Autonomous Learning of Page Flipping Movements via Tactile Feedback

Yi Zheng[1], Filipe Fernandes Veiga[2], Jan Peters[3], Veronica J. Santos[1]

***Abstract*—Robotic manipulation is challenging when both the objects being manipulated and the tactile sensors are deformable. In this work, we addressed the interplay between the manipulation of deformable objects, tactile sensing, and model-free reinforcement learning on a real robot. We showed how a real robot can learn to manipulate a deformable, thin-shell object via feedback from deformable, multimodal tactile sensors. We addressed the learning of a page flipping task using a two-stage approach. For the first stage, we learned nominal page flipping trajectories for two page sizes by constructing a reward function that quantifies functional task performance from the perspective of tactile sensing. For the second stage, we learned adapted trajectories using tactile-driven perceptual coupling, with an intuitive assumption that, while the page flipping trajectories for different task contexts (page sizes) might differ, similar tactile feedback should be expected from functional trajectories for each context. We also investigated the quality of information encoded by two different representations of tactile sensing data: one based on the artificial apical tuft of bio-inspired tactile sensors, and another based on PCA eigenvalues. The results and effectiveness of our learning framework were demonstrated on a real 7-DOF robot arm and gripper outfitted with tactile sensors.**

***Index Terms*—Deformable object manipulation, movement primitive, perceptual coupling, real robot learning, reinforcement learning, tactile sensing**

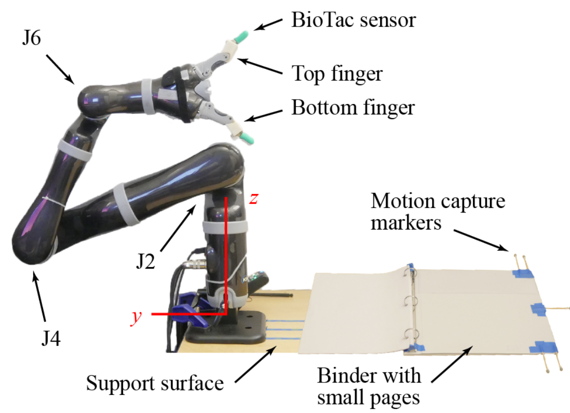


Fig. 1. Three joints (J2, J4, J6) of a 7-DOF Kinova robot arm are controlled to perform page flipping movements in the y-z plane using two fingers outfitted with deformable, multimodal BioTac tactile sensors. Large and small notebook pages (shown) were held by rigid binders placed on a flat support surface parallel to the x-y plane. Passive motion capture markers are attached to the binders for tracking displacement.

I. Introduction

Manipulation skills are important human capabilities. With these skills, humans are able to tackle a wide range of tasks requiring different levels of dexterity, using a large variety of objects, and having distinct desired outcomes. For tasks requiring dexterity, the sense of touch plays a major role in enabling the prediction of key state transitions that occur during manipulation actions. Distinct tactile patterns are associated with transitions such as the making and breaking of contact with an object, or changes in weight during the lift and replacement of a grasped object. Predicting such state transitions allows the human to detect and react to undesired events that produce deviations from the desired task states. This capability is particularly relevant when vision is occluded and the task state cannot be visually inferred [1]. When the sense of touch is taken away, complementary sensory mechanisms such as vision are often insufficient for completing manipulation tasks with the same level of functional performance. This importance of the sense of touch is demonstrated by a human perception study in which a subject, whose sense of touch at the fingertip is temporarily impaired by anesthetization takes much longer to execute a match-lighting task that seems trivial before the anesthesia [2]. Although one could argue that this demonstrates that humans can still perform manipulation tasks using complementary sensory mechanisms, it also demonstrates the importance of tactile sensing for tasks requiring dexterity.

To achieve performance with dexterity comparable to the humans, robots could be equipped with tactile sensors that provide rich information about the contact interactions between themselves and their environment [3], [4]. Once equipped with such sensors, we believe that the ability to complete complex manipulation tasks is dependent on how three learning challenges are addressed: Firstly, the robot needs to learn how to associate specific task state transitions with corresponding sensory events and use this association, or mapping, to detect undesirable states and evaluate the functional performance of the task. Second, once the ability to detect undesirable states through sensory events is acquired, the next problem is to learn which actions to perform in order to compensate for the observed sensory deviations and to return to states associated with acceptable functional task performance. Examples of such corrective actions are incremental adjustments to grasp forces after sensing slippage between the fingertips and the surface of a grasped object, or adjustments of wrist positions to make sure a scraping tool maintains sufficient contact with a tilted surface [5]. Finally, once the previous two learning challenges have been overcome, it is crucial to achieve generalization of

[1]Yi Zheng and Veronica J. Santos are with the Mechanical and Aerospace Engineering Department, University of California, Los Angeles, CA 90095. E-mail: yizheng1992@engineering.ucla.edu, vjsantos@ucla.edu

[2]Filipe Fernandes Veiga is with the Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, USA. E-mail: fveiga@mit.edu

[3]Jan Peters is with the Technische Universität Darmstadt, Darmstadt, Germany, FG Intelligent Autonomous Systems and also with the Max-Planck-Institut für Intelligente Systeme, Tübingen, Germany. E-mail: jan.peters@tuebingen.mpg.de

Manuscript 3682 submitted to 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Received May 21, 2022.

these capabilities to novel contexts or scenarios [6], [7].

In the past, a major bottleneck for solving the three aforementioned learning challenges lay in the limitations of tactile sensing hardware technology. Traditional tactile sensors composed of pressure sensing arrays have low deformation capabilities [8], which may not encode sufficient information for tasks that require detailed knowledge of the finger-surface interactions. In recent years, efforts have been made to develop deformable tactile sensors that can provide multi-modal and high-resolution spatial information, such as the BioTac [9], GelSight [10] and TacTip [11]. Although several successful applications of these sensors exist for tasks such as slip detection [12], [13] or object property classification [14], there have been few examples of their application to more general forms of object manipulation [15], [16], [17], [18]. When the manipulated object is highly deformable (e.g., thin-shell objects such as paper), the application of tactile sensing seems to be even more rarer.

Currently, there are two key challenges that limit the application of state-of-the-art tactile sensors to dexterous object manipulation problems: (i) the dynamics of the interactions between the sensors and the manipulated objects are nontrivial to model, especially for cases where both the sensor and the object are deformable, and (ii) in most cases, it is difficult to quantify the overall functional performance of a manipulation task solely based on tactile information that is inherently localized to finger-object interactions.

To overcome the aforementioned challenges, we propose an approach that begins by deploying a model-free reinforcement learning process seeded via human demonstrations that is then guided by a tactile-based reward function in order to learn a nominal movement trajectory for a specific task context. Using model-free reinforcement learning, we do not require explicit models of the system components (e.g., the tactile sensors, the manipulated objects, and the interactions between them) and are able to extract the necessary information solely from the tactile and proprioceptive data acquired during the process, as shown in other motor skill acquisition tasks using real robots [19], [20]. Once the nominal movement trajectory has been learned, the corresponding sensor readings are considered as the nominal sensing traces. In order to generalize the learned movements to different task contexts, we operate under the assumption that, for tasks where the trajectories required for each context are different, the resultant nominal sensing traces associated with the functional behavior should still be similar. Under this assumption, we can use the differences between sensing traces in order to adapt the movement trajectory to a different context. Such an adaptation is achieved by a separate reinforcement learning process, where adjustments to the nominal trajectory are learned using the differences between the nominal sensing traces and the actual sensor traces, acquired during execution of the learned movement trajectory in a new context. We use our approach to tackle a notebook page flipping task, where both the tactile sensors and manipulated objects are highly deformable. Different contexts result from using different notebook page sizes that require different movement trajectories to flip pages at an acceptable performance level.

Our work contributes to the development of new reinforcement learning approaches for the manipulation of deformable objects while explicitly leveraging state information encoded in tactile sensor data. More specifically, the contributions of this paper are the following: (i) we show that a nominal trajectory with functional behavior can be learned using model-free reinforcement learning and a tactile-based reward function, (ii) we achieve the adaptation of such functional behaviors to a novel context by relying solely on the differences between tactile sensing traces generated by a nominal trajectory and those generated for a novel context, and (iii) our learning approach demonstrates the manipulation of highly deformable thin-shell objects with a real robot.

Section II outlines related work. Section III details the manipulation task and our general methodology. Section IV provides a description of the hardware used in our experiments as well as a discussion of our experimental results. Section V summarizes contributions and limitations of this work, and suggests directions for future work.

II. Related Work

The three topics that interplay in this paper are object deformability, modeling of tactile sensors, and object manipulation in robotics. In this section, we provide brief introductions to related work on each of these topics.

Robotic manipulation of objects (rigid or deformable) has been an active area of research for quite some time. Regarding the manipulation of rigid object, several efforts have focused on grasping or contour-following tasks, with approaches that rely on tactile sensing to enable the detection of salient discrete events (e.g. slip detection, stability estimation, force thresholding) [21], [22], [15], or utilizing tactile information as continuous feedback signals that drive corrective actions [16], [17].

While the manipulation of rigid objects has been extensively investigated, the same cannot be said for the manipulation of deformable objects, especially when considering approaches that leverage tactile sensing. Currently, state-of-the-art approaches to the manipulation of deformable objects rely predominantly on visual sensing [23], with tactile sensing mainly being explored for the classification of object properties [24], [25] or shape estimation [26], [27]. Such limited application of tactile sensing to the manipulation of deformable objects is not surprising, considering that it is extremely difficult to accurately model deformable objects, deformable tactile sensors, and their complex interactions during contact.

Nonetheless, some efforts have been made to model deformable linear objects (DLOs), [28] such as ropes and cables [29], [30], as well as thin-shell objects, such as paper [31] and garments [32]. Still, all of these approaches either require complex models of internal physical states of deformable objects that are difficult to deploy in real robot experiments, or require sufficiently accurate complementary sensing mechanisms (e.g., fixed visual tracking markers) instead of tactile sensing.

A recent work by She, et al. uses tactile sensing to manipulate a deformable object. A deformable, computer vision-based tactile sensor (GelSight [10]) is applied to a cable following task [33]. The authors used a model-based approach for manipulation because the tactile images from the sensor enabled the state of the cable to be continuously observed throughout task execution. In this work, we sought to manipulate a deformable thin-shell object (notebook page). We employed a deformable tactile sensor (BioTac [9]) that does not enable a direct observation of system state for this particular manipulation task, and so we elected to use a model-free reinforcement learning approach.

Typically, endowing a robot with predictive tactile sensing capabilities has been approached as a forward modeling problem, which is nontrivial for deformable tactile sensors that engender soft contact (e.g. BioTac [9], GelSight [10], TacTip [11]). Successful predictive approaches include building latent space dynamics models for the BioTac sensor using deep representation learning to enable object surface servoing [34], and training deep recurrent neural networks to predict sequences of future GelSight tactile images from the current tactile image and applied control actions for the implementation of a model predictive control framework [18]. Note that in both of these examples, the objects being manipulated were rigid and the manipulative actions could be accurately observed and evaluated based on the tactile sensor information.

Recent manipulation approaches attempt to capitalize on the recent successes of reinforcement learning. Some approaches leverage simulation in order to pre-learn policies that are then transferred to real robotic systems [35], [36]. Other approaches begin directly with real robotic systems either by first learning how to evaluate the quality of their actions and using the quality assessments to guide the learning [37], or by focusing on a single manipulation action and only considering one object [38].

