SolarWalk Dataset: Occupant Identification using Indoor Photovoltaic Harvester Output Voltage

Nurani Saoda*
University of Virginia
saoda@virginia.edu

Md Fazlay Rabbi Masum Billah*
University of Virginia
masum@virginia.edu

Victor Ariel Leal Sobral
University of Virginia
sobral@virginia.edu

Bradford Campbell
University of Virginia
bradjc@virginia.edu

ABSTRACT

Occupant identification is paramount for many building applications. Regardless, several practical concerns limit existing solutions to be ubiquitously deployed. Current systems are either intrusive, privacy-invasive, or require obtrusive, maintenance-heavy, and special-purpose infrastructure. As an alternative, the shadow pattern of a person reflected in the output voltage of a photovoltaic harvester power supply in many energy-harvesting devices can be used as a unique person identifying feature. In this paper, we present the first dataset containing the time-series open circuit output voltage traces of indoor photovoltaic cell corresponding to occupant door crossing events to perform occupant identification in smart homes. We collect shadow patterns of five participants from two different doors in two rooms of a building. The dataset consists of a total of 900 door entry and exit events during different hours of the day. We sample the voltage at 50 hz and provide the raw timestamped data. We also pre-process the data to filter the event of interest and label the data with associated occupant id and type of door events. Moreover, we provide insights into future research directions using the dataset. The dataset is available at https://doi.org/10.5281/zenodo.7195748

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**.

KEYWORDS

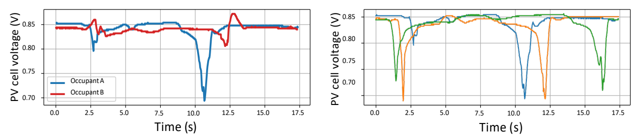
Photovoltaic Harvesters, Occupant Identification

ACM Reference Format:
Nurani Saoda, Md Fazlay Rabbi Masum Billah, Victor Ariel Leal Sobral, and Bradford Campbell. 2022. *SolarWalk* Dataset: Occupant Identification using Indoor Photovoltaic Harvester Output Voltage. In *The 20th ACM Conference on Embedded Networked Sensor Systems (SenSys '22), November 6–9, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3560905.3567773

*Both authors contributed equally to this research.

SenSys '22, November 6–9, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9886-2/22/11.
https://doi.org/10.1145/3560905.3567773



(a) Voltage fluctuations of occupant A and B are different from each other.

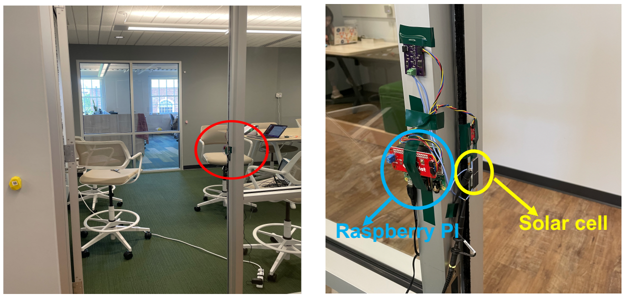
(b) Voltage fluctuations of the same occupant have similar shape.

Figure 1: This figure shows how the output voltage of the solar cell mounted on a doorframe ripples as different occupants pass through the door. First voltage drop corresponds to entering through the door, followed by exiting. The maximum voltage drop and the duration of voltage fluctuations vary differently for occupant A and B . On the other hand, these characteristics remain consistent over multiple trials by the same person.

1 INTRODUCTION

Occupant identification in indoor spaces is a key enabler for many person-specific, human-centered applications including HVAC control, precise water temperature control, occupant-specific energy-metering, and providing time-sensitive critical reminders immediately upon someone entering or leaving home [2, 6]. Such occupant-driven appliance control not only tremendously improves user comfort and convenience, but also plays an instrumental role in resource utilization, reducing energy waste, and better building management [9, 10, 13]. Several solutions exist to accurately identify occupants involving different sensing modalities such as camera/vision audio/acoustic, vibration, infrared, ultrasonic, and RF signals [3, 7, 8, 12, 14]. While all of these approaches have their strengths and drawbacks, we recognize that several limiting challenges still need to be addressed to design an occupant identification system that is non-intrusive, ubiquitous, unobtrusive, and installation-friendly.

