Data in Brief 34 (2021) 106632

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Multi-sensor dataset of human activities in a smart home environment

Gibson Chimamiwa*, Marjan Alirezaie, Federico Pecora, Amy Loutfi

Centre for Applied Autonomous Sensor Systems (AASS), Örebro University, Sweden

ARTICLE INFO

Article history:
Received 2 November 2020
Revised 2 December 2020
Accepted 3 December 2020
Available online 9 December 2020

Keywords:
Activities of daily living
Smart homes
Time series dataset
Activity recognition
Habit recognition

ABSTRACT

Time series data acquired from sensors deployed in smart homes present valuable information for intelligent systems to learn activity patterns of occupants. With the increasing need to enable people to age in place independently, the availability of such data is key to the development of home monitoring solutions. In this article we describe an unlabelled dataset of measurements collected from multiple environmental sensors placed in a smart home to capture human activities of daily living. Various sensors were used including passive infrared, force sensing resistors, reed switches, mini photocell light sensors, temperature and humidity, and smart plugs. The sensors record data from the user's interactions with the environment, such as indoor movements, pressure applied on the bed, or current consumption when using electrical appliances. Millions of raw sensor data samples were collected continuously at a frequency of 1 Hz over a period of six months between 26 February 2020 and 26 August 2020. The dataset can be useful in the analysis of different methods, including data-driven algorithms for activity or habit recognition. In particular, the research community might be interested in investigating the performance of algorithms when applied on unlabelled datasets and not necessarily on annotated datasets. Furthermore, by applying artificial intelligence (AI) algorithms on such data collected over long periods, it is possible to extract patterns that reveal the user's habits as well as detect changes in the habits. This can

* Corresponding author.
E-mail addresses: gibson.chimamiwa@oru.se (G. Chimamiwa), marjan.alirezaie@oru.se (M. Alirezaie), federico.pecora@oru.se (F. Pecora), amy.loutfi@oru.se (A. Loutfi).
Social media: (G. Chimamiwa)

https://doi.org/10.1016/j.dib.2020.106632

benefit in detecting deviations in order to provide timely interventions for patients, e.g., people with dementia.

Specifications Table

Subject	Computer Science
Specific subject area	Smart homes, Time series dataset, Activity recognition, Habit recognition
Type of data	Time series sensor observations from indoor activities of daily living (ADL) performed by the resident. The sensor readings are generated in the form of numeric values.
How data were acquired	The data was acquired from several sensors deployed in the smart home namely, *Mini Photocell Light Sensors, Grove - PIR Motion Sensors, Si7021 Temperature-Humidity Sensor BoB, Force Sensing Resistor (38*38mm), Reed Switches*, and *TP-Link Wi-Fi Smart Plug HS110 with Energy Monitoring. Adafruit Feather HUZZAH ESP8266 WiFi Arduino* micro-controllers were programmed to read data from the sensors. The *Message Queuing Telemetry Transport* (MQTT) (https://mqtt.org/) was used as the publish and subscribe communication protocol for gathering data and sending it to a central database server for storage.
Data format	The data consists of raw sensor values formatted either as integer or floating point data types. Furthermore, each data value is associated with a timestamp (YYYY-MM-DD HH:MM:SS) value to indicate the time point at which the value was recorded.
Parameters for data collection	It was assumed that the data collection is for a typical home environment occupied by a single user. The data collection process was configured to read and send data to the database continuously at a frequency of 1 Hz, as opposed to sending data only when the sensors are triggered. Furthermore, given the real-world nature of the data gathering process, and the need to free the resident from tedious and time consuming annotation process, the data was not labelled.
Description of data collection	The data was collected using non-obtrusive environmental sensors. In order to capture different scenarios within the home environment, 6 different types of sensors were used, namely motion, pressure, light, temperature and humidity, contact, and smart plug sensors. These are comprised of 6 passive infrared, 3 force sensing resistors, 3 reed switches, 3 mini photocell light sensors, 1 temperature and humidity, and 7 smart plugs making a total of 23 sensors. A micro-controller was attached to each sensor, and was configured to read the data and send it over the Internet to a central MySQL database server for storage. The data was collected continuously over a period of six months.
Data source location	Institution: Örebro university City: Örebro Country: Sweden
Data accessibility	Repository name: Mendeley Data Data identification number: https://doi.org/10.17632/t9n68ykfk3.1 Direct URL to data: https://data.mendeley.com/datasets/t9n68ykfk3/1

