



Twomey, N., Diethel, T., Kull, M., Song, H., Camplani, M., Hannuna, S., ... Craddock, I. (2016). The SPHERE Challenge: Activity Recognition with Multimodal Sensor Data. *arXiv*.

Peer reviewed version

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/pure/about/ebr-terms>

The SPHERE Challenge

Activity Recognition with Multimodal Sensor Data

Niall Twomey¹, Tom Diethe¹, Meelis Kull¹, Hao Song¹, Massimo Camplani¹,
Sion Hannuna¹, Xenofon Fafoutis¹, Ni Zhu¹, Pete Woznowski¹, Peter Flach²,
and Ian Craddock¹

¹ Department of Electrical and Electronic Engineering, University of Bristol, UK,

² Department of Computer Science, University of Bristol, UK,

✉ <http://irc-sphere.ac.uk/sphere-challenge/home>

✉ firstname.lastname@bristol.ac.uk

Abstract. This paper outlines the Sensor Platform for HEalthcare in Residential Environment (SPHERE) project and details the SPHERE challenge that will take place in conjunction with European Conference on Machine Learning and Principles and Practice of Knowledge Discovery (ECML-PKDD) between March and July 2016. The SPHERE challenge is an activity recognition competition where predictions are made from video, accelerometer and environmental sensors. Monetary prizes will be awarded to the top three entrants, with €1,000 being awarded to the winner, €600 being awarded to the first runner up, and €400 being awarded to the second runner up. The dataset can be downloaded from the University of Bristol’s data servers ³ [?].

1 Background and Summary

In this section we first describe the SPHERE project, then the SPHERE challenge is described, and outline of the remainder of the paper.

1.1 The SPHERE Project

Obesity, depression, stroke, falls, cardiovascular and musculoskeletal disease are some of the biggest health issues and fastest-rising categories of health-care costs. The financial expenditure associated with these is widely regarded as unsustainable and the impact on quality of life is felt by millions of people in the UK each day. Smart technologies can unobtrusively quantify activities of daily living, and these can provide long-term behavioural patterns that are objective, insightful measures for clinical professionals and caregivers.

To this end the EPSRC-funded “Sensor Platform for HEalthcare in Residential Environment (SPHERE)” Interdisciplinary Research Collaboration (IRC)⁴

³ <https://data.bris.ac.uk/data/dataset/8gccwpx47rav19vk8x4xapcog>

⁴ <http://www.irc-sphere.ac.uk/>

[11,12,3] has designed a multi-modal sensor system driven by data analytics requirements. The system is under test in a single house, and will be deployed in a general population of 100 homes in Bristol (UK). The data sets collected will be made available to researchers in a variety of communities.

Data is collected from the following three sensing modalities:

1. wrist-worn accelerometer;
2. RGB-D cameras (*i.e.* video with depth information); and
3. passive environmental sensors.

With these sensor data, we can learn patterns of behaviour, and can track the deterioration/progress of persons that suffer or recover from various medical conditions. To achieve this, we focus activity recognition over multiple tiers, with the two main prediction tasks of SPHERE including:

1. prediction of Activities of Daily Living (ADL) (*e.g.* tasks such as meal preparation, watching television); and
2. prediction of posture/ambulation (*e.g.* walking, sitting, transitioning).

Reliable predictions of ADL allows us to model behaviour and of residents over time, *e.g.* what does a typical day consist of, what times are particular activities performed *etc.* Prediction of posture and ambulation will complement ADL predictions, and can inform us about the physical well-being of the participant, how mobile/responsive is the participant, how active/sedintary, *etc.*

1.2 The SPHERE Challenge

The task for the SPHERE challenge is to predict posture and ambulation labels given the sensor data from the recruited participants.

We will henceforth refer to posture/ambulation as ‘activity recognition’ for brevity. It is worth noting that the term activity recognition has different interpretations when viewed from accelerometer, video, and environmental sensor perspectives. The definition of activities used in challenge most closely aligns to the terminology used in the field of accelerometer-based prediction.

