



Nurani Saoda &lt;ns8nf@virginia.edu&gt;

## [DATA '22] Accepted submission #96 "SolarWalk Dataset: Occupant..."

1 message

DATA '22 HotCRP &lt;noreply-data22@hotcrp.com&gt;

Thu, Oct 6, 2022 at 11:44 PM

Reply-To: shiweifang@cs.umass.edu, gtferro@mines.edu

To: Brad Campbell <bradjc@virginia.edu>, Md Fazlay Rabbi Masum Billah <masum@virginia.edu>, Md Fazlay Rabbi Masum Billah <mb2vj@virginia.edu>, Nurani Saoda <ns8nf@virginia.edu>, Nurani Saoda <saoda@virginia.edu>, Victor Ariel Leal Sobral <sobral@virginia.edu>, Victor Ariel Leal Sobral <vas4d@virginia.edu>

Dear authors,

The Fifth International SenSys+BuildSys Workshop on Data: Acquisition To Analysis (DATA) (DATA '22) program committee is delighted to inform you that your submission #96 has been accepted to appear in the conference.

\* Title: SolarWalk Dataset: Occupant Identification using Indoor Photovoltaic Harvester Output Voltage

\* Authors: Nurani Saoda (University of Virginia);  
Md Fazlay Rabbi Masum Billah (University of Virginia);  
Victor Ariel Leal Sobral (University of Virginia);  
Bradford Campbell (University of Virginia)

\* Site: <https://data22.hotcrp.com/paper/96?cap=hcav96GiWMrjDVcADUrMfhqTwpJPXC>

Congratulations!

The camera-ready version is due on Oct 15, 2022. This is a HARD DEADLINE. We won't be able to publish if this deadline is missed due to ACM restrictions. Note that this deadline includes the final PDF version, correct name, contact info, and ACM eRight form. We suggest finalize submission before this date.

Please update your final title and contact information on HotCRP within next day as ACM eRight form will be auto generated by the system and we expect them to be ready soon.

Visit the submission site for reviews, comments, and related information. Reviews and comments are also included below.

- DATA '22 Submissions

Review #96A

=====

Reviewer expertise

-----

3. Knowledgeable

Paper summary

-----

The author proposed a method for occupancy detection by identifying the event that personnel entering/leaving a room. The model is based on the output voltage of a PV module, by collecting the timeseries data of the output voltage, and learning the pattern of voltage drops, the model can identify different occupants, and distinguish if they are entering or exiting the room.

Comments for authors

-----

The paper is motivated by the limitation of current occupant identification approach, including requirement for additional devices, or being privacy intrusive. The insight for using a PV module is interesting, but not convincing enough. Infrared sensors or even door/window sensors can also achieve similar functionality, and have low privacy issues. The proposed method cannot show enough novelty or performance improvement comparing with existing work. Furthermore, the PV module for small IoT devices is indeed widely used in smart facilities. But installing it on doorframe can be intrusive and requires additional efforts to connect it with an external hub (like Pi) for reading the voltage data. The usage of collected data can be highly impacted by different installation of the light source and PV module. The speed for occupants entering/leaving the door is highly fixed, which also has low generality.

Were you able to access the dataset at the link provided?

1. Yes

Were you able to manipulate/visualize the dataset? How helpful was the provided notebook/code/description?

Yes. The author proposed two jupyter notebook scripts for processing the data and the model.

Is the dataset description accurate?

Yes.

How does the dataset compare to existing datasets?

The dataset have relatively low compatibility due to the configuration of the experiment.

\*\*\*\*\*

Review #96B

Reviewer expertise

3. Knowledgeable

Paper summary

The paper proposes an interesting task and dataset, which is to use the indoor photovoltaic harvester to identify the occupants in the building. The advantage of photovoltaic harvesters is that they are ubiquitously installed, so no privacy issue involved. The basic idea is to use the difference of temporal voltage patterns caused by shadow of different people when entering and existing the door. The data were collected from five participants with different BMI using Raspberry Pi. The paper also demonstrates the potential usages of the dataset: use machine learning to identify occupants and whether they are entering or existing the room with the voltage trajectories. The initial experiments showed that the accuracy is 88% and 77%.

Comments for authors

I highly appreciate authors' clear illustration of the motivation of dataset and the data collection process, as well as demonstration of the possible future use cases. Also, I have several questions if we would like to put the dataset into practice in the real life (e.g. training machine learning models from the dataset and deploy them in real apartments to achieve the goal of intelligent housing):

1. How large is the inter-person and intra-person voltage difference when they enter and exist the room? If we would like to use the voltage as the feature to differentiate different people, we need to make sure the inter-person variation to be larger than intra-person variation. I think maybe authors can quantify this by measuring the time series similarity between different time series of one person and time series similarity between different persons.
2. In the real world, the cases are more complex than just having occupants with different BMI. For example, even the same person could have different gaits or shadow patterns under different circumstances (e.g. carrying a very heavy bag, or entering and exiting the room with another person so that their shadows are overlapping.) So it would be interesting to consider these "noisy" corner cases if we would like to tackle the occupant identification problem in the real life.
3. Although the dataset is novel and proposes an interesting method to solve the occupant identification problem, the dataset might be small. Even if the machine learning models can achieve a moderate performance with the data collected from five participants, I am not sure how this results could generalize to more people. Also, in the jupyter notebook "Process\_Data.ipynb", a '\_' is supposed to be in the path "data/raw data" and "data/processed data". (i.e. raw\_data, processed\_data)

Were you able to access the dataset at the link provided?

1. Yes

Were you able to manipulate/visualize the dataset? How helpful was the provided notebook/code/description?

-----  
Yes, I have checked both the dataset and executed the notebook, and the results are easy to reproduce.

Is the dataset description accurate?

-----  
Yes

How does the dataset compare to existing datasets?

-----  
The dataset is novel, which is the first dataset of using the indoor photovoltaic cell for solving the occupant identification task.