From Virtual to Real World: Applying Animation to Design the Activity Recognition System

Chengshuo Xia*
Graduate School of Science and Technology
Yokohama, Japan
sugiura@keio.jp

Yuta Sugiura
Faculty of Science and Technology, Keio University
Yokohama, Japan
sugiura@keio.jp

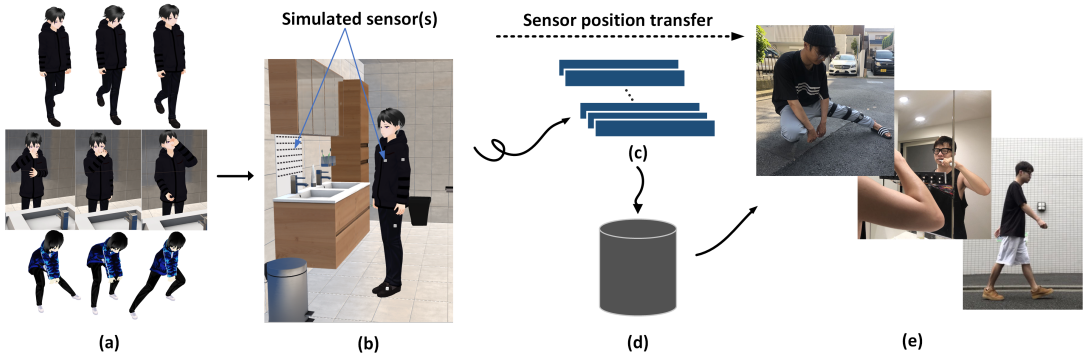

Figure 1: Proposed virtual environment working flow pipeline (a)Input virtual animation.(b) Simulated multi-model sensors placed on the body and environment.(c) Produced virtual dataset. (d) Classifier trained by virtual dataset. (e) Real activity recognized

ABSTRACT

Following the conventional pipeline, the training dataset of a human activity recognition system relies on the detection of the significant signal variation regions. Such position-specific classifiers provide less flexibility for users to alter the sensor positions. In this paper, we proposed to employ the simulated sensor to generate the corresponding signal from human motion animation as the dataset. Visualizing the corresponding items from the real world, the user can determine the sensor's placement arbitrarily and obtain accuracy feedback as well as the classifier interface to get relief from the cost of a conventional training model. With the cases validation, the classifier trained by simulated sensor data can effectively recognize the real-world activity.

CCS CONCEPTS

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

KEYWORDS

Activity recognition, Machine learning, Sensor simulation

ACM Reference Format:
Chengshuo Xia and Yuta Sugiura. 2018. From Virtual to Real World: Applying Animation to Design the Activity Recognition System. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Human activity recognition (HAR) systems have been successfully applied to multiple aspects in human daily life [7, 19, 20]. In general, the classical fashion of the HAR system design started from the collection of multi-modal sensory data on the basis of demand. One of the most significant factors determining the performance of a HAR system is the training process, i.e. the adopted sensor dataset. Conventionally, according to the characteristics of the required system, the deployed sensor is installed at the area where the most significant variation of detected signal is presented. The designer or developer normally follows the experience to decide where the sensor is placed, and thus building the system. The inherent positioning enables the user to avoid selecting the sensor's placement casually. Nevertheless, the fixed position can not totally fit the situation of the user's conditions [3]. For example, for long-term monitoring, providing the selectable positions for wearable sensors can lessen the discomfort caused by wearing. Besides, the installation of non-wearable sensors may also have plenty of potential placement locations indoors, like walls, doors, windows, or shelves,

which can not only affect the system's performance but also interior design and convenience.

Even if more potential positions have been explored from the demand, the re-collection of the dataset in terms of specific position from the real world and retraining process are still the bottlenecks. Various studies have concentrated on the position-aware classifier development to improve the adaptability for the target user [9, 12]. The datasets from different body parts are utilized to pre-train the respective classifiers, and the position determination is executed first before the recognition. Whereas, the flexibility of the designed system depends on the pre-work to a large extent. When the required activities are altered or more sensors are required to improve the accuracy, the complicated training process still needs to be recalled. Therefore, a system that can be flexibly designed to satisfy the user's requirement in terms of sensor position and number without recollecting the sensor data from the real world will be beneficial to the corresponding HAR applications.

