Hz. This visualization provided insight as to how our proposed architecture dealt with dimensionality, learned features to represent multimodal tactile sensor data, and classified features with high accuracy.



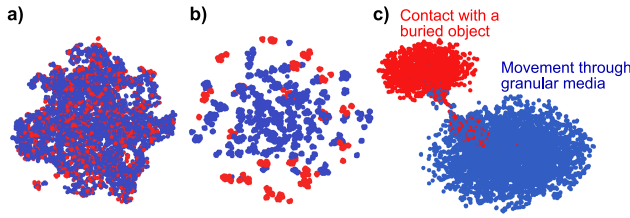


Fig. 6. Using tSNE for visualization, one can observe the effects of different stages of the tactile perception pipeline. (a) Visualized in 3D space, the feature tensors from Figure 2(c) are entangled. (b) The outputs from the sparse coding step in Figure 2(d) are spread out in a high-dimensional feature space. (c) The feature space from the last layer of the model in Figure 2(e) shows that the data are mostly separated into distinct regions according to contact state.

First, the feature tensors from Figure 2(c) ( $8 \text{ IMFs} \times 22 \text{ samples}$  for vibration data,  $5 \text{ IMFs} \times 5 \text{ samples}$  for fluid pressure data, and  $19 \text{ channels} \times 5 \text{ samples}$  for electrode impedance data) were visualized in 3D space as a jumbled cluster of feature tensors with mixed fingertip contact states (Figure 6(a)). Using sparse coding, these entangled feature tensors were fully separated, with feature tensors classified as fingertip contact with a buried object now pushed into the outer regions of an overcomplete 400-dimensional sparse feature space (Figure 6(b)). Finally, the output of the 50-dimensional final layer of the RNN was visualized as two distinct clusters (one cluster per state) with a transitional bridge between the two clusters consisting of a mixture of feature tensors from both classes. As the RNN with GRUs reduced the feature space from 400 dimensions to 50 dimensions, redundant features that were not critical to detecting fingertip contact with a buried object were automatically excluded.

In Tables 2 and 3, we showed how accuracy improves with each pre-processing step in our proposed pipeline. In particular, we illustrated that the use of sparse coding prior to sensor fusion within a fully connected RNN input layer improves the performance of both linear and nonlinear models (“(d)+” cases). The motivation for introducing this unsupervised pre-training step was to improve generalization capabilities and to capture intricate dependencies between parameters. Sparse coding selects model hyperparameters that separate entangled tactile features (Figure 6(a)) such that the features that distinguish different fingertip contact states are as distinct as possible (Figure 6(b)). Given the low off-sample error of the SF-RNN for the test data (Table 3), the unsupervised pre-training step appears to have selected features that are near an optimal solution for high classification accuracy with reduced tendency for overfitting. Furthermore, the process of tuning hyperparameters is greatly simplified when sparse features are employed. The main disadvantage of unsupervised pre-training is that it is time consuming, especially for vibration data (Table 3). If offline pre-training time is not a major concern, then unsupervised pre-training is recommended.

## 6.2 Study of Model Robustness

As described in Section 5.3, for the purposes of a robustness study, we collected additional test data that intentionally varied the granular media types and finger-object contact conditions. We took multimodal models that were originally trained and tested with the cottonwood beads using all three modalities simultaneously and we tested them directly (without any hyperparameter tuning) on the new test dataset. The results of this robustness study are shown in Table 4. The robustness study was only performed on multimodal models that demonstrated decent performance with the original cottonwood bead dataset (i.e., “(d)+” cases that used sparse coding and all three modalities simultaneously in Table 3).

Table 4. Robustness Study: Model Accuracy Is Reported as Mean  $\pm$  Standard Deviation in %

Material	Poppyseeds	Plastic pellets	Plastic BBs	Wooden beads (pine)	Wooden beads (cottonwood)
(d)+Linear SVM	46.0 $\pm$ 0.7	53.7 $\pm$ 0.7	58.5 $\pm$ 0.6	52.4 $\pm$ 0.7	52.2 $\pm$ 0.7
(d)+RNN (SF-RNN)	74.8 $\pm$ 0.5	78.7 $\pm$ 0.5	88.4 $\pm$ 0.4	87.0 $\pm$ 0.4	86.2 $\pm$ 0.4

The models were trained on cottonwood beads but evaluated on the expanded test dataset collected with different granular media types and finger-object contact conditions.

As shown in Table 3, the multimodal models that used sparse coding and all three modalities achieved 99.7% and 98.7% test accuracy for the linear SVM and SF-RNN methods, respectively. Due to the purposeful changes in the granular media types and finger-object contact conditions, the test accuracy of the SF-RNN model decreased, as expected, to as low as 74.8% for poppyseeds and only to 88.4% for plastic BBs. However, the test accuracy of the linear SVM model plummeted, only achieving as high as 58.5% accuracy for plastic BBs (Table 4). This is likely due to the inability of linear models to capture useful nonlinear relationships between rich, multimodal tactile signals and fingertip contact states.

However, nonlinear SF-RNN models were less influenced by variability in the input data. Test accuracy exceeded 74% for all granular media types and exceeded 86% for all granular media having particle diameters greater than or equal to 6 mm. This suggests that the model trained with cottonwood beads was capable of encoding useful features from a range of granular media types. SF-RNN models trained with cottonwood beads data collected with 4 cm/s exploratory movements (Tables 2 and 3) were robust to changes in granular media type and also fingertip speed, which was decreased to 2 cm/s for the test dataset (Table 4).

### 6.3 Belief Map Representation of Tactile Information

The state estimates from the SF-RNN model are simultaneously fed into a belief map as samples of the spatial distribution within the exploration space. Figure 7 shows the evolution of a belief map for a 2D, top-down view of the exploration area as the robot finger moves during a 2-s trial. Upon initial contact with the buried object, the belief map reflects an increased probability of contact with a buried object (red region appears in Figure 7(a)). As the perceived contact persists during the finger movement, the red region of high probability increases in area (Figure 7(b)). Upon a return to free fingertip movement through the granular media, the belief map has been refined to show a region having high probability of contact with a buried object (dark red regions in Figure 7(c)) as well as adjacent regions that are believed to consist of granular media only (dark blue regions in Figure 7(c)).

Upon completion of the first exploratory movement, the width of the dark red region is approximately 10 cm (Figure 7(c)). The cylinder diameter is 9.1 cm and the chord length from first contact to last contact was approximately 5.4 cm. While the belief map could use refinement from additional exploratory movements of the fingertip, the dark red region is of the proper order of magnitude in size and could be used to plan future exploratory movements that reduce uncertainty in the shape and size of the visualization.

