Towards Tactile Texture Classification Using Low Spectral Resolution Analysis with Multi-Probes

Edwin Wu*, Louis Yu†, Shivaji Mallela‡, Aryavardhan Singh‡, Aditya Singh‡, Lei Wu‡, Oleg Favorov§, Olcay Kursun‡
*Montgomery Academy, Auburn University at Montgomery, Montgomery, AL, USA
ewu@aum.edu
†Department of Computer Science, Columbia University, New York, NY, USA
ly2640@columbia.edu
‡Department of Computer Science, Auburn University at Montgomery, Montgomery, AL, USA
{smallel3, lwu, okursun}@aum.edu
§Department of Biomedical Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
favorov@email.unc.edu

***Abstract*—This study explores two methods for classifying a tactile texture dataset: a deep learning approach using convolutional neural networks (CNNs) and a Fast Fourier Transform (FFT)-based method. The CNN model performed well with longer signals, while the FFT-based method proved effective when combined with careful bin selection to reduce dimensionality. Furthermore, multiple probes improved the FFT approach, whereas the CNN maintained high performance even with a single probe. These findings offer practical insights for developing efficient, real-time IoT solutions and assistive tools.**

***Index Terms*—Tactile perception, AI deep learning, Fourier Transform, texture classification, IoT assistive devices**

I. Introduction

Tactile perception is essential for enabling robotics, prosthetics, and assistive devices to interact intelligently with their environments [1]–[4]. Advances in hardware and VLSI technology have enhanced AI-powered IoT and edge devices, expanding their use in various applications, such as telehealth, prosthetics, and robotics [4]–[6]. These devices are particularly valuable for assistive tools that improve the quality of life for the elderly, offering capabilities like remote palpation for skin cancer diagnosis and surface sensing, as well as many other tasks that require accurate, non-contact material perception. This technology provides greater safety and independence for individuals with mobility or sensory impairments, while also benefiting a range of other industries. In this study, we focus on the classification of various textures using tactile signals obtained through sensors attached to a single-point contact probe. These probes can be used to record surface parameters that the elderly, particularly those with disabilities, can analyze before making direct contact. We utilized a previously collected dataset, VibTac-12 [4], and aim to expand it with additional sensors and probes to further develop this technology. This research explores two approaches for classifying these textures: (1) a deep learning approach and (2) a frequency-domain analysis using the Fast Fourier Transform (FFT). We highlight how FFT-based approaches can particularly benefit from multiple probes, making them more advantageous than deep learning in certain configurations. This interplay between multiple probes, each one with sensors with different frequency responses is analogous to hyperspectral imaging, where different wavelengths capture distinct spectral information, enabling richer feature extraction.

This work was supported by National Science Foundation grants 2003740 and 2435093.

II. Materials and Methods

A. Dataset

VibTac-12 consists of vibrotactile signals recorded from 12 different textures, such as sandpapers, Velcro, aluminum foil, and rubber bands. There are 12 accelerometer signals, each corresponding to a different texture, and each signal contains three distinct channels (X, Y, and Z). The signals are captured using a 3D accelerometer sampled at 200 Hz. We extracted a few (5, 10, 20, 40) 2-second snips from each signal for training and a larger set (100) of 2-second snips for testing. The training snips were selected from the first half of the signals (the first 10 seconds), while the test snips were selected from the second half. For 40 snips per texture, this results in a dataset of size $(12 \times 40\,\text{snips}, 2 \times 200\,\text{time steps}, 3\,\text{channels})$. For exploring the performances on a 3-probe version of this dataset (see Figure 1), we created a simulated dataset with a size of $(12 \times 40\,\text{snips}, 1.8 \times 200\,\text{time steps}, 3 \times 3\,\text{channels})$. In this setup, the three probes are spaced 0.1 seconds apart, with each probe rubbing against the surface for 1.8 second. This configuration still covers the same 2-second window (1.8 + 0.1 + 0.1 = 2.0) as in the single-probe version.

B. FFT-Based Approaches

FFT is applied to convert the time-domain signals into the frequency domain [4]. The frequency spectrum was divided into bins, and the magnitudes (square root of powers) within each bin were summed. This binned representation provided a lower-dimensional feature vector for each sample, which was used for classification as a cost-effective alternative to the deep learning approach. Linear SVMs is used for classification [4].

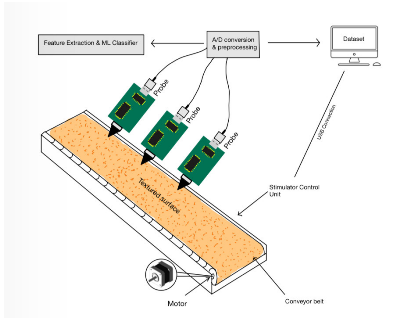

Fig. 1. Simulated 3-probe data based on the real one-probe dataset [4].

C. Deep Learning Based Approaches

In the experiment, two deep learning models, a CNN and an LSTM, were trained for 200 epochs to classify data into 12 categories. The CNN model used two Conv1D layers followed by MaxPooling, GlobalAveragePooling, and Dropout layers, while the LSTM model employed a single LSTM layer with a dropout mechanism and fully connected layers for classification (with the Adam optimizer and categorical crossentropy loss).

III. Experimental Results

As shown in Table I, using longer signals and extracting more snips led to higher accuracy, particularly for the CNN approach. This improvement is likely due to the regularization effect of weight sharing in CNNs, which tends to perform better with larger datasets. For the FFT-based method, we applied binning, and as shown in Figure 2, applying FFT followed by a linear SVM yielded comparable performance. Binning reduced the feature dimensionality from 200 to a smaller number, such as 12, by grouping frequency components into broader ranges and improved the generalization by representing the data without needing fine-grained frequency resolution.

