Towards Effective Tactile Identification of Textures using a Hybrid Touch Approach

Tasbolat Taunyazov[1,2], Hui Fang Koh[3], Yan Wu[1], SMIEEE, Caixia Cai[1], Harold Soh[2]

***Abstract*— The sense of touch is arguably the first human sense to develop. Empowering robots with the sense of touch may augment their understanding of interacted objects and the environment beyond standard sensory modalities (e.g., vision). This paper investigates the effect of hybridizing touch and sliding movements for tactile-based texture classification. We develop three machine-learning methods within a framework to discriminate between surface textures; the first two methods use hand-engineered features, whilst the third leverages convolutional and recurrent neural network layers to learn feature representations from raw data. To compare these methods, we constructed a dataset comprising tactile data from 23 textures gathered using the iCub platform under a loosely constrained setup, i.e., with nonlinear motion. In line with findings from neuroscience, our experiments show that a good initial estimate can be obtained via touch data, which can be further refined via sliding; combining both touch and sliding data results in 98% classification accuracy over unseen test data.**

I. Introduction

Enabling robots to perceive textures through tactile interaction opens up an additional perceptual modality beyond other modes such as vision and audio. Textures often reveal the physical nature of an object or the environment that is interacted with; for example, humans can easily distinguish between wood and patterned laminate by touch, even though the two materials may share a similar visual appearance.

Research on texture classification has been active for more than two decades; a very early study by Mayol-Cuevas *et al.* [1] classified textures using data obtained by sliding an electret microphone on selected fine textures. More recent attempts utilize sophisticated sensors such as the highly-sensitive BioTac® or the high-resolution GelSight [2], and modern deep learning methods such as convolutional and recurrent neural networks [3], [4].

Most prior work has focused on one of the three texture sensing modes of human touch perception: vibration only, passive scanning, and static touch [5] (note that vibration is a key feature used in the passive scanning literature). In this work, we contribute to this fertile area of research by exploring the effectiveness of hybridizing touch perceptual modes. Two key research questions are investigated: do we need prolonged dynamic contact with an object (e.g., a sliding motion) versus a simple spatially-static touch, to accurately classify a texture? Potentially, both touch and sliding reveal different properties of the object. If so, how much sliding data do we actually need, particularly in the presence of noise (e.g., when lower-cost sensors are used or the robot motion is not finely controlled)?

To answer these questions, we collected—and have made publicly available—a tactile dataset obtained via the iCub robot interacting with 23 different textures under *loose movement constraints*. This is unlike other existing datasets which were collected under strict force-controlled *linear* movements [6], [7]. In addition, we include auxiliary data, which includes encoder readings and iCub skin events (estimated force, normal direction, and pressure).

Using this dataset, we tested a texture classification framework (described in Sec. III) for utilizing features from *both* spatially-static touch and sliding data. We contribute three different machine learning (ML) models based on the Support Vector Machine (SVM) that has been widely used in the tactile perception community, and Long-Short-Term-Memory (LSTM) cells and Convolutional Neural Networks (CNNs) that are popular in the deep learning community. Our systematic experiments in Sec. IV clearly show that using both touch and sliding movements are important to accurately determine textures — using both forms of tactile data improved the classification accuracy of all three ML models by as much as 10%. Moreover, only a short window of tactile data is required to enable a highly accurate classification.

In summary, this paper makes three primary contributions:

- A research finding that texture classification accuracy is improved with both sliding and touch movements;
- A texture classification framework with three different ML models comprising statistical machine learning and connectionist approaches;
- A tactile dataset using the iCub humanoid robot with capacitive tactile sensor on 23 textures (a total of 2,852 samples comprising repeated measurements for both touch and sliding movements).

Taking a broader perspective, this work makes clear that multiple perceptual modes are important for texture identification; this warrants future work into combining multiple forms of tactile data for other tasks, e.g., grasp stability prediction and dexterous manipulation.

II. Background and Related Work

Tactile learning is a vibrant field of research and prior work includes tactile-based object classification [8]–[11], grasp stability prediction and enhancement [12]–[14], dexterous manipulation [14]–[16] and texture identification [17]–[19]. Our work focuses on texture classification, which we briefly review in this section.

Authors Affiliation: [1]A*STAR Institute for Infocomm Research, Singapore. [2]National University of Singapore, Singapore. [3]Nanyang Technological University, Singapore.