For this task, accelerometer, RGB-D and environmental data is provided. Accelerometer is sampled at 20 Hz and given in its raw format (see Section 2.1). Raw video is not given in order to preserve anonymity of the participants. Instead, extracted features that relate to the centre of mass and bounding box of the identified persons are provided (see Section 2.2). Environmental data consists of Passive Infra-Red (PIR) sensors, and these is given in raw format (see Section 2.3).

Twenty (posture/ambulation) activities labels are annotated in our dataset, and these are enumerated below together with short descriptions:

- | | |
|--------------------------------|--------------------------------------|
| 1. a_ascend: ascent stairs; | 11. p_stand: standing; |
| 2. a_descend: descent stairs; | 12. t_bend: stand-to-bend; |
| 3. a_jump: jump; | 13. t_kneel_stand: kneel-to-stand; |
| 4. a_loadwalk: walk with load; | 14. t_lie_sit: lie-to-sit; |
| 5. a_walk: walk; | 15. t_sit_lie: sit-to-lie; |
| 6. p_bent: bending; | 16. t_sit_stand: sit-to-stand; |
| 7. p_kneel: kneeling; | 17. t_stand_kneel: stand-to-kneel; |
| 8. p_lie: lying; | 18. t_stand_sit: stand-to-sit; |
| 9. p_sit: sitting; | 19. t_straighten: bend-to-stand; and |
| 10. p_squat: squatting; | 20. t_turn: turn |

The prefix ‘a.’ on a label indicates an ambulation activity (*i.e.* an activity requiring of continuing movement), the prefix ‘p.’ indicate static postures (*i.e.* times when the participants are stationary), and the prefix ‘t.’ indicate posture-to-posture transitions. These labels are the target variables that are to be predicted in the challenge.

The SPHERE challenge will take place in conjunction with the ECML-PKDD⁵ conference, and will consist of two stages:

- **Stage 1. Scripted Data:** The first stage of the challenge uses sensor data that was recorded when the participants performed a pre-defined script. This script is described in later sections of this document. This data will be available from the challenge start to mid-April.
- **Stage 2. Scripted + Real Data:** The scripted data from Stage 1 will be augmented with naturalistic data that was recorded from participants that did not follow a pre-defined script. Both train and test datasets will be augmented with this new data. This data will be available from mid-April to challenge end.

Prizes will be awarded to the first three winners, with €1,000 being awarded to the winner, €600 to the runner up, and €400 to the second runner up. Performance is evaluated using the weighted Brier score (see Section 6 for more details on performance evaluation). Winning participants must participate in the workshop that is associated with the challenge. Participation rules and eligibility criteria are detailed in Section 6.

1.3 Structure of this Paper

The remainder of this paper is structured as follows: in Section 2 the set of sensors used in this dataset are described in greater detail, Section 3 provides details regarding the annotation process, Section 4 describes the primary task of this challenge in greater detail, Section 5 provides a description of the format of the data, and in Section 6 a link to the full set of rules governing this eligibility for participation and winning this challenge is provided.

⁵ <http://www.ecmlpkdd2016.org>

2 Sensors

The following subsections describe the sensing modalities that are found in the smart home. All sensors are synchronised with network time protocol (NTP).

2.1 Accelerometers

Participants wore a device equipped with a tri-axial accelerometer on the dominant wrist, attached using a strap. The device wirelessly transmits data using the Bluetooth Low Energy (BLE) standard to several access points (receivers) positioned within the house. The outputs of these sensors are a continuous numerical stream of the accelerometer readings (units of g , *i.e.* approximately 9.81 m s^{-2}). Accompanying the accelerometer readings are the received signal strength indications (RSSI) that were recorded by each access point (in units of dBm), and these data will be informative for indoor localisation. The accelerometers record data at 20 Hz, and the accelerometer ranges are between $\pm 8 g$. RSSI values are also recorded at 20 Hz, and values are no lower than 110 dBm.

Due to the nature of the sensing platform, there may be missing packets from the data. Recent accelerometer work done by SPHERE researchers on activity recognition with accelerometers includes [4,8].

