Poster Abstract: Realistic Multiuser, Multimodal (IMU, Acoustic) HAR Data Generation through Single User Data Augmentation

Soumyajit Chatterjee, Arun Singh, Bivas Mitra, Sandip Chakraborty
sjituit@gmail.com,erarungwl2013@gmail.com,bivas@cse.iitkgp.ac.in,sandipc@cse.iitkgp.ac.in
IIT Kharagpur, India

ABSTRACT

Multiuser activity recognition has been the core of different context-aware services. However, the development of such services is often plagued by the dearth of multiuser datasets. This paper presents a strategy for generating synthetic multiuser datasets by augmenting existing real-life datasets. The described strategy exploits precise time synchronization and well-known audio augmentation approaches to generate a multimodal activity recognition dataset with locomotive and acoustic signatures.

KEYWORDS

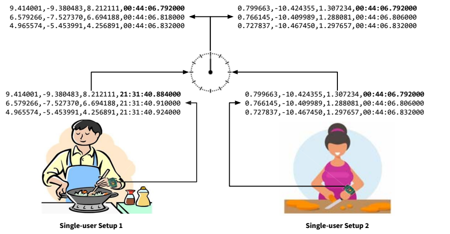
datasets, acoustic signatures, locomotive signatures, human activity recognition, data augmnetation

1 INTRODUCTION AND MOTIVATION

Multimodal sensing for human activity recognition has gathered significant momentum in the past few years. Since then, several publicly available datasets like [4] have gained immense popularity for presenting a rich and diverse repository of multimodal sensing data for complex activities of daily living (ADL). However, most of these datasets are restricted to single-user environments only, and typically for the multiuser activity recognition, the existing datasets have sensing data from unimodal sources like WiFi [5]. This paper presents a strategy to create multiuser datasets from the existing single-user datasets, with multimodal sensing from the locomotive and acoustic sensors. Additionally, we also show that using acoustic virtualization tools like [3], one can simulate physical factors like distance from the microphone, thus generating a rich and diverse dataset. Preliminary evaluations on an in-house collected dataset show the potential of the designed strategy.

2 BROAD IDEA AND METHODOLOGY

For any single-user dataset with IMU and acoustic data, the usual practice is to break the entire data collection duration into separate sessions [4]. In each such session, data from individual users is recorded without the intervention from any other user. Therefore, for any two or more users – (a) the data may not be recorded in the exact same time window, and (b) the environmental context will be completely segregated. On the other hand, in a typical multiuser scenario, with all devices time-synchronized appropriately, the environmental context will be a convolved version of all the acoustic signatures along with random noise. Based on this observation, we design a strategy to mimic these properties of a multiuser environment and adapt the locomotive and acoustic signatures obtained from single-user sources to imitate a virtual multiuser setup. The details follow.



```
9.414001,-9.380483,8.212111,00:44:06.792000
6.579266,-7.527370,6.694188,00:44:06.818000
4.965574,-5.453991,4.256891,00:44:06.832000
```

```
0.799663,-10.424355,1.307234,00:44:06.792000
0.766145,-10.409989,1.288081,00:44:06.806000
0.727837,-10.467450,1.297657,00:44:06.832000
```

```
9.414001,-9.380483,8.212111,21:31:40.884000
6.579266,-7.527370,6.694188,21:31:40.910000
4.965574,-5.453991,4.256891,21:31:40.924000
```

```
0.799663,-10.424355,1.307234,00:44:06.792000
0.766145,-10.409989,1.288081,00:44:06.806000
0.727837,-10.467450,1.297657,00:44:06.832000
```

Figure 1: Synchronization strategy for IMU sensors

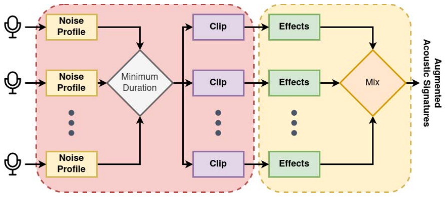


Figure 2: Mixing strategy for acoustic signatures

2.1 IMU Synchronization

The IMU signatures, unlike acoustic signatures, ideally should remain unaltered irrespective of one or more users in the environment. However, there is no guarantee that all of them have been recorded in the same time window. Thus, the generic idea is to choose a new start time common to all the IMU data while maintaining their individual polling intervals (see Figure 1). This makes the new dataset time synchronized within that new time window while their unaltered polling intervals preserve the exact gaps in time where the sensing hardware has observed the changes. Furthermore, while performing this IMU synchronization, one must keep in mind that all the other modalities should also be synchronized with IMU logs which can be achieved by correctly mapping the start times of these modalities with the new synchronized start-time.

2.2 Audio Mixing

In contrast to the IMU data, the acoustic data needs more sophisticated processing to imitate the virtual multiuser setup (Figure 2). We first start with preprocessing steps involving noise profiling and subsequent filtering of individual audio signatures for efficient activity recognition. Subsequently, we clip each audio signal with the length of the shortest audio signature. Although this might not be mandatory, we ensure that the data imitates a multiuser environment for the entire duration. Once this preprocessing is done, the next step is to add the effects like echo and reverb [1].

