Generation of a Partitioned Dataset with Single, Interleave and Multioccupancy Daily Living Activities

Francisco J. Quesada[1], Francisco Moya[1], Javier Medina[1], Luis Martínez[1], Chris Nugent[2], and Macarena Espinilla[1](✉)

[1] Department of Computer Science, University of Jaén, Jaén, Spain
{fqreal,fpmoya,jmquero,martin,mestevez}@ujaen.es
[2] School of Computing and Mathematics, University of Ulster, Jordanstown BT37 0QB, UK
cd.nugent@ulster.ac.uk

Abstract. The advances in electronic devices have entailed the development of smart environments which have the aim to help and make easy the life of their inhabitants. In this kind of environments, an important task is the process of activity recognition of an inhabitant in the environment in order to anticipate the occupant necessities and to adapt such smart environment. Due to the cost to checking activity recognition approaches in real environments, usually, they use datasets generated from smart environments. Although there are many datasets for activity recognition in smart environments, it is difficult to find single, interleaved or multioccupancy activity datasets, or combinations of these classes of activities according to the researchers' needs. In this work, the design and development of a complete dataset with 14 sensors and 9 different activities daily living is described, being this dataset divided into partitions with different classes of activities.

Keywords: Dataset · Activity recognition · Smart environments · Single activities · Interleave activities · Multioccupancy activities

1 Introduction

The advance in the miniaturization of electronic devices in addition to a reduction in their cost, have created an environment whereby we are surrounded by embedded sensing technology, arising the Ambient Intelligence (AmI) concept [14]. The notion of a smart home follows this vision with an environment of embedded technology and processing sensor data with the ability to ascertain the behaviour of its inhabitants.

This kind of environments can help people, especially people with cognitive diseases, in their daily live. Thus, for example, in a smart home it is possible to prompt inhabitants how to finish a given activity daily living (ADL) action, like preparing a tea, remind them where is a particular object, like the sugar, when

J.M. García-Chamizo et al. (Eds.): UCAmI 2015, LNCS 9454, pp. 60–71, 2015.
DOI: 10.1007/978-3-319-26401-1_6

they have to take their medicines, or when they should move after abnormal periods of inactivity [10].

Nowadays, researchers are focusing on their efforts on the improvement of smart homes, in order to be more helpful. So, there are many researching areas with this aim, being one of the most important the area of activity recognition (AR) [3,8,15].

The process of activity recognition aims to recognize the actions and goals of one or more person within the environment based on a series of observations of actions and environmental conditions [1]. It can, therefore, be deemed as a complex process that involves the following steps: (i) to choose and deploy the appropriate sensors to objects within the environment in order to effectively monitor and capture a user's behavior along with the state change of the environment; (ii) to collect, store and process information and, finally, (iii) to infer/classify activities from sensor data through the use of computational activity models.

There are two categories for activity recognition: Data-Driven Approaches (DDA) and Knowledge-Driven Approaches (KDA). The former, DDA, are based on machine learning techniques in which a preexistent dataset of user behaviors is required. A training process is carried out, usually, to build an activity model which is followed by a testing processes to evaluate the generalization of the model in classifying unseen activities [6,9]. Regarding KDA, an activity model is built through the incorporation of rich prior knowledge gleaned from the application domain, using knowledge engineering and knowledge management techniques [2].

In these approaches, activity models must be trained and tested in order to check their performances. Nevertheless, it is very difficult, expensive and complex to test the performance of them in real environments. Thus, it is a challenge to check the performance of activity models by means of datasets in a more economical and easy way. There are many datasets proposed in the literature in which three classes of activities can distinguish:

- Single activity [4,16] is an activity which has been carried out completely before to start the performance of a new one.
- Interleave activity [12] is an activity which is carried out while another activity is performing at the same time.
- Multioccupancy [13] is a class of activity in which some people are performing their activities simultaneously.

The division in these classes allows researchers to train and test their activity models only in one kind of them, not allowing to combine several classes. Nevertheless, there are some cases that can be interesting to train and test activity models the combination of some activity classes. This is difficult because it is not easy to find datasets with these characteristics.