2.2 RGB-D Cameras

Video recordings were taken using ASUS Xtion PRO RGB-Depth (RGB-D) cameras⁶. Automatic detection of humans was performed using the OpenNI library⁷, and false positive detections were manually removed by the organisers by visual inspection. Three RGB-D cameras are installed in the SPHERE house, and these are located in the living room, hallway, and the kitchen. No cameras are located elsewhere in the residence.

In order to preserve the anonymity of the participants the raw video data are not shared. Instead, the coordinates of the 2D bounding box, 2D centre of mass, 3D bounding box and 3D centre of mass are provided.

The units of 2D coordinates are in pixels (*i.e.* number of pixels down and right from the upper left hand corner) from an image of size 640×480 pixels. The coordinate system of the 3D data is axis aligned with the 2D bounding box, with a supplementary dimension that projects from the central position of the video frames. The first two dimensions specify the vertical and horizontal displacement of a point from the central vector (in millimetres), and the final dimension specifies the projection of the object along the central vector (again, in millimetres).

RGB-D data is very valuable in indoor environments as it can facilitate improvement in accuracy for fundamental computer vision tasks such as tracking [2] while also enabling specific analysis at higher levels. RGB-D data collected

⁶ https://www.asus.com/3D-Sensor/Xtion_PRO/

⁷ <https://github.com/OpenNI/OpenNI>

in the SPHERE house has been used for example for specific action recognition [6] and action quality estimation [7].

2.3 Environmental Sensors

The environmental sensing nodes are built on development platforms (Libelium, with CE marking)⁸, powered by batteries or/and 5 V DC converted from mains. PIR sensors are employed to detect presence in the data. Values of 1 indicate that motion was detected, whereas values of 0 mean that no motion was detected. A number of methods that deal with environmental sensor data are detailed in the following [9,10].

2.4 Recruitment and Sensor Layout

Participants were recruited to perform activities of daily living in the SPHERE house while the data recorded by the sensors was logged to a database. Ethical approval was secured from the University of Bristol’s ethics committee to conduct data collection, and informed consent was obtained from healthy volunteers.

In the first stage of data collection, participants were requested to follow a pre-defined script⁹. In the second stage, participants were recorded and data was annotated in a non-scripted naturalistic setting.

Figures 1a and 1b show the floor plan of the ground and first floors of the smart environment respectively.

3 Annotation

A team of 12 annotators were recruited and trained to annotate the set of activities outlined in the introduction. To support the annotation process, a head mounted camera (Panasonic HX-A500E-K 4K Wearable Action Camera Camcorder) recorded 4K video at 25 FPS to an SD-card. This data is not shared in this dataset, and is used only to assist the annotators. Synchronisation between the NTP clock and the head-mounted camera was achieved by focusing the camera on an NTP-synchronised digital clock at the beginning and end of the recording sequences.

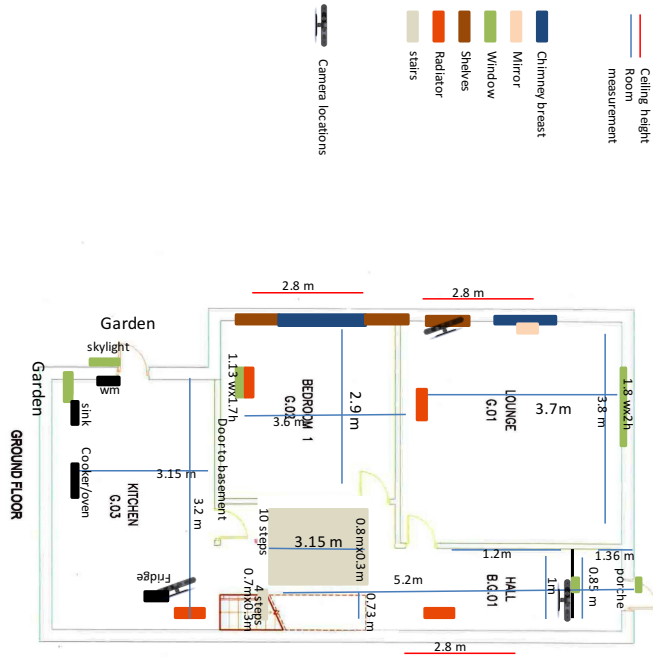
An annotation tool called ELAN¹⁰ was used for annotation. ELAN is a tool for the creation of complex annotations on video and audio resources, developed by the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands.

