

Department of Electrical, Computer, and Software Engineering

Part IV Research Project

Literature Review and
Statement of Research Intent

Project Number: 77

mmWave based human
activity recognition

Samuel Mason


Beck Busch

Supervisor: Kevin I-Kai Wang

28/04/2023

Declaration of Originality

This report is my own unaided work and was not copied from nor written in collaboration with any other person.

Signature: 

Name: Samuel Mason

Abstract

Human activity recognition (HAR) is an area of research that has received increasing attention in recent years. HAR can be applied in many areas, such as human computer interaction (HCI), smart homes, internet of things (IoT), detection of abnormal or suspicious behaviours, fitness trackers and much more. This review explores the various activity segmentation methods and activity types covered, as well as real-time applicability of HAR systems in existing research. Existing research fails to cover the recognition of longer, more complex action sequences, which naturally introduce a delay for recognition, and to the detriment of real-time capabilities. We propose a novel activity recognition method using mmWave radar, one that is capable of smooth, continuous real-time recognition, and outputs estimates of activity classification, which are updated as the action is completed, to satisfy real-time capabilities in complex scenarios.

Introduction

Broadly speaking, human activity can be grouped into the categories of full-body motion, as explored in [1-13] typically involving the movement of multiple limbs, and gestural activity [13, 14-19], which is on a smaller scale and involves arm or hand motions. Posture estimation [20, 21] can also be used to discern human activity by positioning of the joints. For this literature review, involuntary activities are not being considered. For example, monitoring of heart and breathing rate as part of vital signs detection. For our research we only consider line of sight (LOS) conditions.

Sensors

HAR can be realised using numerous technologies, typically consisting of sensing component(s), data processing and pre-processing modules (including activity segmentation), and classification component(s). Classification components are typically realised using machine learning algorithms in existing approaches. Machine learning approaches typically differ based on the specific implementation of HAR systems and can be tailored based on the sensing method. It is for this reason that determining the sensing method of choice is of importance. Some of the different sensing technologies commonly used for HAR are cameras [10], WiFi channel state information (CSI) [6, 7, 9], Lidar [8], sensors that attach to the body, such as accelerometers or gyroscopes [3, 11] and millimetre wave (mmWave) [1, 2, 4, 5, 13-21].

Camera footage was used in [10] for segmenting activity repetitions. However, there has been a shift away from using cameras in HAR applications due to privacy concerns. The advent of newer, commercially available sensing

technologies has made HAR possible without the need for cameras. Camera based methods suffer in performance when lighting conditions are poor, and cannot sense through obstacles, making them unsuitable for a general use case.

Sensors that attach to the body, such as in [3, 11] have seen applications in HAR. However, these types of sensors can be seen as inconvenient and potentially uncomfortable [7], as HAR is not possible unless the user carries the sensors around. There have been efforts to reduce the number of sensors used, as in [11], where accelerometer data from a mobile phone was used to provide HAR, but if a user is not carrying their phone, then HAR is not possible.

Lidar has been explored for its potential application in HAR, albeit in a limited capacity [8]. Lidar has the advantage of not being sensitive to lighting conditions and can be used both indoors and outdoors. However, a major drawback of lidar systems is the cost, as lidar sensors are typically very expensive when compared to electromagnetic wave-based sensing methods. Lidar also requires LOS and cannot sense over longer distances.

Electromagnetic wave-based sensors include WiFi and radar, which have become popular for use in HAR in recent years. WiFi based recognition systems typically employ channel state information (CSI) data to detect human influences on the environment. However, this sensing method is not very precise, as WiFi usually operates between the 2.4GHz and 5GHz bands, resulting in a wavelength in the order of centimetres.

Radar systems, specifically radar within the millimetre wave (mmWave) range are the sensing method of choice for a lot of very recent research. Unlike WiFi CSI, as the name implies, wavelengths are in the order of millimetres. This allows for more fine-grained recognition of HAR tasks. The use of mmWave radar in HAR is a recent development, which has been made possible by the advent of single chip radar.

Both WiFi and radar-based systems are insensitive to lighting conditions, making them preferable over cameras for HAR. Unlike wearable sensor-based recognition systems, only one or two (as in [4]) radars are required for the task of HAR. Radar does not need to be attached to the body, but in applications where system mobility is of importance, radar has been shown to be able to perform HAR tasks. [14] references a radar system embedded within a smartphone, showing that radar-based HAR systems have the ability to be portable.