To achieve this goal, we designed *SolarWalk*, a novel occupant identification system that adopts a small photovoltaic (PV) harvester's output voltage as a sensing modality to identify persons in a smart home context. Since photovoltaic harvesters is used as a power source to many indoor light energy-harvesting devices, *SolarWalk* is non-intrusive, does not require additional sensing hardware, achieves very small form factor to be ubiquitously deployed, and can be peeled-and-sticked in most indoor spaces. The



(a) Hardware deployed in Room 217

(b) Data collection hardware platform

Figure 2: Experimental setup of data collection

output voltage fluctuations of a photovoltaic harvester when a person walks in front of it within a close range (e.g., through a door or a hallway) is a unique identifier of that person due to height, body shape, and gait differences and can be leveraged to distinguish between multiple occupants (as shown in Figure 1). The voltage of the PV cell drops as the person obscures the surface of the cell and restores itself as the person walks away. The amplitude of the ripple voltage is related to the height of the person's shadow and time length of the ripple is associated with someone's gait or walking style. Moreover, the pattern for different entry and exit events are distinguishable and can be used to determine if an occupant entered or exited the room. *SolarWalk* system collects the voltage patterns of occupants from deployed sensors and inputs them as a feature together with occupant and event type labels to a supervised learning-based classifier to distinguish between different occupants.

In this paper, we present the dataset collected as a part of our *SolarWalk* study to investigate the performance of PV cells as person identifying sensors. The dataset contains a total of 150 minutes voltage traces associated with a total of 900 door entry and exit walk events from five different participants collected from two different rooms in a building. We provide the pre-processed voltage trace data with door entry, exit, no events along with labels of occupants and event type (entry/exit) that can be directly used as a feature to a supervised classifier. We describe the data collection methodology and experimental setup to illustrate how the data can be reproduced according to specific use cases. To illustrate an example use case, we also provide a script showing how to use the dataset to train and test simple machine learning models to distinguish participants. Since the shadow pattern of a PV harvester has not been explored as a context-rich sensor, we believe that the dataset will be useful to enable further context-aware applications beyond occupant identification. While several datasets exist for occupant identification using different sensing modalities like RF and vision, to the best of our knowledge, this is the first dataset that captures the raw voltage fluctuations of PV cell associated with human walking events in indoor spaces to identify occupants.

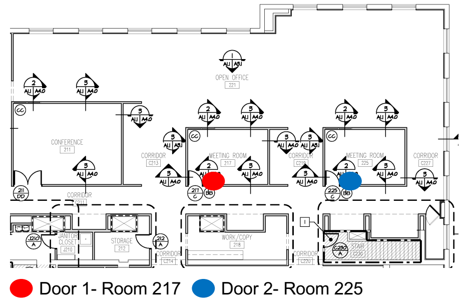


Figure 3: Floor plan showing the installed sensors on two doors of two different rooms.

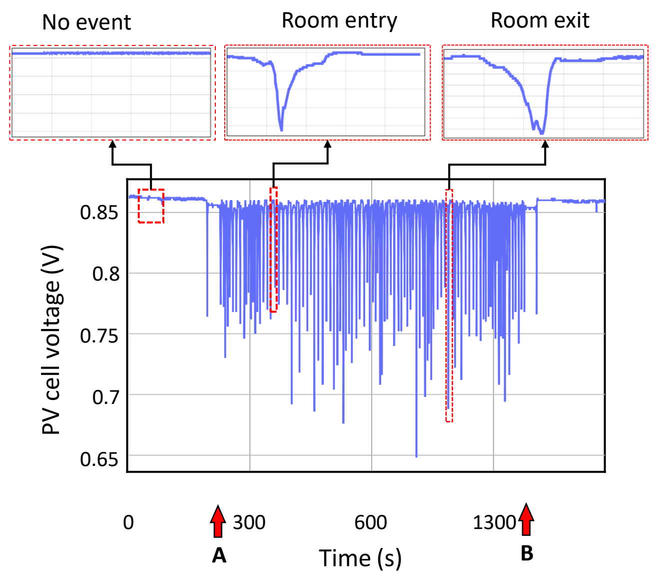
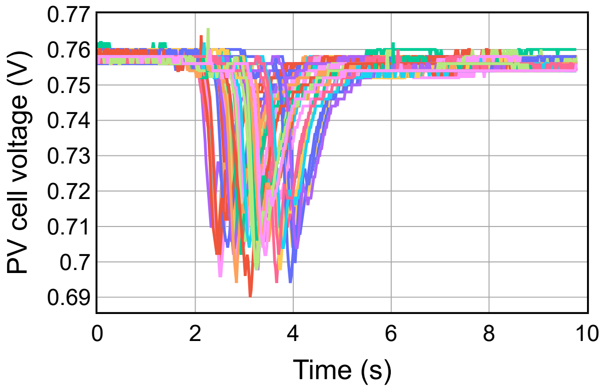


