Seeking Causal, Invariant Structures with Kernel Mean Embeddings in Haptic-Auditory Data from Tool-Surface Interaction

Behnam Khojasteh[1], Yitian Shao[2], and Katherine J. Kuchenbecker[3]

***Abstract*— Causal inference could give future learning robots strong generalization and scalability capabilities, which are crucial for safety, fault diagnosis and error prevention. One application area of interest consists of the haptic recognition of surfaces. We seek to understand cause and effect during physical surface interaction by examining surface and tool identity, their interplay, and other contact-irrelevant factors. To work toward elucidating the mechanism of surface encoding, we attempt to recognize surfaces from haptic-auditory data captured by previously unseen hemispherical steel tools that differ from the recording tool in diameter and mass. In this context, we leverage ideas from kernel methods to quantify surface similarity through descriptive differences in signal distributions. We find that the effect of the tool is significantly present in higher-order statistical moments of contact data: aligning the means of the distributions being compared somewhat improves recognition but does not fully separate tool identity from surface identity. Our findings shed light on salient aspects of haptic-auditory data from tool-surface interaction and highlight the challenges involved in generalizing artificial surface discrimination capabilities.**

I. Motivation

Causality promises to help with current challenges in machine learning such as domain generalization, interpretability and scalability [1]. However, it is not yet always clear how to create causal algorithms that perform well in realistic application domains. High-dimensional mappings such as kernel mean embeddings are an active and promising thrust of this research [2]. Baumann et al. [3] recently incorporated the kernel two-sample test [4] in the non-i.i.d. setting [5] to identify causal structures in dynamical systems. They used the maximum mean discrepancy (MMD) to detect causal changes and demonstrated the effectiveness of their approach on a dual-arm manipulation robot. Another successful extension of Solowjow et al.'s work [5] was recently performed for the task of *multi-user surface recognition*. Specifically, Khojasteh et al. [6] demonstrated a sample-efficient approach for learning to recognize 108 surface textures from multimodal (visual, auditory, and haptic) sensor readings obtained from a public data set recorded by eleven different people. Their data-driven method effectively mitigated speed-, force-, and session-dependent effects during tool-surface interaction by a simple distribution shift of time-series data in order to generalize across human users. This alignment of distribution means was sufficient to boost recognition performance by 9%, presumably because it highlights the distributions' higher-order statistical moments that convey important information about the surface properties. The success of this compensation trick raises the question, whether such an approach succeeds in generalization to unseen sensing tools. Understanding the generalization capabilities of surface perception to unseen sensing tools is of practical interest so that a given pipeline could be deployed on different instruments and different robot body parts (e.g., from the smallest to the largest finger).

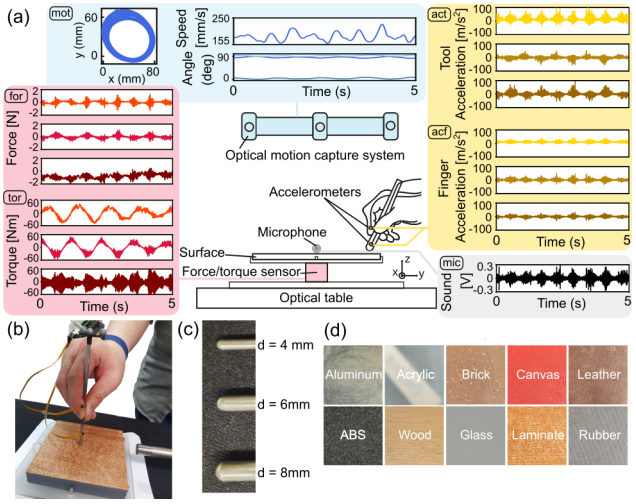

Fig. 1. (a) Design of the test bed and sample recording from each sensor for a 6-mm-diameter steel tool dragging on the wood surface. (b) Test bed being used to record from the 4-mm tool dragging on the laminate surface. (c) Three steel tools with different diameters. (d) Set of ten diverse surfaces.

