Occupancy Sensing and Activity Recognition with Cameras and Wireless Sensors

Yang Zhao
GE Research
Niskayuna, New York
yang.zhao@ge.com

Peter Tu
GE Research
Niskayuna, New York
tu@research.ge.com

Ming-Ching Chang
University at Albany, SUNY
Albany, New York
mchang2@albany.edu

ABSTRACT

We present a system work combining visual cameras and wireless sensors for human occupancy detection and activity recognition. We describe our testbed system, data collected from a human subject study, observations from long-term occupancy experiments, and preliminary analytical results. We apply machine learning algorithms to the human activity recognition data, and identify challenges in applying the state-of-the-art deep learning techniques to wireless sensing of human activity. We find that packet loss due to wireless interference has a significant effect on time series classification. We also find that the convolutional neural networks significantly outperforms the conventional support vector machine method, but further experiments need to be performed to investigate environment-independent classification and the overfitting issue. Finally, we discuss future research topics that can use our testbed of wireless sensors and visual cameras to automate data labeling in deep learning model training.

CCS CONCEPTS

• Computing methodologies → Supervised learning by classification; **• Computer systems organization → Sensor networks**; **• Human-centered computing → Ubiquitous and mobile computing systems and tools**.

KEYWORDS

Occupancy Detection, Activity Recognition, Wireless Sensing

ACM Reference Format:
Yang Zhao, Peter Tu, and Ming-Ching Chang. 2019. Occupancy Sensing and Activity Recognition with Cameras and Wireless Sensors. In *DATA'19 '19: Proceedings of the Second Workshop on Data Acquisition To Analysis, November 10, 2019, New York, NY, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3359427.3361911

1 INTRODUCTION

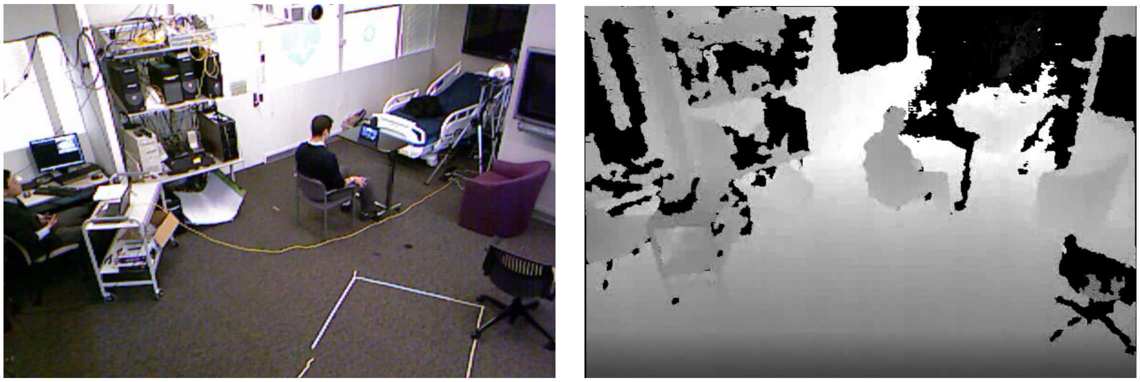
Occupancy sensing and human activity recognition is important in many applications, such as security, healthcare care, and smart facilities. For example, detection of empty beds and rooms can increase asset utilization and speed up hospital workflow [12]. Human activity recognition can provide context information for prompt decision making and operation in hospitals [11]. In our pressure ulcer prevention and care application, patient activity data is critical in analyzing treatment progress for better treatment and care [1].

There are different sensing techniques for occupancy and human activity monitoring. Wearable technique recently becomes a popular way of monitoring people's activity [6], but it requires user cooperation, and wearable sensors need to be attached to the human body, which may cause discomfort to users. Computer vision is also a widely-used technique and it can monitor human activity in a non-cooperative way. However, visual cameras have privacy issues that limit its applicability in many scenarios. As wireless devices are becoming pervasive in recent years, wireless sensor becomes a cost-effective and promising sensing modality for many internet of things (IoT) applications. For example, [14] claims that they can use radio frequency (RF) sensors to estimate human pose accurately through walls. [13] uses channel state information (CSI) from IEEE 802.11 radio chips to perform device-free human activity recognition. In addition, [16] uses Doppler sensors and received signal strength (RSS) measurements to detect occupancy and classify human activity. In this work, we build a sensing system with cameras and two types of wireless sensors to detect occupancy and classify human activities in a non-cooperative way.