Segmentation method

Many of the proposed solutions for HAR in existing research have fixed assumptions about the length of activities being performed. In [14], gesture activities are assumed to fit within 64 frames, amounting to 2.6s. Whereas in [7], researchers limited activity data collection by asking volunteers to perform activities within a specific timeframe. Activity samples are 10s long, with actions being performed between the 4th and 7th second. Volunteers are asked to stay still the rest of the time to prevent overlapping of action data. [16] made use of temporal sampling to align gesture durations, each of which contained 30 frames (1.5s) of heatmaps. [4] performed an analysis on different sample lengths, and found that for their application, 40 frames (3.7s) provided the optimal classification result. Padding is used where necessary, to bring sample length up to 40 frames, by padding with the last frame of an activity.

There are examples of research where activity segmentation is not mentioned at all. [20] and [21] feed single frames into a CNN network, because temporal dependency analysis is not required for posture estimation. [5], which explores full-body activity recognition, decides to buffer data frames for recognition, but makes no mention of specific segmentation methods.

More commonly, however, a sliding window approach is taken for segmentation of activity data. [1] employs a sliding window of fixed length, at 1.2s, and a sliding factor of 0.3s. 12 frames are used in total to make up a feature of dimension $12 \times 50 \times 50 \times 30$, which is then fed into a machine learning model for classification. A sliding window of size (64, 3) is used in [11], which amounts to 3.2s at the chosen 20Hz. An overlap of 50% was used here. [3] involved multiple sliding windows, starting with a 6s window. Data rebalance, cluster analysis and activity segmentation are followed by a sliding window of (2, 3, 4)s with 50% overlap to segment long period activities.

A more complex approach to segmentation is made in [15], where gesture samples, performed in a window of 125 frames (~5s) are segmented with an activity detection module (ADM). The ADM consists of a binary classifier and accumulator, which is trained to detect gesture ends. However, gestures are considered to not exceed 50 frames (~2s). [9] uses a similar approach to detection and introduces an AACA algorithm to segment periods of activity from periods of inactivity. A sliding window related to the WiFi packet transmission rate is used (50 frames in this work) and has a step length of 1. A large CSI variance is taken to be indicative of periods of activity.

Real-time HAR

Investigating the applicability of HAR systems in real-time is only explored by a subset of existing HAR research.

However, the advantages of real-time recognition are apparent, such as in time critical scenarios. Many of previously discussed HAR methods do not have real-time recognition capability [3, 4, 7, 9, 11, 16, 21]. Others, such as [5] claim real-time classification capability, but do not provide performance metrics, specifically regarding model inference time. [1] adopts a DVCNN model to allow for real-time HAR but is not tested for it.

[20] is capable of real-time operation. The proposed model provides an average inference time of 150 μ s, making it capable for real-time use. However, because this work is related to posture estimation, temporal dependencies are not captured, so this inference time is for a single frame. Classifications capturing temporal dependencies are made in [17-19].

[17] is capable of real-time recognition, with a median processing runtime of 10.1ms when gestures are not being detected, and 40.3ms when they are, on a desktop system. The recognition model has also been implemented on an embedded system (Raspberry Pi 4), with a median processing runtime of 31.5ms when gestures are not being detected, and 67.4ms when they are. However, the frame rate that radar data is being captured at must be reduced to allow for real-time recognition. Although recognition is reasonably fast, two frames waiting to be processed need to be discarded in order to achieve real-time performance, which could negatively impact classification results. Also implemented on a Raspberry Pi 4, [18] achieves real-time recognition with the proposed Tesla-V model. This model boasts a low computational complexity, with 0.4GFLOPS and inference time of 0.3s for a batch size of 16. [19] abandons preservation of activity data temporal structure in favour of a lightweight PointNet classifier. The classifier takes 81ms to classify a single gesture on an experimental computer.

Although some of the above research examples have achieved real-time HAR, this capability is only realised because they all assume that activities are short enough to fit within a specified timeframe. In pursuing real-time activity recognition for longer, more complex activity sequences, existing methods will fail.

