Cluster Haptic Texture Database: Haptic Texture Database with Controlled Sliding-Contact Interactions

Michikuni Eguchi[1,2], Yuichi Hiroi[2], *Member, IEEE*, Tomohiro Hayase[2], and Takefumi Hiraki[1,2], *Member, IEEE*

I. Introduction

Human perception integrates information from multiple sensory modalities. In tactile perception, both kinesthesia and touch are crucial for understanding object properties [1], making haptic interfaces valuable for system enhancement. Visual and auditory inputs also influence surface perception, highlighting the potential of datasets that integrate these modalities. Against this background, the creation of comprehensive multimodal haptic datasets [2]–[5] has been actively pursued to advance texture perception research and the development of sophisticated haptic interfaces.

Many prior studies have recorded visual, auditory, and tactile data by freely exploring textured surfaces with rigid probes [2]–[4] or human fingers [5]. However, to our knowledge, no existing dataset simultaneously captures visual, auditory, and tactile modalities with controlled sliding velocity and direction. Sliding velocity and direction significantly influence tactile signals [1]. Precise control of these factors is essential to ensure reliable analysis. Analyzing how acoustic and tactile signals vary with sliding direction and velocity will advance perception research.

Therefore, we propose the Cluster Haptic Texture Database, a multimodal dataset of haptic signals from an artificial fingertip sliding over surfaces (Fig. 1). The dataset includes sound and acceleration data recorded at five velocities, eight directions, and two pressing forces for 118 textures. We conducted classification experiments for texture, velocity, and direction using our dataset to demonstrate that our dataset contains sufficient discriminative features.

II. Methods

A. Texture Data Measurement System

We constructed a sliding-interaction measurement system (Fig. 2a) by repurposing a 3D printer (Ender-3 S1, Creality) as an XYZ stage. We attached a cylindrical urethane rubber fingertip with Shore A15 hardness and 15 mm diameter to the moving part of the XYZ stage. We utilized two nondirectional microphones (M30; Earthworks) for audio recording: main microphone at the fingertip to capture texture-induced sounds and sub microphone on the printer frame to record mechanical noise. Both microphones exhibited a flat frequency response ranging from 10 Hz to 30 kHz and operated at a sampling rate of 44.1 kHz. A load cell (SC616C-1kg; Sensor and Control) measuring contact forces up to 9.8 N with 0.005 N accuracy (using an HX711 ADC) was mounted on the fingertip. A 3-axis MEMS accelerometer (IIS3DWB; STMicroelectronics) capable of measuring ± 16 g with 75 $\mu g/\sqrt{\text{Hz}}$ noise density was also mounted on the fingertip. We implemented a multithreaded Python application that synchronized and formatted sensor streams with kernel-timestamp alignment. The entire setup was enclosed in a soundproof chamber to suppress ambient noise (Fig. 2b).

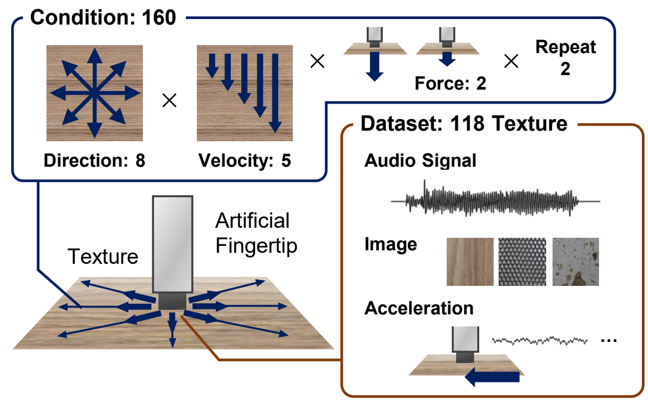

Fig. 1: Concept of our Cluster Haptic Texture Database. Controlled directions and velocities of the sliding artificial fingertip are added to conventional haptic datasets.

B. Recording Procedure

Prior to haptic data recording, we captured texture images (1181×1181 px, 300 dpi) using a flatbed scanner for all materials. During recording, we performed 160 linear traces (80 mm length) for each material, systematically varying five velocities (20-60 mm/s in 10 mm/s increments), eight intercardinal directions, two pressing forces (0.5 N and 1.0 N), and two repeats. Before each measurement block, we adjusted the fingertip height to achieve the target force. After data recording, we performed post-processing including active noise cancellation of audio to remove mechanical noise and temporal synchronization between sensors.

This work was partially supported by JST Moonshot R&D Program “Cybernetic being” Project (Grant number JPMJMS2013) and JSPS KAKENHI Grant numbers JP22H01447, JP23H04328, JP25K00145, and JP25H00722, Japan.

[1]Michikuni Eguchi and Takefumi Hiraki are with University of Tsukuba, 1-2 Kasuga, Tsukuba, Ibaraki, Japan. {eguchi, hiraki}@mvml.slis.tsukuba.ac.jp

[2]Michikuni Eguchi, Tomohiro Hayase, Yuichi Hiroi, and Takefumi Hiraki are with Metaverse Lab, Cluster, Inc., 8-9-5 Nishigotanda, Shinagawa, Tokyo, Japan. {y.hiroi, t.hayase}@cluster.mu

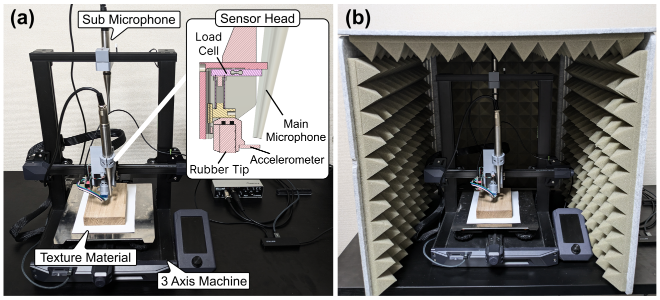

Fig. 2: System overview: (a) Hardware setup for texture data measurement system. (b) The equipment is placed in a soundproof enclosure during data recording.