Value of the Data

- Open access to the raw sensor data can be helpful to further the development of algorithms targeted at smart home environments, for example, in activity recognition. Notably, the Center for Advanced Studies in Adaptive Systems (CASAS) dataset [1] has been used in activity recognition using different machine learning algorithms. However, unlike in this work, their dataset is not based on a real-world setting, as the experimenter was responsible for informing the participants which activity to perform at a given time. In more realistic environments,

users may not always follow the same sequence of activities when performing tasks, e.g., people with dementia who experiences changes along the different stages of the disease [2]. Examples of simple activities that could be recognised using the sensors mentioned above include, *sitting on the couch, watching TV, cooking, sleeping, measuring weight, washing clothes, washing the dishes, making coffee, making a sandwich, making a hot drink, bathing*, and *exiting* or *entering the house*. More complex activities can also be recognised such as *having coffee* while *sitting on the couch* and *watching TV* for 60 minutes before *sleeping* for 8 hours.
- Data collected continuously over a long period of time offers opportunities to test different machine learning algorithms in areas such as habit recognition [2,3] or anomaly detection [4]. For example, considering people with dementia who may experience gradual changes over time [5], spanning several stages [6], it becomes difficult to capture changes in habits from data collected over short periods of time. Examples of habits that could be extracted from the sensor data discussed in this work include, *measuring weight every morning, 10 minutes after getting up from bed*, or *regularly taking a nap for 1 h in the afternoon after watching TV on the couch for more than 2 hours*.
- Since it was collected in an uncontrolled setting, the data can be useful in further investigating practical frameworks for monitoring human activities in their home environments. Testing algorithms on unlabelled data is critical [7], given the rising demand for smart home technologies [8] aimed at enabling elderly people to live independently in their preferred home environments, given the high societal cost associated with keeping them in care centers [9].
- The wide range of sensor devices used in this work also opens avenues for a more holistic analysis of the different aspects of the user's activities of daily living, which could lead to interesting knowledge discovery regarding the user's habits as well as other emerging habits. Such insight would be difficult to obtain by monitoring the user's activities with a more limited range of sensors.
- An additional value of this dataset is that it offers an opportunity for manufacturers to analyse the performance of the sensors for possible enhancements, for example, regarding compatibility with other hardware or software components.

1. Data Description

The gathered sensor data is stored in the E-care@home database [10], which is part of an interconnected set of software components for data collection, labelling, and reasoning tasks. The E-care@home system [11] is a knowledge-driven context recognition system that reads sensor data and sends it to a reasoner equipped with logic to output labelled activity time intervals and other inferred information.

In Fig. 1, we present only the database structure related to this work, namely the *sensor_sample_int, sensor_sample_float, sensor, node*, and *location* tables. The *sensor_sample_int* table stores measurements of integer datatype and this includes data from motion, pressure, light, and contact sensors. On the other hand, the *sensor_sample_float* table contains floating point datatype, which includes data captured by the temperature and humidity sensor and the smart plugs. Furthermore, each record in the *sensor_sample_int* and *sensor_sample_float* tables is linked to the *sensor* table through the *sensor_id* attribute. The datatype of each sensor reading is indicated as the *enum* attribute in the *sensor* table. The *sensor* table stores information such as the id, datatype of the sensor reading, and specific name of a sensor. Each record in the *sensor* table is linked to the *node* table through *node_id*, which is used to identify the object on which a sensor is attached. The *node* table is connected to the *location* table through the *location_id* attribute, which is used to identify a specific smart home where the sensors are deployed. The dataset described in this work is linked to *location_id*, 711, which was specified in the database query used for extracting the dataset for this specific location.