In order to deploy reinforcement learning on real robots and tackle more complex manipulation tasks, we elected to use policy representations with a limited number of parameters to encode the movement of the robot. Several policy representations have been proposed, including deterministic representations such as dynamic movement primitives (DMPs) [39] and probabilistic representations such as probabilistic movement primitives [40] or Gaussian mixture regression (GMR) [41]. In this work, we use the DMP framework mainly due to its successful application to motor skill learning problems with real robots [20].

III. Learning to Manipulate a Thin-shell Object via Tactile Sensor Feedback

We partition the learning challenges into two sub-problems in order to show, first, that a robot can learn a page flipping task using quantitative performance measures based on tactile sensing, and second, that deviations from expected tactile sensor feedback can be used to adapt nominal actions to different contexts. First, we learn nominal trajectories leading to the functional behavior of page flipping. To learn these nominal trajectories efficiently, we bootstrap a model-free reinforcement learning process seeded by human demonstrations via kinesthetic teaching. The reinforcement learning process is guided by a reward function based on tactile signals and motion tracking data. The tactile signals provide information about the contact state between the fingertips and grasped notebook pages. The motion capture data tracks the movement of the notebook for the evaluation of task performance.

Second, after learning a nominal page flipping trajectory, we learn an additional tactile-based feedback term that adapts the nominal trajectory to a different-sized notebook (a different context). The additional feedback term is denoted as a perceptual coupling term [42] and is in fact a separate correction policy. As previously mentioned, the correction policy is learned based on the assumption that, while the nominal movement trajectories for different page sizes might differ, the sensing traces corresponding to functional behaviors should remain similar. Hence, the correction policy for a novel page size should adapt the movement trajectory such that it reproduces the nominal sensing traces corresponding to the functional page flipping behavior generated by learning with the nominal page size. While this correction policy is also learned via a model-free reinforcement learning process, the reward function that guides this process is now purely based on tactile information.

In this section, we provide a brief introduction to the DMPs policy representation (Section III-A) and the reinforcement learning algorithm used for both learning sub-problems (Section III-B). We then describe a qualitative study aimed at establishing the relevance of tactile information to the page flipping task (Section III-C). Leveraging insights from the qualitative study, we describe how we use tactile and marker tracking information to learn the nominal movement trajectories (Section III-D) and how we use tactile sensor feedback exclusively to adapt the nominal trajectory to a different page size (Section III-E). Finally, we describe several alternative representations for tactile information that ensure that maximal tactile information is provided to the reinforcement learning process in a computationally efficient (e.g. low-dimensional) manner (Section III-F).

A. Dynamic Movement Primitives

In order to learn a nominal movement trajectory, the parameters of a trajectory representation are adjusted to reproduce a demonstrated trajectory and fine-tuned by a reinforcement learning algorithm. In this work, we choose the Dynamic Movement Primitives (DMPs) as the parametric representation of a trajectory [39]. A DMP typically consists of (i) a transformation system (trajectory generator), (ii) a phase system, and (iii) a nonlinear forcing function. We also include a gating system to scale the magnitude of the forcing term [43]. We choose a transformation system for discrete movements [44] [43]

$$\begin{bmatrix} \dot{z} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} (\alpha_y(\beta_y(y_g - y) - z) + v f(s))/\tau \\ z/\tau \end{bmatrix}, \quad (1)$$

where y, y_g are the actual position and goal position of a robot movement respectively, α_y is a spring constant, β_y is a damping constant, τ is a temporal scaling factor of the movement duration, and $f(s)$ is a non-linear forcing function of a phase variable s that determines the shape of the robot trajectory.

The forcing function is defined as

$$f(s) = \frac{\sum_{i=1}^{N} \Phi_i(s)\omega_i}{\sum_{i=1}^{N} \Phi_i(s)} s(y_g - y_0), \tag{2}$$

with Gaussian kernel

$$\Phi_i(s) = \exp\left((s - c_i)^2/h_i\right). \tag{3}$$

where c_i and h_i represent the center and width of the Gaussian kernel, respectively. The forcing function is scaled by the difference between the start y_0 and goal position y_g, and by a gating variable. The gating variable v evolves as a sigmoid system [43] scaled by a time constant α_v,

$$\dot{v} = -\alpha_v v(1 - v/v_{\max}). \tag{4}$$

In addition, the forcing function depends on the phase variable s instead of explicitly depending on time. The phase variable evolves as a constant decaying system [43]

$$\dot{s} = -1/\tau. \tag{5}$$

Note that each degree of freedom for the robot has its own transformation system and forcing function. The synchronization of the multiple degrees of freedom is achieved via a shared phase variable.

To encode a demonstration movement y_{demo} as a DMP, the weights $\{\omega\}_i$ associated with the forcing function need to be adapted such that the generated robot movement matches the recorded human demonstration used to seed the learning process. The initial fitting of the DMP weights is achieved by solving the linear regression problem

$$\{\omega\}_i = \operatorname{argmin}_{\{\omega\}_i} \sum_s (f_{\text{target}}(s) - f(s)) \tag{6}$$

where f_{target} is the target forcing function (human demonstration) and is computed by integrating the transformation system (Equation 2) using variables extracted from the demonstration.

B. Model-based Relative Entropy Policy Search

We rely on reinforcement learning to fine-tune the initial movement trajectory that matches the demonstration and then to learn the correction policy to adapt the nominal movement trajectory. Specifically, we use an information-theoretic policy search approach: Model-based Relative Entropy Policy Search (MORE) [45]. By bounding the KL-divergence of two subsequent policy search distributions $KL(\pi(\theta)||q(\theta))$ and the co-variance matrix shrinkage of $\pi(\theta)$, MORE achieves an effective trade-off of exploration and exploitation.

For MORE, the learning problem of maximizing the reward function under the expectation of generated trajectory samples can be formulated as

$$\max_{\pi} J(\pi) = \int R(\theta)\pi(\theta)d\theta \quad \text{s.t.} \tag{7}$$

$$\int \pi(\theta) \log \frac{\pi(\theta)}{q(\theta)} d\theta \leqslant \varepsilon \tag{8}$$

$$-\int \pi(\theta) \log \pi(\theta) d\theta \leqslant \beta \tag{9}$$

$$\int \pi(\theta) d\theta = 1. \tag{10}$$

with the updated policy $\pi(\theta)$, KL-divergnce constraints and entropy bound constraints.

With an additional constraint that $\pi(\theta)$ is a proper probability distribution, the Lagrangian dual for this constraint optimization problem can be obtained in closed form and yields the following solution

$$\pi(\theta) \propto q(\theta)^{\left(\frac{\eta}{\eta+\gamma}\right)} e^{\left(\frac{R(\theta)}{\eta+\gamma}\right)}. \tag{11}$$

The new policy $\pi(\theta)$ is a geometric average of the current policy $q(\theta)$ and an exponential transformation of the reward function. The Lagrangian dual variables η and γ serve as "temperature" parameters that weight each sample drawn using the current policy.

MORE fits a quadratic surrogate model to reward function samples $R_\theta \approx \theta^T R\theta + \theta^T r + r_0$ and assumes that the current policy search distribution is Gaussian $q(\theta) = N(\theta|\mu, \Sigma)$. The new policy search distribution can be obtained in closed form as

$$\pi(\theta) = N(\theta|Ff, F(\eta + \gamma)) \tag{12}$$

where

$$F = (\eta\Sigma^{-1} - 2R)^{-1} \tag{13}$$

$$f = \eta\Sigma^{-1}\mu + r. \tag{14}$$

In practice, η needs to be restricted such that F is positive definite.

C. Relevance of Tactile Information to a Page Flipping Task

To gain insights into which page flipping behaviors are detectable via tactile sensing, multiple sets of DMPs parameters are fitted using different human demonstrations via kinesthetic teaching. The demonstrated page flipping movements can be categorized into three groups (Figure 2):

1) *Semi-circular Trajectories:* In these demonstrations, the robot flips the pages with a relatively semi-circular movement, where the radius of the semi-circle is approximately equal to the width of the page.
2) *Warping Trajectories:* In these demonstrations, the trajectory is either mostly horizontal and parallel to the binder's support surface, or it will begin along a semi-circular path and then move downward toward the support surface prematurely, prior to the page being fully flipped. These trajectories cause the page to warp, leading to "page snapping" as the curvature of the page abruptly changes.
3) *Aggressive Trajectories:* In the initial period of these trajectories, the robot pulls the page excessively. The unnecessarily large movement can undesirably slide the binder along the support surface or lift the binder off the support surface.

We use a deformable, multimodal tactile sensor called the BioTac (SynTouch, Inc., Montrose, CA, USA) to record tactile data ((Figure 4). Each BioTac measures low frequency pressure (P_{dc}), high frequency pressure (P_{ac}), data from 19 impedance electrodes (E), internal temperature (T_{dc}), and temperature flux (T_{ac}). All tactile sensing channels are provided at 100 Hz except for the high frequency pressure data, which

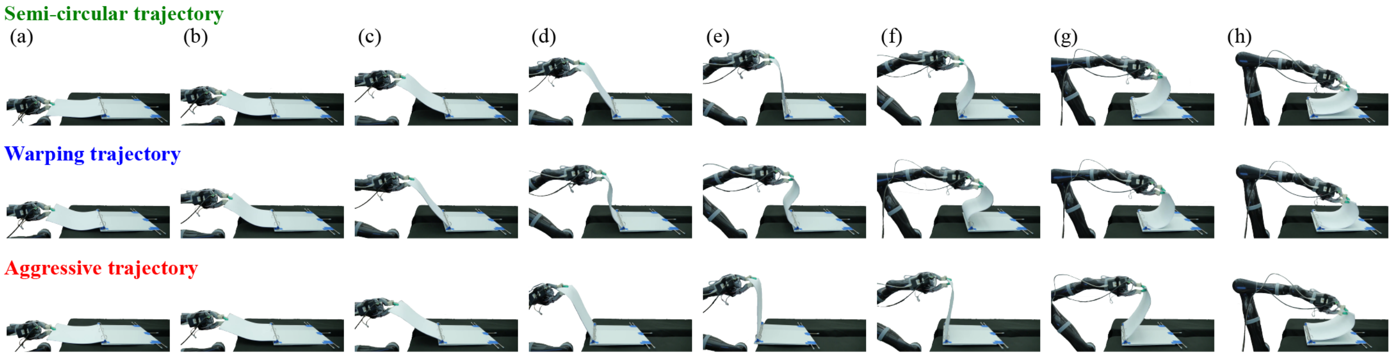


Fig. 2. Representative snapshots of the three categories of page flipping trajectories described in Section III-C are shown in 15 sec increments and the supplemental video. *(Top)* The Semi-circular trajectory represents the ideal page flipping movement. *(Middle)* For the Warping trajectory, page warping is especially pronounced in subfigure *(f)* and page snapping results in the page configuration shown in subfigure *(g)*. *(Bottom)* For the Aggressive trajectory, the binder is pulled aggressively from subfigures *(d)* through *(g)*, resulting in large and numerous displacements of the binder in the y-z plane defined in Figure 1.