Figure 4: Data collection from a single participant at one location. At point A, the participant started walking. The person started with entering the room and exited after 10 seconds. At point B, the participant stopped walking. We collected 50 room entry events and 50 room exit events.

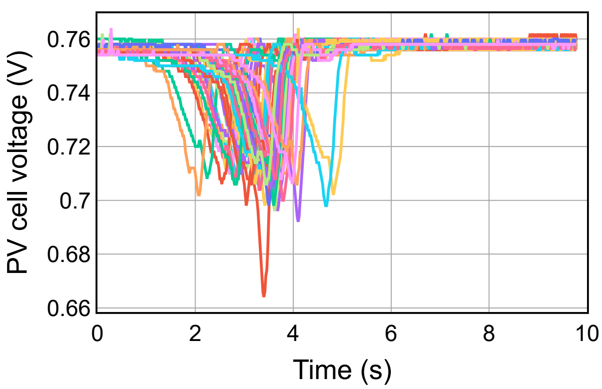
2 DATASET GENERATION METHODOLOGY

2.1 Experimental Setup

To start the data collection, we install a PV cell [5] on two door frames of two rooms in a building. Figure 3 shows the floor plan including the installation points. We place the PV cell approximately halfway above the ground with the surface of the solar cell aligned orthogonal to the floor as shown in Figure 2. The width of both doorways is three feet. We install the device halfway above the floor on the doorframe to cover an optimal range of height. The lower the position of the solar cell, more likely the shadow of a person is going to impact the voltage. However, since solar energy-harvesting sensors usually should be placed as close as possible to a light source, we chose the midway to be the optimum point for deployment. Significant fraction of the light exposure of the room consists of artificial LED lights.



(a) Solar voltage trace of 50 room entry events of a single participant.



(b) Solar voltage trace of 50 room exit events of a single participant.

Figure 5: Data collection step of *SolarWalk* involves each participant walking through the door in every 10 seconds. However, noticeable change in solar cell voltage pattern is observed in the first six seconds, which contains 300 voltage samples. Thus, the dimension of the processed input is 1×300.

2.2 Data Collection Platform

We use the data acquisition platform developed in [11] to record the voltage trace from the PV cell. The data acquisition platform has a Raspberry Pi model 3A+ connected through an I2C interface to both a custom breakout board that contains an ADS1015 [1] analog to digital converter. The ADC samples the open-circuit voltage of a IXYS SMLD121H04L monocrystalline solar cell at 50 Hz. The gain stage of the ADC is configured to 8, resulting on a full-scale resolution of ±4.096 V, and a 2 mV least significant bit size. The platform records and streams data using MQTT protocol to a cloud-hosted database, that we later retrieve to train and evaluate the classifier models. Figure 2(b) shows the set up of the data collection module.

2.3 Data Collection Procedure

Our dataset involves five different participants with different BMI scores. We instructed the participants to enter and exit the door with 10 seconds interval by walking roughly through mid-doorway in their normal walking pace. Some participants were carrying small objects like a cellphone. We collected 200 door entry and exit events from four participants and 100 events from one participant as they walked through the door. Each participants walk a 100 times through each of the doors which involves either entering the room or exiting the room. Figure 4 illustrates our data collection step of a single participant, which started at point A, and ended at point B. In this trial, we collected 50 room entry samples and 50 room exit samples. Each walk spans ten seconds. We performed the data collection throughout different hours of the day including both day and night time to build a robust dataset, since the shadow pattern and the open circuit voltage of the solar cell is expected to change throughout the day.