## 7 DISCUSSION

### 7.1 Notable Trends in the Tactile Data

To further analyze the results in Table 2 and 3, we compared the response of each tactile sensor modality for each contact state (Figures 8 and 9). In Figure 8, average trends in vibration amplitude and internal fluid pressure are shown for 500 trials without a buried object (green) and 250 trials

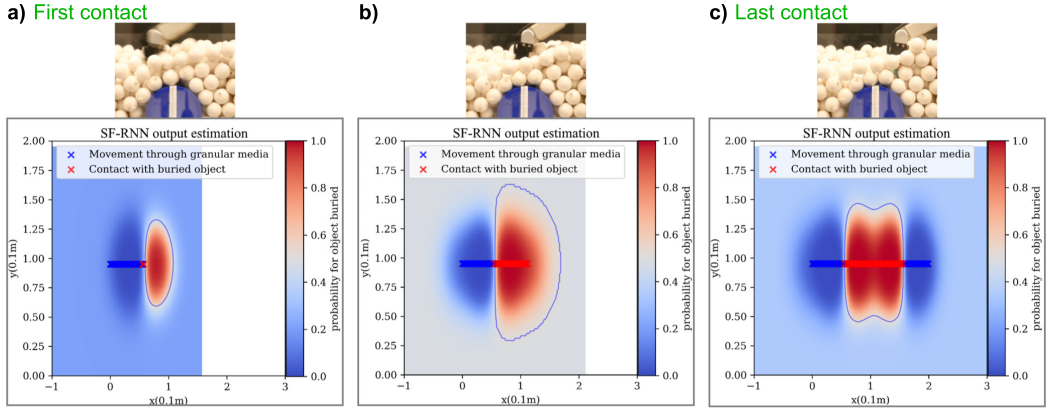


Fig. 7. As the fingertip moves across a buried object (top row), a belief map (bottom row) is updated using fingertip contact state estimates from the SF-RNN model. Snapshots are shown as the fingertip (a) makes first contact with the buried object, (b) passes across the top of the object, and (c) makes last contact with the object and returns to free movement within the granular media.

with a buried object (blue during free movement through granular media and red during contact with a buried object). Since the tactile sensor data are positively skewed, we report median values along with first and third quartile bands in Figures 8 and 9. The interquartile spread reflects inter-trial variability.

**7.1.1 Susceptibility of Different Tactile Modalities to Noise Induced by Movement within Granular Media.** The BioTac sensor uses a fluidic mechanotransduction medium that enables the measurement of vibration and internal fluid pressure via high-pass and low-pass filtered signals, respectively, from a hydro-acoustic pressure sensor. For most tasks conducted in open air environments, the sensitivity of the BioTac’s “fast-adapting” hydrophone receptor to stimuli such as contact transients and vibration is especially useful.

As the robotic finger rakes through the granular media, the BioTac sensor sensitively records vibrations transmitted through the fluid-filled fingerpad caused by repeated interactions with granular media particles. As a result, changes in hydrophone data that might be indicative of key contact events with a buried object get drowned out by noise caused by motion relative to granular media (Figure 8). In addition, the spread between the first and third quartiles of the vibration data increased by a factor of 5 when the fingertip was moved through granular media as opposed to open air. Perhaps with advanced signal processing techniques, changes in vibration magnitudes observed near “first contact” and “last contact” events could be extracted for improved tactile perception.

The BioTac’s “slow-adapting” receptors were also susceptible to noise induced by interactions with granular media. The interquartile spread increased by a factor of approximately 5, 3, and 6 for electrodes #1, #7, and #11 at the fingertip, respectively (Figure 9). The interquartile spread for the internal fluid pressure modality increased the most (by a factor of 15) when the fingertip was moved through granular media as opposed to open air. Nonetheless, larger sustained fingerpad deformations were clearly observed in the “slow-adapting” receptor data when the finger was in contact with the buried object (Figures 8(b) and 9). This was especially true for electrodes at the tip of the finger that would most likely have direct contact with a buried object (electrodes #1, #7, and #11 in Figure 9(b)).

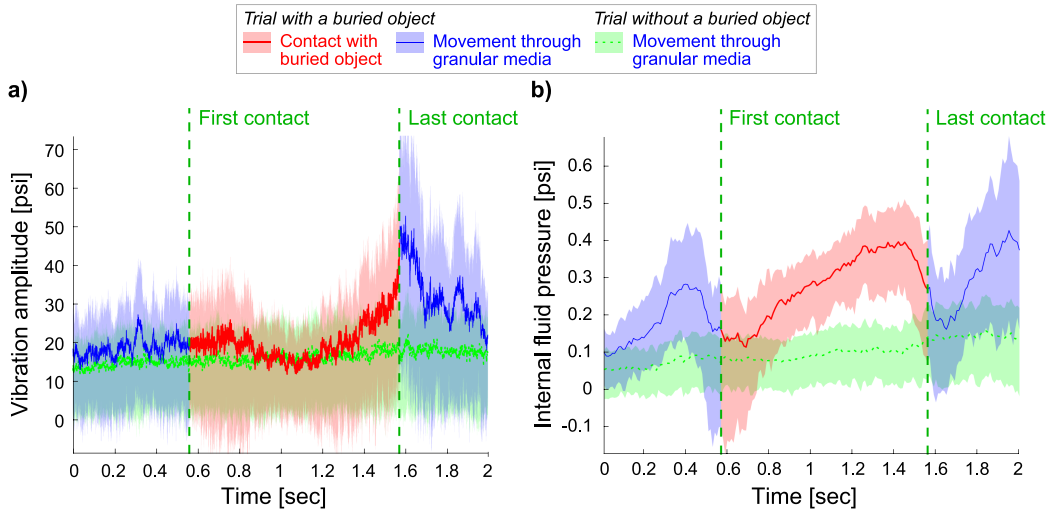


Fig. 8. (a) Vibration amplitude and (b) internal fluid pressure are shown as median values with 1st and 3rd quartile bands for both fingertip contact states during all trials with and without a buried object.

**7.1.2 Information Encoded by Different Tactile Modalities during Movement within Granular Media.** Based on studies of human touch, it is believed that coarse textures and local shape are primarily perceived through spatial cues while fine textures are perceived through vibratory cues [32, 45]. Thus, we hypothesize that spatial information from fingerpad deformation is especially useful for distinguishing between sustained forces induced by a buried object and repetitive contacts with much smaller granular media particles. While changes in fluid pressure and electrode impedance both reflect fingerpad deformation, electrode impedance provides an array of spatial information that contains more information than a single scalar fluid pressure signal for the entire fingertip.

Despite the movement of granular media particles relative to the fingertip, the electrodes are capable of encoding sustained forces, whether due to direct contact with a buried object or granular media jamming preceding direct contact. In our experiments, the maximum height difference between the depth of the object and the depth of the exploratory movement was 1 cm. Thus, the distal aspect of the fingertip made primary contact with the buried object. When contacting the buried object, the fingerpad compresses at the fingertip, as reflected by the increase in impedance of electrode #7 (Figure 9(b)). This results in a bulging of the fingerpad that is reflected by the decrease in impedance of neighboring electrodes #1 and #11 on the sides of the fingertip. Such information could enable advanced behaviors within granular media such as contour-following, obstacle avoidance, and dynamic trajectory planning. Additionally, prior efforts to perceive the size and shape of local geometries using electrode impedance data could potentially be extended from open air scenarios [65] to granular media scenarios.

