Tactile Active Texture Recognition With Vision-Based Tactile Sensors

Alina Böhm[1]
Tim Schneider[1]
Alap Kshirsagar[1]
[1]TU Darmstadt, Germany

Boris Belousov[2]
Jan Peters[1,2,3,4]
Constantin A. Rothkopf[1,3,4]
[2]DFKI, [3]Hessian.AI, [4]CogSci Centre

Lisa Lin[5]
Katja Doerschner[5]
Knut Drewing[5]
[5]University of Giessen

Abstract

This paper investigates active sensing strategies that employ vision-based tactile sensors for robotic perception and classification of fabric textures. We formalize the active sampling problem in the context of tactile fabric recognition and provide an implementation of information-theoretic exploration strategies based on minimizing predictive entropy and variance of probabilistic neural network classifiers. By evaluating our method on a real robotic system, we find that the choice of the active exploration strategy has a relatively minor influence on the recognition accuracy as long as the objects are touched more than once. In a comparison study, while humans achieve 66.9% recognition accuracy, our best approach reaches 90.0%, showing that vision-based tactile sensors are highly effective for fabric recognition.

Classification of fabrics has been tackled with different types of sensors using both supervised and active methods. *Vibration/force-based tactile sensors* of different types—such as the iCub sensors [1, 2, 3], BioTac sensors [2, 3], and custom-designed ones [4, 5, 6]—have been used for supervised texture classification, using spiking neural networks [2], modified RNNs [3], and k-NN classifiers [6]. All these methods rely on high-frequency temporal data, requiring RNNs or spatio-temporal subsampling to keep the input dimensionality low. In contrast, *vision-based tactile sensors* provide high-resolution data but at a lower rate, thereby requiring less history as input. Supervised classification of fabrics was successfully showcased using GelSight heightmap patterns [7] and more advanced spatio-temporal attention features [8]. Furthermore, *active sampling methods* have been developed for GelSight to 'actively' collect data, e.g., repeating touches until a 'good' tactile image is obtained [9], or for material roughness classification [10], where predictions on image patches were weighted by the output variance of a Bayesian CNN to improve the overall label prediction accuracy.

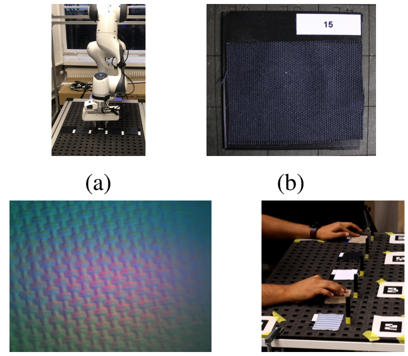
In this paper, we tackle the problem of *tactile active texture recognition* (see Fig. 1): with no pre-training, a robot is given a 'reference' texture and asked to identify it among four comparison textures using as few touches as possible. Unlike [9], we do not want to pre-train on a large dataset but rather quickly adapt on-the-fly, and we do not aim to 'classify' but only 'recognize' fabrics. In contrast to [10], we do not use uncertainty for label prediction but rather for action selection. In the next sections, we formalize the tactile active texture recognition problem, present a general Bayesian decision-theoretic framework for action selection, describe our implementation which leverages probabilistic NNs for uncertainty quantification, and provide extensive empirical studies, including the comparison to human exploration strategies and experiments on a real robot.

1 Problem Setup and Task Formalization

We investigate sample-efficient texture recognition using vision-based tactile sensors such as GelSight Mini [11], Digit [12], or FingerVision [13]. In our setup (Fig. 1), a GelSight Mini sensor is held by a Franka Panda arm [14] and pressed against pieces of fabrics on plastic platforms at predefined locations with randomized amounts of pressure and rotation around the vertical axis to provide more variability in the data. The leftmost platform holds the *reference texture*, while the remaining four platforms hold randomly chosen *comparison textures*, one of which is equal to the reference.

NeurIPS 2023 Workshop on Touch Processing: a new Sensing Modality for AI .