Room occupancy labels are also given in the training sequences. While performance evaluation is not directly affected by room prediction on this, participants may find that modelling room occupancy may be informative for prediction of posture and ambulation.

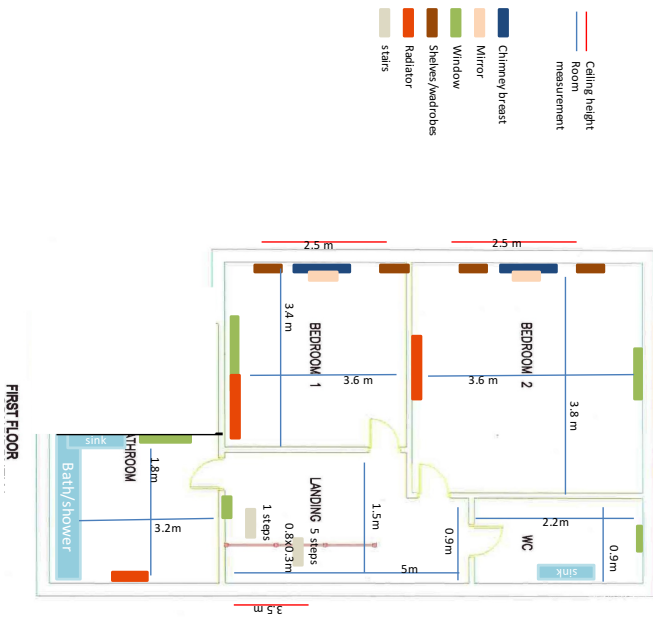
⁸ <http://www.libelium.com/>

⁹ Details of the script can be found at <http://www.irc-sphere.ac.uk/sphere-challenge/script>

¹⁰ <https://tla.mpi.nl/tools/tla-tools/elan/>



(a) Ground floor



(b) First floor

Fig. 1: The floorplan of the SPHERE smart home.

4 Task

The task is to predict posture, ambulation and transition activities for every second of the test data using the sensor data provided. Training data will be in the form of long sequences of data (tens of minutes in length), while test data will be short sequences, typically of length 10-30 seconds. Predictions must be made for every second of data.

4.1 Submission Files

`sample_submission.csv` contains an example submission file¹¹. Each row should follow the Comma Separated Values (CSV) format below. In this enumerated list, the the number indicates the column index of the CSV file.

1. <code>test_record_id</code>	9. $Pr(p_bent)$	17. $Pr(t_lie_sit)$
2. <code>start_time</code>	10. $Pr(p_kneel)$	18. $Pr(t_sit_lie)$
3. <code>end_time</code>	11. $Pr(p_lie)$	19. $Pr(t_sit_stand)$
4. $Pr(a_ascend)$	12. $Pr(p_sit)$	20. $Pr(t_stand_kneel)$
5. $Pr(a_descend)$	13. $Pr(p_squat)$	21. $Pr(t_stand_sit)$
6. $Pr(a_jump)$	14. $Pr(p_stand)$	22. $Pr(t_straighten)$
7. $Pr(a_loadwalk)$	15. $Pr(t_bend)$	23. $Pr(t_turn)$
8. $Pr(a_walk)$	16. $Pr(t_kneel_stand)$	

Here, `test_record_id` is the name of the test directory, `start_time` and `end_time` are time stamps indicating the window over which prediction is to be made (note that the difference between `end_time` and `start_time` will always be 1 sec), and $Pr(.)$ is the predicted probability of each activity during the window. As an example, we would expect 10 predictions for a sequence of length 10 seconds.