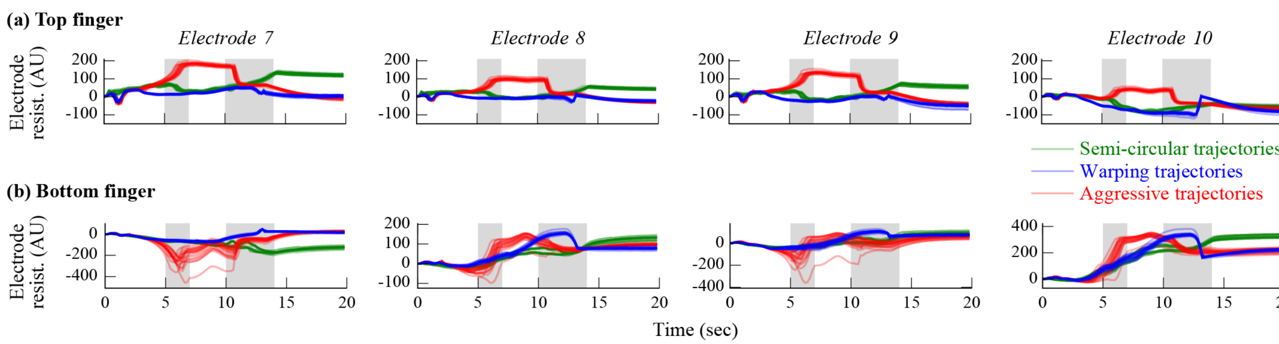


Fig. 3. Representative tactile sensing traces from the artificial apical tuft (electrodes 7, 8, 9, 10 in Fig. 4) are shown in arbitrary units (AU) for the *(a)* top and *(b)* bottom fingers during the flipping of small pages when using the three categories of trajectories (15 rollouts each) described in Section III-C. The semi-circular trajectories (green) generate smoother tactile signals than both warping trajectories (blue) and aggressive trajectories (red). As indicated by the periods shaded in gray, spikes in the tactile signals occur near the end of the page flipping movement for the warping trajectories and near the beginning and end of the movement for the aggressive trajectories.

are provided at 2200 Hz. For each of two fingers, 44 tactile signals are sampled at 100 Hz. In this work, we use the low frequency pressure (P_{dc}) and impedance electrodes (E) only.

After executing the DMPs for each trajectory category, several repeatable patterns can be observed in the tactile sensor data recorded by the deformable, multimodal BioTac sensor [9] used in the experimental evaluation (Figure 3). For semi-circular trajectories, the tactile signals captured during the page flipping movement are relatively smooth, with very few, if any, movements of the binder during the execution of the trajectory. For the warping trajectories, the page snapping causes spikes in the low frequency pressure signal (P_{dc}) and in the electrode voltages (E) provided by the BioTac sensors. As the horizontal gripper trajectory moves closer to the binder's support surface, the more severe the page warping and snapping, making the spikes in the tactile signals more pronounced. For aggressive trajectories, signal spikes are observed when the binder is pulled toward the robot and hits the border of the support surface, when the binder is lifted from and returned to the support surface, and when the gripper moves to flip the page. In addition, aggressive pulling of the notebook pages can tear the page, effectively damaging the notebook.

Based on the above observations, it is clear that undesirable events can be detected as large shifts or transient spikes in several of the tactile sensor data streams. The signal spikes can be interpreted as contact state instabilities during the page flipping movements caused by sub-optimal trajectories. By design, optimal trajectories will attempt to minimize abrupt changes in the tactile signals in order to maintain stable contact throughout the page flipping movement. In order to track undesired gross movement of the binder, which cannot be fully characterized by tactile signals, passive motion capture markers are attached to the binder. Optimal trajectories will also attempt to minimize the movement of the markers, thus minimizing the pulling of the binder.

D. Learning a Nominal Trajectory from Tactile Feedback

For the first learning sub-problem, we learn a functional movement for flipping pages of a notebook by relying on tactile and marker tracking information. We begin by fitting the parameters of a DMP to a demonstration of a warping trajectory, as this category of trajectories exhibited the undesirable behavior during the pilot study presented in Section III-C. We then use MORE to further optimize the trajectory for improved functional performance. The policy distribution parameter θ in the problem formulation specified in Section III-B corresponds to the DMP weights $\{\omega\}_i$.

We design a novel reward function that simultaneously enforces that the contact areas between the fingers and notebook pages remain stable throughout the trajectory, and that the movement of the binder is minimized. Consider $i \in [1, 2, \ldots, n]$ BioTac signal instances recorded during the page flipping trajectory with a sampling frequency of ϕ, and with low frequency pressure channels denoted as P_{dc}, electrode voltages denoted as E, and total displacement distances of the binder markers denoted as D_{markers}, the reward function is defined as

$$R(\tau) = R_{P_{\mathrm{dc}}} + R_E + R_{\mathrm{markers}} \tag{15}$$

where

$$R_{P_{\mathrm{dc}}} = -\alpha(\Gamma_1 + \Gamma_2)^2 \tag{16}$$

penalizes trajectories with large shifts in the P_{dc} channels by considering the maximum shift captured during the trajectory for each of the d fingers, where

$$\Gamma_d = \max_i \left(\left| \frac{{}^{d}P_{\mathrm{dc}}^{i+1} - {}^{d}P_{\mathrm{dc}}^{i}}{\phi} \right| \right), i \in [1, 2, \ldots, n]. \tag{17}$$

In a similar fashion,

$$R_E = -\sigma(\Lambda_1 + \Lambda_2)^2 \tag{18}$$

penalizes large shifts in the electrode values by considering the maximum average shift captured across the 19 electrodes for each of the d fingers, where

$$\Lambda_d = \max_i \left(\frac{1}{19\phi} \sum_{j=1}^{19} |E_j^{i+1} - E_j^{i}| \right), i \in [1, 2, \ldots, n]. \tag{19}$$

Finally, marker movement is also penalized via

$$R_{\mathrm{markers}} = -\lambda D_{\mathrm{markers}}. \tag{20}$$

Note that the pressure, electrode, and marker movement reward terms are scaled by $-\alpha$, $-\sigma$ and $-\lambda$ respectively. The overall reward $R(\tau)$ depends quadratically on the tactile signals of each individual finger and linearly on the displacement distance of the binder.

The reward function defines the task accomplishment via salient tactile features throughout the trajectory. For the page flipping task in this work, a salient tactile signal indicating overall task success/failure at the end of the trajectory is not pronounced. However, there may be other tasks, such as the closure of a ziplock bag, where a salient tactile signal, such as a "click" upon bag closure might exist. For those cases, one could add an additional term to the reward function that acknowledges overall task accomplishment.

E. Adapting Learned Nominal Trajectories to a Novel Context

For the second learning sub-problem, we show that nominal page flipping trajectories learned for the first sub-problem can be adapted to different page sizes while relying solely on tactile information. In contrast to the first learning sub-problem that considered binder displacement in the reward function, the second learning sub-problem does not use any visual feedback related to binder displacement. Specifically, we use the tactile sensing traces produced by executing the optimal page flipping trajectory learned in Section III-D for a specific page size as the nominal tactile sensing traces.

We still require that pages of different sizes be flipped with semi-circular trajectories for functional behavior. As such, we propose that, while the movement trajectory needs to adapt to different page sizes, the tactile sensing traces should remain constant. The robot then learns how to adapt the nominal trajectory to a different page size by trying to match the new tactile sensing traces to the nominal sensing traces. In this manner, we extend learning based on a single demonstration to a different task context.

In order to adapt the nominal trajectories, the nominal DMP needs to be modified to adapt to step-based tactile signals. In this paper, we leverage "Perceptual Coupling Dynamic Movement Primitives" [42], also known as "Associative Skill Memories" [46]. After defining the nominal signal trace instance $S_{nom}(s)$ and the current sensing trace instance $S_{cur}(s)$, the adaptation actions are decided based on the difference between the nominal and current signal traces (i.e., the perceptual coupling term) during the execution of the current page flipping trajectory on pages with a different size. Since concurrent reactions to sensing trace differences depends on which sensing channel diverged from the nominal sensing trace, it is necessary to maintain separate weights for each sensing channel in the adaptation policy.

Therefore, we model the adaptation policy as a mixture of Gaussians that takes the sensing trace differences as inputs and adds the adaptation policy to the nonlinear forcing function of the nominal DMP as shown

$$\hat{f}(s) = f(s) + \sum_{j=1}^{m} \sum_{k=1}^{n} \hat{\omega}_{jk} e^{\frac{(s-c_k)^2}{h_k}} (S_{\mathrm{nom}}^{j}(s) - S_{\mathrm{cur}}^{j}(s))s. \tag{21}$$

Here, m represents the total number of sensing channels used as tactile feedback, n represents the total number of basis functions for each tactile channel, and $\{\hat{\omega}\}_{jk}$ represents the learnable weights of the adaptation policy for a single degree of freedom of the robot arm, which are also trained using the MORE algorithm. Since this task involves three robot joints (described further in Section IV-A), the total dimensionality of the learning problem is $m \times n \times 3$. For simplicity and better synchronization of the nominal DMP and the perceptual coupling element, the centers $\{c\}_k$ and widths $\{h\}_k$ of the Gaussian kernels are set to be identical across all three robot joints. Considering the above, a natural reward function to learn the adaptation policy is the sum of squared differences between the nominal and the current sensing traces over the course of an entire page flipping trajectory. In other words, we set the reward function equal to the square of the perceptual coupling term.

$$R(\tau) = -[S_{\mathrm{nom}}(\tau(s)) - S_{\mathrm{cur}}(\tau(s))]^2 \tag{22}$$

F. Representation of the Tactile Sensing Traces

Our choice of representation of the tactile sensing traces requires a careful balance between richness of tactile information for effective learning and computational tractability for deployment on a real robot. In one extreme case, we

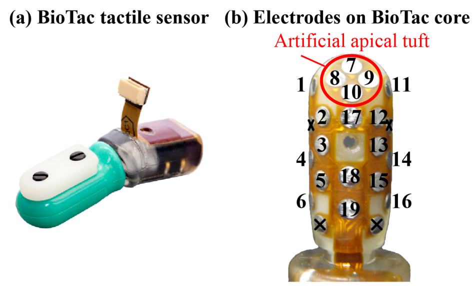


Fig. 4. *(a)* The multimodal BioTac sensor is comprised of a rigid core, elastomeric skin, and fingernail (Image from [47]). *(b)* The rigid core of the BioTac is shown with impedance electrodes individually numbered. The red ellipse highlights the artificial apical tuft (flat region of the distal phalanx), where contact is made with the page in the majority of cases.

could naively use all sensing channels from both BioTacs on the gripper, resulting in 88 total tactile sensing channels (1 P_{dc} channel, 19 E channels, 2 channels associated with temperature, and 22 high frequency pressure values.

Hypothesizing that some of the native BioTac sensing channels would not be necessary for learning the page flipping task, we consider two possible representations for the sensing traces. The first representation that we consider is comprised of a subset of the complete set of BioTac channels. Upon inspection of the page flipping trajectories, we observed that electrodes 7, 8, 9, and 10, located on the artificial apical tuft (flat surface of the distal phalanx) of the BioTac (Figure 4b), are stimulated most strongly and most often during contact with the page in the majority of trials. By focusing only on the P_{dc}, E_7, E_8, E_9, and E_{10} values from each finger, we reduce the dimensionality of the representation of tactile sensing traces from 88 to 10 values.

To further reduce complexity, we average signals over the four electrodes for each individual finger, which results in a total of 4 values (one P_{dc} value and one mean electrode value $\overline{E}_{7-10}$ per finger) sampled at 100 Hz. We believe that averaging over the four apical tuft electrodes is reasonable since we are primarily interested in the average skin deformation of that specific area of the sensor. To appropriately scale P_{dc} and $\overline{E}_{7-10}$, we normalize the data on each signal individually using sensing traces collected from the nominal trajectory learning experiments and assume that they provide a reasonable range for sensing traces that the robot can experience during page flipping.