The agent's goal is to identify the reference among the comparison textures using as few touches as possible. Crucially, the agent has no prior knowledge of the textures, and therefore has to learn about them within one *trial*, i.e., one fixed selection of five fabrics in a particular order. One trial consists of multiple touches and ends after a predefined number of touches in our robot experiments or once the participant has made a decision in our human study. The *action* of the agent is the high-level choice which platform to approach next (the low-level robot control is handled by a Cartesian position controller). We call each step of this action-observation loop a *round*, and we start counting rounds after each object has been touched once, i.e., if the process has terminated after one round, it means the agent has touched all four comparison fabrics and the reference fabric once and then did just one extra touch. Thus, one trial consist of several rounds (up to 20). We do multiple trials with different textures, and multiple *runs* for each trial to reduce the statistical error.



(a) (b)
(c) (d)

Figure 1: The texture recognition task requires identifying a given fabric among four comparison samples: (a) robot arm exploring fabrics; (b) example denim fabric; (c) corresponding tactile image; (d) human participant using index fingers to compare fabric samples.

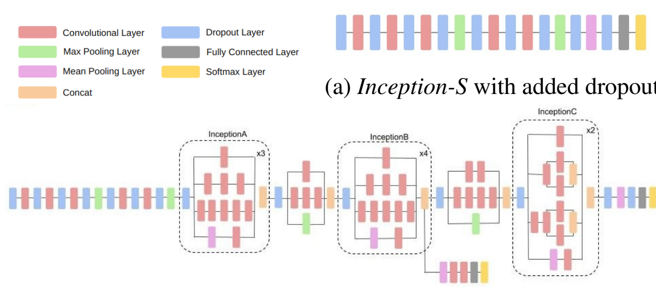
To provide textures for our experiments, we created a dataset of 25 cotton fabrics, chosen to be particularly hard to distinguish by touch, as confirmed by our human study in Sec. 3.2. For each fabric, we collected 200 *samples* with randomly perturbed positions and rotations around the vertical axis. A sample of this dataset can be seen in Fig. 1c. Our complete dataset is available online.

2 Tactile Active Texture Recognition Method

Consider one round of the agent's decision making. Having touched each of the five platforms one or more times, the agent needs to make a decision which platform to touch next. The Bayesian approach to this problem is to build a probabilistic model and to choose the action that provides the most information to support the final decision (i.e., the decision which fabric is identical to the reference) [15]. To implement this approach, we specify the model, describe how it is updated using the new data, and define the *acquisition function*, i.e., the action selection strategy.

2.1 Probabilistic Model Specification

We employ a CNN with dropout layers to implement a probabilistic classifier [16], as dropout was shown to provide a viable approach for uncertainty quantification with neural networks [17]. We consider three CNN variants: i) Inception-v3 [18] pre-trained on ImageNet [19] (*Inception-PT*); ii) randomly initialized Inception-v3 (*Inception-RI*); iii) small unpretrained Inception-v3 (*Inception-S*), which drops all the layers between the first *InceptionA* and the last *InceptionC* blocks (see Fig. 2). Considering these network variants allows us to evaluate the effects of pretraining and the network depth.



(a) *Inception-S* with added dropout

(b) Inception-v3 with added dropout

Figure 2: The considered architectures of the probabilistic classifier: Inception-v3 and small Inception-v3 (*Inception-S*) with dropout.

2.2 Model Update

Once a new tactile image is obtained, the model needs to be updated to incorporate the new evidence. As is common in deep learning, we employ *data augmentation* [20], by generating 10 randomly rotated versions of the same tactile image. Using all the samples collected during the current trial, we retrain the probabilistic NN classifier: the samples of the comparison textures serve as inputs and the respective platform positions serve as labels. The output of the classifier is a probability distribution $p_\theta(i|o)$ over the platform labels $i \in \{1, 2, 3, 4\}$, given an image o and the model parameters θ. Hence,

