Dataset: Occupancy Detection, Tracking, and Estimation Using a Vertically Mounted Depth Sensor

Fabricio Flores*
Carnegie Mellon University
Machine Learning Department
wflores@andrew.cmu.edu

Sirajum Munir*
Bosch Research & Technology Center
Secure and Intelligent IoT Department
sirajum.munir@us.bosch.com

Matias Quintana
National University of Singapore
School of Design and Environment
matias@u.nus.edu

Anand Krishnan Prakash
Lawrence Berkeley National Laboratory
akprakash@lbl.gov

Mario Bergés
Carnegie Mellon University
Civil & Environmental Engineering
mberges@andrew.cmu.edu

ABSTRACT

Occupancy detection, tracking, and estimation has a wide range of applications including improving building energy efficiency, safety, and security of the occupants. As depth sensors are getting cheaper, they offer a viable solution to estimate occupancy accurately in a non-privacy invasive manner. Even though there are publicly available depth datasets, they do not consider placing the sensor in the ceiling looking downwards to estimate occupancy. We deployed four Kinect for XBOX One in four CMU classrooms and conference rooms for a period of four weeks in 2017 and collected over 6 TB of depth data. We annotate this huge dataset by labelling bounding boxes around occupants and release the annotated dataset.

CCS CONCEPTS

• **Computer systems organization** → *Embedded and cyber-physical systems*; • **Human-centered computing**; • **Hardware** → *Sensor applications and deployments*;

KEYWORDS

Datasets, depth data, human detection, occupancy estimation

ACM Reference Format:
Fabricio Flores, Sirajum Munir, Matias Quintana, Anand Krishnan Prakash, and Mario Bergés. 2019. Dataset: Occupancy Detection, Tracking, and Estimation Using a Vertically Mounted Depth Sensor. In *The 2nd Workshop on Data Acquisition To Analysis (DATA'19), November 10, 2019, New York, NY, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3359427.3361916

1 INTRODUCTION

Occupancy detection, tracking, and estimation is primitive to provide efficient energy management of buildings as well as improving safety and security of the occupants. Various sensors have been used for this purpose including PIR motion detectors, RGB cameras, ultrasonic sensors [10], infrared array sensors [8], depth sensors [9], millimeter wave radar [12], radio [3], and thermal images. While each sensing solution has its strength and weakness, with the rapid drop of price, depth sensors have the potential to provide very accurate occupancy estimation in a non-privacy invasive way [9] [7].

There has been several work to detect humans as well as their gesture and pose using depth sensors [2] [11]. However, in most of these works, the depth sensor is placed in front of individuals. There has been very few attempts to collect depth data from a sensor that is mounted in the ceiling looking downwards. To this end, we deployed four Microsoft Kinect for Xbox One (Kinect Version 2.0) into four different classrooms and conference room entrances at the main campus of Carnegie Mellon University (CMU) in 2017. The deployment at each setting is shown in Figure 1. Rooms 1 and 2 correspond to classrooms, while Rooms 3 is a conference room and Room 4 is a student lounge. The size of the doors in each of these rooms varied from 3 feet to 6 feet and the sensors were placed at heights between 9.43 feet to 9.48 feet. Each Kinect was connected to an Intel Next Unit of Computing (NUC) for capturing the depth images. We collected raw depth frames from each door for a period of a work week resulting in over 6TB large scale dataset of depth information. Our dataset captures various realistic scenarios including door opening and closing, students coming with backpacks, gym bags, papers, laptops, bike helmets hanging around the shoulders, carrying jackets in hands, carrying paper scrapbooks, wearing headphones over their heads, wearing sunglasses over their heads, wearing caps, and wearing hoodies. We realize that such one week data collection per deployment does not capture long term occupancy dynamics. However, our long term occupancy data is available in [6], [4], [5]. Also, our dataset does not annotate the objects that are being carried by humans. We release another dataset [7], where various objects are annotated in addition to humans including phone, umbrella, cup, box, gun, laptop, and backpack. Table 1 shows the duration of data collection from each room, the frame rate, an approximate number for the collected frames, the size of the full raw dataset, an approximate number of frames with humans and its respective size.

*Equal contribution.

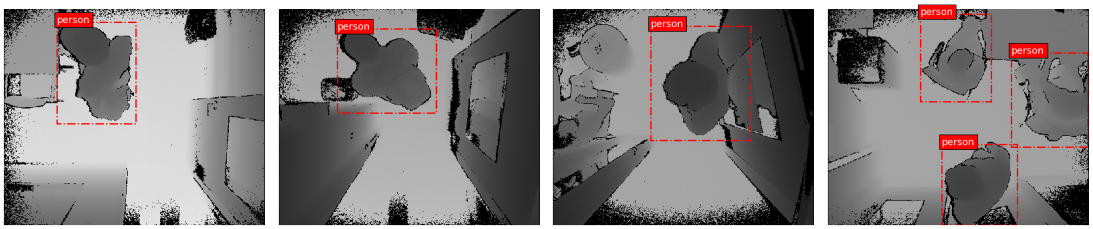
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
DATA'19, November 10, 2019, New York, NY, USA
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6993-0/19/11…$15.00
https://doi.org/10.1145/3359427.3361916