4.2 Performance Evaluation

As both the targets and predictions are probabilistic, classification performance is evaluated with weighted Brier score[1]:

$$BS = \frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C w_c (p_{n,c} - y_{n,c})^2 \quad (1)$$

where N is the number of test sequences, C is the number of classes, w_c is the weight for each class, $p_{n,c}$ is the predicted probability of instance n being from class c , and $y_{n,c}$ is the proportion of annotators that labelled instance n as arising

¹¹ This file is generated with: https://github.com/IRC-SPHERE/sphere-challenge/blob/master/sample_submission.py

from class c . Lower Brier score values indicate better performance, with optimal performance achieved with a Brier score of 0.

For the first phase of the challenge (*i.e.* with scripted data only) the class weights will be uniform, *i.e.* $w_c = 1/C$, $\forall 1 \leq c \leq C$. For the second phase (*i.e.* with both scripted and real data) the weights will be adjusted to place more emphasis on the more difficult tasks. The changes to the weights will be announced on the challenge website and this document will be updated to reflect the changes.

5 Data

This section outlines the format of the data records available from this challenge.

5.1 Directory Structure

Training data and testing data can be found in the ‘train’ and ‘test’ sub-directories respectively. The recorded data are collected under unique codes (each recording will be referred to as a ‘data sequence’). Timestamps are re-based to be relative to the start of the sequences, *i.e.* each sequence always starts from $t = 0$.

Each data sequence contains the following essential files:

1. `targets.csv`
2. `pir.csv`
3. `video_hallway.csv`
4. `video_living_room.csv`
5. `video_kitchen.csv`
6. `meta.json`

Two additional files are also provided in the training sequences:

1. `annotations_*.csv`
2. `location_*.csv`

The data from `annotations_*.csv` is used to create the `targets.csv` file. `locations.csv` is available for participants that want to model the location.

A git repository is publicly available ¹². In this repository a number of scripts for visualisation, benchmarking and data processing are available. (All subsequent sensor images were generated using these scripts.)

5.2 `targets.csv` (train only)

This file contains the probabilistic targets for classification. Multiple annotators may have annotated each sequence, and this file aggregates all of the annotations over 1 sec windows. These files can be created from `annotations_*.csv`,

¹² <https://github.com/IRC-SPHERE/sphere-challenge/>

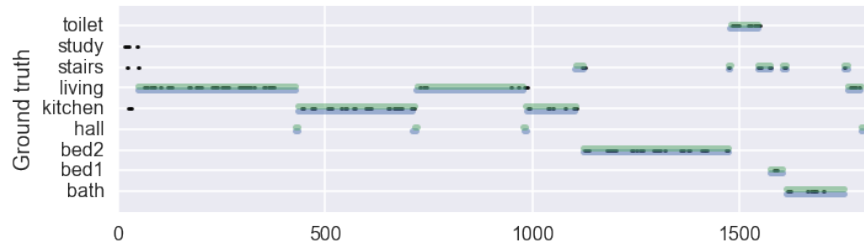


Fig. 2: Example PIR for training record 00001. The black lines indicate the time durations where a PIR is activated. The blue and green wide horizontal lines indicate the room occupancy labels as given by the two annotators that labelled this sequence.

and the following script shows how they are generated: https://github.com/IRC-SPHERE/sphere-challenge/blob/master/gen_targets.py.

The following 20 activities are labelled:

```

annotations = ('a_ascend', 'a_descend', 'a_jump', 'a_loadwalk',
              'a_walk', 'p_bent', 'p_kneel', 'p_lie', 'p_sit',
              'p_squat', 'p_stand', 't_bend', 't_kneel_stand',
              't_lie_sit', 't_sit_lie', 't_sit_stand',
              't_stand_kneel', 't_stand_sit', 't_straighten',
              't_turn')

```

where the prefix naming convention is the same as that described earlier.

`targets.csv` files contain 22 columns:

1. **start**: The starting time of the window (relative to the start of the sequence)
2. **end**: The ending time of the window (relative to the start of the sequence)
3. **targets**: Columns 3-22: the 20 probabilistic targets.