The “fast-adapting” hydrophone receptor appears to be sensitive to transients, especially the punctuated “last contact” event when the fingertip slips off of the buried object (Figure 8(a)). Videos of the experiment suggest that the fingertip, moving at a constant rate in the testbed, vibrates after suddenly losing its sustained contact with the buried object and getting flung into the surrounding granular media particles. Although further investigation is required, we surmise that the sudden release of the fingertip from contact with the buried object caused the temporary peak in vibration amplitude and short-lived increase in internal fluid pressure that were observed shortly after the “last contact” event (Figure 8). Vibration data may be useful for identifying transitions in fingertip

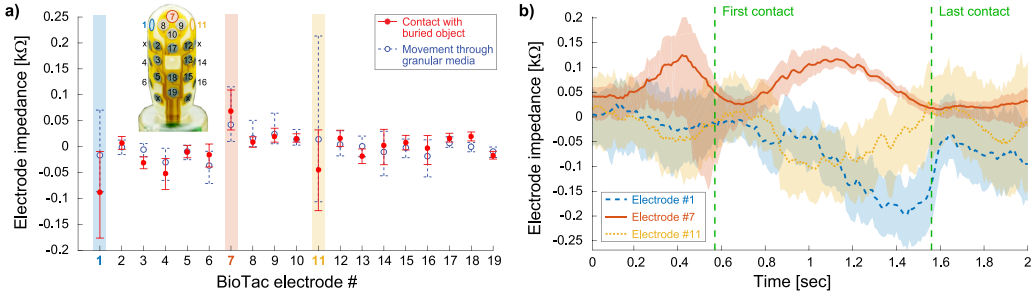


Fig. 9. Electrode impedance is shown as median values with first and third quartile intervals. (a) Trends for all 19 electrodes are shown for both fingertip contact states during all trials with and without a buried object for the same time window (theoretical contact period). (b) Trends for three electrodes at the fingertip (#1, #7, #11) are shown for all trials with a buried object for the entire 2-s period.

contact states, although significant changes in vibration magnitude alone do not seem to occur with the “first contact” event.

**7.1.3 Haptic Sensitivity to Granular Media Jamming.** For transitions of fingertip contact state from “movement through granular media” to “contact with a buried object,” the SF-RNN model estimated the transitions prematurely. The transition to “contact with a buried object” occurred approximately four timesteps (40 ms) prior to the true “first contact” event (Figure 5(d)). These errors are likely caused by our strict labeling policy during the training process. Fingertip contact state labels were assigned according to *direct contact* with a buried object in the absence of granular media (Figures 4(c) and 5). As a result, the “contact with a buried object” labels would not account for indirect contact with a buried object via particle force chains caused by granular media jamming [14]. This premature change in the fingertip contact state estimate just prior to direct “first contact” (theorized from “ground truth” trials in open air) makes intuitive sense when considering granular media jamming effects. In contrast, the change in fingertip contact state occurs in a timely manner at the “last contact” event. Approximately three timesteps (30 ms) were needed for the SF-RNN model to clearly transition away from a state of “contact with a buried object,” but the drop in the state label’s probability was immediate (Figure 5(d)).

Granular media jamming may have contributed to the increase in impedance for electrode #7 at the fingertip just prior to “first contact” (Figure 9(b)), but further investigation is required. Granular media jamming effectively injects uncertainty into the contact state estimate and the resulting spatial belief map. For instance, a buried object might be visualized as larger than its actual size due to premature perception of contact due to granular media jamming. One could reduce this uncertainty by accounting for the effects of jamming, especially how granular media characteristics and fingertip depth may prematurely trigger a transition in the fingertip contact state. The effect of granular media jamming is the subject of a separate investigation and is outside the scope of this work.

## 7.2 Real-time Tactile Perception and Visualization for Human Control of Remote Robotic Systems

The supplemental video includes a real-time demonstration of the belief map update process, for which the graphic rendering was updated every 100–300 ms. During real-time execution, we combine tactile perception outputs from the SF-RNN model (fingertip contact state) with dynamic robot kinematic information (fingertip location) to produce a continuously updated 2D belief map.



Since all prior haptic information is already encoded into the spatial distribution, the updates of the underlying spatial statistics of the belief map can be treated as improvements in the robot's belief with the benefit of new haptic experiences. The long-term memory duration of the belief map can be adjusted according to the needs of the teleoperator.

A belief map conveys environment characteristics as visualizations of spatial probability distributions across a search space, from which a human teleoperator could infer the location, size, and orientation of a buried object. Information from multiple sensorized, semi-autonomous robots that are performing haptic search tasks in parallel could be combined into a single map. This could enable increased productivity for a single human telesupervisor as compared to a single human teleoperator of a leader–follower system. For safety critical tasks, the human may need to manually teleoperate the robot, as in surgical applications [61, 86] or the handling of high-consequence materials [77]. Custom haptic interfaces could be used to control the robot and display contact force information to a teleoperator's hand [76]. In such an instantiation, the belief map would be used to convey remote environment characteristics and little autonomy would be given to the robot.

To increase the level of robot autonomy, the belief framework could be used to construct and update a map of an unknown environment covered by granular media. The ability to perceive within granular media could be used to extend existing haptic [6], and contact SLAM approaches [43]. Moreover, a belief map could be used to generate future exploration trajectories via optimal control [55] or ergodic control [2], which could enable belief maps to be updated more efficiently.

In the shared control or semi-autonomous condition, task execution and control could be shared between humans and robots. For instance, utilizing impedance control, a robot may follow a trajectory supervised by a human teleoperator while still responding naturally to external perturbations [16]. In another example, the robot might be responsible for controlling the orientation of the end-effector while the human controls the position [46]. Alternatively, a human may take the leader role for high-level, large-scale movements toward a region of interest while the robot manages the low-level, precise movements of the end-effector that rely critically upon local sensing [19, 41].

To support human-robot coordination and better utilize the robot, Gombolay et al. recommended to equip robots with as much autonomy as possible and greater robot authority over task allocation [25], such as telling the human operator when each sub-exploration could be started and finished. We posit that trajectory planning methods and algorithms to assess confidence based on the belief map will be necessary to improve the quality of human-robot interaction during exploration within granular media. However, it is important to enact appropriate arbitration between the human leader and robot follower during shared control, as the human domain expert must always retain final authority for safety purposes [16].

## 8 CONCLUSION

Haptic exploration within granular media is an essential task for robots designed for interacting with delicate and/or dangerous buried objects. Locating and retrieving buried devices requires haptic intelligence to carefully maneuver within the granular media, detect, and interact with buried objects. As such, we focused on the fundamental capability of haptically perceiving contact with a buried object and distinguishing such an event from ubiquitous contact with surrounding granular media particles.

### 8.1 Summary of Contributions

In this work, we established a multimodal learning architecture based on Recurrent Neural Networks with Long Short-Term Memory units that incorporates multiple tactile sensing modalities (vibration, internal fluid pressure, and electrode impedance as a proxy for fingerpad deformation).

