How Much Haptic Surface Data Is Enough?

Alex Burka and **Katherine J. Kuchenbecker**
Haptics Group, GRASP Laboratory
ESE, MEAM and CIS Departments
University of Pennsylvania
220 S 33rd St, Philadelphia, PA 19104

Abstract

The Proton Pack is a portable visuo-haptic surface interaction recording device that will be used to collect a vast multimodal dataset, intended for robots to use as part of an approach to understanding the world around them. In order to collect a useful dataset, we want to pick a suitable interaction duration for each surface, noting the tradeoff between data collection resources and completeness of data. One approach frames the data collection process as an online learning problem, building an incremental surface model and using that model to decide when there is enough data. Here we examine how to do such online surface modeling for the initial problem of learning a kinetic friction model. With a long dataset consisting of force, vibration, and speed recorded by a human operator moving a tooling ball end-effector across a flat vinyl surface, we find a good stopping point at 55.4 s.

Introduction

As robots venture out of the lab and begin to interact with humans in unstructured environments, the robots need to understand the physical properties of the objects around them. We focus on identifying properties of flat surfaces from visual inspection, which has applications for robots selecting a gait when walking on varied surfaces, choosing a grasp on an unfamiliar object, and more. In order to build algorithms that can learn the correspondence between visual appearance and these haptic properties, we have designed and built an instrument that we will use to record a large dataset of multimodal data from surface interactions. Learning from this database will identify visual patterns that correspond to haptic properties, while generalizing over purely visual differences such as color and illumination, allowing robots to better understand their surroundings and "feel with their eyes."

Haptic Surface Perception

Our instrument takes inspiration in mission, design and processing strategies from previous texture sensing systems in our lab and in the literature, most directly from the Haptic Camera (Culbertson et al. 2013), a pen-shaped device that includes a force/torque sensor, accelerometers and magnetic tracking. Another similar pen-shaped device is the WHaT (Pai and Rizun 2003), featuring accelerometers and wireless connectivity but no motion tracking (although external visual tracking was added later (Andrews and Lang 2007)). Also notable are Battaglia et al.'s ThimbleSense (Battaglia et al. 2014) that mounts force/torque sensors on robot or human fingers and Xu et al.'s experiments with a SynTouch BioTac mounted on a Shadow Dextrous Hand (Xu, Loeb, and Fishel 2013).

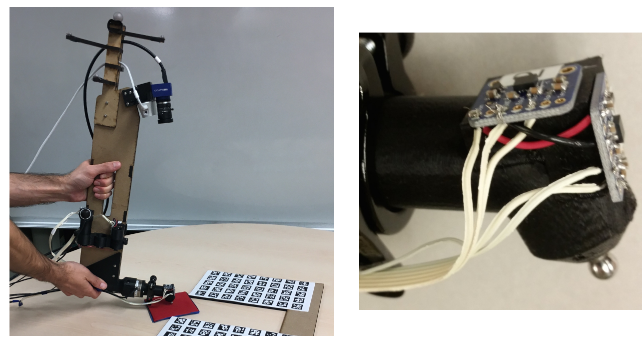

Figure 1: Left: Proton Pack in use. Right: End-effector.

Desiring a portable, multimodal sensing system that can handle a wider range of contact forces while surveying more surfaces in and out of the lab, we introduced the Proton Pack last year (Burka et al. 2016a; 2016b) and have been making modifications since then (Burka et al. 2017) in preparation for collection of a large dataset.

The Proton Pack

The Proton is a sensing system in two parts: a handheld rig that contains multimodal sensors and an interchangeable end-effector, plus a backpack enclosing a battery and computer. This system can be used for untethered data collection, away from power and wi-fi – the backpack broadcasts its own network, and any device with a web browser, such as a smartphone or tablet, can control the system.

Fig. 1 shows an operator holding the Proton Pack with a passive end-effector that terminates in a steel tooling ball, which can be interchanged with similarly-shaped end-effectors that hold an OptoForce force sensor or a SynTouch BioTac. Other haptic sensors that are available regardless of end-effector are an ATI Mini40 six-axis force/torque sensor, a MEMS microphone, an inertial measurement unit (IMU), and dual two-axis high-bandwidth accelerometers mounted directly on each end-effector. At the top of the handheld rig are mounted two cameras: a high-resolution mvBlueFOX3 RGB camera and a Structure depth sensor. A frame of stationary AprilTags (Olson 2011) is placed next to the surface to act as fiducial markers for motion tracking.