The second representation that we consider uses Principal Component Analysis (PCA) to reduce the complete set of BioTac channels to a subset that captures most of the variance in the tactile sensing traces that can be leveraged for learning. Again, we begin by focusing only on the P_{dc}, E_7, E_8, E_9, and E_{10} values from each finger. Upon normalizing the signals as described previously, we pool the tactile signals across both fingers and apply PCA to further reduce the representation of tactile sensing traces from 10 dimensions to 3 dimensions.

Note that different sensing channels on different fingers (2 PDC + 8 electrodes) form the state vector. Prior to performing PCA, the data from each signal channel are individually normalized. This normalization is performed to ensure that the PCA results will not be biased by large magnitude changes resulting from differences in measurement units, measurement ranges, or channel sensitivity.

The first three principal components explain 96% of the total variance in the original 10-D tactile sensing space. The 1st, 2nd, and 3rd principal components explain 45.8%, 38.7%, and 11.5% of the total variance, respectively. We obtain the PCA projection matrix once at the start of the experiment. During runtime, the 10-D normalized BioTac signals are passed through the PCA projection matrix comprised of the first three principal components in order to yield three PCA eigenvalues for learning.

IV. Experimental Procedure and Evaluation

In this section, we present the experimental procedures that were used to evaluate our approach and discuss the results of those experiments. First, we describe the hardware setup used in our experiments in Section IV-A. We then present the training procedure and results for the nominal trajectory learning sub-problem in Section IV-B. In Section IV-C, we present the results of a simplified version of the trajectory adaptation learning sub-problem and assess the impact of the choice of the representation of the tactile sensing traces introduced in Section III-F. In this simplified version of the trajectory adaptation learning sub-problem, some partial knowledge of the novel task context (novel page size) is provided in order to make the adaptation policy learning problem more tractable while we focused on the assesment choice of tactile sensing trace representation. Finally, in Section IV-D, we present the results of the complete trajectory adaptation learning sub-problem, without the benefit of a priori knowledge of novel page size. For this final, complex learning experiment, we used the PCA eigenvalue representation of the tactile signals (Section III-F) generated from the nominal trajectory (Section III-D) learned for large pages.

A. Experimental Set-up

For all experiments, we used a 7 degree-of-freedom (DOF) robot arm (JACO, Kinova, Boisbriand, Quebec, Canada) outfitted with a 4-DOF, three-digit gripper (KG-3, Kinova, Boisbriand, Quebec, Canada) (Figure 1). The ulnar digit was removed from the gripper in order to enable a two-digit precision grip. Each fingertip was equipped with a BioTac tactile sensor, as introduced in Section III-C.

The robot was commanded to grasp and flip two different sizes of notebook pages (small page: 8.5" x 11", large page: 11" x 11"). Retroreflective markers and six T-Series cameras sampled at 100 Hz (Vicon, Culver City, CA, USA) were used to track a rigid binder, containing the notebook pages, that was placed on a support surface parallel to the x-y plane (Figure 1). The binder displacement values in the y-z plane were used by the reward function described in Section III-D for learning a nominal trajectory for the page flipping task.

As shown in Figure 2, the page flipping movement occurs within the y-z plane defined in Figure 1. Through purposeful placement of the robot arm with respect to the binder, we simplify the policy learning problem. Specifically, we operate the robot arm within the y-z plane only. We control only

Manuscript 3682 submitted to 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Received May 21, 2022.

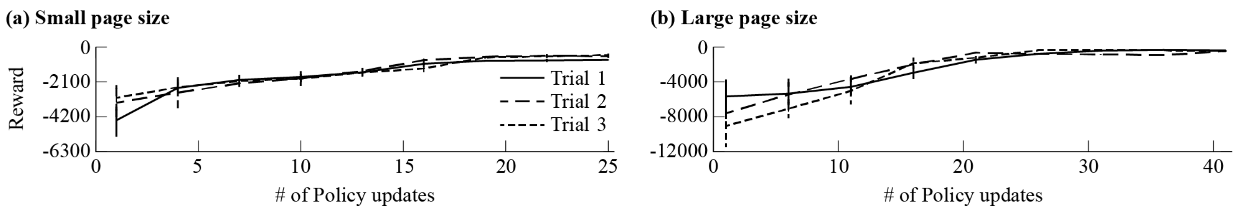


Fig. 5. Learning curves are shown for learning nominal trajectories for flipping *(a)* small pages over the course of 25 policy updates and *(b)* large pages over the course of 41 policy updates. Mean and variance are presented for batches of 10 rollouts.

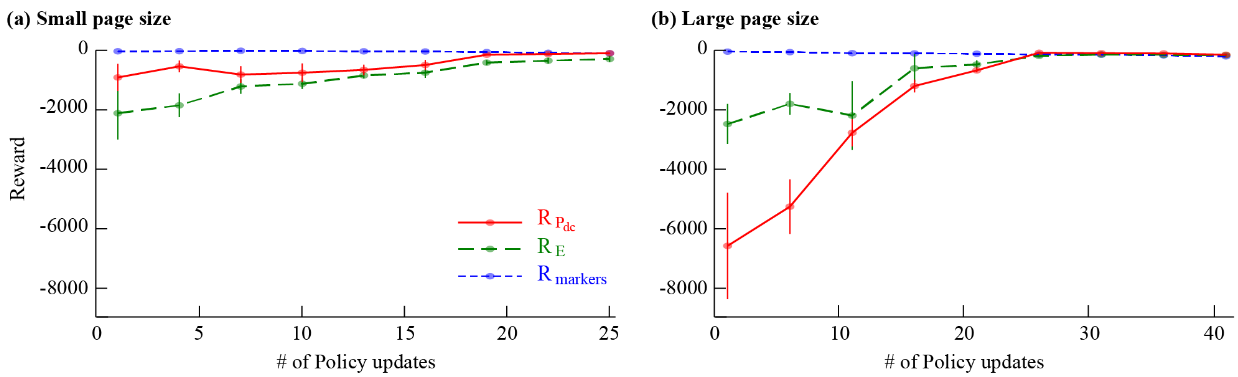


Fig. 6. Individual reward components from Equation 15 are shown for the nominal trajectory learning curves in Figure 5 for flipping *(a)* small pages over the course of 25 policy updates and *(b)* large pages over the course of 41 policy updates. Mean and variance are presented for batches of 10 rollouts. This figure illustrates that the improvement of tactile-related reward components plays a major role in the improvement of the overall reward function as compared to any losses resulting from binder displacement.

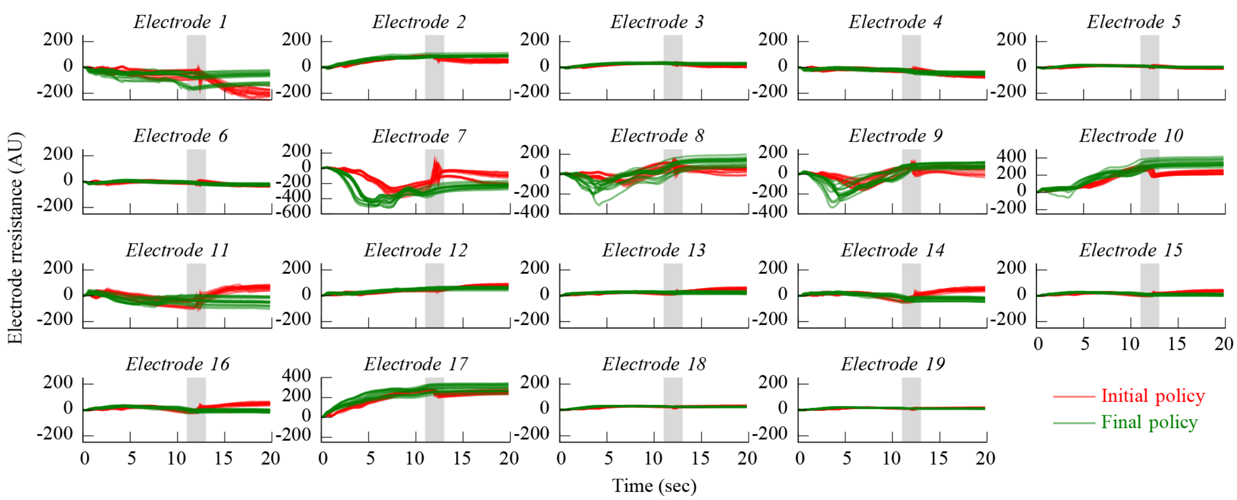


Fig. 7. Data for each of 19 electrodes are shown for the bottom finger for learning a nominal trajectory for flipping a large page. A total of 15 rollouts are shown for each the initial policy after a single policy update (red) and final policy after 41 policy updates (green). The initial policy generates undesired spikes in the electrode signals during the page warping period shaded in gray.

joints 2, 4, and 6 (Figure 1) and constrain all remaining joints, thereby reducing the dimensionality of the policy weights to be tuned during learning.

B. Learning a Nominal Trajectory from Tactile Feedback

Using the proposed framework introduced in Section III-A and Section III-B, along with the reward functions defined from Equation 15 to Equation 20 in Section III-D, we learn nominal trajectories for flipping pages of two different sizes. In each learning trial, the robot first moves to a "home" position and grips a pre-set stack of 20 pages.The home position is determined from a single human demonstration per page size at the start of the experimental session. An experimenter kinesthetically teaches the robot by grasping the robot and guiding it through a suboptimal Warping trajectory, as described in Section III-C. Throughout the kinesthetic teaching, joint angles and joint angular velocities are recorded at 50 Hz. The kinematic data from the human demonstration are used to initialize ten parameters for each of three DMPs (one DMP for each of joints 2, 4, and 6 of the robot arm).

Upon initializing the DMP parameters, a reinforcement learning process, introduced in Section III-B, refines the DMP parameters using the reward function defined in Section III-D. The MORE policy search space has a dimensionality of 30 (three DMPs, each with ten parameters). The MORE ε and β parameters are set to 0.1 and 0.075, respectively. To fill a sample buffer, a total of 40 rollouts are executed and corresponding rewards are generated based on the initial policy distribution, which was defined as a multivariate Gaussian with mean values set equal to the initial DMP weights and a diagonal covariance matrix that was tuned based on preliminary data. Once the sample buffer is filled, the policy distribution is updated every five rollouts.

The learning curves for the nominal trajectory learning experiments are shown in Figure 5 for the small and large pages. Rewards from the updated policies are shown in increments of 3 policy updates (10 rollouts each) for three independent learning trials for the small pages, and in increments of 5 policy updates (10 rollouts each) for three independent learning trials for the large pages.

Since the reward values converge to zero, we see that a policy is successfully refined and learned for each of the two page sizes. The supplemental video shows that the robot learns page flipping trajectories that do not induce undesired page warping or page snapping, thereby avoiding spikes in the tactile signals by design of the reward function.