II. Experiments and Results

Inspired by the recent success of kernel mean embeddings and MMD in multi-user surface recognition, we sought to ascertain *whether surface identity can be separated from tool-dependent effects.* In other words, can an algorithm recognize a known surface when it is explored using a different tool? This question is the haptic equivalent to asking whether a particular object could be recognized when photographed in a new setting by a different camera.

A. Haptic-Auditory Sensing

Our haptic-auditory test bed (Fig. 1(a)) allows us to capture five information sources: contact forces (for) and torques (tor) from a sub-surface force-torque sensor, tool (act) and

B. Khojasteh and K. J. Kuchenbecker are with the Max Planck Institute for Intelligent Systems (MPI-IS), Stuttgart, Germany, and also with the Faculty of Engineering Design, Production Engineering and Automotive Engineering, University of Stuttgart, Stuttgart, Germany. e mail: {khojasteh, kjk}@is.mpg.de

Y. Shao is with the Centre for Tactile Internet with Human-in-the-Loop (CeTI), Technische Universität Dresden, Dresden, Germany, and also with MPI-IS, Stuttgart, Germany. e mail: ytshao@is.mpg.de

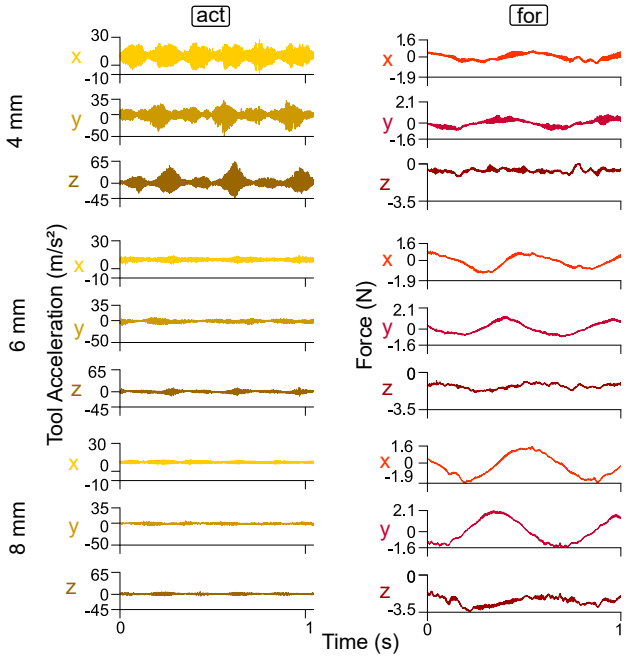

Fig. 2. One second recording on the leather sample with the three tools: (a) three-axis tool accelerations (act) and (b) three-axis contact forces (for).

finger (acf) accelerations sensed at two accelerometer locations (Fig. 1(b)), and contact sounds (mic) from a microphone. Tactile vibrations, in particular, present a promising source of information because they propagate widely [7], [8], offer high temporal resolution for spatial touch-information decoding [9], and make multimodal surface classification robust [6], [9]. We do not consider the measurements from the motion-capture system for this analysis. An experimenter recorded these signals for three solid steel tools of the same length with thermally hardened hemispherical tool tips of 4, 6, and 8 mm diameter (Fig. 1(c)) touching $C = 10$ diverse surface textures (Fig. 1(d)). The surface textures are a subset of the Penn Haptic Texture Toolkit [10] with categories that are representative of prior surface datasets [10], [11], [12]. Their masses are 11.6, 26.1, and 46.4 g, respectively. In total, we have 300 five-second-long recordings of multimodal surface data (3 steel tools $\times$ $C = 10$ surfaces $\times$ 10 trials).