In order to make the dataset publicly accessible, data was extracted from the database and stored in three comma-separated values (csv) files, namely *sensor.csv, sensor_sample_int.csv*, and

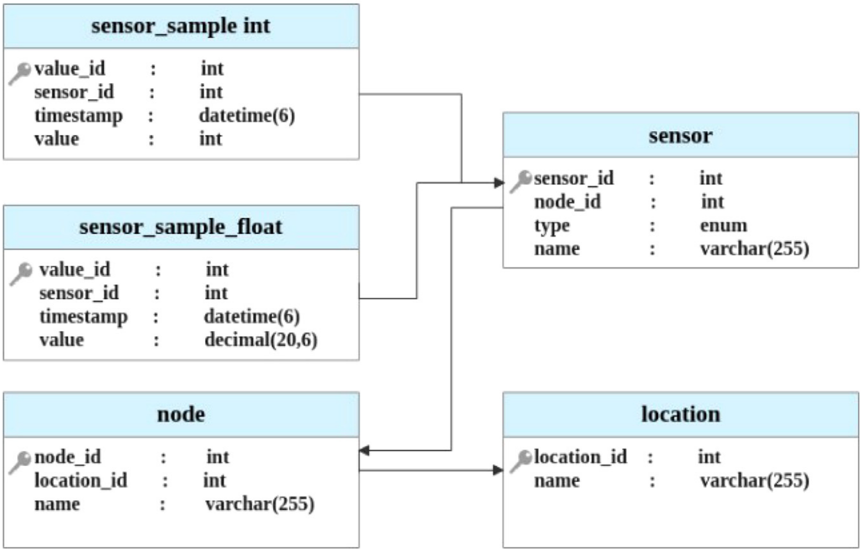

Fig. 1. Database tables for storing the sensor data.

Table 1
Sample of records from *sensor_sample_int* table.

sensor id	timestamp	value	name
5895	2020-05-04 12:55:45	0	bathroom/ambience/motion
5887	2020-07-18 16:00:00	175	kitchen/stove/light
5891	2020-07-28 16:00:00	0	livingroom/ambience/motion
5892	2020-07-28 16:00:00	0	bedroom/ambience/motion
6127	2020-07-28 16:00:00	1024	livingroom/tv/light
5896	2020-07-28 16:00:00	555	bedroom/bed/pressure
6687	2020-07-28 16:00:00	1	bedroom/weightscale/pressure
5889	2020-07-28 16:00:00	6	livingroom/couch/pressure
7125	2020-07-28 16:00:00	1024	bathroom/ambience/light
5893	2020-07-28 16:00:00	0	kitchen/ambience/motion
5888	2020-07-28 16:00:00	0	entrance/door/contact
6686	2020-07-28 16:00:00	0	bedroom/ambience_under_the_bed/motion
6220	2020-07-28 16:00:00	0	balcon/door/contact
5894	2020-07-28 16:00:01	0	corridor/ambience/motion
6253	2020-07-28 16:00:01	0	kitchen/fridge/contact

sensor_sample_float.csv. The data is made available on the Mendeley data repository as provided in the *Data accessibility* section under the *Specifications Table* above.