In this contribution, we present the design and the development of a complete ADL dataset which is partitioned in several subsets depending on the activity class. Thus, the complete dataset has 3 activity classes (single, interleave and multioccupancy) that can be combined in 7 subsets or partitions.

The remainder of the work is structured as follows: in Sect. 2, previous works of datasets that have been developed are reviewed. Section 3 describes, first, the smart environment where the dataset has been developed and the considered ADLs. Then, the generated dataset is described, focusing on their partitions. Finally, in Sect. 4 conclusions and future works are drawn.

2 Related Works

Some datasets have been generated in smart homes or smart environments in order to overcome the difficulties to test and to train activity models in real smart environments. In the literature, some contributions have been presented about the generation of datasets; some of the most important ones are reviewed in this Section.

There are repositories which contain several AmI datasets. One the most well-known repositories is CASAS[1] [4], where there are different datasets developed by Cook et al. and some links to other popular AmI ones. Bellow, some of these datasets are reviewed:

- Tapia et al. [15]. The single activity class dataset is composed of the data collected from 2 apartments inhabited by a man and a woman respectively. The sensor network has 77 sensors in the first apartment and 84 sensors in the second one, gathering data during 14 days. The activities performed during the experiment were labeled by the inhabitants regarding 35 preestablish ADLs like *toileting, grooming, preparing lunch* or *preparing a beverage*.
- Roggen et al. [11] deployed a sensor network with 72 sensors to gather data of 12 common ADLs, performed daily morning, for example, *prepare breakfast*. The single activity class dataset was developed by one person during 25 hours.
- Van Kasteren et. al [16] generated a single activity class dataset locating 14 sensors in an apartment which had an inhabitant during 28 days. The dataset contains 7 ADLs like *prepare breakfast, get drink* or *prepare dinner*.
- Among all of these authors, it is noteworthy the work developed by Cook et al. [4] in the ADL datasets generation. These datasets contain the data that a smart apartment sensor network produces when the inhabitants perform their ADLs. Some of the most representative datasets are the following ones:
  - Cook and Schmitter-Edgecombe [5] developed a single activity class dataset, which was performed by the occupant of an apartment. The sensor network was composed by 37 sensors. In this dataset, there are performed 5 ADLs.
  - Singla et al. [12] developed an interleave activity class dataset. The data was collected from the sensor network, which was composed by 78 sensors, when the participants were performing 8 different ADLs.
  - A multioccupancy class dataset was developed during four months in a student apartment [13]. The data represents two participants in the apartment at the same time performing fifteen ADL activities. The sensor network was composed by 78 sensors, which are the responsible to gather the data.

[1] http://ailab.wsu.edu/casas/datasets/ (last checked on August 27, 2015).

These datasets present the difficulty of combining them in mixed datasets because the structure and the syntax are not the same. Thus, in this work we present the design and development of a complete dataset which has been divided in different partitions which combine activity classes to be used according to the researchers' needs.

3 Dataset Generation

This Section describe the design and development of the partitioned dataset in a smart environment in which daily living activities (single, interleave and multioccupancy) have been carried out. To do so, first, the smart environment, where the dataset has been developed is described and the considered ADLs are indicated. Then, the dataset description is provided, focusing on their partitions.

3.1 Smart Environment and ADLs

Regarding the smart environment, the dataset has been generated in the smartlab of the University of Ulster that consists of a kitchen and a living room, which is illustrated in Fig. 1.