Sensing and perception in a dynamic environment is generally a challenging problem. For occupancy sensing and activity recognition, changes in lighting condition affect the performance of camera-based systems, while RF-based systems are sensitive to the multi-path effect. Multi-modal sensing provides a way to explore the complimentary nature of multiple sensors to achieve high performance. In this work, we demonstrate research effort in this direction. Specifically, we build a multi-modal sensing system by integrating three subsystems: visual tracking system with RGB-D cameras [3], Doppler system with Doppler motion sensors [16], and RF sensor network with IEEE 802.15.4 radio sensors [9]. We perform long-term occupancy sensing experiments and short-term human activity recognition experiments during a three-year human subject study. We present data collected from the study and preliminary analytical results. While various signal processing and machine learning methods can be applied to the dataset and explore different perspectives of the data, we start our analytical work applying deep learning-based and feature-based classification algorithms on the Doppler human activity data. We find that the convolutional neural networks (CNN) significantly outperforms the support vector machine (SVM) method. We also describe observations and challenges from the experiments. For example, the wireless interference problem causes packet loss to the RSS data, which presents a challenge to robust human activity

DATA'19 '19, November 10, 2019, New York, NY, USA

ACM ISBN 978-1-4503-6993-0/19/11…$15.00
https://doi.org/10.1145/3359427.3361911



(a) RGB image **(b) Depth image**

Figure 1: RGB (a) and Depth (b) images from the camera tracking system.

recognition. Finally, we discuss future research topics to further use our dataset and testbed system. The dataset is published on zenodo: https://doi.org/10.5281/zenodo.3454785. We encourage researchers to test their ideas and methods on this dataset.

To summarize, the contribution of this paper includes: 1) integration of a multi-modal sensing system, 2) human subject experiments and data collection, 3) application of CNN on the Doppler human activity data, 4) observations from experiments for future research.

2 SYSTEMS AND EXPERIMENTS

We develop a hybrid wireless [16] and computer vision (CV) system [3] as the testbed and collect four types of sensor data: RGB image, Depth image, Doppler and RSS. Figures 1 and 2 show data examples from experiments of our human subject study (E&I #14053). We also deploy a Vicon motion capturing system so that we can record the ground-truth of the human subject location and activity. We describe the testbed system and experiments next.

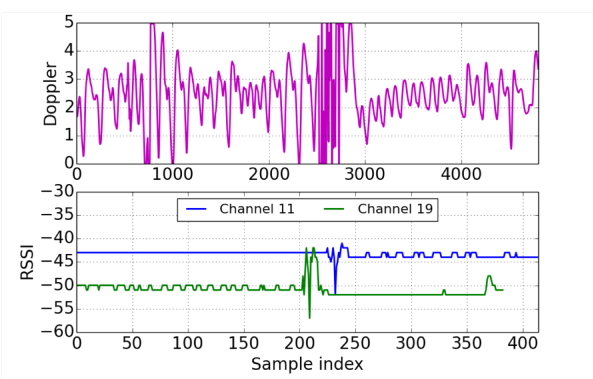


Figure 2: Doppler and RSS measurements from the wireless sensing system.

2.1 Testbed system

The CV system is composed of two Kinect cameras deployed at two corners of a room ceiling to cover the scene of the room with RGB and depth images, as shown in Figure 1. The CV system is able to track multiple people with cameras calibrated at a specific environment [3]. The wireless sensing system includes two types of sensors: a Doppler motion sensor [16] and an RF sensor network [10]. The Doppler sensor is a low-cost dual Doppler sensor modified from a commercial-off-the-shelf range-controlled radar, which operates at 5.8 GHz with two directional antennas [16]. The RF network uses four IEEE 802.15.4 radio nodes (CC2531 from TI) to create a mesh network to measure the RSS between each pair of radio nodes operating on 16 frequency channels at 2.4 GHz. This low-power wireless network is capable of locating a person's location and even estimating the person's respiration rate [10].

A computer with data acquisition software is used to stream video from cameras and record each individual frame, while the data acquisition of the Doppler and RF network is implemented on a BeagleBone embedded computer. The Doppler analog signal is fed into the analog-to-digital converter (ADC) of the BeagleBone board, to which a wireless base station is also connected, to collect the RSS measurements from the RF network. The sampling rates of Doppler and RF sensors are set to be 10 Hz and 3.3 Hz, respectively, by the firmware on the embedded computer. A snapshot of the Doppler and RSS time series is shown in Figure 2.