B. A Causal Lens on Tool-Surface Interactions

The choice of the sensing tool greatly affects the mechanical signals generated during interactions with surfaces, even with controlled surface topology [13], [14]. Both the tool mass and the tool tip geometry contribute to this complexity. For the case of the natural leather surface (Fig 2), we illustrate how the three tools affect the tool accelerations and contact forces. In particular, the mass of the tool has a significant effect on the surface signals. Vibrations (tactile or auditory) from surface contact will have higher signal amplitudes for tools with smaller masses. In contrast, the amplitudes of the contact forces are larger on average for heavier tools. Regarding tool geometry, a smaller tool tip will penetrate more between asperities on the surface than a larger tool [13], thereby giving the haptic-auditory signals higher frequency content and larger vibration magnitudes. Given these observations, performing cross-tool surface classification with sensing tools that vary in both tool mass and tip diameter is challenging. In addition to the choice of the sensing tool, other important factors in tool-surface interactions include the tool speed and applied tool force, the geometry of the interfaces I_1 and I_2 of the contact pair, and the material properties of the surface. We believe that these causal relationships must be taken into account for a robust surface classifier that performs data manipulations to compensate for tool effects. Here, we investigate how distribution mean alignments mitigate these tool effects.

C. Problem Formulation

Our goal is to classify unseen multimodal sensor recordings from physical surface interactions with the same sensing tool (termed 'in-domain') or across sensing tools (termed 'out-of-domain'). From a mathematical perspective, we address this surface-recognition task by focusing exclusively on data distribution differences. We model surface interactions as realizations of a dynamical system [5]. We assume that a set of $C \in \mathbb{N}$ unique surfaces will induce different distributions $\mathbb{P}_1, \ldots, \mathbb{P}_C$, respectively. The classifier compares unlabeled surface trials to the known surface distributions in the library to determine the surface class c from which it most likely came. To infer whether an unseen testing trial and a training trial from a surface library come from identical or non-identical surfaces, we quantify distribution differences between the two trials.

D. Classification Setting

We adopt the setting from recent work with a multimodal multi-user classifier [6]. Given its efficacy in surface recognition [6], we use the squared bias MMD estimator by Gretton et al. [4],

$$\begin{aligned} \mathrm{MMD}_b^2[\mathbb{P}_Y, \mathbb{P}_Z] = \frac{1}{n^2}\sum_{i,j=1}^{n} k(y_i, y_j) \\ + \frac{1}{m^2}\sum_{i,j=1}^{m} k(z_i, z_j) - \frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} k(y_i, z_j), \end{aligned} \tag{1}$$

where $[y_1, ..., y_n]$ and $[z_1, ..., z_m]$ are i.i.d. random variables. In our case, these are $n = m = 100$ samples from Fourier-transformed surface data streams Y_s and Z_s with unknown distributions $\mathbb{P}_{Y_s}$ and $\mathbb{P}_{Z_s}$. For our kernel, we use the squared exponential function, for all statistical tests due to its suitability for visual-haptic-auditory surface data [6]. Instead of optimizing over σ through a grid search, we choose the well-established median heuristic [4]. For multi-source classification, we use the geometric mean to unify the MMD scores of multiple information sources to a overall discrepancy score. This framework uses the arithmetic mean of individual logarithm-transformed MMD values and therefore improves MMD scale-invariance across information sources. Our algorithm leverages the k-nearest neighbors principle to make classification predictions with the global DS scores.