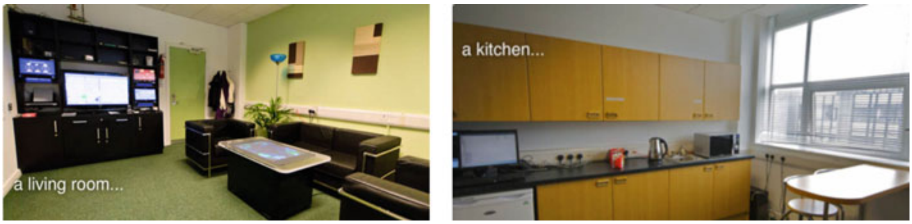

Fig. 1. Smartlab kitchen and the living room

The ADLs that are considered in the generated dataset are the most popular activities that are presented in the literature in such places. Therefore, the activities that we have considered are: *prepare drink*, *call by phone*, *prepare snack*, *watch TV* and *wash dishes*. Nonetheless, it is usual to prepare drinks and snacks in different ways depending on the kind of drink and snack. Thus, we consider that *prepare drink* can be *drink a glass of water*, *prepare a tea with the kettle*, *prepare a hot chocolate in the microwave* or *drink a glass of milk*. On the other hand, the kind of *prepare snack* that are considered in this dataset are *prepare hot snack in the microwave* and *prepare cold snack*. So, the 9 different ADLs, which have been carried out in the smartlab, are described regarding their performance, as follow:

1. **Drink a glass of water**. The inhabitant goes to the kitchen, takes a glass from the glasses cupboard, opens the sink tab and fills the glass with water.

2. **Prepare a tea with the kettle**. The dweller goes to the kitchen, fills the kettle with water and switches it on. After that, he takes the tea bag from the groceries cupboard, the cup from the glasses cupboard and a spoon from the cutlery cupboard. Finally, he puts the tea bag into the cup and pours the hot water from the kettle.
3. **Prepare a hot chocolate in the microwave**. The occupant goes to the kitchen, takes the milk from the fridge and a cup from the glasses cupboard. Later, she pours the milk into the cup and heats it in the microwave. In the meantime, she takes the chocolate from the groceries cupboard and a spoon from the cutlery cupboard. Finally, she puts some chocolate into the cup and saves the resting milk and chocolate in their locations.
4. **Drink a glass of milk**. The resident goes to the kitchen, takes the milk from the fridge and a cup from the glasses cupboard. Later, she pours the milk. Finally, she saves the milk in its location.
5. **Call by phone**. The occupant goes to the living room, picks up the phone and dial. When the call is finished, he hangs up the phone.
6. **Prepare hot snack in the microwave**. The dweller goes to the kitchen, takes a plate from the plates cupboard and the food from the fridge. Afterwards, he heats the food in the microwave and takes the cutlery from the cutlery cupboard.
7. **Prepare cold snack**. The inhabitant goes to the kitchen takes a plate from the plates cupboard and the food from the fridge and the groceries cupboard. After the meal preparation, he saves the remaining food in their locations.
8. **Watch TV**. The resident goes to the living room and switches the TV on. When she finishes, she switches it off.
9. **Washing dishes in the sink**. The occupant goes to the kitchen, open the sink tab and washes all the plates, cutlery, glasses and cups that are dirty. When he has finished, he saves all the tableware in its location.

The sensors' network is composed by contact and pressure sensors. Specifically, we have used Tynetec[2] pressure and contact sensors (see Fig. 2). These sensors use transmit on 169 MHz frequency, and the data is handled by a receiver that is connected to a computer, which saves all the data in an SQL database.

Fig. 2. Pressure and contact sensor

Considering the previous ADLs, several sensors have been located in different places of the smartlab, in order to gather the data from the interactions

[2] www.tynetec.co.uk (last checked on August 27, 2015).

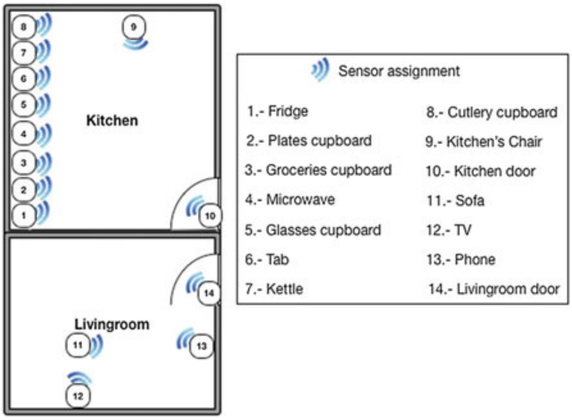

Fig. 3. Sensors' location in the smartlab

between the inhabitants and the environment when they develop the defined ADLs. Therefore, contact sensors have been located on the fridge, the plates cupboard, the groceries cupboard, the microwave, the glasses cupboard, the tab, the kettle, the cutlery cupboard, the kitchen's chair, the kitchen door, the sofa, the TV, the phone and the living room door (see Fig. 3). Furthermore, a couple of pressure sensors has been located on the sofa and the kitchen's chair.