For brevity, we present tactile sensing traces for the nominal trajectory learning sub-problem for large pages only. Reflecting local deformation of the fluid-filled BioTac fingerpad, data from 19 electrodes are presented for the bottom finger in Figure 7. A total of 15 rollouts are shown for two policies: the initial policy after a single policy update and the final policy after 41 policy updates. The corresponding P_{dc} data are presented in Figure 8. The results for learning a nominal trajectory for small pages are similar to Figures 7 and 8 except that the final policies are learned after only 25 policy updates. Figures 7 and 8 show that rollouts of the initial policy, shown in red, generate undesired spikes in the tactile sensing traces during the page warping period shaded in gray. However, the rollouts of the final policy, shown in green, result in much smoother tactile sensing traces characterized by a significant reduction in spiking behaviors. The associated reduction in page warping is also demonstrated in the supplemental video.

Figure 6 and Table I show that tactile state is more relevant and plays a more significant role in the nominal trajectory learning process than binder displacement. Figure 6 illustrates how individual components of the reward function defined in Equation 15 contribute to the learning process and change over the course of learning. The improvement in the tactile reward components ($R_{P_{\mathrm{dc}}} + R_E$) play a major role in the improvement of the overall reward function. We also observe that the tactile reward components for the large pages are worse when compared with those for the small pages. Given that large pages are heavier than small pages, it makes sense that the negative tactile consequences of page warping and page snapping would be more pronounced in the tactile reward components for the large pages.

Table I compares the mean and standard deviation of binder displacement distances before and after the learning process. Binder displacement values are reported from 15 rollouts for each combination of page size (small or large) and learning stage (before or after learning). Although Table I shows that binder displacements are larger after learning, the negative effects of binder displacement on the overall reward function are dwarfed by the significant improvements in the tactile reward components, indicating that the learning process is dominated by tactile state, as desired.

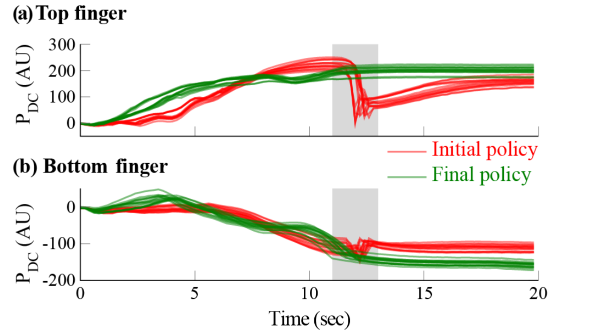


Fig. 8. Low frequency pressure data are shown for the *(a)* top and *(b)* bottom fingers for learning a nominal trajectory for flipping a large page. A total of 15 rollouts are shown for each the initial policy after a single policy update (red) and final policy after 41 policy updates (green). The initial policy generates undesired spikes in the low frequency pressure data during the page warping period shaded in gray.

Tactile state is more relevant for learning nominal trajectories for two main reasons. First, the magnitudes of the changes in tactile data due to page warping and page snapping are greater than those for marker movement resulting from binder displacement. Second, we chose the magnitudes of the scaling factors α, σ, λ in Equations 16, 18, and 20, respectively, such that marker movement would be considered but not heavily weighted in the overall reward function (Equation 15). If we were to increase λ to more heavily weight marker movement than tactile state, we would contradict our definition of what constitutes a functional page flipping behavior, as trajectories that result in page warping and page snapping would be improperly rewarded. Specifically, α was -0.0075 and λ was -2.5 for both page sizes, and σ was -0.0125 for the large page size and -0.05 for the small page size. The scaling constant σ was increased manually for the small page size in order to compensate for the fact that smaller tactile signal spikes result from the snapping of smaller pages.

TABLE I
BINDER DISPLACEMENT DISTANCES ARE REPORTED AS MEAN (STANDARD DEVIATION) FROM 15 ROLLOUTS FOR EACH COMBINATION OF PAGE SIZE AND LEARNING STAGE. "BEFORE LEARNING" REFERS TO THE INITIAL POLICY AFTER A SINGLE POLICY UPDATE. "AFTER LEARNING" REFERS TO THE FINAL POLICY AFTER 25 AND 41 POLICY UPDATES FOR THE SMALL AND LARGE PAGES, RESPECTIVELY.

	Before learning	After learning
Small pages	32.7 (0.5) mm	96.0 (3.7) mm
Large pages	24.0 (5.3) mm	184.9 (9.9) mm

Although we show that it is possible to learn functional page flipping behaviors for different page sizes using tactile information, some limitations were observed. For example, if the initial policy produces a trajectory that is too low and close to the support surface, the pages warp and then contact the binder during the page warping period. As a result, transient perturbations due to page snapping are absorbed by the friction between the binder and the pages, and are not sensed by the BioTacs, whose tactile signals will be smooth. The policy search then gets stuck in a local optimum in which the trajectories appear to maximize the reward function when, in fact, the page flipping behaviors are unacceptable. We acknowledge that the reward function may not capture the tactile consequences of all possible notebook page flipping trajectories. To address this, one could supplement tactile sensing with a complementary sensing modality, such as vision, during learning.

Another example of a limitation is the aforementioned binder displacement after learning. Sometimes, the increased displacement of the binder after learning results in a learned trajectory that is not perfectly semi-circular (as seen in the supplemental video). One possible reason for this result is that perturbations around the set of DMP weights that generate perfectly semi-circular trajectories, can actually cause the robot to move toward the support surface prematurely, resulting in page warping. Page warping would cause the rewards to deteriorate significantly due to the dominant role of the tactile reward components. Thus, page warping is avoided through learned trajectories that are not perfectly semi-circular. Specifically, minor perturbations in DMP weights from semi-circular trajectories can result in task performance and rewards having a large variance. A large variance in the reward function values will be deemed undesirable during policy updates, especially if the values of β and γ, which bound the KL-divergence and entropy reduction constraints of the MORE algorithm, are set to make the learning process risk-averse.

As is commonly done when performing reinforcement learning experiments on a real robot, we tuned hyperparameters in order to ensure that the learning process would converge within a reasonable number of samples. Specifically, the hyper-parameters were tuned such that the initial policy distribution would have sufficient variance to generate samples of DMP weights that would, in turn, generate page flipping movements with different degrees of page warping and snapping. In addition, covariance matrix values were increased for robot joint activations that were observed to be especially sensitive to changes in DMP weights during different phases of the page flipping trajectories. This variance in behavior provides a wide and meaningful range of page flipping behaviors and reward function samples that enable productive policy updates.

C. Impact of Chosen Representation of Tactile Sensing Traces

Once we successfully learned nominal trajectories for the page flipping task for both page sizes, we paused to examine the impact of the choice of representation of the tactile information on a simplified version of the sub-problem for learning adapted trajectories. Using the methods described in Section III-F, we sought to reduce the dimensionality of the tactile sensing traces before attempting the full experiment

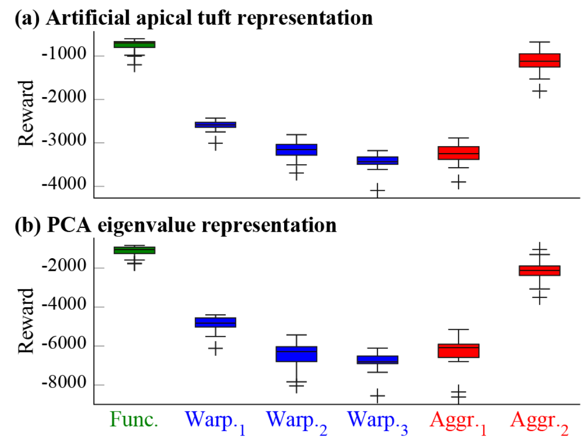


Fig. 9. Distributions of reward function samples are shown for two representations of tactile sensing traces: *(a)* artificial apical tuft, and *(b)* PCA eigenvalues. Fifteen rollouts were performed for each of six trajectories on a binder containing small pages: an ideal semi-circular trajectory (*Functional*), three trajectories that cause page warping and snapping, and two aggressive trajectories. The *Warped_1* trajectory is the nominal trajectory learned for large pages, but purposely applied to small pages.

on the adaptation of the learned nominal trajectories to a novel context. Specifically, we investigated how the values of the reward function samples were affected by two different simplified representations of the tactile sensing traces: (i) one mean P_{dc} and one mean electrode value $\overline{E}_{7-10}$ for the artificial apical tuft, per finger, and (ii) three PCA eigenvalues.

Figure 9 shows the reward function values for the two different representations of the tactile sensing traces. The reward function samples (specified for learning adapted trajectories in Section III-E) are the result of 15 rollouts performed for each of six trajectories on a binder containing small pages. An ideal semi-circular trajectory is denoted as *Functional*. Three trajectories causing page warping and page snapping are denoted as *Warped_1*, *Warped_2* and *Warped_3*. The *Warped_1* trajectory is special in that it is the nominal trajectory learned for large pages, but purposely applied to small pages (a different task context). One aggressive trajectory (*Aggressive_1*) pulls the binder upwards and away from the support surface. Another aggressive trajectory (*Aggressive_2*) pulls the binder toward the base of the robot arm.

Two comments can be made about the similarity in reward distributions between the *Functional* and *Aggressive_2* trajectories. First, the *Functional* trajectory was provided by a human demonstration, which could have resulted in a small degree of aggressiveness since the demonstrator had to manually move the robot arm in order to flip the page. Second, from a utilitarian perspective, a trajectory that is labeled *Aggressive_2* could be acceptable if it is close enough to a trajectory that is deemed *Functional*.

As expected, the desired *Functional* trajectories generate the best reward distributions, regardless of the representation of tactile sensing traces (Figure 9). For both the apical tuft and PCA eigenvalue representations, the reward distributions can be used to distinguish between functional page flipping trajectories and those that cause undesired page warping and snapping. Unfortunately, the reward distributions for the

Functional and *Aggressive_2* trajectories overlap, which make these two categories of trajectories more difficult to be distinguished from one another when using the tactile-based reward function specified in Section III-E with either of the two representations.

Nonetheless, the PCA eigenvalue representation generates a broader reward landscape compared to that generated by the apical tuft representation (Figure 9). In particular, the PCA eigenvalue representation results in a larger difference in mean reward function values between the *Functional* and the *Aggressive_2* trajectories than the apical tuft representation. As a result, functional and non-functional behaviors can be better distinguished when using the PCA eigenvalue representation. It is possible that, by averaging the four electrode measurements across the artificial apical tuft, we lose information that may have encoded differences in page flipping behaviors.

In order to further test the impact of the tactile sensing representations, the first batch of experiments for learning an adaptation policy is conducted under the assumption that the goal position for small pages is known a priori. Specifically, the joint-specific values of the goal position parameter y_g in the perceptual coupling DMPs are set to the goal position values that were obtained from the human demonstration for small pages. Three independent learning trials are conducted with each tactile sensing trace representation. We initialize all DMP weights $\{\omega\}_i$ to the weights learned for large pages and all perceptual coupling feedback weights $\{\hat{\omega}\}_i$ are set to zero. The number of Gaussian basis functions in the perceptual coupling term is set to three. Since three robot joints are subject to control, the learning process explores a 36-D space (4 tactile traces $\times$ 3 basis functions per tactile trace $\times$ 3 robot joints) for the apical tuft representation and a 27-D space ($3 \times 3 \times 3$) for the PCA eigenvalue representation.

Figure 10 shows the learning curves using the apical tuft and PCA eigenvalue tactile sensing representations, respectively. For both tactile sensing representations, the MORE algorithm enables learning, as evidenced by an increase in the mean and the maintenance of a relatively small variance for the distribution of reward function samples. These results suggest that the perceptual coupling term in Equation 22 enables the tactile feedback to drive the adaptation of the initial trajectory intended for flipping large pages toward that necessary for flipping small pages.