2.2 Experiments

We deploy the testbed system described above in a 5.5m by 7.5m room. As shown in the experimental layout in Figure 3, two Kinect cameras are deployed at two diagonal corners of the room ceiling to obtain a full room coverage. The dual Doppler sensor is deployed on a ceiling-mounted rail above the bed with one antenna facing the bed and the other antenna facing the room. Four radio nodes are placed at four corners to form a wireless network.

We have performed two types of experiments: long-term occupancy experiment and short-term activity experiment. For the occupancy experiment, the purpose was to detect if the room is occupied or not, and data was collected with different time duration from one hour to two days. For the activity experiment, we recruited human subjects to perform 42 trials of four activities (each one with two minutes duration): (1) *walking in a room* (10 trials), (2) *sitting in a chair* (10 trials), (3) *lying on a bed* (12 trials), and

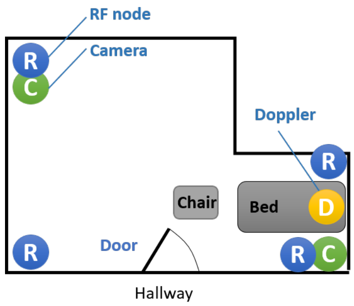


Figure 3: Experimental layout.

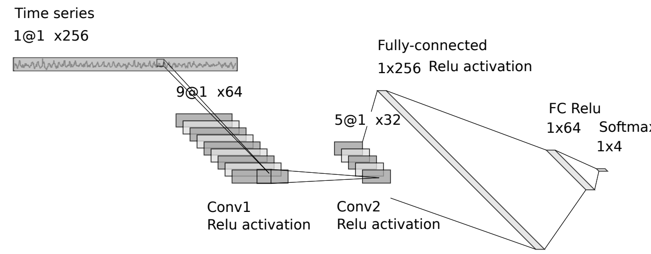


Figure 4: Architecture of CNN networks for human activity recognition.

(4) *body turning on a bed* (10 trials). For the walking activity, the human subjects walk along different paths at different locations in the room. For the sitting activity, the location of the chair is fixed, as shown in Figure 3. For the lying on bed activity, we ask human subjects to breathe normally on bed with three orientations facing upwards, right and left. Finally, for the turning on bed case, human subjects turn their bodies from one side to the other on bed with random time intervals. We also recorded two-minute data of the *empty room* case before and after each human subject trial.

3 DATA ANALYTICS AND CHALLENGES

We analyze data collected from the long-term and short-term experiments, apply machine learning methods, and present our findings and challenges in this section.

3.1 Analysis of Human Activity Data

For the short-term human activity data, we use the Doppler time series data as the example to demonstrate our data analysis approach. All the Doppler data from 42 trials are split into 21 training cases and 21 testing cases, with 1200 samples (2 minutes) in each trial. Using a moving window with a length of 256 samples, we obtained 19845 windowed-samples for training and testing, respectively. We describe the machine learning methods and evaluation results next.

3.1.1 Methods and Results. The deep convolutional neural networks (CNN) are shown to achieve the state-of-the-art performance for both 1-D time series and 2-D image classification [7], thus we apply the CNN on the Doppler time series. We construct two convolutional layers with the rectified linear unit (ReLU) as the activation function, and a fully connected layer with the softmax function before the output layer. The overall architecture of the CNN network is shown in Figure 4. Note that we use a moving window over the time series to generate a large training dataset. The window size should be chosen based on the sampling rate of the sensing system to capture the temporal correlation and features of the time series data. We implement the CNN network using the Keras APIs, and more details of the CNN implementation can be found in [15].

We also use a conventional machine learning method, support vector machine (SVM) with a radial basis function kernel. We choose hand crafted features, such as peak frequency of the Doppler power spectral density, the variance of the time series, as described in [16]. Due to insufficient data in the *sitting in a chair* experiment, we perform the CNN and SVM classification and comparison on the following four classes: *walking in a room*, *lying on a bed*, *body turning on a bed*, and *empty room*. The Doppler time series data for these classes are shown in Figure 5.