Table 1 and Table 2 provide a view of the raw sensor data extracted by joining the *sensor_sample_int* or *sensor_sample_float* with the *sensor* table. All data measurements are recorded at a frequency of 1 Hz and the *timestamp* column represents the time point at which the sensor data was captured. As mentioned already, the *sensor id* column uniquely identifies each sensor record. The *value* column stores the specific sensor data in either binary or analog form. For example, motion and contact sensors record only binary data i.e., either 0 to indicate no motion detected or door closed or 1 to indicate motion detected or door open. The rest of the sensors record analog data whose values change continuously. The *name* column further specifies each sensor record based on the *section* of the house in which the sensor is located, the *object* being measured, and the *property* of the monitored object. For example, given the sensor name

Table 2
Sample of records from *sensor_sample_float* table.

sensor id	timestamp	value	name
6896	2020-07-29 00:00:40	0.000000	kitchen/microwave/current
6632	2020-07-29 16:00:05	0.000000	kitchen/coffeemaker/current
6636	2020-07-29 16:00:05	0.000000	bathroom/washingmachine/current
6633	2020-07-29 16:00:06	0.317000	kitchen/sandwichmaker/current
7139	2020-07-29 16:00:06	0.000000	corridor/ilifeRobot/current
6223	2020-07-29 16:00:06	23.270000	bathroom/ambience/temperature
6635	2020-07-29 16:00:06	0.322000	kitchen/kettle/current
6222	2020-07-29 16:00:06	51.620000	bathroom/ambience/humidity
6634	2020-07-29 16:00:06	0.000000	kitchen/dishwasher/current

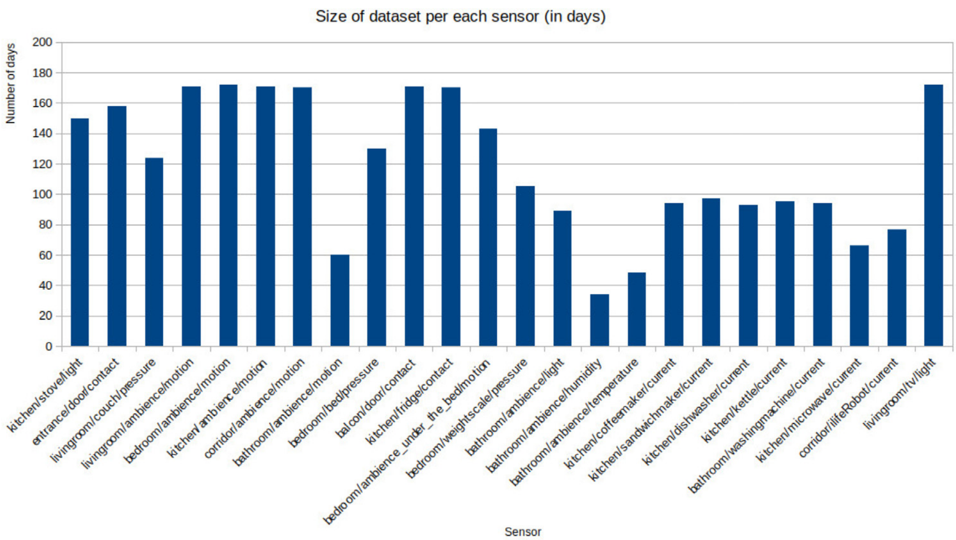

Fig. 2. Size of dataset in days.

livingroom/couch/pressure; livingroom is the section of the house where an activity was captured, *couch* is the object on which action was applied, and *pressure* is the property of the object.

The sensors were deployed on a rolling basis during the six months of data collection. As a result, the size of data collected for the different sensors varies. A summary of the total number of days for which data was collected per each sensor is shown on Fig. 2.

An example of a 24-hour raw sensor data plot for two consecutive days recorded from three sensors placed in the living room, namely motion sensor, light sensor on the TV, and pressure sensor on the couch is shown on Fig. 3.

2. Experimental Design, Materials and Methods

2.1. Environment

The apartment is divided into a living room, kitchen, bedroom, bathroom, corridor, and a balcony. The apartment contains several instruments and objects used by the resident, which could provide useful information on the functional independence of the occupant. These include a bed, couch, TV, fridge, and several electrical appliances such as coffee maker, sandwich maker, dishwasher and others. A pictorial view of the apartment layout is shown on Fig. 4.