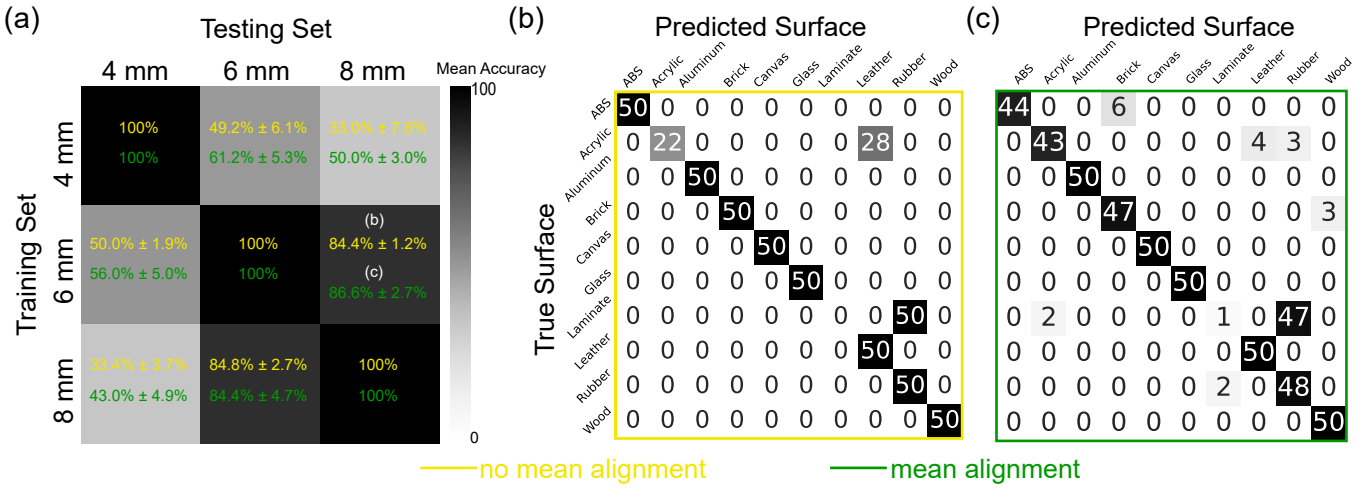

Fig. 3. (a) Surface classification performance of all five information sources combined (for, tor, act, acf and mic) for in- and out-of-domain tool diameters without (yellow) and with (green) the compensation trick; mean alignment facilitates recognition in almost all out-of-domain cases. (b) Exemplary confusion matrix for the case of training on data from the 6 mm tool and testing on data from the 8 mm one without mean alignment. (c) The same confusion matrix with alignment of the distribution means. A confusion matrix showing perfect surface recognition would have a solid black diagonal surrounded by white.

An unlabeled surface trial Z will be classified to the class c in the library with C surfaces according to

$$\min_{c \in C} \mathrm{DS}[Y(c), Z]; \tag{2}$$

it predicts a surface class through the test-train trial pair with smallest global discrepancy distance, i.e., the nearest neighbor. In particular, we consider five trials of each surface in our library (five-shot learning). To reduce the influence of different data distributions on recognition performance, we run our classification pipeline in $R = 5$ repeated iterations for each surface trial of the testing sets.

E. Results

We observe perfect recognition rates for all three in-domain tool settings (diagonal of Fig. 3(a)); the same performance can be achieved with several individual information sources (results not shown). Out-of-domain classification is more challenging with accuracies less than 90% in all six cases. We see relatively symmetric cross-tool performance with the chosen train-test splits from different tools; in particular, generalizing from and to the smallest tool ($d = 4\,\mathrm{mm}$) is most difficult, potentially because its smaller diameter and mass amplify and diversify the contact signals in ways that cannot be predicted from data recorded with larger tools. More penetration between asperities for the smaller tool causes these complex contact signals. This increased unpredictability with smaller tools resembles the complexity observed in cutaneous sliding contact on fine-scale textured surfaces. At smaller spatial scales with multiple contact points, intricate coupling between the contacting pairs resulted in patterns that were challenging to encode, even with higher-order nonlinear techniques [14]. Mean alignment of the distributions improves performance in $5/6$ cases by between 2.2% and 17% mean accuracy. At the same time, in $4/6$ cases this compensation causes an increase in the standard deviation of the accuracy by between 2.0% and 3.1% (Fig. 3(a)). This trend is exemplified by examining the confusion matrices for the case where we train on the 6-mm tool and test on the 8-mm tool (Fig. 3(b) and (c)). To conclude, changing the diameter of the tool with which a surface is explored greatly alters the distribution of the audio-haptic data generated during contact. Shifting the means to align somewhat improves performance but still falls far short of in-domain recognition accuracy. These findings suggest that tool diameter and mass also considerably influence the higher-order statistical moments.