Table 1: Processed Data Format

Data column	Description
event_label	Entry event: 0, Exit event: 1, No event: -1
occupant_label	Occupant's Id ranging [1-5]
collection_time	Data collection time: day/night
voltage_sample_[1-300]	Processed voltage samples from walking events

2.4 Data Pre-processing

Once we collected room entry and exit voltage traces from participants, we analyzed each trace carefully to identify the trigger point of the solar cell. Figure 5(a) illustrates 50 solar cell traces of one participant's entry event. We notice that, although each event spans for 10 seconds, noticeable change in voltage pattern happens in the first six seconds. Similar outcome can be noticed in exit events Figure 5(b). As such, during training and testing our machine learning models, we have taken traces from first six seconds. As our prototype collects data at 50 Hz sampling rate, a single entry event or exit event contains 6×50 samples. Thus, as input feature our ML models take 300 voltage readings.

2.5 Dataset Description

The dataset folder contains a folder named data and two example scripts demonstrating how to filter the raw data and apply machine learning models using the processed data. The data folder has two additional folders named raw_data and processed_data. The raw_data folder contains separate csv files for the participant's 17 minute long timestamped voltage samples (100 walk events) from two doors and another csv file with voltage samples corresponding to no shadow patterns. The naming format of the csv files is occupant_id_door_id_collection_time.csv. The processed_data contains the samples after we performed the filtering explained in Section 2.4 with labels. Table 1 describes the content of each columns of the data.

3 DATASET USE CASE EXAMPLES

3.1 Occupant Identification using *SolarWalk* Dataset

We investigate whether we can distinguish different participants and whether the trace corresponds to an entry or exit event from the collected shadow voltage pattern dataset. As shown in Figure 6,

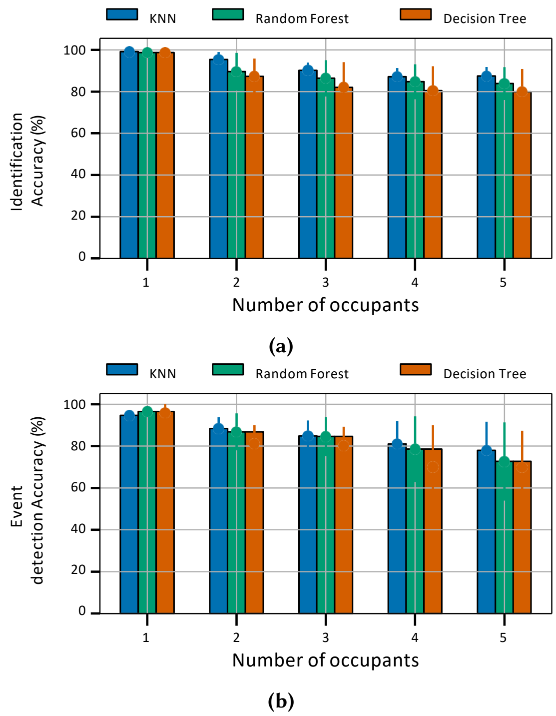


Figure 6: We evaluate *SolarWalk* dataset to identify five occupants from their shadow voltage pattern. With five occupants *SolarWalk*'s KKN classifier achieves 88% accuracy. We also determine whether the participant was entering or exiting the room. Results shows with five occupants the system is 77% accurate to determine the type of events.

with a KNN-based classifier we can identify five participants on average 88% of the time representing a 5-person home and on average 77% of the time, we can determine whether the participants were entering or exiting the room. We compare the performance of two other supervised learning method: decision tree and random forest. To understand how the accuracy is affected with the number of occupants, we evaluate both accuracy with increasing number of occupants. We find that the percentage of accuracy drops from 99% for one occupant to 88% for five occupants.