While the learning curves are similar for both representations (Figure 10), the resultant adapted page flipping trajectories are quite different. Our observation is that the adapted trajectories are more aggressive when learned with the apical tuft representation than with the PCA eigenvalue representation. While avoiding page warping and snapping, the aggressive trajectories pull the binder closer toward the base of the robot before initiating page flipping and can even result in the binder hitting the edge of the support surface. The gentler page flipping trajectories learned with the PCA eigenvalue representation result in smaller displacements of the binder toward the base of the robot. Differences between the trajectories learned using the apical tuft and PCA eigenvalue representations can be seen in the supplementary video. Just as the PCA eigenvalue representation was preferred for distinguishing the functional trajectories from the aggressive trajectories (Figure 9), we conclude that the PCA eigenvalue representation is also preferred for learning adapted trajectories.

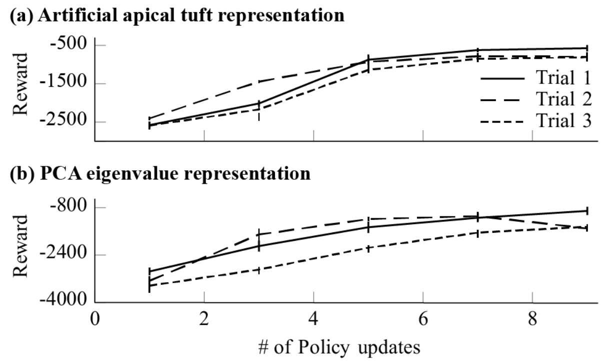


Fig. 10. Learning curves are shown for the simplified experiment on adaptation of the learned nominal trajectories to a novel context using the *(a)* artificial apical tuft and *(b)* PCA eigenvalue representations of tactile sensing traces. The goal position for small pages is known a priori. Mean and variance are presented for batches of 10 rollouts.

Figure 11 shows in greater detail how the tactile sensing traces for both representations change as the adapted trajectory is learned. The nominal tactile sensing traces are taken from 10 rollouts of a nominal trajectory learned for large pages (green). When the nominal trajectory learned for large pages is directly applied to small pages (a different task context), a much different set of tactile sensing traces results before any learning takes place (red). The results from three independent learning trials are shown, with each trial being comprised of 10 rollouts and 9 policy updates.

The adaptation of the tactile sensing traces encouraged by the perceptual coupling term in Equation 22 is most clearly illustrated in Figure 11 for the 1st and 2nd principal components of the PCA eigenvalue representation, which combine to explain 84.5% of the total variance in the original 10-D tactile sensing space. After learning to adapt the initial nominal trajectory for large pages to small pages, the tactile sensing traces for the the learning trials converge toward those for the ideal case in which the nominal trajectory learned for large pages is appropriately applied to large pages. For the apical tuft representation, the adaptation of the tactile sensing traces after learning is most clearly shown for the top finger of the gripper in the P_{dc} and $\overline{E}_{7-10}$ data.

D. Adapting Learned Nominal Trajectories to a Novel Context

Based on the encouraging results described in Section IV-C, we adopted the PCA eigenvalue representation for the tactile sensing traces for the full experiment on adaptation of the learned nominal trajectories to a novel context. For the full experiment, we no longer provide any information about page size. As a result, the goal position y_g now becomes another axis in the policy search space. We show that a nominal trajectory learned for large pages can be successfully adapted to an unknown, novel page size (small, in this case) using perceptual coupling driven by a 3-D PCA eigenvalue representation of tactile feedback. In the simplified version of the sub-problem

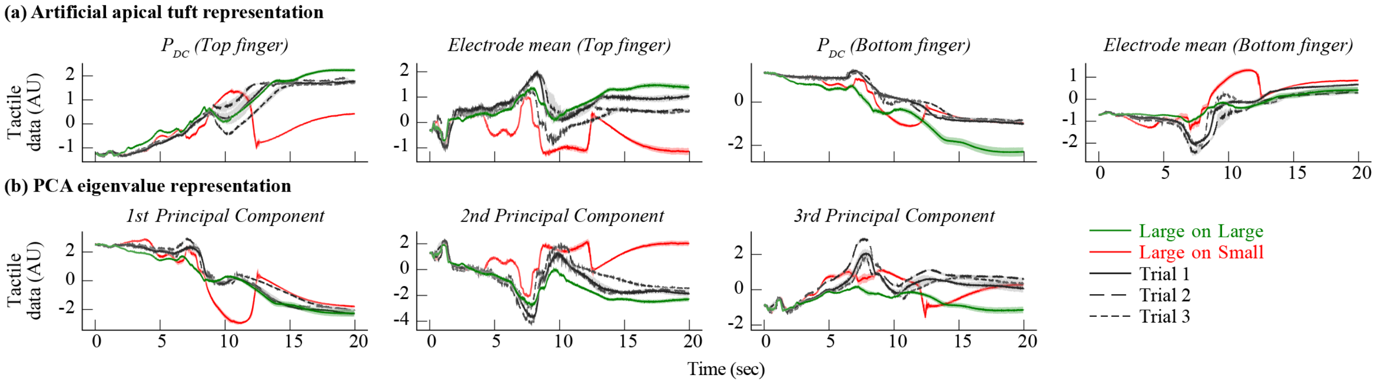


Fig. 11. Distributions of tactile sensing traces are shown for the simplified experiment on adaptation of the learned nominal trajectories to a novel context using the *(a)* artificial apical tuft and *(b)* PCA eigenvalue representations of tactile sensing traces. The goal position for small pages is known a priori. Tactile data are shown for a nominal trajectory learned for large pages and applied to large pages (green) and to small pages prior to adaptation learning (red), for 10 rollouts each. Three independent learning trials (9 policy updates each) show how the tactile sensing traces change as the initial nominal trajectory for large pages is adapted to small pages during adaptation learning.

for learning adapted trajectories, the goal positions that were provided a priori were encoded in joint space. The learning problem becomes much harder when the joint-specific goal positions are no longer provided. Without the provision of such joint-specific constraints, it is possible that naive sampling of trajectories could lead to damage of the robot or movements that do not flip the page at all. To address this issue, we leverage the fact that the page flipping trajectories lie within a 2-D plane. As seen in Figure 1, the z-coordinates for the support surface and binder are constant. Accordingly, we assume that the z-coordinate of the goal position will be constant for the gripper regardless of page size. Since the y-coordinate of the gripper will vary according to page size, we represent different goal positions using the gripper's final y-coordinate.

During learning, the MORE algorithm samples values for the gripper's goal y-coordinate at the end of the page flipping trajectory. The pair of goal (y,z) coordinates for the gripper is then transformed into goal positions in joint space via an inverse kinematics solver. Using the PCA eigenvalue representation of the tactile sensing traces, the learning algorithm searches a 28-D space (3 tactile traces $\times$ 3 basis functions per tactile trace $\times$ 3 robot joints + goal y-coordinate. As before for the simplified experiment on learning adapted trajectories (Section IV-C), all perceptual coupling feedback weights $\{\hat{\omega}\}_i$ are set to zero. We use the final y-coordinate of the gripper from the learned nominal trajectory to generate the initial estimate for y_g in joint space.

As shown by the learning curves in Figure 12, learning of the adapted trajectories was successful for the full experiment in which a nominal trajectory learned for large pages was applied to a novel task context (small pages). The initial trajectory rollouts result in aggressive movements in which the robot lifts the binder off of the support surface and drags the binder farther than necessary for small pages. As the adapted trajectory is learned using tactile-driven perceptual coupling, the distributions of reward function values improve, as reflected by the increase in mean and decrease in variance. After 16 policy updates, the small page is flipped gently, without lifting the binder from the support surface, and with less displacement of the binder.

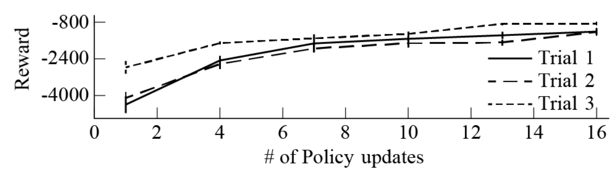


Fig. 12. Learning curves are shown for the full experiment on adaptation of the learned nominal trajectories to a novel context using the PCA eigenvalue representation of tactile sensing traces. The goal position for small pages is not known a priori and must be learned. Mean and variance are presented for batches of 10 rollouts.

Prior to adaptation learning, ten rollouts of a nominal trajectory learned for large pages and applied to small pages result in a mean value of -48.2 mm for the y-coordinate of the goal position. The mean value for the y-coordinate of the goal position is 40.0 mm for the ideal case in which a nominal trajectory learned for small pages is applied small pages. The final policies from three independent learning trials (16 policy updates each) result in mean values for the y-coordinate of the goal position of -40.8 mm, -41.0 mm, and -41.8 mm. In all cases, all standard deviation values were less than 0.03 mm. The mean value of the goal y-coordinate decreases by at least 6 mm (approximately 12%) as the trajectory is adapted from large pages to small pages. By the end of three independent learning trials, the mean goal y-coordinates are most similar to that for the ideal case in which a nominal trajectory learned for small pages is applied to small pages. This illustrates that the y-coordinate of the goal position is also learned and is successfully adapted from a value suited for large pages to a value appropriate for small pages. This demonstrates that a learned nominal trajectory can be successfully adapted to a novel task context using only the tactile sensing traces of a functional behavior as a reference.

Figure 13 shows in greater detail how the tactile sensing traces for the PCA eigenvalue representation change as the adapted trajectory is learned. Tactile sensing trace distributions (mean and variance) are shown in red for 10 rollouts of a

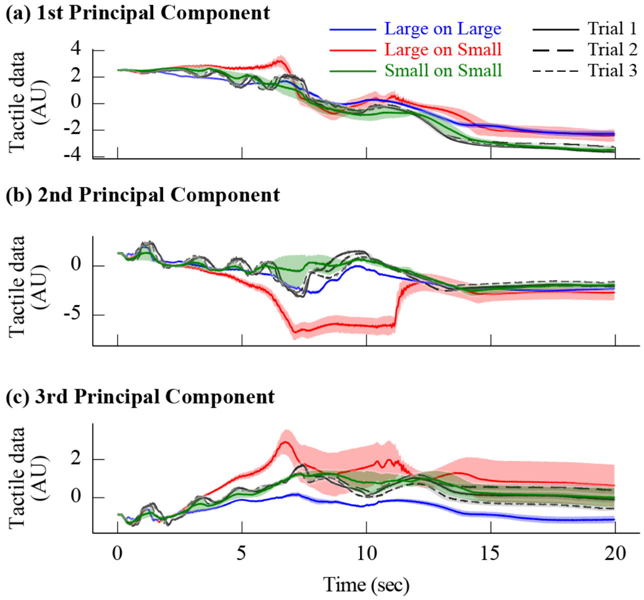


Fig. 13. Distributions are shown for the PCA eigenvalue representation of tactile sensing traces for the full experiment on adaptation of the learned nominal trajectories to a novel context. The goal position for small pages is not known a priori and must be learned. Distributions (mean and variance) are shown for a nominal trajectory learned for large pages and applied to large pages (blue) and to small pages prior to adaptation learning (red), for 10 rollouts each. Three independent learning trials (16 policy updates each) show how the distributions change as the initial nominal trajectory for large pages is adapted for small pages after learning. The distribution of tactile feedback is also shown for the ideal case of a nominal trajectory learned for small pages and applied to small pages (green).

nominal trajectory learned for large pages and naively applied to small pages (a novel task context). The ideal tactile sensing traces are shown in green for a nominal trajectory learned for small pages and applied appropriately to small pages. As desired, after 16 policy updates, the tactile sensing traces for the three independent learning trials converge upon those for the ideal case after learning the adapted trajectory and goal position using tactile-driven perceptual coupling.