Research Table

Refs	Activities	Activity Segmentation Method	Real-time Applicability
[1]	walk, fall, get up, stand to sit, sit to stand, sit to lie, lie to sit	Sliding window of 1.2s, with a sliding factor of 3 frames (0.3s). 12 frames are used in total to make up feature of dimension 12 x 50 x 50 x 30.	Not tested for real-time recognition / inference time not stated.
[3]	walking, walking upstairs, walking downstairs, sitting, standing, laying sit to lie, lie to sit, stand to sit, sit to stand, stand to lie, lie to stand	Sliding window with duration of 6s. Followed by a sliding window of (2, 3, 4)s fixed size with 50% overlap to segment long period activities.	No mention of real-time recognition beyond future research directions.
[4]	walk to room, fall on the floor, stand up from the floor, walk to chair, sit down on chair, stand up from chair, walk to bed, sit down on bed, stand up from bed, get in bed, lie in bed, roll in bed, sit in bed, get out bed	k = 40 frames (3.7 seconds) found as optimal length. Padding is used if the sample length is less than k.	No mention of real-time recognition.
[5]	human walking and vanish from radar, human waving hands when standing or sitting, human sitting to standing and walking transition, human walking back and forth, no micro-Doppler detections, complex detections including all behaviours	No mention of any segmentation method being employed, other than buffering frames.	Real-time classification capability is claimed, no performance metrics for this have been provided.
[7]	Collected dataset: calling, squatting, walking, stand-fall, walk-fall Comparison 1: lie down, fall, walk, run, sit down, stand up Comparison 2: boxing, empty, walking, pushing, waving	Activity samples are 10 seconds long. Actions are performed from the 4th to the 7th second, the volunteers are asked to remain still the rest of the time.	No mention of real-time recognition.
[9]	pushing, waving, kicking, running, falling, boxing, sitting, picking, walking, empty	An AACA algorithm is introduced, to segment periods of activity from periods of inactivity. Sliding window of size 50, step 1 is used.	No mention of real-time recognition.
[11]	walking, jogging, standing, sitting, ascending stairs, descending stairs	Sliding window of size (64, 3) is used. Window contains a fixed 64 points per axis, at 20Hz the time scale is 3.2s. Window has overlap of 50%.	No mention of real-time recognition.
[14]	grab, clockwise turning, tilt left and right, draw to the left, falling, arc, clap, counterclockwise turning, draw to the right, lifting, thumb-index finger, thumb-little finger	Gesture measurement length is set to 64 frames, or about 2.6 seconds. Gestures are assumed to fit within this 2.6s.	Gesture recognition performance is evaluated using a real-time data acquisition board, but real-time classification is not mentioned.
[15]	radial circle, tangential circle, double pinch, single pinch, left-to-right swipe, right-to-left swipe	Gesture samples are performed in a window of 125 frames (~5s). Gestures length is considered to not exceed 50 frames (~2s).	No mention of real-time recognition.

[16]	Positive: push, pull, left swipe, right swipe, clockwise turning, anticlockwise turning Negative: lifting arms, sitting, standing, walking, waving hands	Temporal sampling is used to align gesture durations. Sampled gesture sequences contain 30 frames (1.5s) of heatmaps.	No mention of real-time recognition.
[17]	random movement/walk, pull, push, swipe up, swipe down, swipe right, swipe left, rotate, wave, push-pull	Minimum duration of 200ms before a gesture is detected, total event duration set equal to 50 frames (1s at 50Hz).	Median processing runtime of 10.1ms, 31.5ms when a gesture is not detected, and 40.3ms, 67.4ms when it is detected (on PC and RPi, respectively).
[18]	Used three existing datasets, Pantomime, RadHar, mHomeGes. Included gestures, arm gestures and hand gestures.	Best results were achieved by setting a fixed input of 32 frames for both datasets (~1.07s and 0.533s, 0.8s) for all datasets.	Capable of real-time recognition. Boasts a low computational complexity, with 0.4GFLOPS and inference time of 0.3s for a batch size of 16 on a RPi device.
[19]	One arm: draw a circle clockwise lifting the right arm, draw a circle anticlockwise lifting the right arm, lift both arms then lateral down Mirrored for both arms: lift right arm then down, down right arm then lift, push right arm then pull, pull right arm then push, pull right arm twice, push right arm twice, right arm outward clockwise circle, right arm outward anticlockwise circle	Minimum gesture length of 350ms. Point cloud data is aggregated for the entire duration of an activity.	A lightweight classifier is designed, in order to facilitate real-time operation. Developed PointNet classifier takes 81ms to classify a single gesture on an experimental computer.
[20]	walking, left-arm swing, right-arm swing, both-arms swing	No temporal dependency for posture estimation, therefore, single frames are fed to the CNN network.	Capable of real-time operation. Proposed model provides an average inference time of 150 μ s per frame, making it suitable for real-time implementation.
[21]	lifting left/right arm to the front for 45/90/180 degrees, lifting left/right arm from the side for 45/90/180 degrees, lifting left/right leg for 45/90 degrees, waving hands, walking, random moving	No temporal dependency for posture estimation, therefore, single frames are fed to the CNN network.	No mention of real-time recognition.