Our SF-RNN model architecture learned to fuse sparse representations of multimodal tactile sensor data and performed robustly on novel test data that included variations in granular media type and particle size, fingertip orientation, fingertip speed, and object location along the linear fingertip trajectory. We also introduced a belief map representation to visualize tactile information from an SF-RNN model for a human teleoperator. During real-time implementation, the tactile perception model and belief map were updated every 150 ms and every 100–300 ms, respectively.

Analyses of the different tactile modalities showed that fingerpad deformation was the most informative tactile sensor modality for distinguishing between free movement of the fingertip through granular media and contact with an object buried in granular media. With signal processing, filtered vibration and internal fluid pressure responded to first and last contact sensitively and improved perception accuracy beyond that achieved using electrode impedance data alone.

## 8.2 Limitations and Future Work

Here, we lay the foundation for more complex, semi-autonomous robotic behaviors such as exposure, excavation, and retrieval of buried objects. To improve upon this foundation, we highlight some limitations of the current work. For instance, our robot exploration movements were limited in speed and distance because of the size of our granular media testbed. To protect the exposed joints of the robot gripper, we used a 3D-printed adapter when performing exploratory movements amongst small diameter particles and only plunged the gripper into a testbed of large wooden beads.

Further investigation is needed to assess the effects of experimental variables such as object shape, object rigidity, depth of burial, and more complex, multi-digit exploratory movements on tactile perception. Moreover, accounting for tactile sensitivity to granular media jamming could help to reduce uncertainty in the belief map and better inform teleoperators. In this work, a robot fingertip was lightly raked across a buried object using a preplanned trajectory. Ideally, robots with haptic intelligence could be programmed to conservatively stop and autonomously replan their exploratory movements as soon as contact or granular media jamming is haptically perceived.

Further upgrades of the belief map updating algorithm (Appendix D) could include the use of local length-scale kernels with optimizations such as unsupervised, automatic kernel selections and placement [27, 73]. Fast kernel approximation methods [68] could be implemented to increase computational speed, thereby enabling the visualization of large 3D scenes. With regards to screen-based control interfaces, an attitude-based controller [79] could be integrated to allow teleoperators to use their natural movements and gestures to change the viewing angle of the belief map and command the robot.

## AUTHOR DISCLOSURE

V. Santos serves on the SynTouch Board of Advisors. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of SynTouch.

## APPENDICES

### A EMPIRICAL MODE DECOMPOSITION

The EMD [34] approach is based on the assumption that time-series data consist of simple intrinsic modes of oscillation. Each of these oscillatory modes is represented by an IMF that satisfies two basic conditions to be described below: (i) the number of extrema and the number of zero-crossings may differ at most by one, and (ii) the mean value of the local maxima and local minima envelopes is zero.

In contrast to a Fourier analysis, EMD decomposes a time-series signal  $q(t)$  adaptively by sifting the signal into a series of intrinsic mode functions  $c_j$  until only a small monotonic final residue  $r_n$  remains. For an arbitrary  $q(t)$ , the sifting process starts with the identification of all local extrema. Local maxima are then connected with a cubic spline to create an upper envelope; local minima are similarly connected to produce a lower envelope. The sifting procedure is repeated on the difference between the signal  $q(t)$  and the mean of the upper and lower envelopes until the mean qualifies as an IMF per the two conditions stated previously [34]. Once the iterative sifting process has resulted in a small-enough residue  $r_n$ , the original signal  $q(t)$  can be represented as

$$q(t) = \sum_{j=1}^n c_j + r_n, \quad (3)$$

The original signal  $q(t)$  is thus decomposed in the time domain. The length of each IMF is identical to that of the original signal, which ensures that characteristics of the original signal are preserved, including variations in frequency. This is important for times-series signals that are often the result of multiple, natural processes that may or may not be concurrent. Such data features would remain hidden through the use of Fourier domain or wavelet coefficient analyses that filter out specific frequencies.

## B SPARSE CODING

Sparse coding is widely used for audio signal reconstruction and image processing. The sparsity of nonzero values in the model offers four main advantages: increased storage capacity, representation as explicit signals, faster storage and retrieval of representations, and reduced computational effort [42, 71]. By minimizing an empirical cost, the algorithm learns a dictionary  $D = [d_1, \dots, d_k] \in R^{m \times k}$  and a sparse matrix of coefficients  $\alpha = [\alpha_1, \dots, \alpha_n] \in R^{k \times n}$  that can be used to approximate a signal  $X = [x_1, \dots, x_n] \in R^{m \times n}$  as  $\hat{X} = D\alpha$  in a linearized manner (Equation (2)).

The cost function can be cast as a matrix factorization problem that minimizes errors in the linearized reconstruction of the original signal  $X$  with an additional penalty term that induces sparsity.

$$\min_{D \in C, \alpha \in R^{k \times n}} \frac{1}{2} \|X - D\alpha\|_F + \lambda \|\alpha\|_{1,1}, \quad (4)$$

The operator  $\|\cdot\|_F$  denotes the Frobenius norm. The term  $\|\alpha\|_{1,1}$  denotes the  $l_1$ -norm of the sparse matrix of coefficients  $\alpha$  that is the sum of the magnitude of all coefficients. Per [48], the sparsity-inducing regularization parameter  $\lambda$  was set as  $\lambda = 1.2/\sqrt{m}$ . The sizes of  $D$  and  $k$  are tuned according to a tradeoff between computing time and objective function values (Equation (4)). To prevent  $D$  from having arbitrarily large values, its columns  $d_1, \dots, d_k$  are constrained to have an  $l_2$ -norm less than or equal to one. This is achieved by defining  $C$  as a convex set of matrices satisfying the constraint:  $C \triangleq \{D \in R^{m \times k} \text{ s.t. } \forall j = 1, \dots, k, d_j^T d_j \leq 1\}$ .

## C RNN WITH GRUS

Standard RNNs take temporal sequences of data as inputs and yield high-level representations or labels as outputs. The feedforward neural network structure is enhanced by “recurrent edges” between adjacent timesteps that consider temporal information both forward and backward in time. Such internal loops across timesteps allow temporal information to persist within the network. At time  $t$ , nodes with recurrent edges receive input from current data point  $x_t$  as well as input from a previous timestep ( $t - 1$ ) from hidden node  $h_{t-1}$ . In this manner, the previous input  $x_{t-1}$  may influence the softmax output layer  $y_t$  through the recurrent connection.

$$h_t = \sigma(W_h \cdot [h_{t-1}, x_t] + b_h), \quad (5)$$

$$y_t = \text{softmax}(W_y h_t + b_y), \quad (6)$$

where  $\sigma$  represents an activation function (sigmoid function) and  $W_*$  and  $b_*$  are weight and bias parameters, respectively.