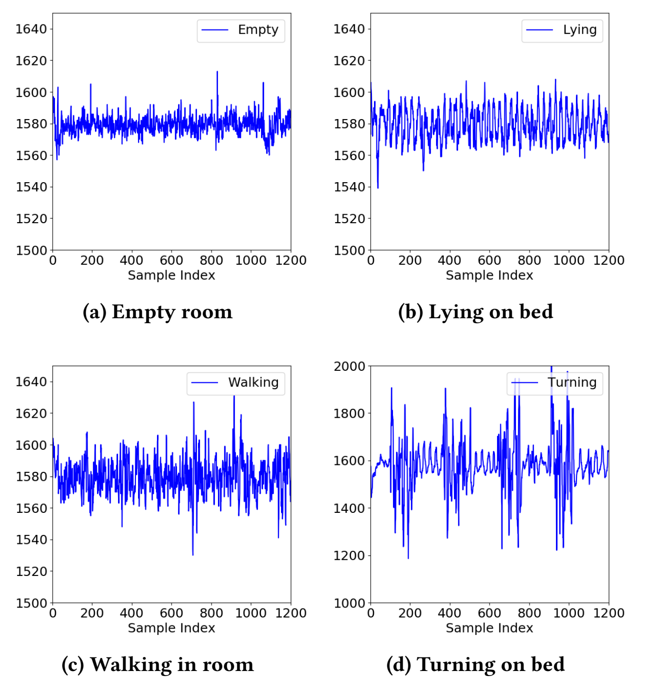


Figure 5: Doppler time series for four classes (a) empty room, (b) lying on bed, (c) walking in room and (d) turning on bed.

Finally, we evaluate the performance of the CNN and SVM classifiers with two evaluation methods: trial-based evaluation and window-based evaluation. For the trial-based evaluation, we perform evaluation for 21 testing trials. That is, we have one classification result for each of the two-minute (1200 samples) testing trials. For the window-based evaluation, the classifiers are evaluated on all 19845 testing samples. In another word, we have one classification result for each time window with 256 samples. Tables 1 and 2 show the confusion matrices of the CNN-based classifier from these two

	Empty	Lying	Turning	Walking
Empty	5	0	0	0
Lying	0	6	0	0
Turning	0	0	5	0
Walking	1	0	0	4

Table 1: Confusion matrix from trial-based evaluation.

	Empty	Lying	Turning	Walking
Empty	4595	0	0	130
Lying	0	5288	266	116
Turning	0	106	4298	321
Walking	532	75	0	4118

Table 2: Confusion matrix from window-based evaluation.

evaluation methods. We see that only one trial from 21 trials is misclassified for the trial-based evaluation. For the window-based evaluation, the classification rates are 97.2%, 93.3%, 90.9% and 87.1% for the empty room, lying, turning and walking activities, respectively. Note that the SVM classifier was evaluated in [16], where the average classification rate is 55% when using the same Doppler time series data. Thus, the CNN-based classifier outperforms the SVM-based classifier by 37% on average.

3.1.2 Challenges and Discussion. We find that the RSS measurements suffer from the wireless interference problem, which is a challenge for multi-modal sensor fusion and time series classification. As shown in Figure 2, we see high variations in the RSS when the human subject turned from one orientation to another orientation during the lying-in-bed experiment. However, the wireless link on Channel 19 has over 40 packets loss during the two-minute experiment, and thus the high variation of RSS on Channel 19 appeared earlier (around index 210 in Figure 2) than that of Channel 11 and the Doppler time series. Without using data imputation for missing data, the wireless interference problem will significantly affect time synchronization for sensor fusion.

In addition, the packet loss issue will also affect the temporal correlation of the time series data and thus the performance of time series classification. We find that if we omit the missing values of the RSS data and directly apply CNN on RSS, the average classification rate is 77%. We can use various data imputation methods, such as the mean values or the K-nearest neighbors, to fill in the missing values due to packet loss. We can also use recurrent neural networks, such as the gated recurrent unit (GRU)-D method as proposed in [2] to jointly predict the missing values and perform time series classification. As future work, we plan to further investigate the effect of packet loss on the performance of various time series classification methods.

3.2 Analysis of Occupancy Data

For the data collected from the long-term occupancy experiment, we use the CV system to automatically label the data as occupied or empty room. We find that the variance of the Doppler time series in a time window can be used to detect room occupancy. We also find that the Doppler time series from the bed sensor and the room sensor have different probabilistic distributions, but both have long ending tails.