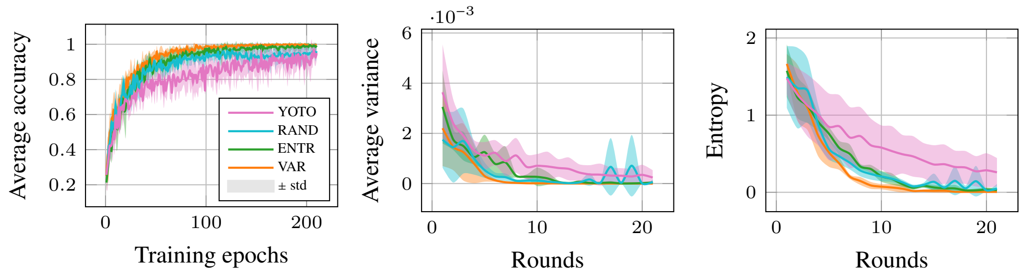


Figure 3: Comparison of the exploration strategies on the *tactile active texture recognition* task. Average prediction accuracy, average variance, and entropy of the predictions are shown. *Inception-S* is used in all experiments. The *Variance* strategy achieves the highest accuracy, closely followed by *Entropy* and *Random*. Interestingly, the *Variance* strategy leads to a faster entropy decrease even than *Entropy* (rightmost plot).

the model learns to map the texture samples to the platform labels. When queried with the reference texture (unseen during training), the model outputs a 'probability distribution' over the labels. To obtain a more robust estimate, we apply the model to $n_{\text{ref}} = 10$ randomly rotated copies o_k^{ref} of the reference image and average the probabilities, $i^* = \arg\max_i \frac{1}{n_{\text{ref}}} \sum_{k=1}^{n_{\text{ref}}} p_\theta(i|o_k^{\text{ref}})$.

2.3 Active Sample Selection Strategy

The decision which platform to explore next is made based on the model uncertainty. As described in Sec. 2.1, we add dropout layers to Inception-v3 (see Fig. 2) to model the epistemic uncertainty [17]. By querying the dropout network with the same input multiple times, we obtain different output samples and can gauge the uncertainty by their distribution. We compare four sample selection strategies. i) *Random* strategy is a non-active baseline that selects the next texture to touch according to a uniform distribution. ii) *Variance* strategy selects the platform for which the variance of the class probability predictions is the highest, $i_{\text{next}} = \arg\max_i \frac{1}{n_{\text{ref}}} \sum_{k=1}^{n_{\text{ref}}} \text{Var}_{m\sim p(m)}[p_\theta(i|o_k^{\text{ref}}, m)]$, where $p(m)$ is the distribution of the dropout masks. iii) *Entropy* strategy selects the platform that contributes the most to the class distribution entropy for the reference object $i_{\text{next}} = \arg\max_i \frac{1}{n_{\text{ref}}} \sum_{k=1}^{n_{\text{ref}}} \mathbb{E}_{m\sim p(m)}[-p_i^k \ln p_i^k]$ where $p_i^k \coloneqq p_\theta(i|o_k^{\text{ref}}, m)$. iv) *You Only Touch Once (YOTO)* is a trivial baseline that does not sample at all after the initial five touches, i.e., each object touched once. This baseline provides a reference to quantify the 'value' of the actively gathered data.

3 Experimental Results

Our experiments aim at identifying what components of the algorithmic architecture matter for active texture recognition with vision-based tactile sensors. For that, we evaluate model architectures and the active sample selection strategies, and subsequently present a human study which investigates human exploration strategies on the same task.

3.1 Active Texture Recognition

We compare the three network architectures (Sec. 2.1) and the four active sampling strategies (Sec. 2.3). Each model is trained for 210 epochs, 10 epochs after creating a baseline and then 10 more epochs after resampling in each of the 20 rounds. For all three models, we collect the results of five subsets of fabrics using the four different strategies and average the performance of each model. Figure 4 shows that the models perform similarly. We believe the reason why pre-training is not advantageous in this case is the retrospective addition of dropout layers, which the model was not trained for. Since *Inception-S* shows strong performance, we employ it in our further experiments. In Fig. 3, the influence of the exploration strategies is investigated. The model performs best using *Variance*, closely followed by *Entropy* and *Random* strategies. While *YOTO* performs quite well on the training data after 20 rounds, its test prediction accuracy is only 80% on average.