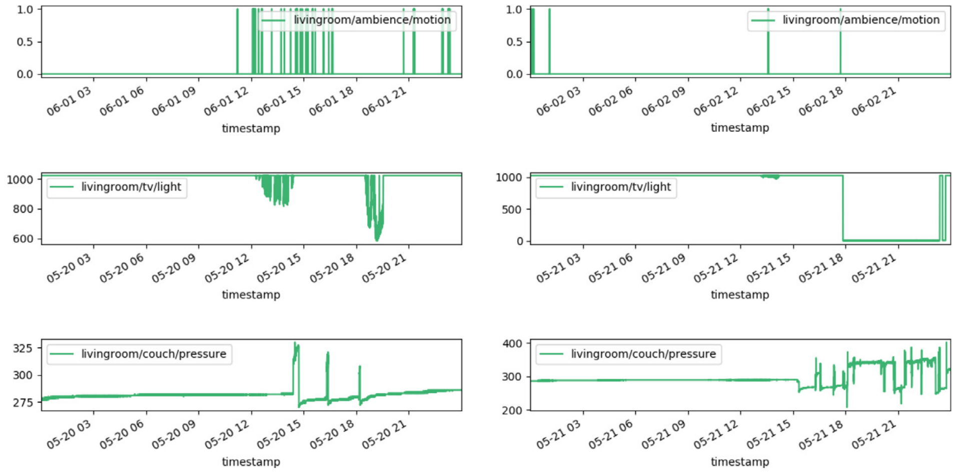

Fig. 3. Plot of data from motion, TV light, and couch pressure sensors.

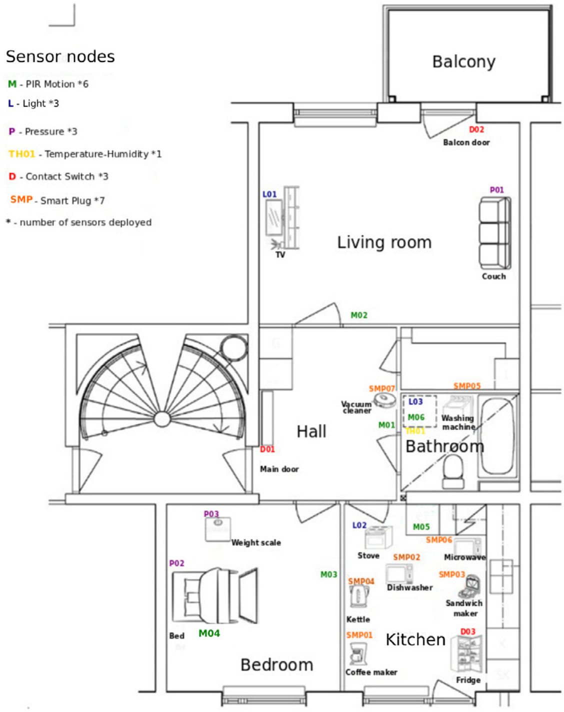

Fig. 4. Sensor distribution in the experimental home environment.

2.2. Materials

The choice of sensors used in this experiment was motivated by the desire to monitor important ADL that can help to provide sufficient assessment on the functional independence of a user in the home environment [12]. These activities include basic ADL, e.g., bathing, sleeping,

Table 3
Summary of sensors, object monitored, and location.