5.3 `pir.csv` (train + test)

This file contains the start time and duration for all PIR sensors in the smart environment. PIR sensors are located in the following rooms:

```

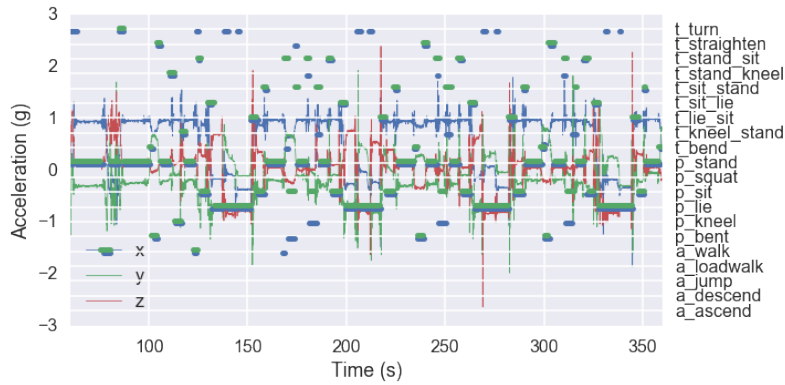
pir_locations = ('bath', 'bed1', 'bed2', 'hall', 'kitchen',
                'living', 'stairs', 'study', 'toilet')

```

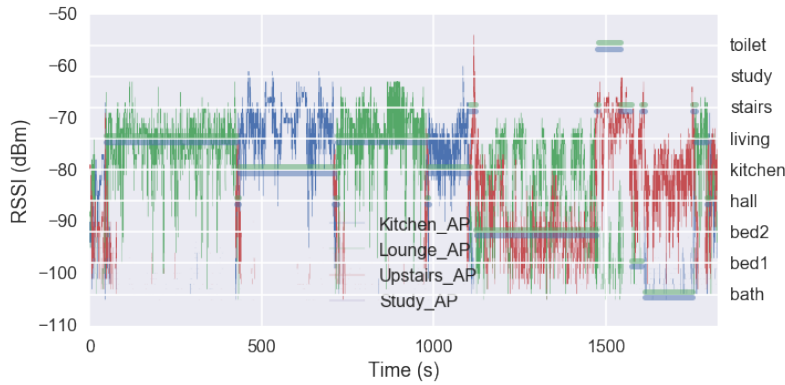
The columns of this CSV file are:

1. **start**: the start time of the PIR sensor (relative to the start of the sequence)
2. **end**: the end time of the PIR sensor (relative to the start of the sequence)
3. **name**: the name of the PIR sensor being activated (from the `pir_locations` list)
4. **index**: the index of the activated sensor from the `pir_locations`

Example PIR sensor activations and ground truth locations are overlaid in Figure 2. Note, the PIR sensors can be noisy in nature.



(a) Acceleration signal trace shown over a 5 minute time period. Annotations are overlaid.



(b) RSSI values from the four access points. Room occupancy labels are shown by the the horizontal lines.

Fig. 3: Example acceleration and RSSI signals for training record 00001. The line traces indicate the accelerometer/RSSI values recorded by the access points. The horizontal lines indicate the ground-truth as provided by the annotators (two annotators annotated this record, and their annotations are depicted by the green and blue traces respectively).

5.4 meta.json (train + test)

This file contains meta-data regarding the data records regarding the duration of the sequence. This data can be inferred from the sensor data but is provided for convenience.

5.5 acceleration.csv (train + test)

The acceleration file consists of eight columns:

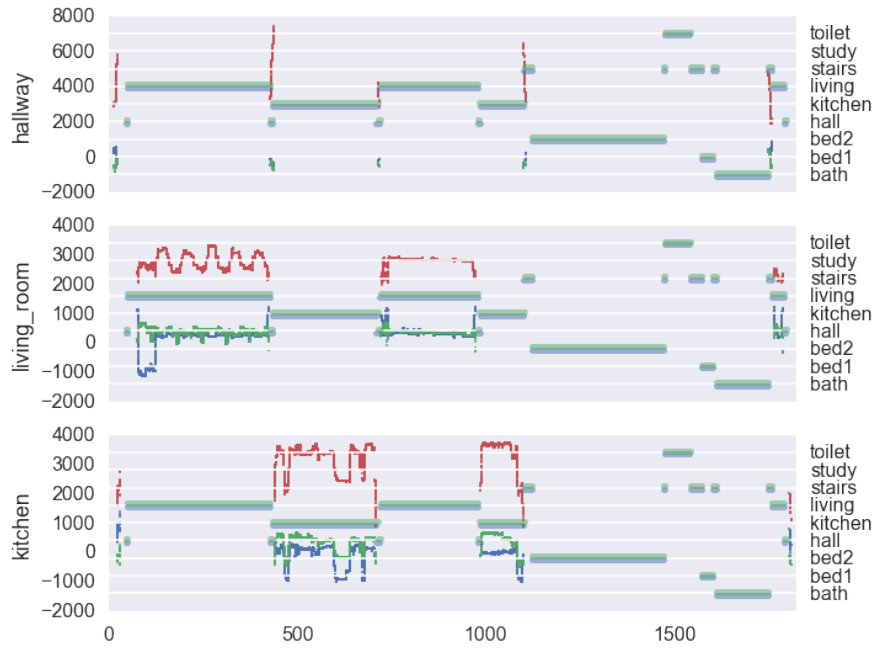


Fig. 4: Example `centre_3d` for training record 00001. The horizontal lines indicate annotated room occupancy. The blue, green and red traces are the x , y , and z values for the 3D centre of mass.

1. **t**: this is the time of the recording (relative to the start of the sequence)
2. **x/y/z**: these are the acceleration values recorded on the x/y/z axes of the accelerometer.
3. **Kitchen_AP/Lounge_AP/Upstairs_AP/Study_AP**: these specify the RSSI of the acceleration signal as received by the access kitchen/lounge/upstairs access points. Empty values indicate that the access point did not receive the packet.

Sample acceleration and RSSI with overlaid annotations are shown in Figure 3. Figure 3a shows the acceleration signals overlaid with activity labels, and Figure 3b shows the RSSI signal information with the overlaid room occupancy.

5.6 video_*.csv (train + test)

The following columns are found in the `video_hallway.csv`, `video_kitchen.csv` and `video_living_room.csv` files:

1. **t**: The current time (relative to the start of the sequence)
2. **centre_2d_x/centre_2d_y**: The x- and y-coordinates of the centre of the 2D bounding box.

3. **bb_2d_br_x/bb_2d_br_y**: The x and y coordinates of the bottom right (br) corner of the 2D bounding box
4. **bb_2d_tl_x/bb_2d_tl_y**: The x and y coordinates of the top left (tl) corner of the 2D bounding box
5. **centre_3d_x/centre_3d_y/centre_3d_z**: the x, y and z coordinates for the centre of the 3D bounding box
6. **bb_3d_brb_x/bb_3d_brb_y/bb_3d_brb_z**: the x, y, and z coordinates for the bottom right back (brb) corner of the 3D bounding box
7. **bb_3d_ftl_x/bb_3d_ftl_y/bb_3d_ftl_z**: the x, y, and z coordinates of the front left top (ftl) corner of the 3D bounding box.

Example 3D centre of mass data is plotted for the hallway, living room and kitchen cameras in Figure 4. Room occupancy labels are overlaid on these, where we can see strong correspondence between the detected persons and room occupancy.

5.7 Supplementary Files

The following two sets of file need not be used for the challenge, but are included to facilitate users that wish to perform additional modelling of the sensor environment. For example, indoor localisation can be modelled with the `locations.csv` file, *c.f.* [5].