3.2 Dataset Description

The dataset is a CSV file which gather the interactions in the smart environment between the inhabitants and the sensors' network.

Each dataset line is an event that is a state variation in a sensor of the smart environment. Regarding each event, the following information is saved: date, time, sensor identifier and sensor value. Table 1 shows the possible values of each sensor, once the raw sensor data has been processed.

The dataset has been labeled manually with the ADLs that are carried out in order to know in which event an activity begins or ends. The label which marks the beginning is $Begin\_[IdActivity]$ and the mark which means the end of the activity is $End\_[IdActivity]$. Furthermore, in the multioccupancy dataset, it is necessary to indicate the inhabitants labels, $P\_[IdOccupant]$.

Depending on each person and his/her culture and experience, it is usual that the same ADL be performed in a different way. Thus, depending on the ADL performance the sensors' interactions can be different. Table 2 shows the sensors that interact during a particular ADL. There are sensors that are compulsory (Y), optional (O) and unnecessary (N). For example, in the *watch TV* ADL, there are compulsory that the *living room door* and the *TV* sensors have been activated, being optional the *sofa* sensor activation.

Regarding the dataset performance, it has been developed by a man during one week. He interacts with the smart environment, carrying out his daily activ-

Table 1. Value of the set of sensors

IdSensor	Name	Values
D01	Kitchen door sensor	(OPEN / CLOSE)
D02	Living room door senor	(OPEN / CLOSE)
D03	Cutlery cupboard sensor	(OPEN / CLOSE)
D04	Dishes cupboard sensor	(OPEN / CLOSE)
D05	Glasses and cups cupboard sensor	(OPEN / CLOSE)
D06	Pantry cupboard sensor	(OPEN / CLOSE)
D07	Microwave sensor	(OPEN / CLOSE)
D08	Frigde door sensor	(OPEN / CLOSE)
M01	Chair sensor	(AUSENCE / PRESENCE)
M02	Sofa sensor	(AUSENCE / PRESENCE)
TV	Television sensor	(ON / OFF)
PH	Phone sensor	(PICK_UP / HANG_UP)
WT1	Water sensor	(OPEN / CLOSE)
KT	Kettle sensor	(ABSENT / PRESENT)

Table 2. Sensors which interact in each activity

Id	Activity	D01	D02	D03	D04	D05	D06	D07	D08	M01	M02	TV	PH	WT1	KT
1	Drink a glass of water	Y	N	N	N	Y	N	N	N	O	N	N	N	Y	N
2	Prepare a tea	Y	N	O	N	Y	N	N	N	O	N	N	N	O	Y
3	Prepare a hot chocolate	Y	N	O	N	Y	Y	Y	Y	O	N	N	N	N	N
4	Drink a glass milk	Y	N	N	N	Y	N	N	Y	O	N	N	N	N	N
5	Call by phone	N	Y	N	N	N	N	N	N	N	O	N	Y	N	N
6	Prepare hot snack	Y	N	O	Y	N	N	Y	Y	O	N	N	N	N	N
7	Prepare cold snack	Y	N	O	Y	N	Y	N	Y	O	N	N	N	N	N
8	Watch TV	N	Y	N	N	N	N	N	N	N	O	Y	N	N	N
9	Washing dishes	Y	N	O	O	O	N	N	N	O	N	N	N	Y	N

ity routines in single and interleave ways. Besides, sometimes a woman interacts with the smart environment at the same time that the man, carrying out some activities simultaneously (multioccupancy). In Fig. 4 is illustrated some events about the three different classes of activities. Then some details about each activity classes are provided.

- **Single activity class**. There are 364 single ADLs, being *drink a glass of water* and *watch TV* the most frequent activities and *prepare hot chocolate* the fewer one. Figure 5 depicts the single activity class distribution.
- **Interleave activity class**. There are 24 interleave ADLs. Some activities that have been performed in a interleave manner are *call by phone* while the occupant is *preparing a hot chocolate* or *drink a glass of water* while the inner is *preparing hot snack*. The interleave activity class distribution is showed in Fig. 6.