For comparison, tactile sensing traces are shown in blue for a nominal trajectory learned for large pages that is applied appropriately to large pages. First, we see that task context does affect the tactile feedback, as exemplified by the slight differences between the tactile sensing traces for the rollouts that do not require adaptation, but are learned for different page sizes (blue for large pages, green for small pages). Nonetheless, the tactile feedback for the rollouts that do not require adaptation (blue, green) are more similar to one another than to the tactile feedback for the rollouts that do require adaptation (red). This supports our assumption that, while the page flipping trajectories for different page sizes might differ, similar tactile sensing traces should be expected from functional trajectories for each of the page sizes. Second, we see that the learning trials that adapt to small pages lead to tactile sensing traces that are most similar to those from rollouts for small pages that do not require adaptation (ideal green case). This trend is most clearly visible for the 3rd principal component.

Figure 14 compares distributions of reward function values for different cases of trajectory rollouts. Reward function samples are shown in red for 10 rollouts of a nominal trajectory learned for large pages and naively applied to small pages (a novel task context). Reward function samples are shown in green for the ideal case in which a nominal trajectory learned for small pages is applied appropriately to small pages.

After 16 policy updates, the reward function samples for the three independent learning trials generally converge upon those for the ideal case after learning the adapted trajectory and goal position. The improvement in reward function values is most clearly seen in the boxplots for the 2nd and 3rd principal components. For the 1st principal component, it was initially surprising to see little improvement in reward function values with learning. We believe this may be caused by the fact that the trajectory is being adapted from a nominal trajectory learned for a different task context. Some of the undesired properties of the initial trajectory may remain prevalent in the adapted trajectory and are reflected in the 1st principal component of the tactile feedback representation.

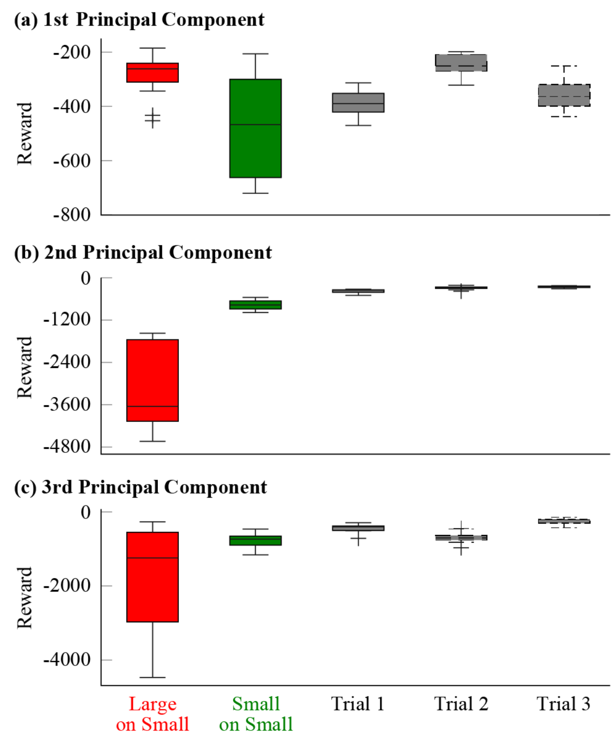


Fig. 14. Distributions are shown for reward function samples using the PCA eigenvalue representation of tactile sensing traces for the full experiment on adaptation of the learned nominal trajectories to a novel context. Goal position is not known a priori and must be learned. Distributions are shown for a nominal trajectory learned for large pages and applied to small pages prior to adaptation learning (red), for 10 rollouts each. Three independent learning trials (16 policy updates each) show how the distributions change as the initial nominal trajectory for large pages is adapted for small pages after learning. The distribution of reward function samples is also shown for the ideal case of a nominal trajectory learned for small pages and applied to small pages (green).

V. Conclusion

With experiments on real robots, we demonstrated a learned manipulation of deformable, thin-shell objects via a page flipping task. We showed that the functional performance of the task can be quantified from the perspective of tactile sensing. We also verified our intuitive assumption that there exist tactile features that can be used to adapt learning to novel task contexts for the manipulation of deformable objects. This insight could facilitate the design of tactile-based controllers for more complex manipulation tasks involving deformable objects and deformable tactile sensors.

A. Summary of Contributions

In this paper, we demonstrated the ability for a real robot to learn how to manipulate a deformable thin shell and adapt the learned functional behavior to other task contexts. More specifically, we demonstrated that a real robot can learn a page flipping task via tactile information. We addressed the learning of this task using a two-stage approach. For the first learning sub-problem, we learned nominal page flipping trajectories by constructing a reward function that quantifies functional task performance and is driven by tactile feedback. Nominal trajectories were learned specifically for small or large pages using human demonstrations via kinesthetic teaching.

For the second learning sub-problem, we learned adapted trajectories by constructing a reward function that used tactile-driven perceptual coupling. We assumed that, while the page flipping trajectories for different task contexts (page sizes) might differ, similar tactile feedback should be expected from functional trajectories for each of the contexts. We performed a simplified experiment on adaptation of the learned nominal trajectories to a novel context in which the goal position for small pages was known a priori. Using this simplified case, we compared two different representations of tactile sensing traces and concluded that a PCA eigenvalue representation encodes essential tactile information to enable learning. Finally, we performed a full experiment on adaptation of the learned nominal trajectories to a novel context in which the goal position for small pages had to be additionally learned. We showed that functional behaviors for different task contexts shared features in the tactile feedback that enabled successful learning of adapted trajectories via tactile-driven perceptual coupling.

B. Limitations and Future Work

One limitation of this work is that, for practical purposes, we reduced the control of the 7-DOF robot arm to three joints such that the page flipping movement would be constrained to a 2-D plane. If all 7 DOFs of the robot were enabled, the learning algorithm might encounter regions of the policy parameter space associated with unnecessarily complex robot motions, such as the twisting of notebook pages through wrist rotation. Defining an effective reward function based on raw tactile sensor data becomes very challenging for such complex scenarios.

Another limitation is that this work does not address additional factors, such as object texture, that might affect the generalizability of a tactile-driven policy. Further investigations are needed to assess the applicability of the "tactile invariance" notion to other tasks and scenarios, when more factors that might affect tactile sensor signals are introduced. Specifically, we believe that one interesting direction is to investigate tactile invariance not only in the context of a specific task, but rather taking it to a higher level of abstraction. For example, many contact manipulation tasks can be decomposed into a sequence of different subtasks (primitives). If they exist, intra-subtask tactile invariances could be used to compose a skill with varying task context, or speed up the learning of a new skill [48], [49].

Another interesting line of investigation is the use of multiple sensing modalities to capture task-relevant features across different task contexts. If low-level representations that encode information related to "task invariance" could be extracted from high-dimensional multimodal sensory data, the representations could enable the generalization of learned policies to new task instances more efficiently [50].

ACKNOWLEDGMENTS

We thank Casey Asaoka for assistance with the experimental set-up. This work was supported in part by Office of Naval Research Award #N00014-18-1-2814. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily represent the official views, opinions, or policies of the funding agencies.

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.

REFERENCES

[1] R. S. Johansson and J. R. Flanagan, "Coding and use of tactile signals from the fingertips in object manipulation tasks," *Nature Reviews Neuroscience*, vol. 10, no. 5, p. 345, 2009.

[2] R. Johansson. The effects of anesthesia on motor skills - youtube. [Online]. Available: https://www.youtube.com/watch?v=0LfJ3M3Kn80

[3] H. Yousef, M. Boukallel, and K. Althoefer, "Tactile sensing for dexterous in-hand manipulation in robotics—a review," *Sensors and Actuators A: Physical*, vol. 167, no. 2, pp. 171 – 187, 2011, solid-State Sensors, Actuators and Microsystems Workshop.

[4] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensing—from humans to humanoids," *IEEE Trans Robot*, vol. 26, no. 1, pp. 1–20, 2010.

[5] G. Sutanto, Z. Su, S. Schaal, and F. Meier, "Learning sensor feedback models from demonstrations via phase-modulated neural networks," in *IEEE Intl Conf on Robotics and Autom (ICRA)*, 2018, pp. 1142–1149.

[6] A. Kupcsik, M. P. Deisenroth, J. Peters, A. P. Loh, P. Vadakkepat, and G. Neumann, "Model-based contextual policy search for data-efficient generalization of robot skills," *Artif Intell*, vol. 247, pp. 415–439, 2017.

[7] V. Tangkaratt, H. V. Hoof, S. Parisi, G. Neumann, J. Peters, and M. Sugiyama, "Policy search with high-dimensional context variables," in *Proc AAAI Conf on Artificial Intelligence*, 2017, p. 2632–2638.

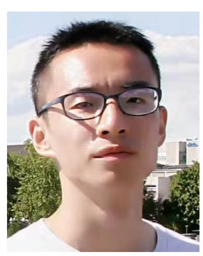
[8] N. Jamali, M. Maggiali, F. Giovannini, G. Metta, and L. Natale, "A new design of a fingertip for the icub hand," in *2015 IEEE/RSJ Intl Conf on Intelligent Robots and Systems (IROS)*, 2015, pp. 2705–2710.

[9] N. Wettels, V. J. Santos, R. S. Johansson, and G. E. Loeb, "Biomimetic tactile sensor array," *Adv Robot*, vol. 22, no. 8, pp. 829–849, 2008.

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

[11] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. F. Lepora, "The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies," *Soft Robotics*, vol. 5, no. 2, pp. 216–227, 2018.