978-1-6654-9624-7/22/$31.00 ©2022 IEEE
DOI 10.1109/IPSN54338.2022.00067

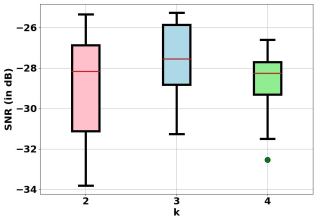


Figure 3: Variations in SNR with increasing k

Finally, the next step is to convolute all the acoustic signatures to generate the mixed ambient environment.

3 PRELIMINARY RESULTS

In this section, we conduct some preliminary experiments to analyze the overall quality of the generated dataset.

3.1 Datasets and Implementation

We apply this entire strategy to a single user dataset collected from a workshop environment. We had 4 participants for the data collection, each wearing a Moto 360 smartwatch (capturing IMU data at 50Hz). Additionally, the environment had a microphone (OnePlus3 smartphone), sampled at 44.1kHz.

For the implementation, the IMU synchronization is implemented in `python3.x`. On the other hand, the audio processing pipeline is implemented on Audacity [2]. Additionally, we also simulate the distance using [3]. The entire codebase can be found here[1]. The repository contains unmixed and augmented data samples, simulating a virtual environment with a dual occupancy for validation.

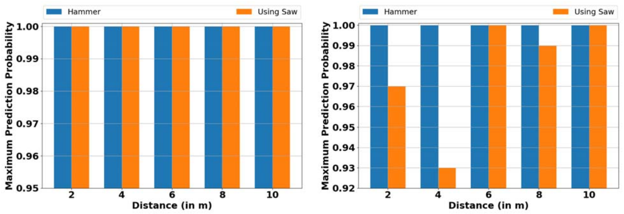
3.2 Volume and Validity

Undoubtedly, the given strategy can significantly increase the overall volume of data. For example, from a single-user dataset with n users, where any user performs all the a activities, one can create $\binom{n}{k} \times a^k$ $(k \leq n)$ unique virtual multiuser setups with k users.

However, volume is not the primary concern; instead, it is the impact of the augmentation on the final data quality. The designed strategy keeps the IMU signatures completely unaltered, albeit the audio signatures get convoluted, rendering the final output noisy and unuseful. From Figure 3, we can see that even with the increasing number of acoustic sources, the output signal's quality does not degrade significantly. Thus, the designed strategy preserves the quality necessary for standard acoustic signal processing.

3.3 Simulating Inequality in Distance

Subsequently, we analyze the impact inequality of distance between the users and the microphone. For this, we use [3] to create a virtual room with two sources and a central microphone. Next, we analyze the audio signatures generated as the output from this virtual room setup, using [1] to check whether the generated signatures still stay meaningful or not for any state-of-the-art audio-based recognition to understand. From Figure 4a and Figure 4b, we can see that even with unequal distance from the single microphone, the final rendered signal is still capable of capturing meaningful signatures for HAR with none of the activities getting predicted with less than 0.90 confidence.



(a)

(b)

Figure 4: Maximum prediction probability of Ubicoustics [1] on the audio generated from the virtual setup (room dimension $50 \times 50 \times 50\mathrm{m}^3$) with unequal distance from the fixed microphone. (a) Source of hammer varying but saw fixed at 2m, (b) Source of saw varying but hammer fixed at 2m.

4 CONCLUSION AND FUTURE WORKS

This paper provides a strategy to generate a multiuser dataset from the existing single-user dataset. We evaluate the strategy on a real-life dataset. Evaluation using state-of-the-art activity recognition models shows that this strategy can be applied for generating multiuser datasets with locomotive and acoustic signatures. As future work, we intend to exploit this strategy to generate complex multiuser scenarios with multimodal sensing, which can then be used to develop models that recognize complex ADL(s).

REFERENCES

[1] Gierad Laput, Karan Ahuja, Mayank Goel, and Chris Harrison. 2018. Ubicoustics: Plug-and-play acoustic activity recognition. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 213–224.

[2] D. Mazzoni. 1999–2019. Audacity software is copyright 1999-2019 Audacity Team. The name Audacity is a registered trademark of Dominic Mazzoni. https://www.audacityteam.org/, Last Accessed: February 19, 2022.

[3] Robin Scheibler, Eric Bezzam, and Ivan Dokmanić. 2018. Pyroomacoustics: A python package for audio room simulation and array processing algorithms. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 351–355.

[4] Ekaterina H Spriggs, Fernando De La Torre, and Martial Hebert. 2009. Temporal segmentation and activity classification from first-person sensing. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 17–24.

[5] Sheng Tan, Linghan Zhang, Zi Wang, and Jie Yang. 2019. MultiTrack: Multi-user tracking and activity recognition using commodity WiFi. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[1] https://github.com/stilllearningsoumya/data_augmentation_strategy