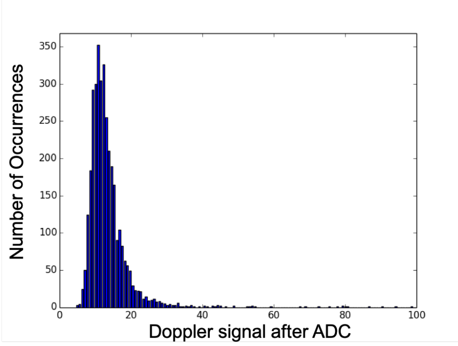
Figure 6 shows the histograms of the Doppler variance (from 100 digitized Doppler samples in a window) from the Doppler sensors facing the bed and room, during a 12-hour experiment without any occupants in the room. Observe that the bed Doppler sensor has a narrower distribution of variance than the room sensor. This difference comes from the different noise statistics of the sensor hardware, and it is also due to the intrinsic motion of the environment that we discuss next. From Figure 6 we also see that both histograms have long tails, with the variance values up to 90 and 100 (note that the Doppler samples are digitized after the AD converter, and thus have no unit). However, the variance of the Doppler sensor from an occupied room can be as low as 80. Thus, a Doppler variance threshold-based detection method cannot achieve high detection rate.

An interesting observation from our long-term experiment is that the Doppler signal can be affected by the environmental intrinsic motion. From a two-day experiment, we find that the time series from the bed Doppler sensor showed sinusoidal-shape pattern even when no occupant was present in the room. We also find that this pattern disappeared after 6pm Friday, which was the time that the building facility turned off the heating, ventilation and air conditioning (HVAC) system in that room. It turns out that the plastic mattress of the hospital bed in the room shrinks and expands periodically due to the HVAC system. Since the directional antenna of the bed Doppler sensor is only two meters away from the mattress surface, the bed Doppler sensor is sensitive enough to capture the deformation of the bed mattress caused by the HVAC system. To make the Doppler sensing system more robust to the environmental noise, a state machine [5] can be used as a low-pass filter to achieve better detection performance.

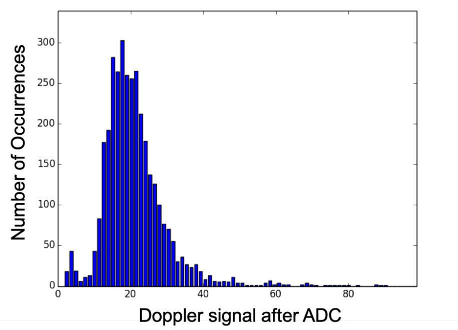
3.3 Future Work

As future work, we plan to design experiments to further explore the complimentary nature of the CV system and the wireless sensing systems for occupancy detection, location estimation and activity recognition. One challenge in applying deep learning algorithms is the requirement of a large training dataset. We propose to use our testbed system to achieve automatic data labeling for the wireless sensing systems. Since the camera-based tracking system achieves high performance under good light conditions, we can collect data from a controlled environment and use the CV system to label the Doppler and RSS data, instead of manually labeling the ground-truth. As described in Section 3.2, we used cameras to automatically label data as occupied or empty-room for occupancy detection with Doppler and RSS data. However, for human activity recognition, we had to manually label each two-minute trial. In future, we can ask participants to perform specific gestures before performing certain activity training, and the CV system can recognize predefined gestures and then automatically label collected data.

We also plan to perform further experiments to investigate the overfitting issue and the missing data issue. We currently collect data from only one environment. Although the classification rate from the CNN algorithm is over 90% for the Doppler time series classification, we need to perform more experiments from different



(a) Bed Doppler Sensor



(b) Room Doppler Sensor

Figure 6: Histogram of variance from (a) Doppler sensor facing bed and (b) Doppler sensor facing room.

environments to investigate the possible overfitting issue and explore environment-independent recognition capability. With more data collected, we can also further investigate the missing data issue caused by the wireless interference, which is a unique problem in wireless sensing and many IoT applications. As an example, we plan to investigate the recurrent neural networks framework proposed in [2] to jointly perform data imputation and time series classification.

4 RELATED WORK

For the human activity recognition problem, various wireless sensing techniques have been proposed and developed recently [8]. [13] uses channel state information (CSI) of wireless devices and proposed a sparse representation classification-based method to recognize lying, sitting, standing and walking activities. [16] fuses the received signal strength (RSS) measurements from a wireless network with the Doppler signal and uses the support vector machine (SVM) method to classify four activities. In [4], four wireless testbeds including WiFi, ultrasound, mmWave and visible light are used to extract environment-independent features for device-free human activity recognition.