We chose to have a human operator, rather than a robot, hold the sensing rig for several reasons. First, and most importantly, it avoids contamination of our recorded surface vibration data with actuator vibrations that would be inherent in any robotic system. Consumers of our dataset who compare against robot-gathered data will need to account for their own actuator noise, but the reduction of noise in the dataset itself will simplify such comparisons. Second, though a robot's motions would be more precise and repeatable, by using a human operator we can take advantage of already-learned human intuition as to which motions will be most informative when exploring a new surface, and we avoid having to create and tune a motion controller for robotic contact interaction with unknown surfaces. Third, a human-operated system is more portable and flexible than a robotic one, facilitating extensive data capture inside and outside the lab.

Friction Modeling

Data collected by the Proton can be used to extract various visual and mechanical surface properties. In this preliminary work we focus on friction, a basic mechanical property that has been the subject of much modeling effort.

Armstrong-Hélouvry, Dupont and De Wit survey a wide range of friction models (Armstrong-Hélouvry, Dupont, and Canudas De Wit 1994). Their review culminates in a combined seven-parameter model, including Coulomb friction, viscous friction, the Stribeck effect, rising static friction, frictional memory, and presliding displacement.

Bilinear model In this work we use a simple model for friction, including Coulomb friction and viscous friction, but neglecting second-order effects for simplicity. The model is given by the following equation:

$$\vec{F}_t(t) = -\Big(\alpha||\vec{F}_n(t)|| + \beta||\vec{v}_t(t)||\Big)\frac{\vec{v}_t(t)}{||\vec{v}_t(t)||} \quad (1)$$

The time-varying friction force $\vec{F}_t$ acts in opposition to the tangential velocity $\vec{v}_t$, with a magnitude linearly related to that of the normal force $\vec{F}_n$ and the magnitude of the tangential velocity, where α and β are coefficients fit from experimental data.

Online Learning

Machine learning typically runs over a complete dataset with few constraints on efficiency or processing power required, since the goal is only to learn a model that represents the data well. Typically the model itself is designed to run quickly, but the training process is not (e.g., evaluating vs. training a neural network). However, substantial literature exists in online learning, where a model is updated continuously, often in real time, while the full dataset is not known.

A key question is at what point the learned model can be considered done, and the learning process can stop incorporating new data. One answer is a technique called Stabilizing Predictions (Bloodgood and Vijay-Shanker 2009). In that work, which considers the problem of word sense disambiguation (a discrete classification task), a small unsupervised test set is reserved from the input data, and learning is considered done when the (unchecked) predictions on the test set have stabilized. In this work, we adapt this technique to a continuous-time domain, fitting a first-order model to data and inspecting the stability of the model coefficients, which completely determine the model's predictions on unseen data.

Methods

We collected 120 seconds of data while rubbing the steel tooling ball end-effector on a vinyl surface at varying speeds and normal forces. Modalities recorded were force data at 3000 Hz and pose measurements at 15 Hz. Using the processes detailed in our earlier publications (Burka et al. 2016a; 2016b; 2017), we rotated the force measurements into the world frame (relative to the fiducial markers), compensated for contributions from the force of gravity, and decomposed the force and velocity measurements into their respective tangential and normal components, yielding tangential force opposing the direction of motion $F_t(t)$, normal force $F_n(t)$, and end-effector tip speed $v_t(t)$. These data streams were smoothed using a low-pass filter with a cutoff frequency of 15 Hz, then segmented into 100 ms windows (300 samples each). To avoid attempting to fit a kinetic friction model to data reflecting the effect of static friction, windows with an average tip speed of less than 25 mm/s were not considered.

We would like to find the point at which the data encompasses a representative sample of interactions with the vinyl surface, which is a good stopping point for data collection. To analyze progress toward this ideal during the vinyl interaction, we simulated an observation by adding each window in turn to running sets of input data (normal force and tip speed) and output data (tangential force). At each iteration a least-squares fit was used to find the best coefficients of the bilinear model for the friction process (see (1)) based on all data observed up to that point. This model was then used to predict the tangential force in the next window.

In theory, the prediction error should decrease to near zero as the model increases in accuracy. However, random noise disrupts the data and makes the simple bilinear model inaccurate on short time scales. Therefore, instead of tracking the incremental prediction error, we looked at the variation of the model coefficients α and β at each window k. This approach mimics the Stabilizing Predictions method reviewed above (Bloodgood and Vijay-Shanker 2009), which tracks the stability of unsupervised predictions on a test set.