TABLE I
Accuracy Across Different Classifiers and Number of Snips Per Class for Snip durations of 1 and 2 seconds

Length	Snips	*CNN*	*LSTM*	*FFT+SVM*	*FFT+RF*
1 sec	5	$78\% \pm 3\%$	$57\% \pm 5\%$	$75\% \pm 4\%$	$60\% \pm 2\%$
1 sec	10	$88\% \pm 2\%$	$70\% \pm 4\%$	$82\% \pm 2\%$	$68\% \pm 2\%$
1 sec	20	$92\% \pm 1\%$	$79\% \pm 3\%$	$86\% \pm 1\%$	$76\% \pm 2\%$
1 sec	40	$95\% \pm 1\%$	$82\% \pm 3\%$	$87\% \pm 1\%$	$79\% \pm 2\%$
2 sec	5	$86\% \pm 3\%$	$54\% \pm 6\%$	$85\% \pm 3\%$	$64\% \pm 6\%$
2 sec	10	$95\% \pm 1\%$	$73\% \pm 3\%$	$89\% \pm 2\%$	$74\% \pm 3\%$
2 sec	20	$96\% \pm 0\%$	$82\% \pm 4\%$	$93\% \pm 1\%$	$79\% \pm 2\%$
2 sec	40	$98\% \pm 0\%$	$88\% \pm 2\%$	$94\% \pm 1\%$	$83\% \pm 2\%$

IV. Conclusions

In this study, we evaluated two approaches for classifying tactile data from the VibTac-12 dataset: a deep learning model using a CNN and an FFT-based method followed by linear SVM. Our results show that for longer signals and larger datasets, the CNN approach consistently achieved higher accuracy, benefiting from the inherent regularization provided by weight sharing. In fact, for single-probe data with longer snips, deep learning was able to squeeze the maximum performance,

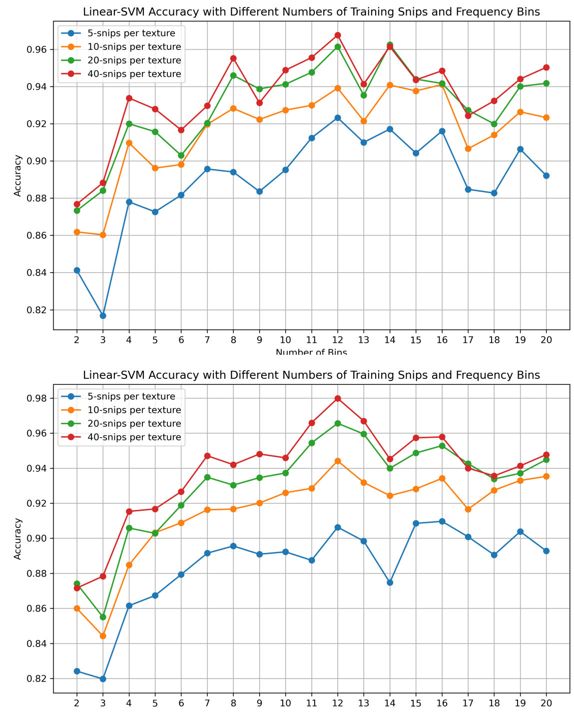

Fig. 2. SVM accuracies for various bin sizes: Top - Single-probe data; Bottom - Simulated 3-probe data.

achieving high classification accuracy without the need for additional probes. However, for shorter snips or when fewer examples are available, the FFT approach with careful bin selection proved to be a strong alternative. By reducing the dimensionality of the frequency representation, FFT-based classification provided a simpler yet effective solution in these cases. FFT benefits from the increased spatial resolution provided by multiple probes and combining it with the domain knowledge embedded in selecting optimal frequency bins can offer a more computationally efficient option for smaller datasets or when real-time processing is needed.

References

[1] M. Pulcinelli, L. Zoboli, F. De Tommasi, C. Massaroni, V. Altomare, A. Grasso, A. Gizzi, E. Schena, and D. Lo Presti, “A multi-sensor tactile system based on fiber bragg grating sensors for soft tissue palpation,” *IEEE Sensors Journal*, vol. 24, no. 16, pp. 25 205–25 214, 2024.

[2] H. Park, K. Park, S. Mo, and J. Kim, “Deep neural network based electrical impedance tomographic sensing methodology for large-area robotic tactile sensing,” *IEEE Transactions on Robotics*, vol. 37, no. 5, pp. 1570–1583, 2021.

[3] M. S. K, S. Malayil, D. AV, F. R. TP, and M. P. Abdul Kabeer, “Affordable navigation aid: A smart walking stick for the visually impaired,” in *2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 2024, pp. 1–5.

[4] O. Kursun and A. Patooghy, “An embedded system for collection and real-time classification of a tactile dataset,” *IEEE Access*, vol. 8, p. 97462–73, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2996576.

[5] Z. Hu, L. Lin, W. Lin, Y. Xu, X. Xia, Z. Peng, Z. Sun, and Z. Wang, “Machine learning for tactile perception: Advancements, challenges, and opportunities,” *Advanced Intelligent Systems*, p. 2200371, 2023.

[6] O. Kursun, A. Patooghy, P. Poursani, and O. V. Favorov, “Cg-cnn: Self-supervised feature extraction through contextual guidance and transfer learning,” *IEEE Access*, vol. 12, pp. 155 851–155 866, 2024.