As the deep neural networks (DNN)-based machine learning methods have achieved the state-of-the-art performance in the computer vision related applications, researchers have used convolutional neural networks (CNN) and other deep learning methods in the human activity wireless sensing application [4, 8]. In this work, we present preliminary results in applying the CNN algorithm on the Doppler time series data.

5 CONCLUSION

We have integrated a computer vision tracking system with two wireless sensing systems to collect four types of sensor data for occupancy detection and human activity recognition. We describe our observations and findings from a three-year human subject study. We discuss challenges, problems and solutions in applying machine learning algorithms in human activity recognition. Finally, we propose future topics and directions on how to further use the dataset and the multi-modal system.

ACKNOWLEDGMENTS

This work was funded by the Department of Veterans Affairs under Grant No VA118-13-C-0049.

REFERENCES

[1] Ming-Ching Chang, Ting Yu, Jiajia Luo, Kun Duan, Peter Tu, Yang Zhao, Nandini Nagraj, Vrinda Rajiv, Michael Priebe, Elena A Wood, et al. 2017. Multimodal sensor system for pressure ulcer wound assessment and care. *IEEE Transactions on Industrial Informatics* 14, 3 (2017), 1186–1196.

[2] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. 2018. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports* 8, 1 (2018), 6085. https://doi.org/10.1038/s41598-018-24271-9

[3] Jixu Chen, Ming-Ching Chang, Tai-Peng Tian, Ting Yu, and Peter Tu. 2015. Bridging computer vision and social science: A multi-camera vision system for social interaction training analysis. In *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 823–826.

[4] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, Wenyao Xu, and Lu Su. 2018. Towards Environment Independent Device Free Human Activity Recognition. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom '18)*. ACM, New York, NY, USA, 289–304. https://doi.org/10.1145/3241539.3241548

[5] Christopher Donald Johnson, Brandon Stephen Good, Andrew Phelps Day, David S Toledano, Yang Zhao, and Jeffrey Richardson Terry. 2017. Methods and systems to sense situational awareness with a dual doppler and control for optimized operations. Patent Application No. 14885169.

[6] Oscar D Lara and Miguel A Labrador. 2012. A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials* 15, 3 (2012), 1192–1209.

[7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (27 5 2015), 436–444. https://doi.org/10.1038/nature14539

[8] Yongsen Ma, Gang Zhou, and Shuangquan Wang. 2019. WiFi Sensing with Channel State Information: A Survey. *ACM Comput. Surv.* 52, 3, Article 46 (June 2019), 36 pages. https://doi.org/10.1145/3310194

[9] Neal Patwari and Joey Wilson. 2010. RF sensor networks for device-free localization: Measurements, models, and algorithms. *Proc. IEEE* 98, 11 (2010), 1961–1973.

[10] Neal Patwari, Joey Wilson, Sai Ananthanarayanan, Sneha K Kasera, and Dwayne R Westenskow. 2013. Monitoring breathing via signal

strength in wireless networks. *IEEE Transactions on Mobile Computing* 13, 8 (2013), 1774–1786.

[11] Dairazalia Sánchez, Monica Tentori, and Jesús Favela. 2008. Activity recognition for the smart hospital. *IEEE intelligent systems* 23, 2 (2008), 50–57.

[12] Bex George Thomas, Srinivas Bollapragada, Kunter Akbay, David Toledano, Peter Katlic, Onur Dulgeroglu, and Dan Yang. 2013. Automated Bed Assignments in a Complex and Dynamic Hospital Environment. *Interfaces* 43, 5 (2013), 435–448. http://www.jstor.org/stable/43699330

[13] Bo Wei, Wen Hu, Mingrui Yang, and Chun Tung Chou. 2015. Radio-based Device-free Activity Recognition with Radio Frequency Interference. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks (IPSN '15)*. ACM, New York, NY, USA, 154–165. https://doi.org/10.1145/2737095.2737117

[14] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. 2018. Through-Wall Human Pose Estimation Using Radio Signals. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[15] Yang Zhao, Zhaoyuan Yang, and Weizhong Yan. 2019. *Adversarial Attack on Time Series Classification in Human Activity Sensing*. Technical Report 2019GRC0211. GE Research, General Electric, Niskayuna, NY.

[16] Yang Zhao, Ting Yu, and Jeff Ashe. 2015. Poster: Non-invasive Human Activity Monitoring Using a Low-cost Doppler Sensor and an RF Link. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems (SenSys '15)*. ACM, New York, NY, USA, 397–398. https://doi.org/10.1145/2809695.2817882