Project Scope

From the above discussion and table representation, it can be seen that there is a gap in existing research with regard to the recognition of activity sequences in more complex scenarios. The recognition of complex activity sequences is not currently possible, or at the very least, has not been tested for in existing literature, to the best of our knowledge.

Smooth HAR is being used to term scenarios in which actions are being performed sequentially, thus the realisation of a smoothly operating recognition system (from an observer's point of view) is required to classify these actions in real-time.

Our designed system should be able to accurately classify an activity given less than half of a (discrete) activity constituting a longer activity sequence, or, in the case of a continuous activity, within half of the periodic motion subsumed by the activity. New activities should be detectable within 0.5s, that is, following an activity transition, the system should start working on classification of the new activity. Our system can improve its estimate of an activity given more time but will output its estimate regardless. In this way, real-time HAR can be achieved, without the need for complete activity sequences to be fed to classification models.

Applications of smooth HAR recognition systems include HCI, monitoring and surveillance systems, fitness trackers, smart homes, and lab safety. The improved ability to recognise complex action sequences allows for the seamless operation of these systems, where complex sequences can arise. Below is an outline for how we plan to implement our proposed smooth HAR recognition system, with tentative deadlines for progress.

Project Steps

1. Design of System Architecture

Careful consideration of the overall architecture of our system is required before proceeding with subsequent steps. This is because we want to avoid negative flow-on consequences caused by earlier decision-making.

Suggested deadline: Week 10, Semester 1.

2. Radar Configuration

This step involves the purchase of the radar and other components that will be needed for data collection. To allow for the delivery of these components, the radar needs to be ordered before the inter-semester break.

Suggested deadline: Week 11, Semester 1.

3. Construction of Pre-processing Pipeline

The pre-processing pipeline needs to be figured out before data collection can happen, because it will affect how we collect and sort radar data. The pre-processing of radar data is necessary to ensure that inputs to classification algorithms are consistent with the required format. Suggested deadline: Last week of inter-semester break.

4. Collection of Activity Data

Key decisions that need to be made in this step are the types of activity sequences that we will be collecting, as well as the collection method(s). We are planning to gain ethics approval for data collection. Suggested deadline: Week 8, Semester 1 for submission of ethics form, Week 2, Semester 2 for data collection.

5. Creation of Classification Algorithm(s)

In this step, we will design machine learning model(s) for the classification of activity data collected in the previous step. Preliminary testing to evaluate the effectiveness of different prospective algorithms will be carried out. Once we have settled on a classification model, then further testing can be done. Suggested deadline: Week 8, Semester 2.

6. System Finalisation

Once all the preceding steps have been completed, the overall system performance and robustness will be evaluated. This step may include an ablation study to gauge the effectiveness of our specific implementation, depending on the nature of our implementation. Suggested deadline: Week 10, Semester 2.