Sensor	Sensor Type	Object Monitored	Location
M01	Motion	Ambience	Corridor
M02	Motion	Ambience	Living room
M03	Motion	Ambience	Bedroom
M04	Motion	Ambience	Bedroom
M05	Motion	Ambience	Kitchen
M06	Motion	Ambience	Bathroom
L01	Light	TV	Living room
L02	Light	Stove	Kitchen
L03	Light	Ambience	Bathroom
P01	Pressure	Couch	Living room
P02	Pressure	Bed	Bedroom
P03	Pressure	Weight scale	Bedroom
TH01	Temperature & Humidity	Ambience	Bathroom
D01	Reed switch	Door contact	Entrance
D02	Reed switch	Door contact	Balcon
D03	Reed switch	Fridge Door contact	Kitchen
SMP01	Smart plug	Coffee maker	Kitchen
SMP02	Smart plug	Dishwasher	Kitchen
SMP03	Smart plug	Sandwich maker	Kitchen
SMP04	Smart plug	Kettle	Kitchen
SMP05	Smart plug	Washing machine	Bathroom
SMP06	Smart plug	Microwave	Kitchen
SMP07	Smart plug	Vacuum cleaner	Corridor

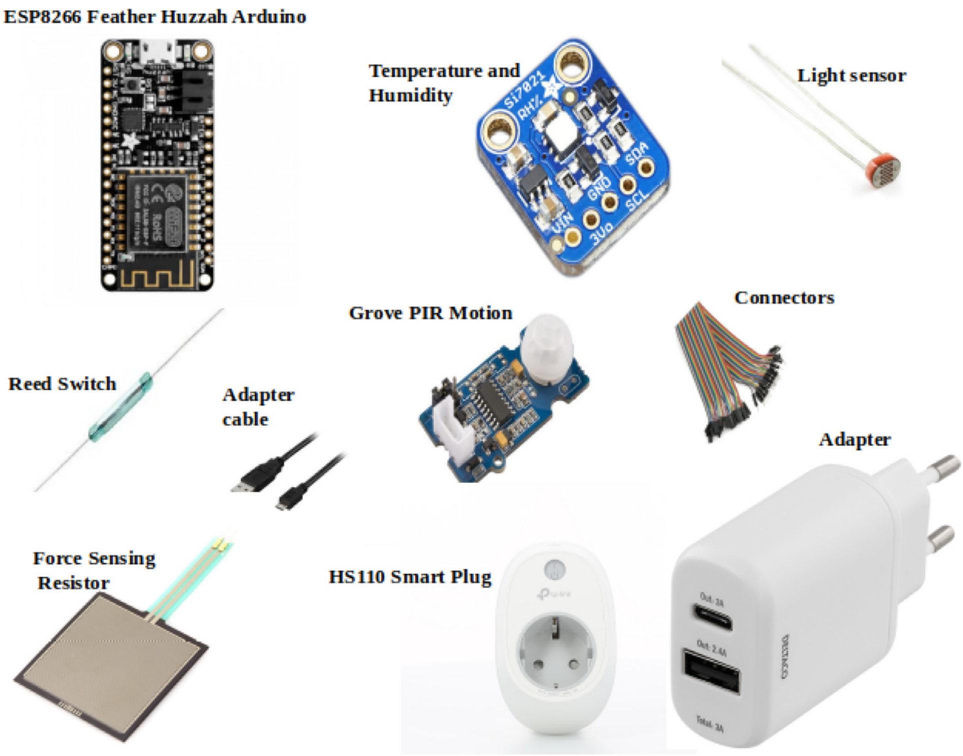

Fig. 5. Instruments used in the data gathering process.