C. Textures

We prepared 118 materials in nine categories (Wood, Stone, Polymer, Metal, Glass, Composite, Ceramic, Biological Leather, and Cloth) following previous haptic databases [2], [5], and mounted them on a magnetic sheet.

III. Experiments

We conducted three classification experiments using the proposed dataset to investigate whether our dataset contains sufficient discriminative features: texture (118 classes), velocity (5 classes), and direction (8 classes) classification using audio and acceleration signals.

A. Setup

1) Data Preprocessing: We converted both audio and acceleration signals into log-mel spectrograms, which provides a frequency representation suitable for machine learning.

2) Classifiers: We used pre-trained CNN (ResNet34), Vision Transformer (ViT small), and SVC as classifiers. The CNN and ViT treat log-mel spectrograms as 2D image-like data. We flattened the spectrograms for SVC into 1D feature vectors following previous haptic studies [5]. We also developed a multimodal classifier combining audio and acceleration data. The architecture uses separate CNN/ViT encoders for each modality, followed by transformer layers for feature fusion and fully connected layers for classification. For classical approaches, we concatenated flattened log-mel spectrograms from both modalities [5].

3) Training and Evaluation: We split the texture dataset into training (70%), validation (10%), and test (20%) sets. We evaluated classifiers using accuracy metrics, averaging results over five runs with different random seeds.

B. Experiment I: Texture Classification

Figure 3 shows the texture classification results. The CNN-based classifier achieved the highest accuracy (96.0% with multimodal data) among all models. This indicates that our dataset's 118 textures have distinct features that CNNs can effectively capture. The lower performance of acceleration-based classification compared to audio-based suggests that acceleration signals may contain less distinctive texture features than frictional sounds.

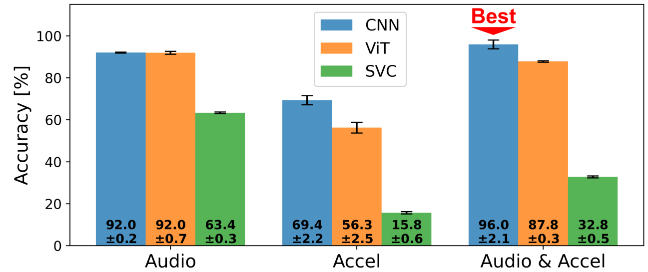

Fig. 3: Texture classification performance across three machine learning models, showing mean accuracy values from five-fold cross-validation with standard deviation error bars.

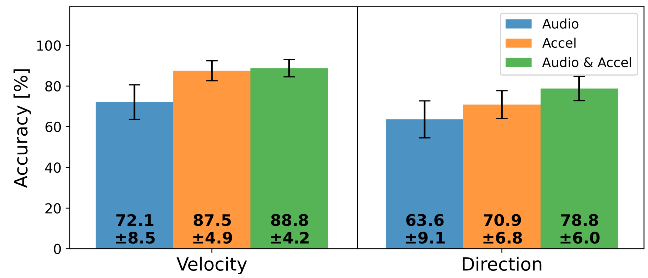

Fig. 4: Velocity and direction classification performance across all textures, showing mean accuracy and standard deviation across all textures.

C. Experiment II: Velocity and Direction Classification

Based on Experiment I's results, we used the best-performing CNN-based model for velocity and direction classification.

Figure 4 shows the classification results. The multimodal approach achieved the highest accuracy (88.8% for velocity, 78.8% for direction). The high accuracies, well above chance level, confirm that our dataset contains discriminative features for sliding velocity and direction. The large standard deviations indicate that the distribution of motion-dependent features varies significantly across different textures.

IV. Conclusion

We proposed the Cluster Haptic Texture Database, comprising 118 textures explored at five traversal velocities, eight directions, and two force levels, and capturing synchronized visual, and tactile signals using an XYZ stage. We aim to make our dataset publicly available by the conference date.

References

[1] S. J. Lederman *et al.*, "Haptic perception: A tutorial," *Attention, Perception & Psychophysics*, vol. 71, no. 7, pp. 1439–1459, Oct. 2009.

[2] H. Zheng *et al.*, "Deep learning for surface material classification using haptic and visual information," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2407–2416, Dec. 2016.

[3] B. Khojasteh *et al.*, "Robust surface recognition with the maximum mean discrepancy: Degrading haptic-auditory signals through bandwidth and noise," *IEEE Trans. Haptics*, vol. 17, no. 1, p. 58–65, Jan. 2024.

[4] M. Strese *et al.*, "A haptic texture database for tool-mediated texture recognition and classification," in *Proceedings of 2014 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*, Oct. 2014, pp. 118–123.

[5] A. Devillard *et al.*, "Concurrent haptic, audio, and visual data set during bare finger interaction with textured surfaces," in *Proceedings of 2023 IEEE World Haptics Conference (WHC)*, 2023, pp. 101–106.