In this work, we developed a virtual platform combined with sensor simulation to design the non-visual HAR systems. In such a virtual environment, the human motion animation is identified as the real activity, and with sensor simulated as shown in the real world gathers the information. The user could casually indicate their interested or uninterested location to place the sensor. And with motion re-conducting, the simulated sensor units are able to sense the signal and form the dataset. The virtual platform can provide the accuracy as the feedback of examined positions via cross-validation. Also the classifier interface trained by simulation signal can be generated forwarding to the real world and recognize the actual activity. The working flow pipeline is presented in Figure 1.

2 BACKGROUND

With an advanced deep learning network, a vision-based HAR system is provided with an abundant dataset and high accuracy [4, 5]. Furthermore, the other devices based systems have also been paid close attention. To design a portable system or be combined with consuming electronics for convenience, such a system is closer to real life, for example the electrocardiogram signal [1], the foot pressure [10], ambient light [17], Wi-Fi signal [14], distance sensor[18], etc. Developed systems would be effectively at the designated sensor placement. While in a practical sense, the selection of a placement necessitates both considerations of space design and system performance. Therefore, an inherent balance between sensor positioning and system accuracy emerges.

From the HAR related human-computer interaction community, close attention has been placed at what feedback can the classified activities provide and what interaction approaches can be applied, such as MotionMA [16] and Zsensor [8]. While, few researches have been conducting on the sensor simulation. Young et. al [22] developed a simulation environment to simulate the inertial signal including the acceleration, rotation, and magnetic intensity via the given IMU trajectory model. Another work of Shingo et. al [13] designed a simulation tool to realize multi-sensor setting activity recognition via the motion capture data as well. Kwon et. al [6] transferred the video object into virtual body motion model and then calculated the acceleration data to expand the current dataset. Such systems realize the simulation of acceleration data based on the motion capture (MoCap) data stream which contains the initial information of relevant coordinate variation of determined markers.

Hence, we aimed to develop a simulation platform that can provide a realistic virtual environment to be helpful to HAR design. Rather than the initial motion capture dataset, our approach directly processes the humanoid model with motion animation. Referring to the real world sampling procedure, the proposed system retained the facticity of signal acquisition and reduced the requirement of input to allow more possible applications based on virtual animation.

3 METHODOLOGY

The simulation environment is established based on the Game Engine Unity3D. Unity3D has advantages to visualize the daily items and characteristics in the virtual environment which enables the simulation to be as close as possible to the real world.

3.1 Input Data Stream

Previous work simulated multi-modal simulated sensor data all based on MoCap dataset containing coordinate value and skeleton hierarchy data. In our work, we applied the entire segment motion information to the humanoid model, only use the coordinate in the virtual world. User can be modelled by a specific humanoid model with different characteristics as well as mapping the real size of subject to a virtual model. We adopted the Xsens as our input device to generate the motion data via the MVN system.

3.2 Sensor Simulation

After user-defined motion data input, then the corresponding sensors need to be simulated. Combined with the characteristics of the developed environment, the accelerometer and distance sensor are likely to be simulated to enable both the wearable and non-wearable HAR system.

For distance data sampling, *Raycast* function is utilized to simulate the infrared signal and detect the distance between the obstacle and emission source. To produce the acceleration signal, we referred to the structure of the actual accelerometer, i.e. spring-mass-damper structure. As shown in Figure 2, the cube simulates the mass of the accelerometer. Adding the dual-spring to establish the detector of the acceleration of three-axes. The whole simulated accelerometer moves following the referenced joint during the activity. Therefore, when an external force is applied on the sensor, the mass will generate the acceleration according to Newton's second law with analysing the situation of stress of proof mass, the inertial force can be calculated by Equation 1.

$$m\ddot{x} + c\dot{x} + kx = ma \tag{1}$$

Where c is a damper coefficient of the spring and k is a constant factor of spring. The x represents the displacement of the mass. Following this way, the relative displacement between the mass and lower boundary board can be recorded and according to differential formula the simulated accelerometer can generate the acceleration signal, which is similar to a person wearing related sensors on the body.