Email: e0348851@u.nus.edu, khuifang2202@gmail.com, wuy@i2r.a-star.edu.sg, cai_caixia@i2r.a-star.edu.sg, harold@comp.nus.edu.sg

To our knowledge, the first attempt at automated texture classification was performed in the 1990s [1]; a human-controlled electret microphone was slid across the surfaces of 18 different materials to collect sound signals. These signals were transformed into frequency features, and then classified (with 94% accuracy) using Learning Vector Quantization [20].

More recent work has explored autonomous texture classification by robots. For example, Jamali and Sammut [6] conducted a comprehensive study of texture-classification techniques. A robot finger (attached to a robot arm) with a tactile array on the fingertip was slid over six different textures with constant velocity and force. The collected signals were analyzed in frequency spectrum and the five frequencies with highest amplitudes are used for classification. A larger study by Fishel and Loeb [7] demonstrated texture classification of 117 materials with 95.4% accuracy using the BioTac sensor, a highly sensitive multimodal tactile sensor. The sensor was attached to a robotic arm with precise control on force and traction, and a Bayesian classification technique was used on three features: traction from motor current, roughness, and fineness. A Bayesian approach was also used in [19] to achieve a classification accuracy of 99% using BioTac sensor data collected from ten different textures.

In contrast to the sliding-based approaches above, Holscher *et al.* [21] showed that *static features* (e.g., temperature and thermal flow) were more predictive of textures compared to frequency features obtained using a BioTac sensor. Unfortunately, prolonged acquisition times for certain static features rendered the approach unsuitable for many interaction tasks.

Recent developments in deep learning have prompted their use in texture classification. For example, convolutional neural networks (CNNs) were used in [18] to classify data (6 different textures) from a tactile array sensor with 97.3% accuracy. Very recent work has applied Deep Maximum Covariance Analysis (DMCA) and Long-Short-Term Memory (LSTM) [3] on static touch tactile data gathered using the GelSight [2], a highly-dense optical tactile sensor array.

In summary, the prior work above has shown that highly-accurate texture identification from tactile data is possible using a variety of machine learning techniques on different tactile/touch sensors. However, the datasets used in the sliding-based classification studies were typically gathered under strict constraints, e.g., linear movements and force control. Tactile data obtained "in the wild" will likely to be from sensors with varying quality and under less controlled circumstances, and as a result, more noisy.

In this work, we hypothesize that a combination of spatially-static touch (vertical downward movement to contact the object) and prolonged dynamic contact (sliding) is more informative when data is noisy. Indeed, human beings use a combination of multiple perceptual modes [5] — both spatial patterns and scanning vibration — to perceive textures [22]. In the following sections, we discuss our approach to test this hypothesis and our findings.

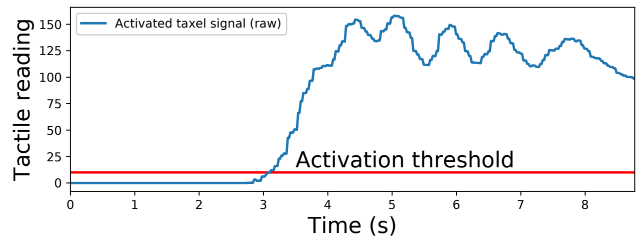

Fig. 1: An example of the raw tactile data generated by a single taxel during a sliding motion. The activation threshold was used to filter inactive taxels.

III. Methodology and Comparison Framework

To test our hypothesis, we developed and evaluated three alternative machine learning (ML) methods, and compared whether using *either* (i) touch or (ii) sliding data, or (iii) a *combination* of both, attained the best performance on a texture classification task. The compared methods comprise both statistical and connectionist techniques, and were chosen to be representative of approaches that have been applied to texture prediction [4], [6], [18]; in particular, we use the Support Vector Machine (SVM), Long-Short-Term Memory network (LSTM) and Convolutional Neural Network (CNN). Note however, that prior approaches have focused on either touch or sliding data; our methods are *able to utilize both*. In the following, we detail the touch/sliding features, and methods developed in our study.

A. Raw Tactile Features

Each model was trained to predict one of C texture classes using tactile data. There are two basic sets of raw tactile features, i.e., those gathered during touch or sliding movements. The raw tactile data obtained from each sensor/taxel i is represented as a discrete time signal $x_i(t)$ (see Fig. 1). Often, tactile sensors comprise multiple taxels and hence, the raw data reading is a set of time-series $X = \{x_i(t)\}_{i=1}^{N_T}$. Depending on the model, the signals were either used in raw form or pre-processed into features, and used to classify C textures. The ground truth texture label for each sample i is denoted y_i.