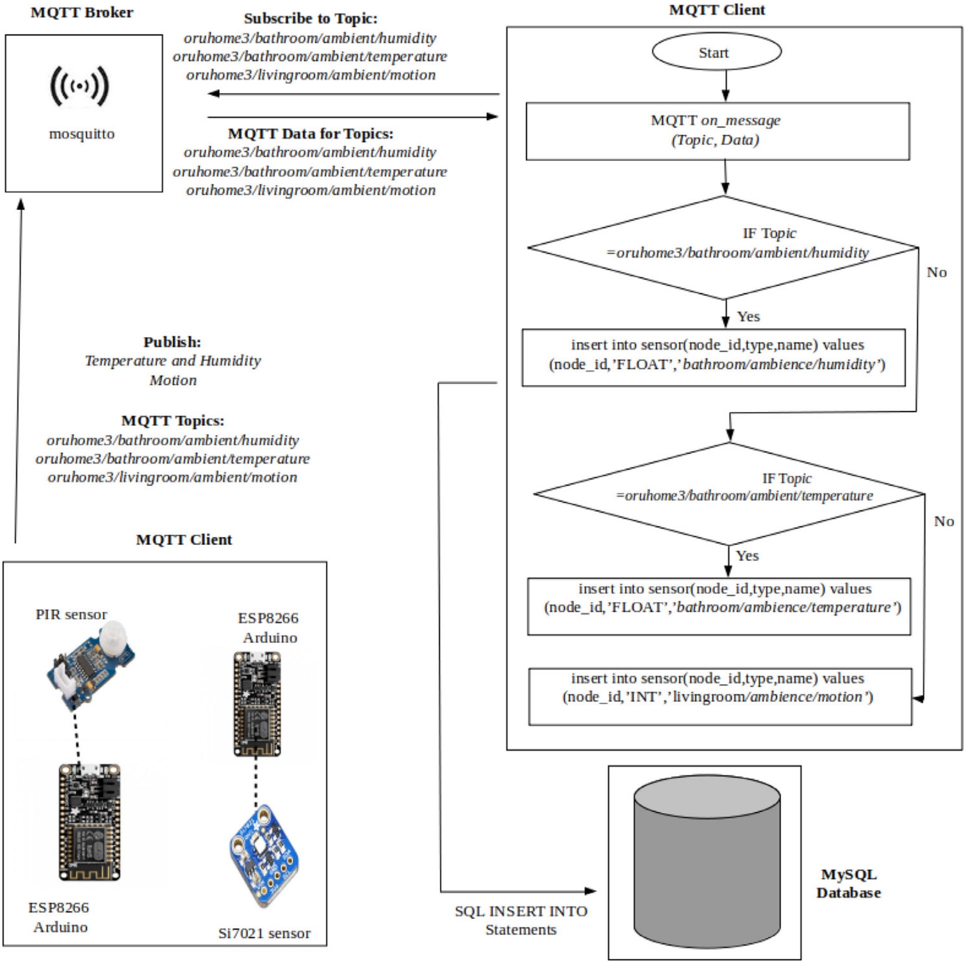

Fig. 6. Architecture of the design experiment.

sitting, and functional mobility as well as instrumental ADL such as preparing food or washing the dishes. A detailed summary of the sensor label, sensor type, object monitored, and the location of the sensor in the home is shown on Table 3.

2.2.1. Ambient sensors:

We used the Mini Photocell Light Sensor to measure the light intensity when the TV or the stove is switched on, the Grove - PIR Motion Sensor to detect the presence of the occupant in specific locations of the home, the Si7021 Temperature and Humidity Sensor BoB to measure temperature and humidity of the air in the bathroom.

2.2.2. Object sensors:

We also used sensors attached to objects in the house, including Force Sensing Resistor (38*38mm) to detect amount of pressure applied on the bed, couch and weight scale, the Reed Switch to detect contact when the entrance and exit door as well as fridge door are opened or closed. In addition, several TP-Link Wi-Fi Smart Plugs HS110 with Energy Monitoring were used to measure current usage from different electric home appliances including coffee maker, microwave, sandwich maker, electric kettle, dishwasher, washing machine, and vacuum cleaner (ilifeRobot).

2.2.3. Micro-controller:
Each sensor device is connected to the Adafruit Feather HUZZAH ESP8266 WiFi arduino micro-controller. The micro-controller is programmed to read the data from the sensor using the Arduino IDE programmable interface.

The collection of sensors and instruments used in the data gathering experiment is shown on Fig. 5.