Different from other simulation fashion, our approach aims to maintain the fidelity to produce the sensor data via recurrenting

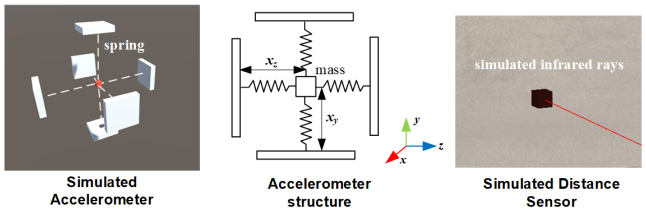

Figure 2: Built simulated accelerometer

the sampling process. Only the motion animation combined with coordinate information in a virtual environment is exploited without any prior knowledge of motion position sequence. Therefore, the humanoid model in activity resembles a real subject being in an action. However, the executed human motion in Unity is based on frame-by-frame. The procedure of generating the simulation data can be recognized as the re-sampled process. However, the operation of Unity is affected by the frame rate and script's complexity seriously, which normally causes the repeated frame. Therefore, we decreased the sampling slot at first (i.e. down sampled), which is able to eliminate the influence of frame repetition. The down-sampled frequency is the half of required sampling frequency. Subsequently, interpolating the sampled data into required sampling frequency.

3.3 HAR system

Various types of ML models have been successfully utilized in HAR with respect to identifying different kinds of activity. Due to the data characteristic and less database for wearable devices, we selected the Support Vector Machine (SVM) model (with an RBF kernel, C=1000) as the classifier [11]. Additionally, to improve the ability to classify data sensed from the real world, data augmentation (DA) was adopted to expand the dataset from individual. Referencing the work of [15], the DA for wearable sensor data can be performed in the original dataset or feature domain. We adopted the approach described as follows.

- Permutation: Perturb the time location of input data;
- Time warping: Distort temporal locations;
- Magnitude warping: Warp the signal's magnitude;

However, for non-wearable types, HAR is performed based on the grayscale figure, which we select to change the humanoid's size to enrich the data samples, especially the shoulder and chest width, to simulate different body types.

- Change the size of the model: Multiplied by a factor of 1.2 and 0.8 to alter the body type

Regarding extracted features, for a wearable case, the signal from the body is used to calculate the time and frequency domain. We selected the *mean*, *variance*, *standard variance*, *75th percentile*, and *inter-percentile* as time-domain features. In addition, the *mean*, *median value of the power spectrum*, and *Fourier coefficient* are adopted as frequency features [21]. Calculated features are subsequently normalized to eliminate the effects of the amplitudes of different input signals. For the non-wearable case, we converted the signal from a specific position into a pixel of a grayscale figure. Through segmentation, each figure contains the signal information of all

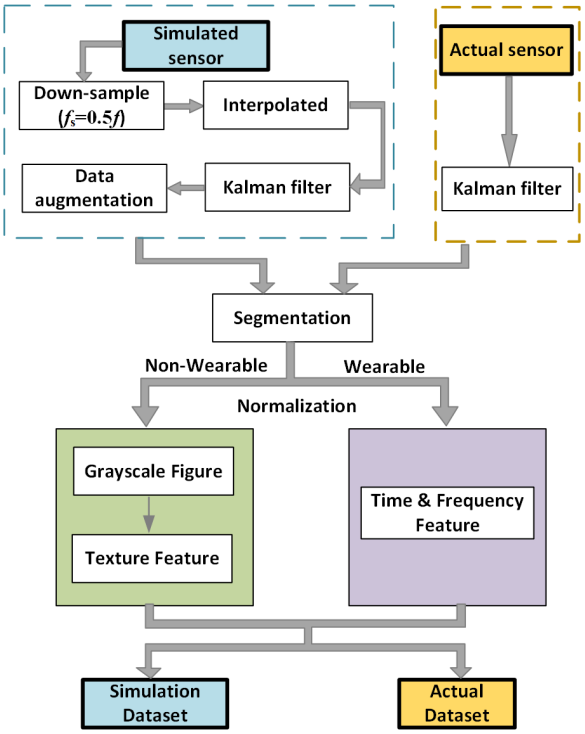

Figure 3: The workflow of activity recognition. The simulation data and actual data are both segmented to extract the features according to different types of system. And finally obtain the simulation and acutal dataset

sensors in an interval. The texture feature is extracted via the Gabor filter and input into the classifier. The corresponding work process of feature generation is given in Figure 3.

4 CASE STUDY

We invited nine participants to take part in the experiments. We showed disparate scenarios and presented the output from the platform for establishing a HAR system. The detailed validation protocol is presented as follows.