[12] F. Veiga, J. Peters, and T. Hermans, “Grip stabilization of novel objects using slip prediction,” *IEEE Trans Haptics*, vol. 11, no. 4, pp. 531–542, 2018.
[13] J. W. James, N. Pestell, and N. F. Lepora, “Slip detection with a biomimetic tactile sensor,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3340–3346, 2018.
[14] S. Luo, J. Bimbo, R. Dahiya, and H. Liu, “Robotic tactile perception of object properties: A review,” *Mechatronics*, vol. 48, pp. 54 – 67, 2017.
[15] F. Veiga, B. Edin, and J. Peters, “Grip stabilization through independent finger tactile feedback control,” *Sensors*, vol. 20, no. 6, 2020.
[16] N. F. Lepora, A. Church, C. de Kerckhove, R. Hadsell, and J. Lloyd, “From pixels to percepts: Highly robust edge perception and contour following using deep learning and an optical biomimetic tactile sensor,” *IEEE Robotics and Autom Letters*, vol. 4, no. 2, pp. 2101–2107, 2019.
[17] S. Dong and A. Rodriguez, “Tactile-based insertion for dense box-packing,” in *IEEE/RSJ Intl Conf on Intelligent Robots and Systems (IROS)*, 2019, pp. 7953–7960.
[18] S. Tian, F. Ebert, D. Jayaraman, M. Mudigonda, C. Finn, R. Calandra, and S. Levine, “Manipulation by feel: Touch-based control with deep predictive models,” in *IEEE Intl Conf on Robotics and Automation (ICRA)*, 2019, pp. 818–824.
[19] J. Kober, J. A. Bagnell, and J. Peters, “Reinforcement learning in robotics: A survey,” *Int J Robot Res*, vol. 32, no. 11, pp. 1238–1274, 2013.
[20] M. P. Deisenroth, G. Neumann, and J. Peters, “A survey on policy search for robotics,” *Foundations and Trends in Robotics*, 2013.
[21] M. Li, Y. Bekiroglu, D. Kragic, and A. Billard, “Learning of grasp adaptation through experience and tactile sensing,” in *2014 IEEE/RSJ Intl Conf on Intelligent Robots and Systems*, 2014, pp. 3339–3346.
[22] B. Sundaralingam, A. S. Lambert, A. Handa, B. Boots, T. Hermans, S. Birchfield, N. Ratliff, and D. Fox, “Robust learning of tactile force estimation through robot interaction,” in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 9035–9042.
[23] J. Sanchez, J. Corrales, B. Bouzgarrou, and Y. Mezouar, “Robotic manipulation and sensing of deformable objects in domestic and industrial applications: a survey,” *Int J Robot Res*, vol. 37, no. 7, pp. 688–716, 2018.
[24] W. Yuan, M. A. Srinivasan, and E. H. Adelson, “Estimating object hardness with a gelsight touch sensor,” in *2016 IEEE/RSJ Intl Conf on Intelligent Robots and Systems (IROS)*, 2016, pp. 208–215.
[25] W. Yuan, Y. Mo, S. Wang, and E. H. Adelson, “Active clothing material perception using tactile sensing and deep learning,” in *2018 IEEE Intl Conf on Robotics and Automation (ICRA)*, 2018, pp. 4842–4849.
[26] J. Sanchez, C. M. Mateo, J. A. Corrales, B. Bouzgarrou, and Y. Mezouar, “Online shape estimation based on tactile sensing and deformation modeling for robot manipulation,” in *2018 IEEE/RSJ Intl Conf on Intelligent Robots and Systems (IROS)*, 2018, pp. 504–511.
[27] R. Wang, S. Wang, S. Du, E. Xiao, W. Yuan, and C. Feng, “Real-time soft body 3d proprioception via deep vision-based sensing,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3382–3389, 2020.
[28] M. Moll and L. E. Kavraki, “Path planning for deformable linear objects,” *IEEE Trans Robot*, vol. 22, no. 4, pp. 625–636, 2006.
[29] T. Bretl and Z. McCarthy, “Quasi-static manipulation of a kirchhoff elastic rod based on a geometric analysis of equilibrium configurations,” *Intl J of Robotics Research*, vol. 33, no. 1, pp. 48–68, 2014.
[30] A. Shah, L. Blumberg, and J. Shah, “Planning for manipulation of interlinked deformable linear objects with applications to aircraft assembly,” *IEEE Trans Autom Sci Engin*, vol. 15, no. 4, pp. 1823–1838, 2018.
[31] C. Elbrechter, R. Haschke, and H. Ritter, “Folding paper with anthropomorphic robot hands using real-time physics-based modeling,” in *IEEE-RAS Intl Conf on Humanoid Robots (Humanoids)*, 2012, pp. 210–215.
[32] V. Petrík, V. Smutný, P. Krsek, and V. Hlaváč, “Physics-based model of a rectangular garment for robotic folding,” in *IEEE/RSJ Intl Conf on Intelligent Robots and Systems (IROS)*, 2016, pp. 951–956.
[33] Y. She, W. Shaoxiong, D. Siyuan, N. Sunil, A. Rodríguez, and E. Adelson, “Cable manipulation with a tactile-reactive gripper,” in *Proc Robotics: Science and Systems*, 2020.
[34] G. Sutanto, N. Ratliff, B. Sundaralingam, Y. Chebotar, Z. Su, A. Handa, and D. Fox, “Learning latent space dynamics for tactile servoing,” in *IEEE Intl Conf on Robotics and Autom (ICRA)*, 2019, pp. 3622–3628.
[35] F. F. Veiga, “Towards dexterous in-hand manipulation through tactile sensing,” Ph.D. dissertation, Technische Univ., Darmstadt, July 2018.
[36] S. H.Huang, M. Zambelli, J. Kay, M. F. Martins, Y. Tassa, P. M. Pilarski, and R. Hadsell, “Learning gentle object manipulation with curiosity-driven deep reinforcement learning,” *arXiv:1903.08542*, 2019.
[37] Y. Chebotar, K. Hausman, Z. Su, G. S. Sukhatme, and S. Schaal, “Self-supervised regrasping using spatio-temporal tactile features and reinforcement learning,” in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 1960–1966.
[38] H. van Hoof, T. Hermans, G. Neumann, and J. Peters, “Learning robot in-hand manipulation with tactile features,” in *IEEE-RAS Intl Conf on Humanoid Robots (Humanoids)*, 2015, pp. 121–127.
[39] S. Schaal, *Dynamic Movement Primitives–A Framework for Motor Control in Humans and Humanoid Robotics*. Tokyo: Springer Tokyo, 01 2006, pp. 261–280.
[40] A. Paraschos, C. Daniel, J. Peters, and G. Neumann, “Probabilistic movement primitives,” in *Proc Intl Conf on Neural Information Processing Sys - Vol. 2*, 2013, p. 2616–2624.
[41] S. Calinon, F. Guenter, and A. Billard, “On learning, representing, and generalizing a task in a humanoid robot,” *IEEE Trans Sys, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 2, pp. 286–298, 2007.
[42] J. Kober, B. Mohler, and J. Peters, “Learning perceptual coupling for motor primitives,” in *IEEE/RSJ Intl Conf on Intelligent Robots and Systems*, 2008, pp. 834–839.
[43] F. Stulp and G. Raiola, “Dmpbbo: A versatile python/c++ library for function approximation, dynamical movement primitives, and black-box optimization,” *J Open Source Software*, vol. 4, no. 37, p. 1225, 2019.
[44] H. Hoffmann, P. Pastor, D. Park, and S. Schaal, “Biologically-inspired dynamical systems for movement generation: Automatic real-time goal adaptation and obstacle avoidance,” in *IEEE Intl Conf on Robotics and Automation*, 2009, pp. 2587–2592.
[45] A. Abdolmaleki, R. Lioutikov, J. Peters, N. Lau, L. P. Reis, and G. Neumann, “Model-based relative entropy stochastic search,” in *Advances in Neural Information Processing Systems 28*, 2015, pp. 3537–3545.
[46] P. Pastor, M. Kalakrishnan, L. Righetti, and S. Schaal, “Towards associative skill memories,” in *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, 2012, pp. 309–315.
[47] J. Fishel, G. Lin, and G. Loeb, *BioTac® Product Manual, V21*. SynTouch LLC, 2018.
[48] T. Narita and O. Kroemer, “Policy blending and recombination for multimodal contact-rich tasks,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2721–2728, 2021.
[49] Z. Su, O. Kroemer, G. E. Loeb, G. S. Sukhatme, and S. Schaal, “Learning manipulation graphs from demonstrations using multimodal sensory signals,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 2758–2765.
[50] M. A. Lee, Y. Zhu, P. Zachares, M. Tan, K. Srinivasan, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, “Making sense of vision and touch: Learning multimodal representations for contact-rich tasks,” *IEEE Transactions on Robotics*, vol. 36, no. 3, pp. 582–596, 2020.



Yi Zheng received the B.S. degree in reliability engineering from Beihang University, Beijing, China, in 2014 and the M.S. degree in mechanical and aerospace engineering from the University of California at Los Angeles, Los Angeles, CA, USA, where he is currently working toward the Ph.D. degree in mechanical and aerospace engineering in the Biomechatronics Laboratory, under the supervision of Dr. Veronica Santos. His research interests include tactile manipulation, reinforcement learning, and robotics control systems.

Filipe Fernandes Veiga received the M.Sc. degree from the Instituto Superior Técnico, Lisbon, Portugal, in 2012, and the Ph.D. degree in machine learning and robotics from the Technische Universität Darmstadt, Darmstadt, Germany, in 2018.

He is a Postdoctoral Associate with the Computer Science & Artificial Intelligence Lab, the Massachusetts Institute of Technology, Cambridge, MA, USA, working with Prof. Edward Adelson in the development and usage of GelSight fingertip sensors for improving the dexterous manipulation skills of robots. His current research focuses on exploring biomimetic control approaches in conjunction with state-of-the-art machine learning in order to learn real-time state estimation and control systems.

Jan Peters (Fellow, IEEE) is a full professor (W3) for Intelligent Autonomous Systems at the Computer Science Department of the Technische Universitaet Darmstadt. Jan Peters has received the Dick Volz Best 2007 US PhD Thesis Runner-Up Award, the Robotics: Science & Systems - Early Career Spotlight, the INNS Young Investigator Award, and the IEEE Robotics & Automation Society's Early Career Award as well as numerous best paper awards. In 2015, he received an ERC Starting Grant and in 2019, he was appointed as an IEEE Fellow and in 2020 an ELLIS fellow.

Despite being a faculty member at TU Darmstadt only since 2011, Jan Peters has already nurtured a series of outstanding young researchers into successful careers. These include new faculty members at leading universities in the USA, Japan, Germany, Austria, Finland and Holland, postdoctoral scholars at top computer science departments (including MIT, CMU, and Berkeley) and young leaders at top AI companies (including Amazon, Google and Facebook).

Jan Peters has studied Computer Science, Electrical, Mechanical and Control Engineering at TU Munich and FernUni Hagen in Germany, at the National University of Singapore (NUS) and the University of Southern California (USC). He has received four Master's degrees in these disciplines as well as a Computer Science PhD from USC. Jan Peters has performed research in Germany at DLR, TU Munich and the Max Planck Institute for Biological Cybernetics (in addition to the institutions above), in Japan at the Advanced Telecommunication Research Center (ATR), at USC and at both NUS and Siemens Advanced Engineering in Singapore. He has led research groups on Machine Learning for Robotics at the Max Planck Institutes for Biological Cybernetics (2007-2010) and Intelligent Systems (2010-2021).

Veronica J. Santos (Member, IEEE) received the B.S. degree in Mechanical Engineering with a Music minor from the University of California, Berkeley, in 1999, and the M.S. and Ph.D. degrees in Mechanical Engineering with a Biometry minor from Cornell University in 2004 and 2007, respectively.

From 2007 to 2008, she was a postdoctoral researcher with the University of Southern California, where she contributed to the development of a bio-inspired tactile sensor for prosthetic hands. From 2008 to 2014, she was an Assistant Professor of Mechanical and Aerospace Engineering with Arizona State University. She is currently a Professor of Mechanical and Aerospace Engineering with the University of California, Los Angeles, and the Director of the UCLA Biomechatronics Lab. Her research interests include hand biomechanics, human-machine systems, tactile sensing and perception, and prosthetics/robotics for grasp and manipulation.

Dr. Santos was selected for an NSF CAREER Award, numerous teaching awards, and a U.S. National Academy of Engineering Frontiers of Engineering Education Symposium. She has served as an IEEE International Conference on Robotics and Automation Editor, an ASME Journal of Mechanisms and Robotics Associate Editor, an ACM Transactions on Human-Robot Interaction Associate Editor, the 2020 IEEE Haptics Symposium Editor-in-Chief, and as a General Co-Chair for the 2022 IEEE Haptics Symposium. She will also serve as a General Co-Chair for the 2024 IEEE Haptics Symposium.