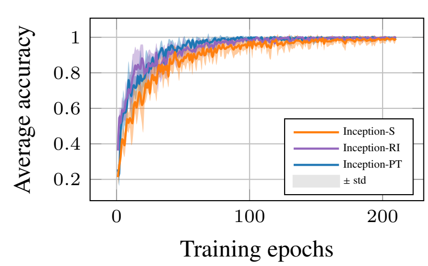


Figure 4: Comparing different Inception-v3 models on the active texture recognition task. Notably, the small *Inception-S* network performs as well as the larger *Inception-PT*.

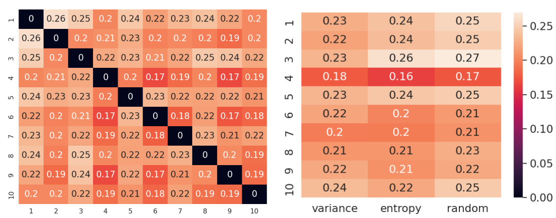
3.2 Human Study

To find out how well the proposed tactile active recognition method performs compared to humans, we carry out an experiment with ten human participants. We attempt to characterize human exploration strategies using information-theoretic metrics, and see whether insights from humans can be transferred to robots. During a trial, the participants are blindfolded and can only use their fingertips to explore the fabric surfaces (see Fig. 1d). They indicate their response by resting their finger on the chosen fabric. The data of this experiment and the code for analysis are available online. Table 1 compares the prediction accuracies of the human participants and the robot. For fair comparison, we only allow the robot to use the same number of touches that humans used.

Humans	*Variance*	*Entropy*	*Random*	*YOTO*
66.88% ±16.93%	**90.00%** ±15.24%	88.13% ±14.24%	89.38% ±14.35%	80.63% ±22.42%

Table 1: Comparison of the final accuracies achieved by the different exploration strategies. Notably, all robotic strategies are superior to humans, showing that the vision-based tactile sensor alone provides an advantage over the human touch in this task. The exploration strategy plays a relatively minor role.

3.3 Behavior Comparison Between Participants and the Robot



(a) inter-participant (b) participant-robot

Figure 5: Comparing the exploration strategies among participants and against the information-theoretic strategies. The numbers indicate the JSD between the distributions of time spent over objects, averaged over trials. The inter-subject variability is comparable to the subject-robot variability, therefore no uniform judgement about what strategy all humans use can be made.

To compare the exploration strategies, we formalize the problem by normalizing the time spent by the human participants on each object per trial, to get a distribution of relative times per fabric. Subsequently, we employ the symmetric Jensen-Shannon divergence (JSD) to compare these time distributions, thereby comparing human and robotic strategies at least in this restricted sense. The JSD takes values in the range $[0, 1]$. The resulting distances of comparing the robotic strategies to each other are 0.14 between *Variance* and *Entropy*, 0.12 between *Entropy* and *Random*, and 0.16 between *Variance* and *Random*. Thus, the *Entropy* strategy appears more similar to *Random* than to *Variance*. While the distances between the robot strategies are in the range 0.12–0.16, the inter-participant (Fig. 5a) and the participant-robot (Fig. 5b) distances are around 0.2 and higher.

4 Discussion & Conclusion

We have investigated the performance of a Bayesian approach to active sampling for fabric texture recognition with vision-based tactile sensors. *First*, on our four-class recognition task, where the network needs to adapt quickly with only a handful of training samples, we found that a smaller model performs as good as the ImageNet-pretrained Inception-v3 (see Fig. 4). Thus, we conclude that big pretrained networks are not necessary for few-shot recognition tasks. *Second*, we did not find a significant difference between the considered exploration strategies. The random sampling strategy showing similar performance suggests that our texture recognition task is relatively straightforward for the vision-based tactile sensors, as opposed to humans using only one finger. These results are in agreement with [10], where GelSight was shown to achieve a higher accuracy on material roughness classification than human participants. *Third*, in the human study, we found that there is a significant variability among the participants with regards to the exploration strategy (see Fig. 5). The inter-participant variability was found to be similar to the participant-robot variability, meaning that there is no universal exploration strategy that all participants have followed. On average, human exploration behavior was closer to the information-theoretic strategies than to random exploration.