B. SVM with Hand-Crafted Features

The Support Vector Machine (SVM) is a widely-used supervised learning model that leverages on the kernel trick to perform nonlinear classification [23]. In brief, the SVM learns a representative subset of training points ("support vectors") that maximize the *margin*, i.e., the smallest distance between any sample and the decision boundary induced by the support vectors.

Our first model a multi-class SVM [24] with a standard radial basis function (RBF) kernel. Since the raw tactile features are temporal in nature and have varying lengths, we hand-crafted (HC) fixed-length feature representations similar to those used in prior work [7], [19]:

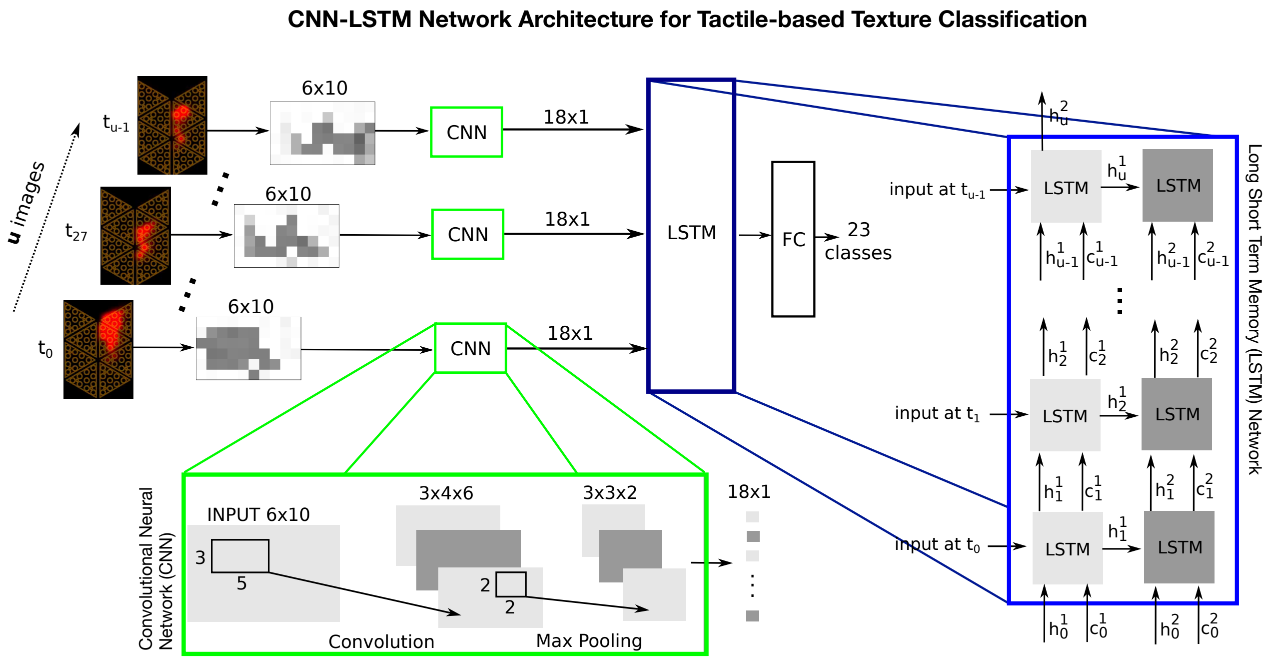

Fig. 2: CNN-LSTM architecture. The tactile readings are used to form "tactile images" that are passed through convolutional layers to obtain a learned intermediate feature representation. These features are fed into an LSTM network at each time step and the last output of a sequence is taken as the texture prediction. Note that the same LSTM network is used in the SVM-LSTM model.

- **Touch-HC** features capture the (approximate) slope of activated taxels during touch movement. The change of the i-th taxel for touching movement over time is

$$\delta_i = \frac{x_i(t_s) - x_i(t_e)}{T}$$

  where t_s and t_e represent the start and end of the touch [19]. We use the empirical mean and standard deviation of δ_i across all the taxels as the touch features.
- **Sliding-HC** features comprise frequency-based and statistical features. More precisely, we compute the Fourier transform of the raw tactile signal and obtain the roughness, fineness, and frequency at maximum intensity [7]. The statistical features are the average mean $\mu = \frac{1}{z}\sum_{i=0}^{i=z}\mu_i$ and standard deviation $\sigma = \frac{1}{z}\sum_{i=0}^{i=z}\sigma_i$ across all activated pixels, where μ_i and σ_i are mean and standard deviation of readings from the i-th taxel during the sliding phase.