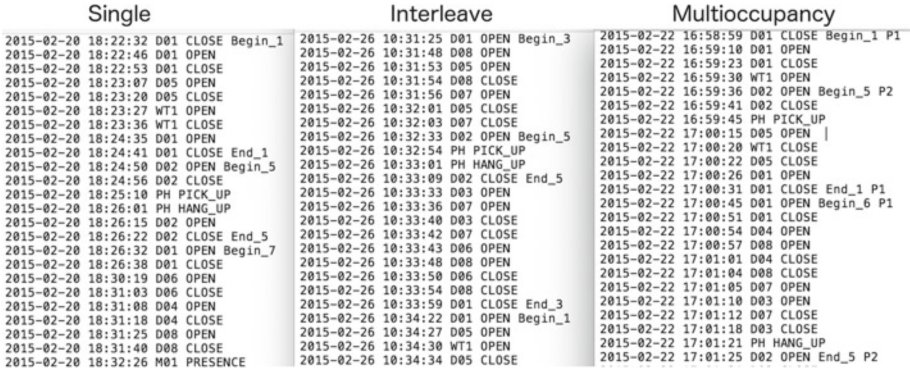

Fig. 4. Single, Interleave and Multioccupancy dataset classes

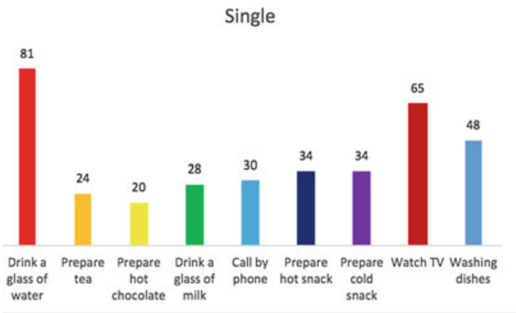

Fig. 5. Single activity distribution

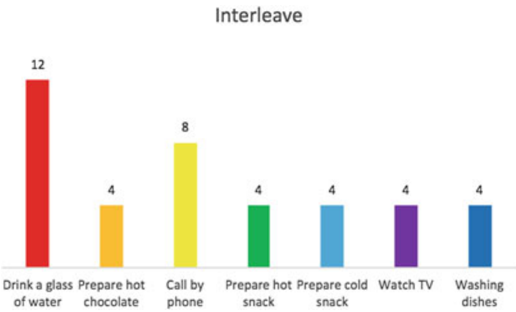

Fig. 6. Interleave activity distribution

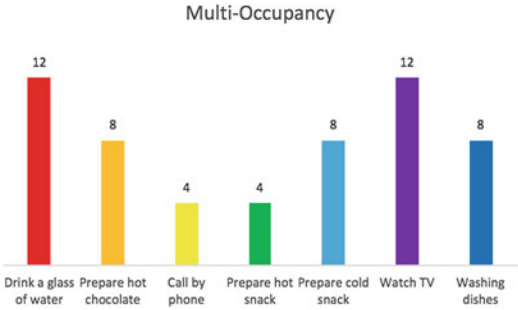

Fig. 7. Multioccupancy distribution

- **Multioccupancy class**. There are 56 ADLs with 36 ones performed in a multioccupancy way. For example, when someone is *drinking a glass of water*, the other one is *calling by phone*. Another example is when someone is *watching TV*, the other one is *preparing a cold snack*. Figure 7 depicts the multioccupancy data distribution.

In order to adapt the dataset to the research needs, the complete dataset has been split into different partitions. So, 3 activity classes (single, interleave and multioccupancy) have been considered and the 7 possible combinations of them have been made.

Furthermore, due to the excellent results obtained by Jurek et al. [7] using a numeric representation as well as binary representation in a dataset of a smart environment with single activities, our single dataset has been converted into binary representation and numeric representation.