The GRU is a simplified LSTM unit and only consists of two gates: a reset gate  $r$  and an update gate  $z$ . Each gate is composed of a sigmoid neural net layer and pointwise multiplication operation. The reset gate  $r_t$  (Equation (7)) decides whether the previous hidden state is ignored and a  $\tanh$  layer produces a vector of new candidate values  $\tilde{C}_t$  (Equation (8)) that could be added to the current state (Equation (10)). The update gate  $z_t$  (Equation (9)) combines the forget and input gate, and enables the GRU to add or remove (i.e., remember or forget) information with respect to the memory cell or cell state  $C_*$  (Equation (10)), which is equivalent to the hidden state  $h$ , since the GRU merges the cell state and the hidden state. In this formulation, the cell state is hard-bounded and helps to store long-term temporal dependencies within the RNN.

$$r_t = \sigma(W_r \cdot [C_{t-1}, x_t] + b_r), \quad (7)$$

$$\tilde{C}_t = \tanh(W_c \cdot [r_t * C_{t-1}, x_t] + b_c), \quad (8)$$

$$z_t = \sigma(W_z \cdot [C_{t-1}, x_t] + b_z), \quad (9)$$

$$h_t = C_t = (1 - z_t) * C_{t-1} + z_t * \tilde{C}_t. \quad (10)$$

## D BELIEF MAP REPRESENTATION

When exploring an unstructured environment in which an object may be buried, there will be instances in which there is uncertainty in the fingertip contact state. Within a given state space,  $X \subset R^n$ , the probability of whether an object is buried in a particular region can be represented as a spatial distribution  $\phi_t(x)$  for all  $x_t \in X$  that may change over time. The spatial distribution is initially unknown, but is presumed to be a uniform distribution until fingertip contact state estimates are generated.

During haptic exploration within the space  $X$ , the robot was assumed to be able to measure the exact values from the spatial distribution  $\phi_t(x)$ . In other words, the fingertip contact state and fingertip location were assumed to be true when used to create and update the belief map. The  $\phi_t(x)$  distribution was updated through a probabilistic classification method that considers the fingertip contact state estimates  $y_t \in [0, 1]$  and their spatial locations  $x_t \in R^2$  along a given exploration trajectory. Specifically, we applied SVC with a RBF kernel. Spatial information was stored in an RBF kernel  $K$  that maps the influence of contact state  $y_t$  at its spatial location  $x_t$  into an infinite dimensional feature space. The spatial reach, or sphere of influence, of the radial basis function was set by  $\gamma$  in (Equation (11)).

The belief map is also capable of keeping track of the states it has previously sampled. Prior state information is encoded in an SVC score function (Equation (12)) [13], which sums learned mappings *from all previous timesteps*. Finally, the SVC scores were calibrated by Platt Scaling (Equation (13)) [64]. In our case, the spatial distribution  $\phi_t(x)$  was transformed into a probability distribution, indicating whether an object is buried at a specific location ( $P(y_t = 1|x)$ ), since the fingertip contact state  $y_t = 1$  was assumed to be true during contact with a buried object,

$$K(x, x_t) = \exp(-\gamma \|x - x_t\|^2), \quad (11)$$

$$\phi_t(x) = \sum_t \alpha_t y_t K(x, x_t) + b, \quad (12)$$

$$P(y_t = 1|x) \approx P_{A,B}(\phi_t) = \frac{1}{1 + e^{A\phi_t(x)+B}}, \quad (13)$$

The variables  $\alpha_i$  and  $b$  represent the SVC weights and bias, respectively. The Platt scaling parameters  $A$  and  $B$  are determined by solving a regularized maximum likelihood problem [44].

## E TEST OF SF-RNN MODEL PERFORMANCE

Table 5. Effects of Multimodality on SF-RNN Model Performance:  
Model Accuracy Is Reported as Mean in %

Model description	Test accuracy (%)		
SF-RNN trained with 1 modality	Trained and tested with	Accuracy	
	Vibration	95.7	
	Fluid Pressure	85.8	
	Electrode impedance	<b>97.6</b>	
SF-RNN trained with 2 modalities	Tested with	Trained with	
		Vibration, fluid pressure	Vibration, electrode impedance
	Vibration	<b>96.6</b>	86.6
	Fluid pressure	91.0	N/A
	Electrode impedance	N/A	<b>96.9</b>
	2 modalities	97.1	98.5
SF-RNN trained with 3 modalities	Tested with		Accuracy
	1 modalities	Vibration	90.0
		Fluid pressure	83.1
		Electrode impedance	<b>97.8</b>
	2 modalities	Vibration, fluid pressure	95.2
		Vibration, electrode impedance	<b>98.6</b>
		Fluid pressure, electrode impedance	98.3
	3 modalities	Vibration, fluid pressure, electrode impedance	98.7

Best performing models are shown in bold. The models were trained and evaluated with cottonwood beads.

## ACKNOWLEDGMENTS

The authors thank Chad Jenkins and Rod Grupen for discussions on occupancy maps and belief maps, Randall Hellman and Kenneth Gutierrez for insightful discussion and guidance, and Eunsuk Chong and Lionel Zhang for assistance with editing.

## REFERENCES

- [1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, and others. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation (OSDI'16)*. USENIX Association, Savannah, GA, USA, 265–283.
- [2] Ian Abraham, Ahalya Prabhakar, Mitra J. Z. Hartmann, and Todd D. Murphey. 2017. Ergodic Exploration Using Binary Sensing for Nonparametric Shape Estimation. *IEEE Robot. Autom. Lett.* 2, 2 (2017), 827–834.
- [3] Jeffrey Aguilar and Daniel I. Goldman. 2016. Robophysical study of jumping dynamics on granular media. *Nat. Phys.* 12, 3 (Mar. 2016), 278–283. <https://doi.org/10.1038/nphys3568>