Since a majority of taxels do not come into contact with the object and remain inactive during a given motion, we first filtered the taxels using an activation threshold; only taxels with readings above the threshold for a period of 1s were used to compute the features above (see Fig. 1).

C. SVM for Touch and LSTM for Sliding

Unlike touch motions that are relatively quick ($\approx$1s), sliding generally produces far longer sequences ($\approx$7s in our experiments) that may not be well-represented by the hand-crafted features above. As such, our second model replaces the SVM (using Sliding-HC features) with an LSTM network [25], a popular neural-network model designed specifically for time-series data.

Compared to the classic recurrent neural network (RNN), the LSTM learns to control three internal structures (input, output, and forget gates) that modulate what it remembers and forgets. As such, LSTMs can operate over arbitrary time intervals [26]. We used a two-layer LSTM network with 50 LSTM units in each layer. The network was trained to predict the texture classes via a softmax layer from *raw activated* taxel data. We combine both the SVM and LSTM into a single SVM-LSTM model using a heuristic: the SVM is first used to narrow down the potential classes by selecting the top-k classes ordered by predictive posterior probability [27]. The winning class is the one with the highest LSTM output score among the top-k classes ($k = 6$ in our experiments).

D. Representation Learning: CNN and LSTM

Our third model is representative of the modern deep-learning approach, i.e., instead of using any hand-crafted features, we learn feature representations in an end-to-end manner via neural networks. We use the raw taxel data directly as "tactile images" and apply a CNN-LSTM network; each tactile image is a $m \times n$ matrix where each element is the tactile reading from a single taxel (in our experiments, a tactile image was of size 6×10 and represented readings from 60 taxels).

An overview of our network architecture is shown in Fig. 2. To elaborate, our CNN consists of one convolutional layer

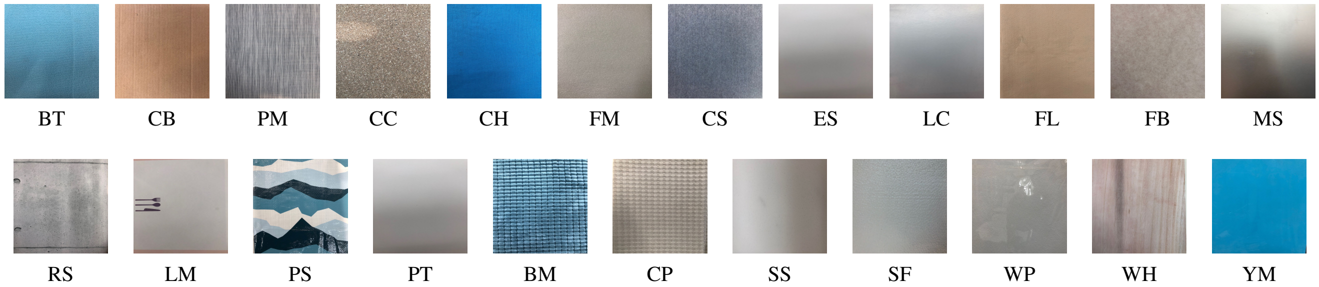

Fig. 3: Snapshots of the 23 materials used in this study: bath towel (BT), cardboard (CB), place sheet (PM), cork (CC), cotton material (CH), cushion foam (FM), denim (CS), Ethylene-vinyl acetate sheet (ES), fake leather (LC), felt (FL), fiberboard (FB), metal sheet (MS), remake sheet(RS), luncheon mat (LM), polypropylene picnic sheet (PS), polypropylene cutting board (PT), bathroom mat (BM), carpet (CP), sponge (SS), styrofoam (SF), thin polypropylene wrapping paper (WP), hard wood (WH), and yoga mat (YM).

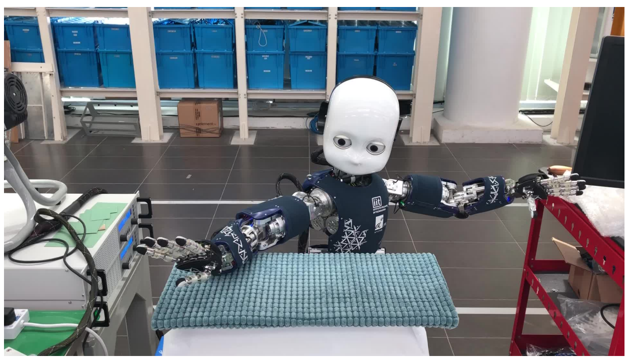

Fig. 4: Experiment setup using the iCub robot. The material is fixed on a non-deformable metal surface and taxels on the iCub forearm are used to collect the tactile data.

