Surface-Assisted In-Air Gesture Recognition

Minhyeok Kim, Donghun Lee, Hyosu Kim
School of Computer Science and Engineering
Chung-Ang University
Seoul, South Korea
{alsgur7859, ldh019, hskimhello}@cau.ac.kr

***Abstract*—In-air hand gestures can be effectively utilized to support immersive interactions in an online meeting. In this study, we propose a novel system to support in-air hand gestures by leveraging the vibrations introduced by the gestures. When a user makes hand gestures while putting her elbow on a table, vibrations are produced due to the gestures and travel through the table's surface. Based on this observation, we collect gesture-driven vibrations and use them for gesture recognition. Our evaluations with real-world users demonstrate that we can precisely recognize in-air hand gestures (e.g., an average accuracy of 96.02%) using only commercial devices, such as smartphones and smartwatches.**

I. Introduction

With the emergence of COVID-19, numerous classes and meetings have shifted to online formats, giving rise to various activities such as presentations and document sharing in remote settings. As this situation prolongs and people are increasingly participating in online interactions, the necessity to alleviate inconveniences in online meeting interactions has grown. As a response to the demand, many studies have attempted to leverage hand gestures to support seamless and efficient interactions in online environments [5], [6]. Imagine that in virtual meetings, one could smoothly navigate presentation slides by simply gesturing with their hands instead of using a keyboard or mouse. However, the existing works have limitations, especially in terms of deployability. This is because they mainly rely on dedicated hardware, such as Infrared cameras [7] and specialized wearable devices [8].

In this paper, we explore the potential of recognizing gestures by leveraging the gesture-driven vibrations. Figure 1 illustrates the usage scenario of our proposed system. We first assume that a user puts her elbow on a table. When she makes a in-air hand gesture, such as a clap, vibrations are produced due to the impact introduced by the gesture and propagate through the surface of the table. We then sense the vibrations using the smartphone placed on the table and the smartwatch worn by the user and eventually recognize the gesture. That is, we leverage only commercial mobile and wearable devices to realize hand gesture interfaces.

II. Related Works

Previously, several attempts have been made to interact with devices using vibrations on flat surfaces like tables. SurfaceSight [1] enhances IoT devices with LIDAR to perceive actions occurring on the surface, objects on the surface, and the presence of individuals on the surface. Vibwrite [2] utilizes unique vibration patterns for each user during actions like pin number input, pattern input, and gesture input on various surfaces for authentication. SurfaceLink [3] detects swipe actions occurring on a surface with multiple devices using acoustic sensors, leveraging this for interactions among the devices. Chen, M. et al. [4] detects sounds generated while writing characters on a surface using acoustic sensors, enabling the reconstruction of the written characters. In summary, the existing studies turns nearby surfaces into an input space by recognizing actions made on the surfaces. Unlike these works, we propose a in-air gesture recognition system, which can support more natural interactions in online conversations.

Fig. 1. Usage scenario of our proposed system. A user puts his elbow on a table, while wearing a smartwatch in his left hand and placing a smartphone on the right side.

III. System Design

In this section, we introduce a novel method for in-air hand gesture recognition (see Figure 2). It consists of two phases; the *training* and *interaction* phases. In the training phase, we first ask users to perform specific hand gestures multiple times (e.g., 10 20 times), while putting their elbow on a table. We then collect vibration data from the smartphone placed on the table and the smartwatch and finally constructs a classification model using the collected training data. In the interaction phase, the users interact with others in online by making the trained gestures on the surface. We then recognize the gestures

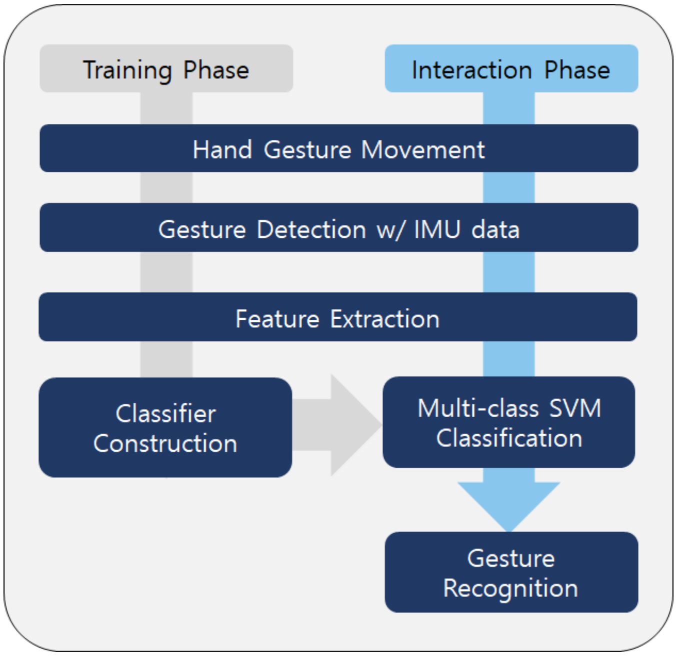

Fig. 2. Overall Workflow

by using the vibration data obtained during the interactions and the classification model.