To find a suitable stopping point, we calculated the relative differences between each successive pair of model coefficients to get $\Delta\alpha$ and $\Delta\beta$ (see (2) and (3)) and smoothed

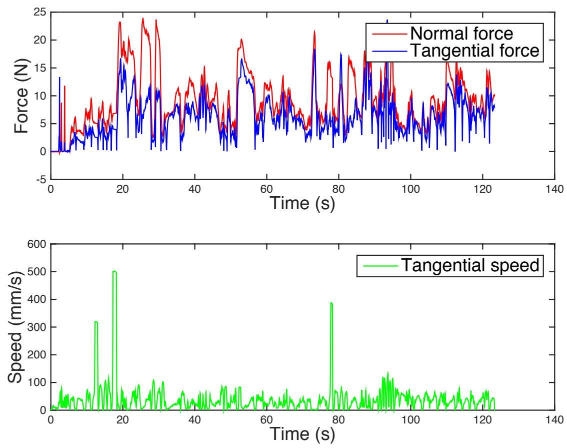

Figure 2: Inputs and outputs of the friction model. Top: normal and tangential components of contact force. Bottom: Tangential speed of the end-effector. The friction model takes normal force and tangential speed as inputs and predicts tangential force.

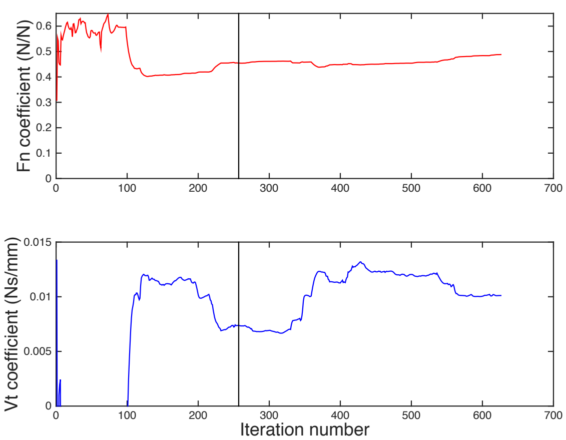

Figure 3: Coefficients of each incremental friction model (see (1)). Top: α, the normal force coefficient. Bottom: β, the tangential speed coefficient. Vertical lines mark the chosen stopping point.

these differences using a moving average filter (window size 10), resulting in $\Delta\alpha^s$ and $\Delta\beta^s$. We set the stopping point to the first iteration at which the average of $\Delta\alpha^s(k)$ and $\Delta\beta^s(k)$ drops below a tunable threshold of 0.001.

$$\Delta\alpha(k) = \frac{|\alpha(k) - \alpha(k-1)|}{\alpha(k-1)} \tag{2}$$

$$\Delta\beta(k) = \frac{|\beta(k) - \beta(k-1)|}{\beta(k-1)} \tag{3}$$

Results and Discussion

Fig. 2 shows the recorded contact force and speed, which formed the inputs and output to the online learning process. The coefficients of each incremental model are shown

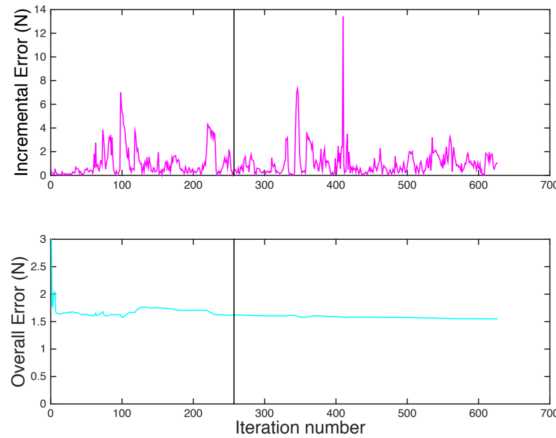

Figure 4: Model error during the online learning process. Top: Incremental performance of each model on the next unseen time window. Bottom: Overall performance of each model on the entire dataset. Vertical lines mark the chosen stopping point.