- **Input**: user(s) conducts the defined activity to generate the motion animation with 90 seconds;
- **Data processing**: the animation is replayed in the virtual environment with indicated sensors. And the simulation data is segmented by a two-second window;
- **Feedback (in virtual environment)**: accuracy calculated by 5-fold cross-validation of simulation dataset;
- **Output**: the classifier trained by simulation dataset with designated sensor positions
- **Validation**: the sensors are placed at the mapped position in the real world and user are asked to conduct the activity with 60s and process data with the same method;
- **Result**: use the trained classifier to calculate the accuracy of recognition of real activity;

Figure 4: Activity type and actual wearing of case-a (a) User with MoCap suit for input generation.(b)Virtual activity

- **Environment**: ThinkPad, X1 Carbon; i7-8565U with Unity3D (2018.4.14) and Python (3.7);

4.1 Case description

a. ***Wearable acceleromters system for daily activity recognition***

In this part, we leveraged platform to generate the relevant result regarding several types of daily activity recognition with an accelerometer worn on the body. We recruited three participants (two males and one female; age: 21/26/24) to wear the accelerometers. The user's choice of the sensor's position and number as well as recognized activities are presented as in Table 1. As the virtual accelerometer adopted in this part, we configure the related parameters of designed spring-mass system, i.e., the accelerometer. According to the real sensor design, the damper coefficient is requested to be large smaller than the constant factor to ensure a rapid response [2]. The sampling rate is 60Hz and thus 30Hz-sampling is executed in virtual environment. The animation capture as well as produced virtual motion is presented in Figure 4.

b. ***Wearable distance sensor system for exercises recognition***

In this part, the platform is applied to design a wearable distance sensor system which the sensor is attached on the lower limbs of an individual. The sensors attached to each body part have multiple directions and transmission angles and the HAR system is utilized to recognize exercises. We invited three people (males; age: 19/26/27) to test this application. Three types of exercises, the *heel up/down, squat, and hip stretch*, are proposed as recognized activities.

To build the prototype of the distance sensor-based HAR, we adopted the infrared distance sensor (GP2Y0A21YK0F, Sharp),which has an effective detection range from 10 to 80 cm. Additionally, the Arduino chip (ARDUINO PRO MINI) is used for data acquisition and transmission. Each subject must execute activities lasting for 60 seconds with 50Hz sampling rate for the real-world testing. Design process is shown in Figure 5.

c. ***Distance sensing system for motion recognition in a bathroom***

Ambient sensors can also be employed to capture the signal while the body part can be liberated without any device. In this part, we referred to conventional dense sensing scene to design a HAR system oriented to the activity recognition in a bathroom. The user can determine where to place the infrared distance sensor and how many sensors according to his/her preference in virtual environment. We invited user G (male; height 175; weight: 69kg) who lived in to test the experiment and two participants (user H and I) to test the robustness of the classifier.

We use the virtual platform to give the scene, that in a bathroom, activities like "*washing hands*", "*washing face*" and "*brushing teeth*" need to be recognized. The applied area is assumed to be in front of the washbasin. The user would like to install the sensors on the mirror above the sink, facing the person. In the virtual environment, the user tested different types of sensor combination and finally decided use ten sensors arranged in a straight line. Subsequently, we examined the performance of classifier trained by simulation classifier. We utilized the infrared distance sensor (GP2Y0A21YK0F, Sharp) and the Arduino chip (ARDUINO PRO MINI) for data sampling and validation. Design process is presented in Figure 6.

4.2 Result

The testing result is presented in Table 2. According to the case-a, for some lower body parts, the accuracy distinction between the simulation and actual data is relatively large. The reason is that the lower body part generally causes relatively huge movements during daily motion, like running. The larger movement will make the spring-damper system inaccurate to an extent. Following this point, to classify the activity that the subject interacts with floors higher than the ground, such as climbing stairs, the huge movement in a vertical direction enables the worst classification of the simulated classifier. The distance sensor does not entail the calculation of the signal that the data is directly returned. To improve the accuracy of a simulated classifier, it is crucial to ensure the similarity of the two waveforms from simulated and real situations. Moreover, applying a distance sensor both to wearable cases and mounted at external ambient, mapping from the virtual environment to the real world is necessary. For case-c, we also invited another two subjects (males; height: 166cm/171cm; weight: 61kg/64kg) to test the performance of the designed system. We used the body type data to alter the relevant humanoid model in simulation and generate the simulation data. With the adopted sensor location the simulation accuracy is decreased to 78.21%. Because of differences in body size of subjects, the signal from limited channels cannot train the model with good

Table 1: Users demand of case-a

Subject	Sensor position demand	Recognized activity
A	Number less than 2 and dislikes wearing it on the waist	standing/walking/running/squatting
B	Number less than 3 and prefer to wear on the chest at first	walking/running/lying/sitting
C	No restriction	standing/walking/running/going upstairs/downstairs