- [4] Jeffrey Aguilar, Tingnan Zhang, Feifei Qian, Mark Kingsbury, Benjamin McInroe, Nicole Mazouchova, Chen Li, Ryan Maladen, Chaohui Gong, Matt Travers, Ross L. Hatton, Howie Choset, Paul B. Umbanhowar, and Daniel I. Goldman. 2016. A review on locomotion robotics: the study of movement at the intersection of robotics, soft matter and dynamical systems. *Rep. Progr. Phys.* 79, 11 (Nov. 2016), 110001. <https://doi.org/10.1088/0034-4885/79/11/110001>
- [5] Karlin Bark, Jason Wheeler, Pete Shull, Joan Savall, and Mark Cutkosky. 2010. Rotational skin stretch feedback: A wearable haptic display for motion. *IEEE Trans. Hapt.* 3, 3 (Jul. 2010), 166–176. <https://doi.org/10.1109/TOH.2010.21>
- [6] Feryal M. P. Behbahani, Guillem Singla-Buxarrais, and A. Aldo Faisal. 2016. Haptic SLAM: An ideal observer model for Bayesian inference of object shape and hand pose from contact dynamics. In *Haptics: Perception, Devices, Control, and Applications*, Fernando Bello, Hiroyuki Kajimoto, and Yon Visell (Eds.). Vol. 9774. Springer International Publishing, Cham, 146–157. [https://doi.org/10.1007/978-3-319-42321-0\\_14](https://doi.org/10.1007/978-3-319-42321-0_14)
- [7] Bedřich Beneš, Enkhtuvshin Dorjgotov, Laura Arns, and Gary Bertoline. 2006. Granular material interactive manipulation: Touching sand with haptic feedback. In *Proceedings of the International Conference in Central Europe on Computer Graphics and Visualization (WSCG'06)*. UNION Agency-Science Press, Plzen, Czech Republic, 7.
- [8] Daniel Bideau, John A. Dodds, and Centre de physique des Houches. 1991. *Physics of Granular Media*. Nova Science, Commack, NY.
- [9] J. D. Brown, A. Paek, M. Syed, M. K. O'Malley, P. A. Shewokis, J. L. Contreras-Vidal, A. J. Davis, and R. B. Gillespie. 2013. Understanding the role of haptic feedback in a teleoperated/prosthetic grasp and lift task. In *Proceedings of the 2013 World Haptics Conference (WHC'13)*. IEEE, 271–276. <https://doi.org/10.1109/WHC.2013.6548420>
- [10] Maya Cakmak and Andrea L. Thomaz. 2012. Designing robot learners that ask good questions. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'12)*. ACM Press, 17. <https://doi.org/10.1145/2157689.2157693>
- [11] Giorgio Cannata, Simone Denei, and Fulvio Mastrogiovanni. 2010. Tactile sensing: Steps to artificial somatosensory maps. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'10)*. IEEE, 576–581.
- [12] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*. Association for Computational Linguistics, Doha, Qatar, 1724–1734. <https://doi.org/10.3115/v1/D14-1179>
- [13] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Mach. Learn.* 20, 3 (Sep. 1995), 273–297. <https://doi.org/10.1007/BF00994018>
- [14] Eric I. Corwin, Heinrich M. Jaeger, and Sidney R. Nagel. 2005. Structural signature of jamming in granular media. *Nature* 435, 7045 (Jun. 2005), 1075–1078. <https://doi.org/10.1038/nature03698>
- [15] Sudeep Dasari, Frederik Ebert, Stephen Tian, Suraj Nair, Bernadette Bucher, Karl Schmeckpeper, Siddharth Singh, Sergey Levine, and Chelsea Finn. 2020. RoboNet: Large-Scale Multi-Robot Learning. In *Proceedings of the Conference on Robot Learning (Proceedings of Machine Learning Research)*, Leslie Pack Kaelbling, Danica Kragic, and Komei Sugiura (Eds.). Vol. 100. PMLR, 885–897. <http://proceedings.mlr.press/v100/dasari20a.html>
- [16] Agostino De Santis, Bruno Siciliano, Alessandro De Luca, and Antonio Bicchi. 2008. An atlas of physical human-robot interaction. *Mechan. Mach. Theory* 43, 3 (Mar. 2008), 253–270. <https://doi.org/10.1016/j.mechmachtheory.2007.03.003>
- [17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'09)*.
- [18] Siyuan Dong, Wenzhen Yuan, and Edward H. Adelson. 2017. Improved GelSight tactile sensor for measuring geometry and slip. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'17)*. 137–144. <https://doi.org/10.1109/IROS.2017.8202149>
- [19] Anca D. Dragan and Siddhartha S. Srinivasa. 2013. A policy-blending formalism for shared control. *Int. J. Robot. Res.* 32, 7 (Jun. 2013), 790–805. <https://doi.org/10.1177/0278364913490324>
- [20] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. 2010. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.* 11, (Feb. 2010), 625–660.
- [21] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *J. Mach. Learn. Res.* 9 (Aug. 2008), 1871–1874. <http://www.jmlr.org/papers/v9/fan08a.html>
- [22] Jeremy Fishel, Gary Lin, and Gerald E. Loeb. 2013. *SynTouch LLC BioTac Product Manual*, v.16.
- [23] Jeremy A. Fishel and Gerald E. Loeb. 2012. Bayesian exploration for intelligent identification of textures. *Front. Neurobot.* 6 (2012), 4. <https://doi.org/10.3389/fnbot.2012.00004>
- [24] James J. Gibson. 1962. Observations on active touch. *Psychol. Rev.* 69, 6 (1962), 477.
- [25] Matthew C. Gombolay, Reymundo A. Gutierrez, Shanell G. Clarke, Giancarlo F. Sturla, and Julie A. Shah. 2015. Decision-making authority, team efficiency and human worker satisfaction in mixed human-robot teams. *Auton. Robots* 39, 3 (Oct. 2015), 293–312. <https://doi.org/10.1007/s10514-015-9457-9>
- [26] Nick Gravish, Paul B. Umbanhowar, and Daniel I. Goldman. 2010. Force and flow transition in plowed granular media. *Phys. Rev. Lett.* 105, 12 (Sep. 2010). <https://doi.org/10.1103/PhysRevLett.105.128301>