(with convolutions of size 3×5) and a max pooling layer (2×2). The resultant output activations are vectorized and passed through Rectified Linear Units (ReLU) to derive a 18×1 vector for *each* tactile image. The output of the CNN is then fed to the LSTM (with the same architecture described in III-C) as the sequence progresses. Unlike the SVM-LSTM model, the CNN and LSTM layers operate together and thus, the same structure was applied regardless of the dataset (touch, slide or both). We take the output at the end of the sequence as the class prediction.

E. Implementation, Training and Evaluation

All our methods were developed in Python; the SVM was implemented using the `scikit-learn` library [28], whilst the LSTM and CNN were developed using PyTorch [29]. SVM hyperparameters were selected using a grid search to maximize the k-fold cross-validation accuracy scores on the training set [30].

The LSTM and CNN were trained to minimize the multi-class cross-entropy loss,

$$\mathscr{L} = -\sum_{i}\sum_{c=1}^{C} \mathbb{I}(c, y_i)\log(p_{i,c})$$

where y_i is ground truth class for sample i, $\mathbb{I}(\cdot,\cdot)$ is a binary 0-1 indicator function, and $p_{i,c}$ is predicted probability of the class c for sample i (obtained from the softmax layer). We use the Adam optimizer [31] for a maximum of 3000 epochs, with a batch size of 23 touch/sliding sequences. A dropout of 0.8 was used in the neural-network models to prevent overfitting.

For each of the aforementioned methods, we trained three different variants: touch data only, sliding data only, and a combination of both touch and sliding data. Note that for the SVM-LSTM model, only the SVM was used to classify touch data, and only the LSTM for sliding. We used accuracy score as our primary comparison measure:

$$\text{Accuracy} = \frac{1}{N}\sum_{i}^{N} \mathbb{I}(\hat{y}_i, y_i) \tag{1}$$

where $\hat{y}_i$ and y_i are the predicted and ground truth class labels for sample i, respectively.

IV. Robot Experiment and Results

In this section, we detail our experiment using an iCub humanoid robot. A wide range of 23 materials (Fig. 3) were selected to better evaluate our key research questions. We describe the iCub tactile sensor, the experimental setup, materials tested, and our principal findings.

A. iCub Tactile Sensor

The iCub is an open source humanoid robot with tendon based actuation and tactile skin on its parts. It contains 18 patches of tactile sensors on its hand, forearm, upper arm and torso. A patch is made of triangular modules, and each module consists of 10 taxels. Each taxel is a capacitive sensor; the dielectric deforms when pressure is applied [32]. Note that the distances between taxels are unequal.

In this work, we used the 60 taxels on the iCub forearm, which are distributed across a curved surface in a loose lattice structure. We chose the forearm to gather sufficient tactile data; the fingertips have far fewer sensors, and the palm does not come into contact well with flat objects due to its concave shape. In our experiment, approximately 20 taxels become active during contact with materials.

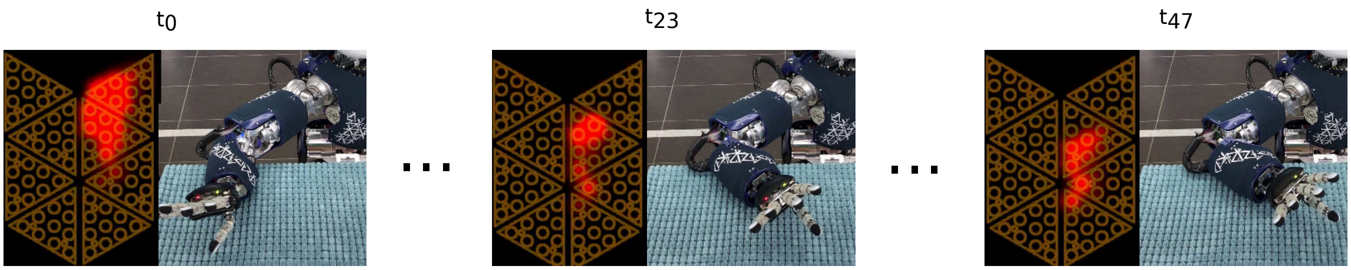

Fig. 5: Tactile sensor readings (red dots,left) during a sliding motion on the bathroom mat (BM) during three time points. Different sets of taxels are activated during the motion, which can provide additional data about the underlying texture.