The main idea in both representations is that each instance of the dataset is a vector with $n+1$ components, the first n components corresponds to the value of the n sensors involved in the smart environment and the last component, $n+1$ corresponds to the activity performed. In the numeric representation (see Fig. 8), the value of sensor indicates how many times there are a state change of such sensor in the activity. In the binary representation (see Fig. 9), the value of sensor is 1, if there is a state change in the sensor during activity, and 0, if no change.

The dataset files are accessible[3]. Each link has associated a .zip file which contains a CSV and a RTF file. The CSV file has the set of events, where the *begin* and the *end* of each activity is labeled, and the RTF file includes the dataset description that contains relevant information. Moreover, the *Sensors' Layout* link, is a PNG file that depicts the smart environment sensors' distribution.

So, in the previous URL, the generated dataset is available in the seven combinations of classes and, furthermore, the dataset with single activities is available with the binary representation and numeric representation, following the scheme of the web page is presented:

[3] http://ceatic.ujaen.es/smartlab (last checked on August 27, 2015).

D01	D02	D03	D04	D05	D06	D07	D08	KT	M01	M02	PH	TV	WT1	Activity
5	0	0	0	2	0	0	0	0	0	0	0	0	2	1
0	4	0	0	0	0	0	0	0	0	0	2	0	0	5
4	0	0	2	0	2	0	2	0	2	0	0	0	0	7
0	4	0	0	0	0	0	0	0	0	3	0	2	0	8
4	0	0	2	2	0	0	0	0	0	0	0	0	1	9
4	0	0	0	2	2	0	0	2	0	0	0	0	4	2
0	4	0	0	0	0	0	0	0	0	0	2	0	0	5
4	0	0	2	0	2	0	2	0	2	0	0	0	0	7
0	4	0	0	0	0	0	0	0	0	3	0	2	0	8

Fig. 8. Numeric representation of the single dataset

D01	D02	D03	D04	D05	D06	D07	D08	KT	M01	M02	PH	TV	WT1	Activity
1	0	0	0	1	0	0	0	0	0	0	0	0	1	1
0	1	0	0	0	0	0	0	0	0	0	1	0	0	5
1	0	0	1	0	1	0	1	0	1	0	0	0	0	7
0	1	0	0	0	0	0	0	0	0	1	0	1	0	8
1	0	0	1	1	0	0	0	0	0	0	0	0	1	9
1	0	0	0	1	1	0	0	1	0	0	0	0	1	2
0	1	0	0	0	0	0	0	0	0	0	1	0	0	5
1	0	0	1	0	1	0	1	0	1	0	0	0	0	7
0	1	0	0	0	0	0	0	0	0	1	0	1	0	8

Fig. 9. Binary representation of the single dataset

- Sensors' Layout
  - *SensorsLayout.png*

1. Single
   - *Single.zip*
     - *DataSingle.csv*
     - *ReadmeSingle.rtf*
   - *SingleNumeric.csv*
   - *SingleBinary.csv*
2. Interleave
   - *Interleave.zip*
     - *DataInterleave.csv*
     - *ReadmeInterleave.rtf*
3. Multioccupancy
   - *Multioccupancy.zip*
     - *DataMultioccupancy.csv*
     - *ReadMultioccupancy.rtf*
4. Single and Interleave
   - *SingleInterleave.zip*
     - *DataSingleInterleave.csv*
     - *ReadmeSingleInterleave.rtf*
5. Single and Multioccupancy
   - *SingleMultioccupancy.zip*
     - *DataSingleMultioccupancy.csv*
     - *ReadmeSingleMultioccupancy.rtf*

6. Interleave and Multioccupancy
   – *Interleave and Multioccupancy.zip*
     - *DataInterleaveMultioccupancy.csv*
     - *ReadmeInterleaveMultioccupancy.rtf*
7. Single, Interleave and Multioccupancy
   – *SingleInterleaveMulti.zip*
     - *DataSingleInterleaveMulti.csv*
     - *ReadmeSingleInterleaveMulti.rtf*

4 Conclusions and Future Works

This contribution has presented the generation of a complete smart home dataset which gather the data from the sensors' network when some of the most common daily living activities are performing. The dataset is divided in 7 partitions that combine three different classes of ADLs: single activity, interleave activities and multioccupancy. These classes can be used individually or together, depending on the researchers' needs. Furthermore, the numeric and binary representation has been used for the single partition. Regarding future works, we are focused on incrementing the number of different dataset activities and the frequency of them. Furthermore, we will address the issue of how the sensor data can be recorded automatically without human interaction using NFC (Near Field Comunication) technology. Finally, we are focused on using the dataset with ambient intelligence algorithms.