- [27] Vitor Guizilini and Fabio Ramos. 2018. Towards real-time 3D continuous occupancy mapping using Hilbert maps. *Int. J. Robot. Res.* 37, 6 (May 2018), 566–584. <https://doi.org/10.1177/0278364918771476>
- [28] A. Gupta and M.K. O'Malley. 2006. Design of a haptic arm exoskeleton for training and rehabilitation. *IEEE/ASME Trans. Mechatr.* 11, 3 (Jun. 2006), 280–289. <https://doi.org/10.1109/TMECH.2006.875558>
- [29] James C. Gwilliam, Zachary Pezzementi, Erica Jantho, Allison M. Okamura, and Steven Hsiao. 2010. Human vs. robotic tactile sensing: Detecting lumps in soft tissue. In *Proceedings of the IEEE Haptics Symposium*. IEEE, 21–28. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5444685](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5444685).
- [30] Randall B. Hellman, Cem Tekin, Mihaela van der Schaar, and Veronica J. Santos. 2018. Functional Contour-following via Haptic Perception and Reinforcement Learning. *IEEE Trans. Hapt.* 11, 1 (Mar. 2018), 61–72. <https://doi.org/10.1109/TOH.2017.2753233>
- [31] Katharina Hertkorn, Maximo A. Roa, and Christoph Borst. 2013. Planning in-hand object manipulation with multi-fingered hands considering task constraints. In *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*. IEEE, Los Alamitos, CA, 617–624. <https://doi.org/10.1109/ICRA.2013.6630637>
- [32] M. Hollins, S. J. Bensmaïa, and E. A. Roy. 2002. Vibrotactile and texture perception. *Behav. Brain Res.* 135, 1–2 (2002), 51–56.
- [33] Won Hong. 2001. *Modeling, Estimation, and Control of Robot-Soil Interactions*. Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge, MA.
- [34] Norden Eh Huang. 2014. *Hilbert-Huang Transform and Its Applications*. Vol. 16. World Scientific.
- [35] Norden E. Huang, Ching C. Chern, Kang Huang, Liming W. Salvino, Steven R. Long, and Kuang Lung Fan. 2001. A new spectral representation of earthquake data: Hilbert spectral analysis of station TCU129, Chi-Chi, Taiwan, 21 September 1999. *Bull. Seismol. Soc. Am.* 91, 5 (2001), 1310–1338.
- [36] Heinrich M. Jaeger, Sidney R. Nagel, and Robert P. Behringer. 1996. Granular solids, liquids, and gases. *Rev. Mod. Phys.* 68, 4 (1996), 1259. <https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.68.1259>.
- [37] Ashesh Jain, Avi Singh, Hema S. Koppula, Shane Soh, and Ashutosh Saxena. 2016. Recurrent neural networks for driver activity anticipation via sensory-fusion architecture. In *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA'16)*. IEEE, 3118–3125.
- [38] Shengxin Jia and V. J. Santos. 2017. Multimodal haptic perception within granular media via recurrent neural networks. In *Proceedings of the RSS Workshop on Tactile Sensing for Manipulation: Hardware, Modeling, and Learning*.
- [39] Keehoon Kim, J. E. Colgate, J. J. Santos-Munne, A. Makhlin, and M. A. Peshkin. 2010. On the design of miniature haptic devices for upper extremity prosthetics. *IEEE/ASME Trans. Mechatr.* 15, 1 (Feb. 2010), 27–39. <https://doi.org/10.1109/TMECH.2009.2013944>
- [40] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR'15), San Diego, CA, USA, May 7-9, 2015*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1412.6980>
- [41] Ayse Kucukylmaz, Tevfik Metin Sezgin, and Cagatay Basdogan. 2013. Intention recognition for dynamic role exchange in haptic collaboration. *IEEE Trans. Hapt.* 6, 1 (2013), 11.
- [42] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng. 2007. Efficient sparse coding algorithms. In *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. C. Platt, and T. Hoffman (Eds.). MIT Press, 801–808.
- [43] Li Zhang, Siwei Lyu, and Jeff Trinkle. 2013. A dynamic Bayesian approach to real-time estimation and filtering in grasp acquisition. In *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*. IEEE, Los Alamitos, CA, 85–92. <https://doi.org/10.1109/ICRA.2013.6630560>
- [44] Hsuan-Tien Lin, Chih-Jen Lin, and Ruby C. Weng. 2007. A note on Platt's probabilistic outputs for support vector machines. *Mach. Learn.* 68, 3 (Aug. 2007), 267–276. <https://doi.org/10.1007/s10994-007-5018-6>
- [45] Jack M. Loomis, David Katz, and Lester E. Krueger. 1991. The World of Touch. *Am. J. Psychol.* 104, 1 (1991), 147. <https://doi.org/10.2307/1422857>
- [46] Dylan P. Losey, Craig G. McDonald, Edoardo Battaglia, and Marcia K. O'Malley. 2018. A review of intent detection, arbitration, and communication aspects of shared control for physical human-robot interaction. *Appl. Mech. Rev.* 70, 1 (Feb. 2018), 010804. <https://doi.org/10.1115/1.4039145>
- [47] Shan Luo, Joao Bimbo, Ravinder Dahiya, and Hongbin Liu. 2017. Robotic tactile perception of object properties: A review. *Mechatronics* 48 (2017), 54–67. <https://doi.org/10.1016/j.mechatronics.2017.11.002>
- [48] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. 2010. Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.* 11 (Jan. 2010), 19–60. <http://www.jmlr.org/papers/v11/mairal10a.html>.
- [49] T. S. Majmudar and R. P. Behringer. 2005. Contact force measurements and stress-induced anisotropy in granular materials. *Nature* 435, 1079 (Jun. 2005), 1079–1082. <https://doi.org/10.1038/nature03805>
- [50] R. D. Maladen, Y. Ding, C. Li, and D. I. Goldman. 2009. Undulatory swimming in sand: Subsurface locomotion of the sandfish lizard. *Science* 325, 5938 (Jul. 2009), 314–318. <https://doi.org/10.1126/science.1172490>

- [51] Ryan D. Maladen, Yang Ding, Paul B. Umbanhowar, Adam Kamor, and Daniel I. Goldman. 2011. Biophysically inspired development of a sand-swimming robot. Georgia Institute of Technology. Retrieved from <https://smartech.gatech.edu/handle/1853/44542>.
- [52] H. Marvi, C. Gong, N. Gravish, H. Astley, M. Travers, R. L. Hatton, J. R. Mendelson, H. Choset, D. L. Hu, and D. I. Goldman. 2014. Sidewinding with minimal slip: Snake and robot ascent of sandy slopes. *Science* 346, 6206 (Oct. 2014), 224–229. <https://doi.org/10.1126/science.1255718>
- [53] Nicole Mazouchova, Paul B. Umbanhowar, and Daniel I. Goldman. 2013. Flipper-driven terrestrial locomotion of a sea turtle-inspired robot. *Bioinspir. Biomimet.* 8, 2 (Apr. 2013), 026007. <https://doi.org/10.1088/1748-3182/8/2/026007>
- [54] Stephen P. Molner. 1999. The art of molecular dynamics simulation (Rapaport, D. C.). *J. Chem. Educ.* 76, 2 (1999), 171. <https://doi.org/10.1021/ed076p171>
- [55] Rowland O’Flaherty and Magnus Egerstedt. 2015. Optimal exploration in unknown environments. In *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS’15)*. IEEE, 5796–5801.
- [56] Corey S. O’Hern, Stephen A. Langer, Andrea J. Liu, and Sidney R. Nagel. 2001. Force Distributions near Jamming and Glass Transitions. *Phys. Rev. Lett.* 86, 1 (Jan. 2001), 111–114. <https://doi.org/10.1103/PhysRevLett.86.111>
- [57] George Y. Onoda and Eric G. Liniger. 1990. Random loose packings of uniform spheres and the dilatancy onset. *Phys. Rev. Lett.* 64, 22 (1990), 2727. <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.64.2727>.
- [58] Claudio Pacchierotti, Domenico Prattichizzo, and Katherine J. Kuchenbecker. 2016. Cutaneous feedback of fingertip deformation and vibration for palpation in robotic surgery. *IEEE Trans. Biomed. Eng.* 63, 2 (Feb. 2016), 278–287. <https://doi.org/10.1109/TBME.2015.2455932>
- [59] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. In *Conference on Neural Information Processing Systems (NIPS) Workshop on Autodiff*.
- [60] Fabian Pedregosa, Gaëll Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12 (Oct. 2011), 2825–2830.
- [61] Brian S. Peters, Priscila R. Armijo, Crystal Krause, Songita A. Choudhury, and Dmitry Oleynikov. 2018. Review of emerging surgical robotic technology. *Surg. Endosc.* 32, 4 (Apr. 2018), 1636–1655. <https://doi.org/10.1007/s00464-018-6079-2>
- [62] Z. Pezzementi, E. Plaku, C. Reyda, and G. D. Hager. 2011. Tactile-o recognition from appearance information. *IEEE Trans. Robot.* 27, 3 (Jun. 2011), 473–487. <https://doi.org/10.1109/TRO.2011.2125350>
- [63] Andrea Pigorini, Adenauer G. Casali, Silvia Casarotto, Fabio Ferrarelli, Giuseppe Baselli, Maurizio Mariotti, Marcello Massimini, and Mario Rosanova. 2011. Time-frequency spectral analysis of TMS-evoked EEG oscillations by means of Hilbert-Huang transform. *J. Neurosci. Methods* 198, 2 (Jun. 2011), 236–245. <https://doi.org/10.1016/j.jneumeth.2011.04.013>
- [64] John C. Platt. 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*. MIT Press, 61–74.
- [65] Ruben D. Ponce Wong, Randall B. Hellman, and Veronica J. Santos. 2014. Spatial asymmetry in tactile sensor skin deformation aids perception of edge orientation during haptic exploration. *IEEE Trans. Hapt.* 7, 2 (Apr. 2014), 191–202. <https://doi.org/10.1109/TOH.2013.56>
- [66] Václav Purchart. 2009. Interactive manipulation of sand on a TIN terrain model for virtual reality. In *Central European Seminar on Computer Graphics (CESCG)*. Budmerice, Slovakia, 105–111.
- [67] Feifei Qian, Tingnan Zhang, Chen Li, Pierangelo Masarati, Aaron M. Hoover, Paul Birkmeyer, Andrew Pullin, Ronald S. Fearing, and Daniel I. Goldman. 2013. Walking and running on yielding and fluidizing ground. In *Robotics: Science and Systems VIII*, Nicholas Roy, Paul Newman, and Siddhartha Srinivasa (Eds.). The MIT Press, 345–352. <https://doi.org/10.7551/mitpress/9816.003.0049>
- [68] Fabio Ramos and Lionel Ott. 2016. Hilbert maps: Scalable continuous occupancy mapping with stochastic gradient descent. *Int. J. Robot. Res.* 35, 14 (Dec. 2016), 1717–1730. <https://doi.org/10.1177/0278364916684382>
- [69] Patrick Richard, Mario Nicodemi, Renaud Delannay, Philippe Ribiere, and Daniel Bideau. 2005. Slow relaxation and compaction of granular systems. *Nat. Mater.* 4, 2 (Feb. 2005), 121–128. <https://doi.org/10.1038/nmat1300>
- [70] Benjamin A. Richardson and Katherine J. Kuchenbecker. 2019. Improving haptic adjective recognition with unsupervised feature learning. In *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA’19)*. IEEE, Los Alamitos, CA, 3804–3810. <https://doi.org/10.1109/ICRA.2019.8793544>
- [71] Edmund T. Rolls and Alessandro Treves. 1990. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Netw.: Comput. Neur. Syst.* 1, 4 (Jan. 1990), 407–421. [https://doi.org/10.1088/0954-898X\\_1\\_4\\_002](https://doi.org/10.1088/0954-898X_1_4_002)
- [72] Connor Schenck, Jonathan Thompson, Sergey Levine, and Dieter Fox. 2017. Learning robotic manipulation of granular media. In *Proceedings of the 1st Annual Conference on Robot Learning (Proceedings of Machine Learning Research)*,