III. Future Work

To increase the robustness and accuracy of haptic surface classification in out-of-domain tool settings, we need to further investigate into how tool geometry and mass causally affect contact signals. A potential future direction could include systematic analysis of the relationship between tool speed, applied normal force, and the resulting contact signals for each selected tool across diverse surfaces. In particular, contact vibrations and forces may provide complementary surface information from the tool interaction due to characteristic signal properties in the low and high frequency ranges. Alternative cross-tool compensation strategies that provide more flexible data manipulation beyond just fitting the distribution means are also worth exploring. Understanding causal structural invariance in tool-surface interaction is complex but would pave the way for both recognition and synthesis of an arbitrary amount of data from physical tool-surface interactions.

Acknowledgments

The authors thank Jonathan Fiene and Julian Martus for developing the accelerometer boards; Farimah Fazlollahi and Bernard Javot for helpful feedback on the test bed; Dominika Lisy for helping with data acquisition; the German Research Foundation (DFG) Project EXC 2050/1, 390696704, CeTI Cluster of Excellence at TU Dresden for supporting Y. Shao; and the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting B. Khojasteh.

REFERENCES

[1] B. Schölkopf, "Causality for machine learning," in *Probabilistic and Causal Inference: The Works of Judea Pearl*, 2022, pp. 765–804.

[2] K. Muandet, K. Fukumizu, B. Sriperumbudur, B. Schölkopf *et al.*, "Kernel mean embedding of distributions: A review and beyond," *Foundations and Trends® in Machine Learning*, vol. 10, no. 1-2, pp. 1–141, 2017.

[3] D. Baumann, F. Solowjow, K. H. Johansson, and S. Trimpe, "Identifying causal structure in dynamical systems," *Transactions on Machine Learning Research*, vol. 2022, no. 7, 2022.

[4] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 723–773, 2012.

[5] F. Solowjow, D. Baumann, C. Fiedler, A. Jocham, T. Seel, and S. Trimpe, "A kernel two-sample test for dynamical systems," *arXiv preprint arXiv:2004.11098*, 2020.

[6] B. Khojasteh, F. Solowjow, S. Trimpe, and K. J. Kuchenbecker, "Multimodal multi-user surface recognition with the kernel two-sample test," *IEEE Transactions Automation Science and Engineering*, pp. 1–16, 2023.

[7] Y. Shao, V. Hayward, and Y. Visell, "Spatial patterns of cutaneous vibration during whole-hand haptic interactions," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 4188–4193, 2016.

[8] Y. Shao, H. Hu, and Y. Visell, "A wearable tactile sensor array for large area remote vibration sensing in the hand," *IEEE Sensors Journal*, vol. 20, no. 12, pp. 6612–6623, 2020.

[9] Y. Shao, V. Hayward, and Y. Visell, "Compression of dynamic tactile information in the human hand," *Science Advances*, vol. 6, no. 16, p. eaaz1158, 2020.

[10] H. Culbertson, J. J. L. Delgado, and K. J. Kuchenbecker, "One hundred data-driven haptic texture models and open-source methods for rendering on 3D objects," in *Proceedings of the IEEE Haptics Symp.*, 2014, pp. 319–325.

[11] M. Strese, "Haptic material acquisition, modeling, and display," Ph.D. dissertation, Technische Universität München, 2021.

[12] A. Burka, A. Rajvanshi, S. Allen, and K. J. Kuchenbecker, "Proton 2: Increasing the sensitivity and portability of a visuo-haptic surface interaction recorder," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 439–445.

[13] C. G. McDonald and K. J. Kuchenbecker, "Dynamic simulation of tool-mediated texture interaction," in *Proceedings of IEEE World Haptics Conference*, Daejeon, South Korea, Apr. 2013, pp. 307–312.

[14] B. Khojasteh, M. Janko, and Y. Visell, "Complexity, rate, and scale in sliding friction dynamics between a finger and textured surface," *Scientific Reports*, vol. 8, no. 1, p. 13710, 2018.