2.3. Method

The experiment was designed in such a way as to be both simple and easy to set up. The micro-controller contains the configurations for sensor devices to send data to topics defined on the MQTT server through the publish and subscribe architecture. The configurations are written using the Arduino IDE interface and uploaded to the micro-controller using a USB cable. Once the code is uploaded, the micro-controller together with the attached sensor are placed in the Wi-Fi enabled home environment and sensor data is published to the MQTT broker. MQTT clients can then subscribe to the topics in order to receive the published data. A JavaScript Object Notation (JSON) file is configured to link the MQTT topics to specific sensor names in the database [10]. In the case of smart plugs, the code used to read and publish data to the MQTT server runs separately on a laptop connected to the same Wi-Fi network. On Fig. 6, we illustrate how the different components are connected.

Ethics Statement

Prior to conducting the experiment all ethical guidelines were followed including obtaining consent from the apartment resident.

CRediT Author Statement

Gibson Chimamiwa was responsible for setting up the data collection experiment and gathering the data. Gibson Chimamiwa, Marjan Alirezaie, Federico Pecora, and Amy Loutfi contributed to the writing of the manuscript. Marjan Alirezaie, Federico Pecora, and Amy Loutfi contributed to the supervision of the project. Amy Loutfi was responsible for the project administration and funding.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have influenced the work reported in this article.

Acknowledgments

This work has been supported by both the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 754285, and the distributed research environment E-care@home funded by the Swedish Knowledge Foundation (KKS), 2015–2019.

References

[1] D.J. Cook, M. Schmitter-Edgecombe, Assessing the quality of activities in a smart environment, Methods Inf. Med. 48 (2019) 363–369.
[2] G. Chimamiwa, M. Alirezaie, H. Banaee, U. Köckemann, A. Loutfi, Towards habit recognition in smart homes for people with dementia, in: Proceedings of the European Conference on Ambient Intelligence, Springer, 2019, pp. 363–369.
[3] J. Lee, N. Melo, Habit representation based on activity recognition, Sensors 20 (7) (2020) 1928.
[4] D. Arifoglu, A. Bouchachia, Detection of abnormal behaviour for dementia sufferers using convolutional neural networks, A.I. Med. 94 (2019) 88–95.
[5] A.L. Fymat, On dementia and other cognitive disorders, Clin. Res. Neurol. 2 (1) (2019) 1–14.
[6] B. Reisberg, S.H. Ferris, M.J. de Leon, T. Crook, The global deterioration scale for assessment of primary degenerative dementia, Psychiatry J. 139 (1982) 1136–1139.
[7] S. Ranasinghe, F. Al Machot, H.C. Mayr, A review on applications of activity recognition systems with regard to performance and evaluation, Int. J. Distrib. Sens. Netw. 12 (8) (2016). 1550147716665520
[8] M. Amiribesheli, H. Bouchachia, A tailored smart home for dementia care, J. Ambient. Intell. Humaniz Comput. 9 (6) (2018) 1755–1782.
[9] M. Prince, A. Wimo, M. Guerchet, G. Ali, Y. Wu, M. Prina, World alzheimer report 2015–the global impact of dementia, an analysis of prevalence, incidence, cost and trends, Alzheimer's Dis. Int. 17 (2015) 2016.
[10] U. Köckemann, M. Alirezaie, J. Renoux, N. Tsiftes, M.U. Ahmed, D. Morberg, M. Lindén, A. Loutfi, Open-source data collection and data sets for activity recognition in smart homes, Sensors 20 (3) (2020) 879.
[11] M. Alirezaie, J. Renoux, U. Köckemann, A. Kristoffersson, L. Karlsson, E. Blomqvist, N. Tsiftes, T. Voigt, A. Loutfi, An ontology-based context-aware system for smart homes: E-care@home, Sensors 17 (7) (2017) 1586.
[12] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, A. Bauer, Monitoring activities of daily living in smart homes: understanding human behavior, IEEE Signal Process. Mag. 33 (2) (2016) 81–94.