- Sergey Levine, Vincent Vanhoucke, and Ken Goldberg (Eds.), Vol. 78. PMLR, 239–248. <http://proceedings.mlr.press/v78/schenck17a.html>.
- [73] Ransalu Senanayake, Anthony Tompkins, and Fabio Ramos. 2018. Automorphing kernels for nonstationarity in mapping unstructured environments. In *Proceedings of Machine Learning Research*, Aude Billard, Anca Dragan, Jan Peters, and Jun Morimoto (Eds.), Vol. 87. 443–455. <http://proceedings.mlr.press/v87/senanayake18a.html>.
  - [74] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 3104–3112.
  - [75] C. P. Tan, Y. H. Zweiri, K. Althoefer, and L. D. Seneviratne. 2005. Online Soil Parameter Estimation Scheme Based on Newton-Raphson Method for Autonomous Excavation. *IEEE/ASME Trans. Mechatron.* 10, 2 (Apr. 2005), 221–229. <https://doi.org/10.1109/TMECH.2005.844706>
  - [76] Hanns Tappeiner, Roberta Klatzky, Patrick Rowe, Jorgen Pedersen, and Ralph Hollis. 2013. Bimanual haptic teleoperation for discovering and uncovering buried objects. In *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*. IEEE, Los Alamitos, CA, 2380–2385. <https://doi.org/10.1109/ICRA.2013.6630900>
  - [77] Office of Environmental Management U.S. Department of Energy. 2018. *Research and Technology Roadmap: Robotics and Remote Systems for Nuclear Cleanup*. Technical Report. Retrieved from <http://rrsd.ans.org/wp-content/uploads/2018/06/DOE-EM-Robotics-Roadmap-March-2018-e-version.pdf>.
  - [78] Laurens Van Der Maaten. 2014. Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* 15, 1 (2014), 3221–3245. <http://www.jmlr.org/papers/volume15/vandermaaten14a/source/vandermaaten14a.pdf>.
  - [79] Amber Michelle Walker, David P. Miller, and Chen Ling. 2015. User-c design of an attitude aware controller for ground reconnaissance robots. *J. Hum.-Robot Interact.* 4, 1 (Jan. 2015), 30. <https://doi.org/10.5898/JHRI.4.1.Walker>
  - [80] Julie M. Walker, Amy A. Blank, Patricia A. Shewokis, and Marcia K. O'Malley. 2015. Tactile feedback of object slip facilitates virtual object manipulation. *IEEE Trans. Hapt.* 8, 4 (Oct. 2015), 454–466. <https://doi.org/10.1109/TOH.2015.2420096>
  - [81] Chester K. Wentworth. 1922. A Scale of Grade and Class Terms for Clastic Sediments. *J. Geol.* 30, 5 (Jul. 1922), 377–392. <https://doi.org/10.1086/622910>
  - [82] Xiaobin Xiong, Aaron D. Ames, and Daniel I. Goldman. 2017. A stability region criterion for flat-footed bipedal walking on deformable granular terrain. In *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'17)*. IEEE, 4552–4559.
  - [83] Akihiko Yamaguchi and Christopher G. Atkeson. 2016. Differential dynamic programming for graph-structured dynamical systems: Generalization of pouring behavior with different skills. In *Proceedings of the 16th IEEE-RAS International Conference on Humanoid Robots (Humanoids'16)*. IEEE, Los Alamitos, CA, 1029–1036. <https://doi.org/10.1109/HUMANOIDS.2016.7803398>
  - [84] Akihiko Yamaguchi and Christopher G. Atkeson. 2016. Stereo vision of liquid and particle flow for robot pouring. In *Proceedings of the 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids'16)*. IEEE, Los Alamitos, CA, 1173–1180. <https://doi.org/10.1109/HUMANOIDS.2016.7803419>
  - [85] Akihiko Yamaguchi, Christopher G. Atkeson, and Tsukasa Ogasawara. 2015. Pouring skills with planning and learning modeled from human demonstrations. *Int. J. Human. Robot.* 12, 03 (Sep. 2015), 1550030. <https://doi.org/10.1142/S0219843615500309>
  - [86] Guang-Zhong Yang, James Cambias, Kevin Cleary, Eric Daimler, James Drake, Pierre E. Dupont, Nobuhiko Hata, Peter Kazanzides, Sylvain Martel, Rajni V. Patel, Veronica J. Santos, and Russell H. Taylor. 2017. Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy. *Sci. Robot.* 2, 4 (Mar. 2017), eaam8638. <https://doi.org/10.1126/scirobotics.aam8638>

Received January 2020; revised February 2021; accepted February 2021