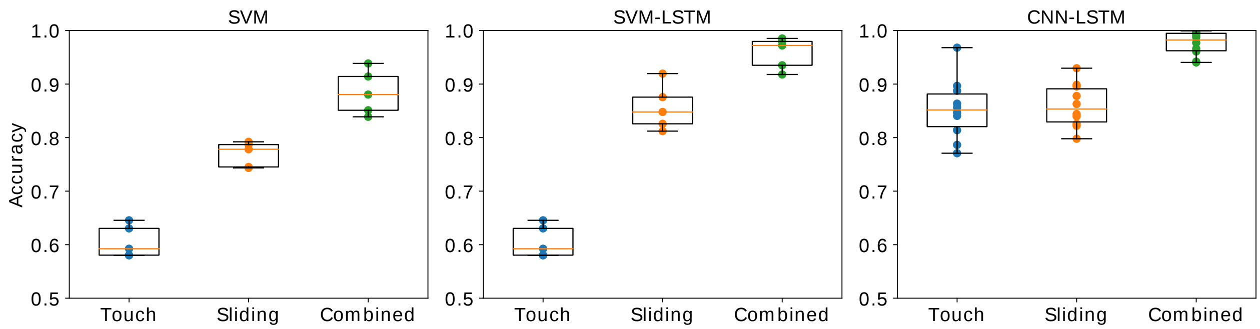

Fig. 6: Boxplots comparing the three methods (SVM, SVM-LSTM, CNN-LSTM) trained on touch data, sliding data, and a combination of both. The models trained with a combination of data perform significantly better relative to the models trained solely on touch or sliding data. The CNN-LSTM achieves the best accuracy scores overall.

B. Experimental Setup

Our experimental setup is illustrated in Fig. 4; each material was laid and secured on top of a non-deformable metal surface. This ensured that the material remained fixed when the forearm touched or slid across the material's surface. Instead of using linear movements via inverse kinematics, movements were performed by varying one degree-of-freedom (DoF) in the iCub joint space. As the iCub is a compliant robot, the movements were not precise and varied between trials; we argue that the data collected in this manner is more realistic and representative of tactile readings during real interactions.

For touch movements, we varied the shoulder joint angle to exert a static touch onto the material; the joint angle was altered from 87° to 93° with angular velocity $1°/s$. For sliding movements, the elbow joint angle changed from 90° to 30° with angular velocity $5°/s$; the forearm slides as the elbow moves. Both backward and forward movements were recorded. Fig. 5 illustrates the activated taxels on the forearm during a sliding movement.

The touch and sliding movements were conducted independently due to technical limitations; the touch movements sometimes caused tendons to snap, which interrupted the experiment. Because the touch and sliding data do not come from the same action movement, the statistical dependency between the two stages is weaker. We posit that models trained under this circumstance should perform equally well (or better) for data collected during a single touch-slide motion. For both movements, data was collected for a single material before moving on to the next. Tactile sensors were calibrated before each procedure using `SkinManager` [33] and the readings were recorded at 50 Hz.

In total, we performed 62 touch movements and 62 sliding movements (31 forwards and 31 backwards) *per material*. As such, dataset comprises 2,852 samples where each sample is a sequence of tactile sensor readings, encoder values, and force estimates obtained from iCub's torque sensor.

TABLE I: Accuracy scores for the three ML models trained with Touch, Sliding, and a Combination of Touch and Sliding data. Standard deviations are shown in brackets.

Method	Touch	Sliding	Combination
SVM	0.61 (0.028)	0.77 (0.019)	0.88 (0.037)
SVM-LSTM	0.61 (0.028)	0.86 (0.035)	0.96 (0.028)
CNN-LSTM	0.85 (0.054)	0.86 (0.038)	0.98 (0.022)

C. Results

Fig. 6 and Tbl. I summarize the results obtained by applying each of the ML models to the final dataset; each model was repeated 10 times on random train-test (80-20) splits. For all three models, the models that use a combination of touch and sliding data achieve the best scores. In fact, the accuracy difference is $\approx 10\%$, which is significant performance gain when using both types of data. These results suggest that different texture properties are obtained during touch and sliding.