Limitations. Comparing human and robotic tactile perception is limited due to the different nature of the sensors. However, this concern was addressed in [10], where human performance using touch was compared to using GelSight images for material roughness classification. Humans were found to be much better at classification using their sense of touch rather than vision. Our results further suggest that the internal representations may be more important than the exploration strategy. Lastly, we only tested our method on four comparison fabrics. However, we expect it to scale well to more textures.

Acknowledgements. We acknowledge the grants *Einrichtung eines Labors des Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) an der Technischen Universität Darmstadt* and *The Adaptive Mind* of the Hessisches Ministerium für Wissenschaft und Kunst. The authors gratefully acknowledge the computing time provided to them on the high-performance computer Lichtenberg at the NHR Centers NHR4CES at TU Darmstadt. This is funded by the Federal Ministry of Education and Research, and the state governments participating on the basis of the resolutions of the GWK for national high performance computing at universities (`www.nhr-verein.de/unsere-partner`).

References

[1] T. Taunyazov, H. F. Koh, Y. Wu, C. Cai, and H. Soh, "Towards effective tactile identification of textures using a hybrid touch approach," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 4269–4275.

[2] T. Taunyazov, Y. Chua, R. Gao, H. Soh, and Y. Wu, "Fast texture classification using tactile neural coding and spiking neural network," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9890–9895.

[3] R. Gao, T. Tian, Z. Lin, and Y. Wu, "On explainability and sensor-adaptability of a robot tactile texture representation using a two-stage recurrent networks," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1296–1303.

[4] B. M. R. Lima, V. P. da Fonseca, T. E. A. de Oliveira, Q. Zhu, and E. M. Petriu, "Dynamic tactile exploration for texture classification using a miniaturized multi-modal tactile sensor and machine learning," in *2020 IEEE International Systems Conference (SysCon)*. IEEE, 2020, pp. 1–7.

[5] S. Huang and H. Wu, "Texture recognition based on perception data from a bionic tactile sensor," *Sensors*, vol. 21, no. 15, p. 5224, 2021.

[6] S.-a. Wang, A. Albini, P. Maiolino, F. Mastrogiovanni, and G. Cannata, "Fabric classification using a finger-shaped tactile sensor via robotic sliding," *Frontiers in Neurorobotics*, vol. 16, p. 10, 2022.

[7] 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 (CVPR)*, 2013, pp. 1241–1247.

[8] G. Cao, Y. Zhou, D. Bollegala, and S. Luo, "Spatio-temporal attention model for tactile texture recognition," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9896–9902.

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

[10] A. Amini, J. I. Lipton, and D. Rus, "Uncertainty aware texture classification and mapping using soft tactile sensors," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 4249–4256.

[11] "GelSight Mini System - GelSight," Apr. 2023, [Online; accessed 1. Jun. 2023]. [Online]. Available: https://www.gelsight.com/product/gelsight-mini-system

[12] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer *et al.*, "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3838–3845, 2020.

[13] A. Yamaguchi and C. G. Atkeson, "Recent progress in tactile sensing and sensors for robotic manipulation: can we turn tactile sensing into vision?" *Advanced Robotics*, vol. 33, no. 14, pp. 661–673, 2019.

[14] "Franka Emika - Next Generation Robotics." May 2023, [Online; accessed 1. Jun. 2023]. [Online]. Available: https://www.franka.de

[15] B. Settles, "Active learning literature survey," *Machine Learning*, vol. 15, no. 2, pp. 201–221, 1994.

[16] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, 2021.

[17] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *International Conference on Machine Learning (ICML)*. PMLR, 2016, pp. 1050–1059.

[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2009, pp. 248–255.

[20] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.