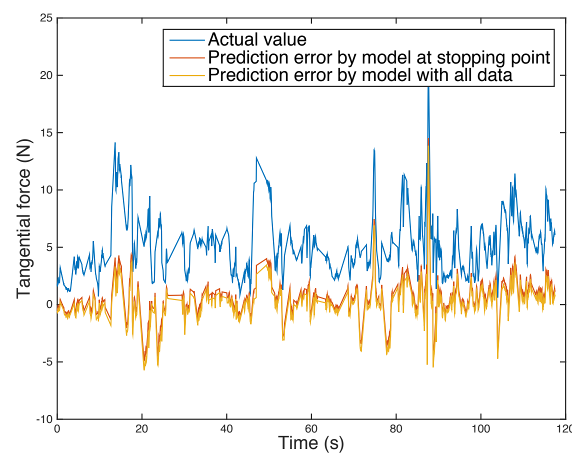

Figure 5: Model comparison. The ground truth tangential force is plotted along with the prediction error for a model trained on data up to the chosen stopping point (root-mean-square error = 1.67 N), and one trained on the entire dataset (root-mean-square error = 1.61 N).

in Fig. 3. We measured the root mean square (RMS) error of the tangential force predictions $\hat{F}_t(t)$ against the ground truth $F_t(t)$, both for a single window w starting at time t_0^w, and for the entire dataset of N samples:

$$e_{inc}(w) = \sqrt{\frac{1}{300}\sum_{i=1}^{300}\left|\hat{F}_t\left(t_0^w + \tfrac{i}{3000}\right) - F_t\left(t_0^w + \tfrac{i}{3000}\right)\right|} \tag{4}$$

$$e_{tot} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\hat{F}_t\left(\tfrac{i}{3000}\right) - F_t\left(\tfrac{i}{3000}\right)\right|} \tag{5}$$

The two error metrics are plotted in Fig. 4. The incremental prediction error (top) represents the performance of each model on the next time window that was not used in its training, while the overall error (bottom) shows each model evaluated against the entire dataset. Application of the coefficient stabilization criterion, as described in the previous section, implies a stopping point after 257 windows, or 55.4 seconds of data. As can be seen in Fig. 5, the predictions generated by the model at the stopping point are nearly indistinguishable from those generated by the final model. Our idea is that this algorithm could run during data collection and alert the operator when enough data has been gathered.

Strangely, the overall error shown in Fig. 4 (bottom) trends downward but does not drop significantly during the learning process. This lack of reduction may imply that the data is too noisy for our simple model of friction to account for all effects, or that the choice of 100 ms windows focuses too narrowly, so that the evaluation of model predictions on a single window reflects random noise rather than model accuracy. However, the overall error, which tests the predictions of each incremental model over the entire 120 second dataset (thereby including data that the model was not given for learning), does stabilize at about 1.5 N RMS after initial fluctuations. The point at which the overall error stabilizes is close to the point at which the model coefficients stabilize. This correlation implies that coefficient stabilization can be used as a heuristic, indicating that the overall error (which cannot be evaluated in an online setting) is stabilizing as well.

There is clearly a danger, when using a threshold, of stopping too early (perhaps the incremental prediction error drops very low after a particularly uneventful stretch of time), stopping too late (because the chosen threshold is too conservative), or getting stuck in a local minimum, where the learned model is good enough for the data collected so far but misses possible generalization. Indeed, Fig. 3 clearly shows that the best fit value of the tangential speed coefficient changed significantly after the stopping point (though the predictions in Fig. 5 appear largely unaffected). More surfaces should be evaluated in order to fine-tune the stabilization criterion.

Although the chosen stopping point reflects stabilization of the model coefficients, it is also informative to analyze the comprehensiveness of the distribution of normal forces and speeds. Figs. 6 and 7 show visualizations of the distribution of normal forces and tangential speeds, respectively, after 24 seconds, 60 seconds, and the full 120 seconds. Clearly, the distribution at 60 seconds is of the same shape as the final distribution, mainly different in overall magnitude, but the distribution at 24 seconds has noticeable regions of missing data. If it is possible to guide the data collection process in order to achieve such a distribution earlier in time, we expect the model coefficients to stabilize sooner as well. Future directions may include hinting the human operator so as to encourage a uniform distribution of applied normal forces and tangential speeds.

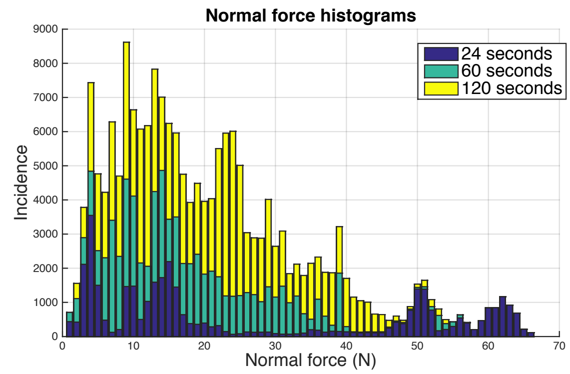

Figure 6: Histograms of the distribution of normal force at three points during dataset collection.