Comparing the different approaches, we see that CNN-

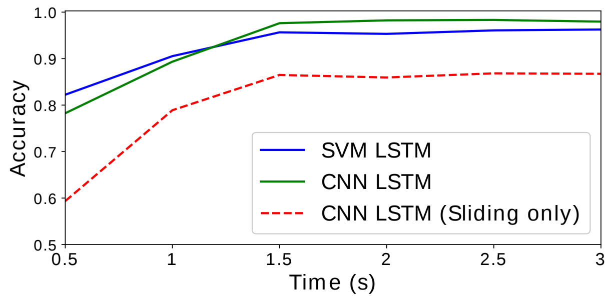

Fig. 7: Model accuracy at different time points. The "sliding only" dashed line represents the CNN-LSTM model using *only* the sliding data. The SVM-LSTM and CNN-LSTM lines denote models trained using both touch and sliding data. Touch data provides an initial boost to performance which is sustained throughout the time sequence.

LSTM model performs the best overall, particularly when trained/tested using only touch data. This finding echoes other recent work (e.g., in computer vision [34], [35]) that learned representations are able to dramatically outperform hand-engineered features. Interestingly, the CNN works well even though the tactile images are generated from taxels that are not spatially uniform. It appears that sufficient structure was present for the model to derive good intermediate feature representations.

Fig. 7 shows the performance of the SVM-LSTM and CNN-LSTM over time. Both models are comparable, with the CNN-LSTM achieving the higher result by the end of the sliding motion. Interestingly, we see the SVM-LSTM achieves marginally higher scores in the beginning; future work could explore why exactly this occurred. Nevertheless, both models achieve peak accuracy early in the sequence ($\approx$ 1.5s), suggesting that short sliding motions are sufficient for accurate texture identification.

The "sliding only" line shows the CNN-LSTM model trained and tested only using sliding data; without the information provided by touch, the model is unable to achieve the performance of the other two models. Again, this suggests that important distinguishing information is sensed by the touch motion that is different from that obtained via sliding.

V. Conclusions and Future work

In this paper, we demonstrated that accurate and robust texture classification for variety of materials is feasible using our proposed hybrid mode of touch and sliding movements. This suggests that tactile data gathered during each of the movements reveals different underlying textural properties.

Within our comparison framework, three models were developed and compared from a performance standpoint; the connectionist model that uses both the CNN and LSTM achieved a high accuracy of 98% on our dataset of 23 textures. Further analysis revealed that only a short sliding motion was required to isolate the correct texture class. We have made our tactile dataset (comprising tactile sensor readings, encoder data, and force estimates) publicly available online at `https://github.com/crslab/TactileLearning`.

Moving forward, we posit that other tasks may also benefit from a hybridization of tactile modes, e.g., object identification and grasp stability prediction. Future work may also examine alternative sensors and models for accurate texture identification; although the CNN-LSTM achieved excellent scores, we believe further improvement is possible and future developments would bring robots closer to (or surpass) human-level performance.

Acknowledgements

This research is partially supported by the Agency for Science, Technology and Research (A*STAR) under its AME Programmatic Funding Scheme (Project #A18A2b0046) and the SINGA(A*STAR) Int'l. Award. We acknowledge partial funding from the NRF White Space: National Robotics Programme of Singapore (Grant #1722500063).

References

[1] W. W. Mayol-Cuevas, J. Juarez-Guerrero, and S. Munoz-Gutierrez, "A first approach to tactile texture recognition," in *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218)*, vol. 5. IEEE, 1998, pp. 4246–4250.

[2] R. Li and E. H. Adelson, "Sensing and recognizing surface textures using a gelsight sensor," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1241–1247.

[3] E. A. A. G. C. R. F. Shan Luo, Wenzhen Yuan, "Vitac: Feature sharing between vision and tactile sensing for cloth texture recognition," in *Robotics and Automation (ICRA), 2018 IEEE International Conference on*, 2018, pp. 2722–2727.

[4] S. W. E. A. Wenzhen Yuan, Yuchen Mo, "Active clothing material perception using tactile sensing and deep learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 4842–4849.

[5] S. Kuroki, M. Sawayama, and S. Nishida, "Haptic texture perception on 3d-printed surfaces transcribed from visual natural textures," in *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 2018, pp. 102–112.

[6] N. Jamali and C. Sammut, "Majority voting: Material classification by tactile sensing using surface texture," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 508–521, 2011.

[7] J. A. Fishel and G. E. Loeb, "Bayesian exploration for intelligent identification of textures," *Frontiers in neurorobotics*, vol. 6, p. 4, 2012.