A. Gesture-driven Vibration Data Collection

In practice, we have no prior information about when users make gestures, causing difficulties in collecting gesture-driven vibration data. We address this challenge using the Constant False Alarm Rate (CFAR) algorithm, which detects sudden variations in a given input sequence. That is, because the vibration recordings obtained by both smartphone and smartwatch show significant differences between the presence and absence of the gestures, we can precisely detect the gestures using the method and collect the vibration data used for model construction or gesture recognition.

B. Feature Extraction

Once detected, we extract features optimized based on our observations on the gesture-driven vibration data. We first filters out high frequency components ($>$ 300 Hz). This is because hand gestures introduces low frequency vibrations. After that, we obtain several classes of features from the filtered data. For example, we use statistical features including integral absolute value; variance; root mean square; standard deviation; and median absolute deviation. We also use fast Fourier transform and power spectral density. The selected features are combined into one feature vector to use a classifier.

C. Classification

We use a multi-class Support Vector Machine (SVM) because it works well even with a small amount of data. In the training phase, we first build the SVM classifier using the features extracted from the training vibration data set. We then leverage the constructed model to recognize gestures.

TABLE I
HAND GESTURE CLASSIFICATION ACCURACY OF 6 USERS VIA A 10-FOLD CROSS VALIDATION.

User	U1	U2	U3	U4	U5	U6	Average
Accuracy	99.33%	99.33%	95.67%	95.38%	95.05%	91.36%	96.02%

IV. EXPERIMENT

In this study, we collected hand gesture data from 6 users (4 males, 2 females). Each user was requested to provide a total of 300 hand gesture data samples, 30 times for each of the 10 different hand gestures. The users were instructed to place a Google Pixel 4 smartphone on a fixed position of a table and keep their elbows rest on the table's fixed positions as the default setup. Additionally, a Google Watch 4 smartwatch was worn on their left hand. We then measured the gesture recognition accuracy of our proposed system for each individual user via a 10-fold cross validation.

The accuracy of the 6 users can be seen in Table 1. Their average accuracy is 96.02%, which means when performing a single gesture 30 times, only 1-2 misclassifications occur. This level of accuracy demonstrates that within the default settings, the 10 different hand gestures are distinct and well distinguished from each other. Among them, the user with the highest accuracy achieved 99.33% (U1 and U2), misclassifying only 2 out of 300 tested gestures. The user with the lowest accuracy achieved 91.36% (U6), with approximately 25 out of 300 gestures misclassified. This suggests that relatively similar gestures are not yet entirely distinct for all users. Overall, looking at our research results, we confirmed the ability to uniquely distinguish hand gestures under default settings. This indicates the potential to distinguish user hand gestures uniquely in a wider range of environments.

V. CONCLUSION

In this work, we introduced a new in-air hand gesture recognition system using smart devices, such as smartphones and smartwatches, and nearby object's surfaces. In particular, we use vibrations produced by a user's gestures to recognize the gesture. Our experimental results showed high accuracy in gesture recognition, thus allowing us to provide seamless and natural interactions.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF-2022R1C1C1012664).

REFERENCES

[1] Laput, G. and Harrison, C. (2019) 'SurfaceSight', Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3290605.3300559.
[2] Liu, J. et al. (2017) 'VibWrite', Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. https://doi.org/10.1145/3133956.3133964.

[3] Goel, M. et al. (2014) 'SurfaceLink', Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/2556288.2557120.

[4] Chen, M. et al. (2019) 'Your table can be an input panel', Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(1), pp. 1–21. https://doi.org/10.1145/3314390.

[5] J. Wang, J. Chen, Y. Qiao, J. Zhou and Y. Wang, "Online Gesture Recognition Algorithm Applied to HUD Based Smart Driving System," 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Beijing, China, 2019, pp. 289-294, doi: 10.1109/ISMAR-Adjunct.2019.00-26.

[6] Y. Fang, K. Wang, J. Cheng and H. Lu, "A Real-Time Hand Gesture Recognition Method," 2007 IEEE International Conference on Multimedia and Expo, Beijing, China, 2007, pp. 995-998, doi: 10.1109/ICME.2007.4284820.

[7] Yi Li, "Hand gesture recognition using Kinect," 2012 IEEE International Conference on Computer Science and Automation Engineering, Beijing, China, 2012, pp. 196-199, doi: 10.1109/ICSESS.2012.6269439.

[8] S. Kallio, J. Kela and J. Mantyjarvi, "Online gesture recognition system for mobile interaction," SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483), Washington, DC, USA, 2003, pp. 2070-2076 vol.3, doi: 10.1109/ICSMC.2003.1244189.