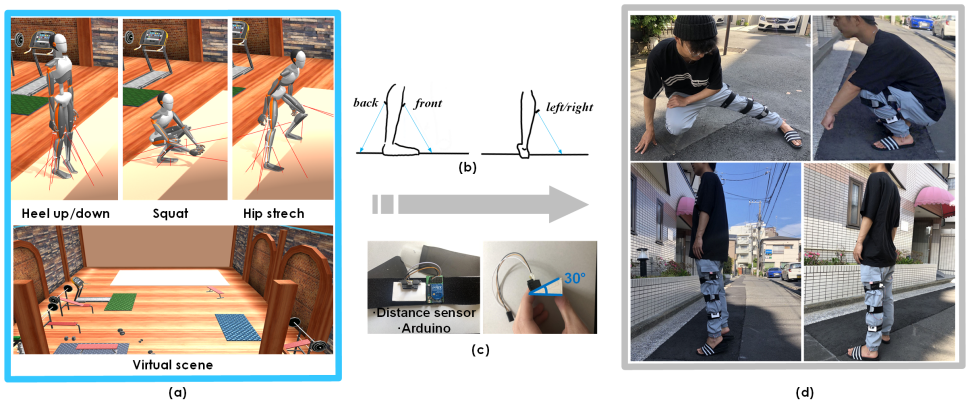

Figure 5: The application scene of case-b. (a) Designed HAR system in virtual environment. (b) Sensor concept. (c) Built real sensor. (d) Real-world testing

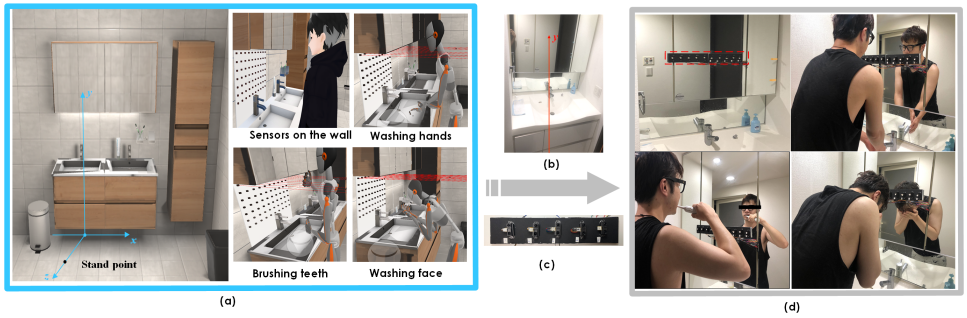

Figure 6: The application scene of case-c. (a) Designed HAR system in virtual environment. (b) Actual applied scene. (c) Built sensor component. (d) Real-world testing

robustness. Increasing the number and expanding the sensor area can address the issue, but more space will be utilized.

5 CONCLUSION AND FUTURE WORK

In this paper, we designed a virtual simulation environment to produce the simulated multi-modal sensor data. The sensor position is identified as an interface between the user and designed HAR system. We proposed to drop back the position choice to the end-user side. With fidelity of virtual environment, the humanoid model as well as applied scene can both be customized to fit the actual situation. Moreover, our method offer more choices and enable the user more arbitrarily to place the sensor in the virtual scenario. As the virtual dataset generated, the accuracy is provided as feedback to help the user to determine whether the current position is satisfied and adjust the subsequent sensor placement and numbers.

However, in this work we only conducted the small amount case studies rather than large number of user study, because our focus is related to each individual's preference. For different subjects, their choices of sensor positioning are supposed to be various. With future work, more cases are able to be involved. As our design is based on the application level, i.e. the animation file, more ubiquitous sensing systems based on a virtual simulator of sensors can also be designed. Combining with VR/AR to install sensors in virtual environments from serious/somatosensory game may also help to explore applications combined with the real world.

Table 2: Result of different cases

Type		Virtual		Real world	
		Position	Accuracy of cross validation	Position	Accuracy of real testing
Case a	A	Chest	96.81%	Chest	96.11%
	B	Chest, right foot, left lower leg	99.81%	Chest, right foot, left lower leg	97.79%
	C	Chest, right shoulder, head	97.38%	Chest, right shoulder, head	89.95%
Case b		Left lower leg-left, Left upper leg-back	100%	Left lower leg-left, Left upper leg-back	91.25%
Case c		from (-0.225, 1.40) to (0.225, 1.40) with interval of 0.05	91.67%	from (-22.5cm, 142cm) to (22.5cm, 142cm) with interval of 5cm	90.69%

REFERENCES

[1] Sung-Gwi Cho, Masahiro Yoshikawa, Kohei Baba, Kazunori Ogawa, Jun Takamatsu, and Tsukasa Ogasawara. 2016. Hand motion recognition based on forearm deformation measured with a distance sensor array. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4955–4958.