[8] H. Soh and Y. Demiris, "Incrementally learning objects by touch: Online discriminative and generative models for tactile-based recognition," *IEEE transactions on haptics*, no. 4, pp. 512–525, 2014.

[9] A. Vásquez, Z. Kappassov, and V. Perdereau, "In-hand object shape identification using invariant proprioceptive signatures," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 965–970.

[10] H. Soh and Y. Demiris, "Spatio-temporal learning with the online finite and infinite echo-state gaussian processes," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 3, pp. 522–536, 2015.

[11] Z. Kappassov, D. Baimukashev, O. Adiyatov, S. Salakchinov, Y. Massalin, and H. A. Varol, "A series elastic tactile sensing array for tactile exploration of deformable and rigid objects," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 520–525.

[12] Z. Kappassov, J.-A. Corrales, and V. Perdereau, "Tactile sensing in dexterous robot hands," *Robotics and Autonomous Systems*, vol. 74, pp. 195–220, 2015.

[13] Y. Su, Y. Wu, K. Lee, Z. Du, and Y. Demiris, "Robust grasping for an under-actuated anthropomorphic hand under object position uncertainty," in *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*. IEEE, 2012, pp. 719–725.

[14] Y. Wu, Y. Su, and Y. Demiris, "A morphable template framework for robot learning by demonstration: Integrating one-shot and incremental learning approaches," *Robotics and Autonomous Systems*, vol. 62, no. 10, pp. 1517–1530, 2014.

[15] Y. Bekiroglu, J. Laaksonen, J. A. Jørgensen, V. Kyrki, and D. Kragic, "Assessing grasp stability based on learning and haptic data," *IEEE Transactions on Robotics*, vol. 27, no. 3, p. 616, 2011.

[16] Y. Su, Y. Wu, H. Soh, Z. Du, and Y. Demiris, "Enhanced kinematic model for dexterous manipulation with an underactuated hand," in *Proceedings of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS 2013.*, Nov 2013, pp. 2493–2499.

[17] Z. Yi, Y. Zhang, and J. Peters, "Bioinspired tactile sensor for surface roughness discrimination," *Sensors and Actuators A: Physical*, vol. 255, pp. 46–53, 2017.

[18] S. S. Baishya and B. Bäuml, "Robust material classification with a tactile skin using deep learning," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 8–15.

[19] D. Xu, G. E. Loeb, and J. A. Fishel, "Tactile identification of objects using bayesian exploration," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3056–3061.

[20] T. Kohonen, "Learning vector quantization," in *Self-organizing maps*. Springer, 1995, pp. 175–189.

[21] J. Hoelscher, J. Peters, and T. Hermans, "Evaluation of tactile feature extraction for interactive object recognition." in *Humanoids*, 2015, pp. 310–317.

[22] A. I. Weber, H. P. Saal, J. D. Lieber, J.-W. Cheng, L. R. Manfredi, J. F. Dammann, and S. J. Bensmaia, "Spatial and temporal codes mediate the tactile perception of natural textures," *Proceedings of the National Academy of Sciences*, vol. 110, no. 42, pp. 17 107–17 112, 2013.

[23] B. Schölkopf, A. J. Smola, F. Bach *et al.*, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[24] J. Weston and C. Watkins, "Multi-class support vector machines," Citeseer, Tech. Rep., 1998.

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[26] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," in *IET Conference Proceedings*. IET, 1999, pp. 850–855(5).

[27] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

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

[29] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS 2017 Autodiff Workshop*, 2017.

[30] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler *et al.*, "Api design for machine learning software: experiences from the scikit-learn project," *arXiv preprint arXiv:1309.0238*, 2013.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[32] P. Maiolino, M. Maggiali, G. Cannata, G. Metta, and L. Natale, "A flexible and robust large scale capacitive tactile system for robots," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3910–3917, 2013.

[33] A. Del Prete, F. Nori, G. Metta, and L. Natale, "Control of contact forces: The role of tactile feedback for contact localization," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 4048–4053.

[34] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, "Action recognition in video sequences using deep bi-directional lstm with cnn features," *IEEE Access*, vol. 6, pp. 1155–1166, 2018.

[35] S. Yeung, O. Russakovsky, N. Jin, M. Andriluka, G. Mori, and L. Fei-Fei, "Every moment counts: Dense detailed labeling of actions in complex videos," *International Journal of Computer Vision*, vol. 126, no. 2-4, pp. 375–389, 2018.