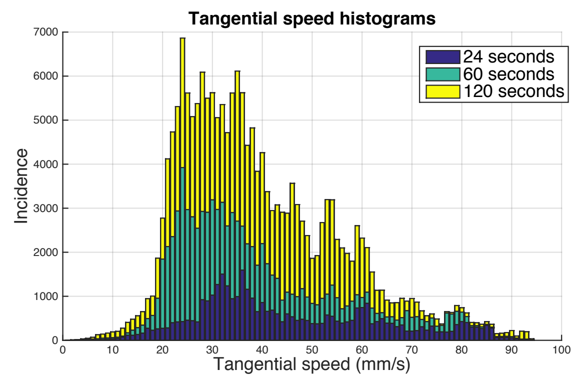

Figure 7: Histograms of the distribution of tangential speed at three points during dataset collection.

Conclusions and Future Work

In a case study with a vinyl surface, we have shown that metrics based on incremental model building are feasible for automatically determining the stopping point for data collection. This analysis is preliminary, as we have examined only a single surface (vinyl rubber) and a single surface property (kinetic friction). We must try the method on other surfaces to see whether it scales, and we must determine suitable values for the window length and coefficient stabilization threshold.

Going forward, we will use the method presented here to inform the parameters of our large dataset collection program. We may decide to collect a varying amount of data per surface, or we might collect pilot data on a representative set of surfaces and use incremental models to find a common data size for all surfaces. Other options include using the normal force and tangential speed distributions, either in an online setting to help the human operator collect good data or in post-processing to select representative regions of a long recording.

References

Andrews, S., and Lang, J. 2007. Interactive scanning of haptic textures and surface compliance. In *Proc. IEEE International Conference on 3-D Digital Imaging and Modeling*, 99–106.

Armstrong-Hélouvry, B.; Dupont, P.; and Canudas De Wit, C. 1994. A survey of models, analysis tools and compensations methods for the control of machines with friction. *Automatica* 30(7):1083–1138.

Battaglia, E.; Grioli, G.; Catalano, M. G.; Santello, M.; and Bicchi, A. 2014. ThimbleSense: an individual-digit wearable tactile sensor for experimental grasp studies. In *Proc. IEEE International Conference on Robotics and Automation*, 2728–2735.

Bloodgood, M., and Vijay-Shanker, K. 2009. A method for stopping active learning based on stabilizing predictions and the need for user-adjustable stopping. In *Proc. 13th Conference on Computational Natural Language Learning*, 39–47.

Burka, A.; Hu, S.; Helgeson, S.; Krishnan, S.; Gao, Y.; Hendricks, L. A.; Darrell, T.; and Kuchenbecker, K. J. 2016a. Design and implementation of a visuo-haptic data acquisition system for robotic learning of surface properties. In *Proc. IEEE Haptics Symposium*.

Burka, A.; Hu, S.; Helgeson, S.; Krishnan, S.; Gao, Y.; Hendricks, L. A.; Darrell, T.; and Kuchenbecker, K. J. 2016b. Proton: A visuo-haptic data acquisition system for robotic learning of surface properties. In *Proc. IEEE International Conference on Multisensor Fusion and Integration*.

Burka, A.; Rajvanshi, A.; Allen, S.; and Kuchenbecker, K. J. 2017. Proton 2: Increasing the sensitivity and portability of a visuo-haptic surface interaction recorder. In *IEEE International Conference on Robotics and Automation*. Under review.

Culbertson, H.; Unwin, J.; Goodman, B. E.; and Kuchenbecker, K. J. 2013. Generating haptic texture models from unconstrained tool-surface interactions. In *Proc. IEEE World Haptics Conference*, 295–300.

Olson, E. 2011. AprilTag: A robust and flexible visual fiducial system. In *Proc. IEEE International Conference on Robotics and Automation*, 3400–3407.

Pai, D. K., and Rizun, P. 2003. The WHaT: a wireless haptic texture sensor. In *Proc. IEEE Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 3–9.

Xu, D.; Loeb, G. E.; and Fishel, J. A. 2013. Tactile identification of objects using Bayesian exploration. In *Proc. IEEE International Conference on Robotics and Automation*, 3056–3061.