Acknowledgments. This contribution was supported by Research Projects TIN-2012-31263, CEATIC-2013-001, UJA2014/06/14 and by the Doctoral School of the University of Jaén. Invest Northern Ireland is acknowledge for partially supporting this project under the Competence Centre Program Grant RD0513853 - Connected Health Innovation Centre.

References

1. Chen, L., Hoey, J., Nugent, C.D., Cook, D.J., Yu, Z.: Sensor-based activity recognition. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **42**(6), 790–808 (2012)
2. Chen, L., Nugent, C.: Ontology-based activity recognition in intelligent pervasive environments. Int. J. Web Inf. Syst. **5**(4), 410–430 (2009)
3. Chen, L., Nugent, C.D., Wang, H.: A knowledge-driven approach to activity recognition in smart homes. IEEE Trans. Knowl. Data Eng. **24**(6), 961–974 (2012)
4. Cook, D., Schmitter-Edgecombe, M., Crandall, A., Sanders, C., Thomas, B.: Collecting and disseminating smart home sensor data in the casas project. In: Proceedings of the CHI Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research, pp. 1–7 (2009)
5. Cook, D.J., Schmitter-Edgecombe, M., et al.: Assessing the quality of activities in a smart environment. Methods Inf. Med. **48**(5), 480–485 (2009)
6. Gu, T., Wang, L., Wu, Z., Tao, X., Lu, J.: A pattern mining approach to sensor-based human activity recognition. IEEE Trans. Knowl. Data Eng. **23**(9), 1359–1372 (2011)

7. Jurek, A., Nugent, C., Bi, Y., Wu, S.: Clustering-based ensemble learning for activity recognition in smart homes. Sensors **14**(7), 12285–12304 (2014)
8. Lepri, B., Mana, N., Cappelletti, A., Pianesi, F., Zancanaro, M.: What is happening now? detection of activities of daily living from simple visual features. Pers. Ubiquit. Comput. **14**(8), 749–766 (2010)
9. Li, C., Lin, M., Yang, L.T., Ding, C.: Integrating the enriched feature with machine learning algorithms for human movement and fall detection. J. Supercomput. **67**(3), 854–865 (2014)
10. Moshtaghi, M., Zukerman, I., Russell, R.: Statistical models for unobtrusively detecting abnormal periods of inactivity in older adults. User Model. User-Adap. Inter. **25**(3), 231–265 (2015)
11. Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Forster, K., Troster, G., Lukowicz, P., Bannach, D., Pirkl, G., Ferscha, A., et al.: Collecting complex activity datasets in highly rich networked sensor environments. In: 2010 Seventh International Conference on Networked Sensing Systems (INSS), pp. 233–240. IEEE (2010)
12. Singla, G., Cook, D.J., Schmitter-Edgecombe, M.: Tracking activities in complex settings using smart environment technologies. Int. J. Biosci. Psychiatry Technol. (IJBSPT) **1**(1), 25 (2009)
13. Singla, G., Cook, D.J., Schmitter-Edgecombe, M.: Recognizing independent and joint activities among multiple residents in smart environments. J. Ambient Intell. Humanized Comput. **1**(1), 57–63 (2010)
14. Streitz, N., Nixon, P.: The disappearing computer. Commun. ACM **48**(3), 32–35 (2005)
15. Tapia, E.M., Intille, S.S., Larson, K.: Activity recognition in the home using simple and ubiquitous sensors. In: Ferscha, A., Mattern, F. (eds.) PERVASIVE 2004. LNCS, vol. 3001, pp. 158–175. Springer, Heidelberg (2004)
16. Van Kasteren, T., Noulas, A., Englebienne, G., Kröse, B.: Accurate activity recognition in a home setting. In: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 1–9. ACM (2008)