`annotations_*.csv` (train only):

This file provides the individual annotations as provided by the annotators. The target variables are the same as for `targets.csv`. The `annotations_*.csv` contains the following four columns:

1. **start**: the start time of the activity (relative to the start of the sequence)
2. **end**: the end time of the activity (relative to the start of the sequence)
3. **name**: the name of the label (from the `annotations` list)
4. **index**: the index of the label name starting at 0

`location_*.csv` (train only):

This file provides the annotation labels for room occupancy. The following rooms are labelled:

```
location_names = ('bath', 'bed1', 'bed2', 'hall', 'kitchen',
                  'living', 'stairs', 'study', 'toilet')
```

`location.csv` contains the following four columns:

1. **start**: the time a participant entered a room (relative to the start of the sequence)
2. **end**: the time the participant left the room (relative to the start of the sequence)
3. **name**: the name of the room (from the `location_names` list)
4. **index**: the index of the room name starting at 0

6 Eligibility and Rules

The full set of rules for the challenge can be found at:
<http://irc-sphere.ac.uk/sphere-challenge/rules>

Acknowledgements

This work was performed under the SPHERE IRC funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/K031910/1.

References

1. Brier, G.W.: Verification of forecasts expressed in terms of probability. *Monthly weather review* 78(1), 1–3 (1950)
2. Camplani, M., Hannuna, S., Mirmehdi, M., Damen, D., Paiement, A., Tao, L., Burghardt, T.: Real-time rgb-d tracking with depth scaling kernelised correlation filters and occlusion handling. In: Xianghua Xie, M.W.J., Tam, G.K.L. (eds.) *Proceedings of the British Machine Vision Conference (BMVC)*. pp. 145.1–145.11. BMVA Press (September 2015), <https://dx.doi.org/10.5244/C.29.145>
3. Diethe, T., Flach, P.A.: Smart-homes for eHealth: Uncertainty management and calibration. In: *NIPS 2015 Workshop on Machine Learning in Healthcare (2015)*
4. Diethe, T., Twomey, N., Flach, P.: Active transfer learning for activity recognition. In: *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*
5. Fafoutis, X., Mellios, E., Twomey, N., Diethe, T., Hilton, G., Piechocki, R.: An rssi-based wall prediction model for residential floor map construction. In: *Proceedings of the 2nd IEEE World Forum on Internet of Things (WF-IoT) (2015)*
6. Tao, L., Hannuna, S., Burghardt, T., Camplani, M., Paiement, A., Aldamen, D., Mirmehdi, M., Craddock, I.: A Comparative Home Activity Monitoring Study using Visual and Inertial Sensors (2015)
7. Tao, L., Paiement, A., Aldamen, D., Mirmehdi, M., Hannuna, S., Camplani, M., Burghardt, T., Craddock, I.: A comparative study of pose representation and dynamics modelling for online motion quality assessment. *Computer Vision and Image Understanding* (11 2015)
8. Twomey, N., Diethe, T., Flach, P.: Bayesian active learning with evidence-based instance selection. In: *Workshop on Learning over Multiple Contexts, European Conference on Machine Learning (ECML'15) (2015)*
9. Twomey, N., Diethe, T., Flach, P.: Unsupervised learning of sensor topologies for improving activity recognition in smart environments. To appear in *Neurocomputing* (2016)
10. Twomey, N., Flach, P.: Context modulation of sensor data applied to activity recognition in smart homes. In: *Workshop on Learning over Multiple Contexts, European Conference on Machine Learning (ECML'14) (2014)*
11. Woznowski, P., Fafoutis, X., Song, T., Hannuna, S., Camplani, M., Tao, L., Paiement, A., Mellios, E., Haghighi, M., Zhu, N., Hilton, G., Damen, D., Burghardt, T., Mirmehdi, M., Piechocki, R., Kaleshi, D., Craddock, I.: A multi-modal sensor infrastructure for healthcare in a residential environment. In: *Int. Conf. Communications (ICC) Workshops*, in press (2015)

12. Zhu, N., Diethe, T., Camplani, M., Tao, L., Burrows, A., Twomey, N., Kaleshi, D., Mirmehdi, M., Flach, P.A., Craddock, I.: Bridging e-health and the internet of things: The SPHERE project. *IEEE Intelligent Systems* 30(4), 39–46 (2015), <http://doi.ieeecomputersociety.org/10.1109/MIS.2015.57>