3.2 Future Directions

Our study shows that shadow pattern on a PV cell can be a unique attribute of a person to distinguish them from other individuals in a small smart home population. One future direction is to investigate how accurately we can determine whether a person is walking or rushing or running by their reflection on the PV cell voltage. Such activity monitoring can provide useful analytics without requiring the user to wear any devices. Moreover, we could estimate the walking speed of a person from the time series properties of their shadow pattern. Previous study shows that a person's gait and walking speed can be an indicator of their mental state and linked to anxiety, depression, and dementia [4, 15]. Therefore such information could benefit many individuals.

4 CONCLUSION

Future sensors will vastly benefit from the contextual ques of their installation location. Simultaneously, the ubiquitous nature of computing demands computers to be simple, unobtrusive, and pervasively-deployable. Taking a step towards this vision, in this paper, we introduce *SolarWalk* dataset that enables occupant identification using small photovoltaic voltage traces. We believe the dataset will help the community to explore further into this research directions and lead to potential applications beyond occupant identification.

5 ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their valuable insights and helpful feedback on improving this paper. This work is supported in part by the National Science Foundation (NSF) under grants CNS-1823325 and CNS-2144940, and the Strategic Investment Fund at the University of Virginia.

REFERENCES

[1] ADS101x 12-bit ADC. 2018. . https://www.ti.com/lit/ds/symlink/ads1015.pdf?ts=1663602271273&ref_url=https%253A%252F%252Fwww.google.com%252F.

[2] Bharathan Balaji, Jian Xu, Anthony Nwokafor, Rajesh Gupta, and Yuvraj Agarwal. 2013. Sentinel: occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. 1–14.

[3] Shiwei Fang, Tamzeed Islam, Sirajum Munir, and Shahriar Nirjon. 2020. Eyefi: Fast human identification through vision and wifi-based trajectory matching. In *2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 59–68.

[4] Ruth A Hackett, Hilary Davies-Kershaw, Dorina Cadar, Martin Orrell, and Andrew Steptoe. 2018. Walking speed, cognitive function, and dementia risk in the English longitudinal study of ageing. *Journal of the American Geriatrics Society* 66, 9 (2018), 1670–1675.

[5] IXOLAR High Efficiency SolarMD. 2016. IXYS. https://ixapps.ixys.com/DataSheet/SLMD121H04L_Nov16.pdf.

[6] Josiah Jee, Aveek K Das, Parth H Pathak, and Prasant Mohapatra. 2016. Motionsync: personal energy analytics through motion tags and wearable sensing. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. 65–74.

[7] Nacer Khalil, Driss Benhaddou, Omprakash Gnawali, and Jaspal Subhlok. 2017. Sonicdoor: scaling person identification with ultrasonic sensors by novel modeling of shape, behavior and walking patterns. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. 1–10.

[8] Shijia Pan, Ningning Wang, Yuqiu Qian, Irem Velibeyoglu, Hae Young Noh, and Pei Zhang. 2015. Indoor person identification through footstep induced structural vibration. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*. 81–86.

[9] Alessandra Parisio, Damiano Varagnolo, Daniel Risberg, Giorgio Pattarello, Marco Molinari, and Karl H Johansson. 2013. Randomized model predictive control for HVAC systems. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. 1–8.

[10] Juhi Ranjan, Erin Griffiths, and Kamin Whitehouse. 2014. Discerning electrical and water usage by individuals in homes. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. 20–29.

[11] Victor Ariel Leal Sobral, John Lach, Jonathan L Goodall, and Bradford Campbell. 2021. Thermal Energy Harvesting Profiles in Residential Settings. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 520–523.

[12] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. 2003. Silhouette analysis-based gait recognition for human identification. *IEEE transactions on pattern analysis and machine intelligence* 25, 12 (2003), 1505–1518.

[13] Wenpeng Wang, Jianyu Su, Zackary Hicks, and Bradford Campbell. 2020. The Standby Energy of Smart Devices: Problems, Progress, & Potential. In *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 164–175.

[14] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. 2016. WiWho: WiFi-based person identification in smart spaces. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 1–12.

[15] Nan Zhao, Zhan Zhang, Yameng Wang, Jingying Wang, Baobin Li, Tingshao Zhu, and Yuanyuan Xiang. 2019. See your mental state from your walk: Recognizing anxiety and depression through Kinect-recorded gait data. *PLoS one* 14, 5 (2019), e0216591.