References

- [1] C. Yu, Z. Xu, K. Yan, Y. -R. Chien, S. -H. Fang and H. -C. Wu, "Noninvasive Human Activity Recognition Using Millimeter-Wave Radar," in *IEEE Systems Journal*, vol. 16, no. 2, pp. 3036-3047, June 2022.
- [2] J. Wang, Y. Zhao, X. Ma, Q. Gao, M. Pan and H. Wang, "Cross-Scenario Device-Free Activity Recognition Based on Deep Adversarial Networks," in *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 5416-5425, May 2020.
- [3] J. -H. Li, L. Tian, H. Wang, Y. An, K. Wang and L. Yu, "Segmentation and Recognition of Basic and Transitional Activities for Continuous Physical Human Activity," in *IEEE Access*, vol. 7, pp. 42565-42576, 2019.
- [4] Bhavanasi, G., Werthen-Brabants, L., Dhaene, T. *et al.* Patient activity recognition using radar sensors and machine learning. *Neural Comput & Applic* 34, 16033–16048 (2022).
- [5] R. Zhang and S. Cao, "Real-Time Human Motion Behavior Detection via CNN Using mmWave Radar," in *IEEE Sensors Letters*, vol. 3, no. 2, pp. 1-4, Feb. 2019, Art no. 3500104.
- [6] D. Wang, J. Yang, W. Cui, L. Xie and S. Sun, "Multimodal CSI-Based Human Activity Recognition Using GANs," in *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17345-17355, 15 Dec.15, 2021.
- [7] X. Cheng, B. Huang and J. Zong, "Device-Free Human Activity Recognition Based on GMM-HMM Using Channel State Information," in *IEEE Access*, vol. 9, pp. 76592-76601, 2021.
- [8] F. Luo, S. Poslad and E. Bodanese, "Temporal Convolutional Networks for Multiperson Activity Recognition Using a 2-D LIDAR," in *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7432-7442, Aug. 2020.
- [9] H. Yan, Y. Zhang, Y. Wang and K. Xu, "WiAct: A Passive WiFi-Based Human Activity Recognition System," in *IEEE Sensors Journal*, vol. 20, no. 1, pp. 296-305, 1 Jan.1, 2020.
- [10] S. -H. Cheng, M. A. Sarwar, Y. -A. Daraghmi, T. -U. İk and Y. -L. Li, "Periodic Physical Activity Information Segmentation, Counting and Recognition From Video," in *IEEE Access*, vol. 11, pp. 23019-23031, 2023.
- [11] J. Huang, S. Lin, N. Wang, G. Dai, Y. Xie and J. Zhou, "TSE-CNN: A Two-Stage End-to-End CNN for Human Activity Recognition," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 292-299, Jan. 2020.
- [12] T. Huynh-The, C. -H. Hua, N. A. Tu and D. -S. Kim, "Physical Activity Recognition With Statistical-Deep Fusion Model Using Multiple Sensory Data for Smart Health," in *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1533-1543, 1 Feb.1, 2021.
- [13] P. Zhao, C. X. Lu, B. Wang, N. Trigoni and A. Markham, "CubeLearn: End-to-end Learning for Human Motion Recognition from Raw mmWave Radar Signals," in *IEEE Internet of Things Journal*.

- [14] J. -T. Yu, L. Yen and P. -H. Tseng, "mmWave Radar-based Hand Gesture Recognition using Range-Angle Image," *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, Antwerp, Belgium, 2020, pp. 1-5.
- [15] A. Ali *et al.*, "End-to-End Dynamic Gesture Recognition Using MmWave Radar," in *IEEE Access*, vol. 10, pp. 88692-88706, 2022.
- [16] Baiju Yan, Peng Wang, Lidong Du, Xianxiang Chen, Zhen Fang, Yirong Wu, mmGesture: Semi-supervised gesture recognition system using mmWave radar, *Expert Systems with Applications*, Volume 213, Part B, 2023.
- [17] A. Ninos, J. Hasch and T. Zwick, "Real-Time Macro Gesture Recognition Using Efficient Empirical Feature Extraction With Millimeter-Wave Technology," in *IEEE Sensors Journal*, vol. 21, no. 13, pp. 15161-15170, 1 July1, 2021.
- [18] D. Salami, R. Hasibi, S. Palipana, P. Popovski, T. Michoel and S. Sigg, "Tesla-Rapture: A Lightweight Gesture Recognition System from mmWave Radar Sparse Point Clouds," in *IEEE Transactions on Mobile Computing*.
- [19] H. Xie, P. Han, C. Li, Y. Chen and S. Zeng, "Lightweight Midrange Arm-Gesture Recognition System From mmWave Radar Point Clouds," in *IEEE Sensors Journal*, vol. 23, no. 2, pp. 1261-1270, 15 Jan.15, 2023.
- [20] A. Sengupta, F. Jin, R. Zhang and S. Cao, "mm-Pose: Real-Time Human Skeletal Posture Estimation Using mmWave Radars and CNNs," in *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10032-10044, 1 Sept.1, 2020.
- [21] Cong Shi, Li Lu, Jian Liu, Yan Wang, Yingying Chen, Jiadi Yu, mPose: Environment- and subject-agnostic 3D skeleton posture reconstruction leveraging a single mmWave device, *Smart Health*, Volume 23, 2022.