(a) Room 1 (b) Room 2 (c) Room 3 (d) Room 4

Figure 1: Deployment of Kinects at the ceiling near doorways in four deployment settings

	Start Date	End Date	Frames per Second	Total Frames	Total Size	Selected Frames	Selected Frames Size
Room 1	JUN-14-2017	JUL-20-2017	28.08	6.2M	1.78 TB	14,000	3.0 GB
Room 2	MAY-08-2017	MAY-15-2017	28.90	6.1M	1.1 TB	8,000	1.7 GB
Room 3	JUN-27-2017	JUL-03-2017	28.33	6.5M	1.7 TB	30,000	5.5 GB
Room 4	MAY-26-2017	MAY-31-2017	29.39	6.2M	1.55 TB	18,000	3.5 GB

Table 1: Description of the entire data collection.



(a) Room 1 (b) Room 2 (c) Room 3 (d) Room 4

Figure 2: Sample images of the annotated depth frames from each deployment setting.

2 DATA

Our dataset consists of around 70,000 depth images after discarding the depth frames that do not contain any humans. Each depth image has the resolution of 512×424 (single channel). Each pixel is a 16 bit unsigned integer that provides the distance from the sensor to the nearest object in millimeters. We annotate each of these frames by labelling bounding boxes around each person in the frame. An example depth frame from each deployment setting with annotated bounding box is shown in Figure 2. Note that the depth images have been pre-processed for ease of visibility in Figure 2. We save the annotations as XML files in PASCAL VOC format, which provide the location of each person within an image and the corresponding label. This is the first dataset that captures such diverse nature of movement of people using vertically mounted depth sensors.

After the annotation, we divide the dataset into training, validation, and testing. The approximate number of depth images in each partition is show in Table 2. Such a separation will help future works to perform benchmarking comparison. When creating this split, we maintain the temporal sequence of frames, i.e., the frames with increasing file numbers are in chronological order instead of randomly shuffled so that they can be used to track and estimate occupancy. The dataset is available for download from here [1].

	Training	Validation	Testing
Images	42,000	14,000	14,000

Table 2: Number of annotated depth images for training, validation, and testing.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Building Technologies Office Award Number DE-EE0007682. The views expressed in this paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Fabricio Flores is supported by Programa 'Becas de Excelencia', Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT), Ecuador.

REFERENCES

[1] 2019. Occupancy detection dataset. http://doi.org/10.5281/zenodo.3404204

[2] Kanad K Biswas and Saurav Kumar Basu. 2011. Gesture recognition using microsoft kinect®. In *The 5th international conference on automation, robotics and applications*. IEEE, 100–103.

[3] Chen-Yu Hsu, Rumen Hristov, Guang-He Lee, Mingmin Zhao, and Dina Katabi. 2019. Enabling Identification and Behavioral Sensing in Homes using Radio Reflections. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 548.

[4] KS Liu, S Munir, J Francis, C Shelton, and S Lin. 2017. Poster Abstract: Long term occupancy estimation in a commercial space: an empirical study. IPSN.

[5] KS Liu, S Munir, S Lin, and C Shelton. 2018. Understanding Occupancy Patterns in a Commercial Space. *IEEE Technical Committee on Cyber-Physical Systems* 3, 1.

[6] Kin Sum Liu, Elvin Vindel Pinto, Sirajum Munir, Jonathan Francis, Charles Shelton, Mario Berges, and Shan Lin. 2017. COD: a dataset of commercial building occupancy traces. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 50.

[7] Niluthpol Chowdhury Mithun, Sirajum Munir, Karen Guo, and Charles Shelton. 2018. ODDS: real-time object detection using depth sensors on embedded GPUs. In *Proceedings of the 17th ACM/IEEE International Conference on Information Processing in Sensor Networks*. IEEE Press, 230–241.

[8] Hessam Mohammadmoradi, Sirajum Munir, Omprakash Gnawali, and Charles Shelton. 2017. Measuring people-flow through doorways using easy-to-install ir array sensors. In *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 35–43.

[9] Sirajum Munir, Le Tran, Jonathan Francis, Charles Shelton, Ripudaman Singh Arora, Craig Hesling, Matias Quintana, Anand Krishnan Prakash, Anthony Rowe, and Mario Berges. 2017. FORK: Fine Grained Occupancy estimatoR Using Kinect on ARM Embedded Platforms. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys '17)*. ACM, New York, NY, USA, Article 33, 2 pages. https://doi.org/10.1145/3137133.3141461

[10] Oliver Shih and Anthony Rowe. 2015. Occupancy estimation using ultrasonic chirps. In *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*. ACM, 149–158.

[11] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. 2013. Real-time human pose recognition in parts from single depth images. *Commun. ACM* 56, 1 (2013), 116–124.

[12] Peijun Zhao, Chris Xiaoxuan Lu, Jianan Wang, Changhao Chen, Wei Wang, Niki Trigoni, and Andrew Markham. 2019. mID: Tracking and Identifying People with Millimeter Wave Radar. In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 33–40.