[2] RA Dias, LA Rocha, L Mol, RF Wolffenbuttel, and E Cretu. 2010. Time-based micro-g accelerometer with improved damper geometry. In *2010 IEEE Instrumentation & Measurement Technology Conference Proceedings*. IEEE, 672–675.

[3] Francine Gemperle, Chris Kasabach, John Stivoric, Malcolm Bauer, and Richard Martin. 1998. Design for wearability. In *digest of papers. Second international symposium on wearable computers (cat. No. 98EX215)*. IEEE, 116–122.

[4] Shanyan Guan, Shuo Wen, Dexin Yang, Bingbing Ni, Wendong Zhang, Jun Tang, and Xiaokang Yang. 2019. Human Action Transfer Based on 3D Model Reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 8352–8359.

[5] Ahmad Jalal, Yeon-Ho Kim, Yong-Joong Kim, Shaharyar Kamal, and Daijin Kim. 2017. Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern recognition* 61 (2017), 295–308.

[6] Hyeokhyen Kwon, Catherine Tong, Harish Haresamudram, Yan Gao, Gregory D Abowd, Nicholas D Lane, and Thomas Ploetz. 2020. IMUTube: Automatic extraction of virtual on-body accelerometry from video for human activity recognition. *arXiv preprint arXiv:2006.05675* (2020).

[7] Gierad Laput and Chris Harrison. 2019. Sensing fine-grained hand activity with smartwatches. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[8] Gierad Laput, Walter S Lasecki, Jason Wiese, Robert Xiao, Jeffrey P Bigham, and Chris Harrison. 2015. Zensors: Adaptive, rapidly deployable, human-intelligent sensor feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1935–1944.

[9] Andrea Mannini, Angelo M Sabatini, and Stephen S Intille. 2015. Accelerometry-based recognition of the placement sites of a wearable sensor. *Pervasive and mobile computing* 21 (2015), 62–74.

[10] Ayumi Ohnishi, Tsutomu Terada, and Masahiko Tsukamoto. 2018. A Motion Recognition Method Using Foot Pressure Sensors. In *Proceedings of the 9th Augmented Human International Conference*. 1–8.

[11] Charissa Ann Ronao and Sung-Bae Cho. 2016. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert systems with applications* 59 (2016), 235–244.

[12] Timo Sztyler, Heiner Stuckenschmidt, and Wolfgang Petrich. 2017. Position-aware activity recognition with wearable devices. *Pervasive and mobile computing* 38 (2017), 281–295.

[13] Shingo Takeda, Tsuyoshi Okita, Paula Lago, and Sozo Inoue. 2018. A multi-sensor setting activity recognition simulation tool. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 1444–1448.

[14] 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.

[15] Terry T Um, Franz MJ Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 2017. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 216–220.

[16] Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2013. MotionMA: motion modelling and analysis by demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1309–1318.

[17] Raghav H Venkatnarayan and Muhammad Shahzad. 2018. Gesture recognition using ambient light. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–28.

[18] Tianna-Kaye Woodstock, Richard J Radke, and Arthur C Sanderson. 2016. Sensor fusion for occupancy detection and activity recognition using time-of-flight sensors. In *2016 19th International Conference on Information Fusion (FUSION)*. IEEE, 1695–1701.

[19] Chi-Jui Wu, Steven Houben, and Nicolai Marquardt. 2017. Eaglesense: Tracking people and devices in interactive spaces using real-time top-view depth-sensing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3929–3942.

[20] Jason Wu, Chris Harrison, Jeffrey P Bigham, and Gierad Laput. 2020. Automated Class Discovery and One-Shot Interactions for Acoustic Activity Recognition. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[21] Chengshuo Xia and Yuta Sugiura. 2020. Wearable Accelerometer Optimal Positions for Human Motion Recognition. In *2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech)*. IEEE, 19–20.

[22] Alexander D Young, Martin J Ling, and Damal K Arvind. 2011. IMUSim: A simulation environment for inertial sensing algorithm design and evaluation. In *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks*. IEEE, 199–210.