Let the Objects Tell What You are Doing

Gabriele Civitarese
University of Milan
Via Comelico 39, Milano
gabriele.civitarese@unimi.it

Claudio Bettini
University of Milan
Via Comelico 39, Milano
claudio.bettini@unimi.it

Stefano Belfiore
University of Milan
Via Comelico 39, Milano
stefano.belfiore@studenti.unimi.it

Ubicomp/ISWC'16 Adjunct , September 12-16, 2016, Heidelberg, Germany

DOI: http://dx.doi.org/10.1145/2968219.2968285

Abstract

Recognition of activities of daily living (ADLs) performed in smart homes proved to be very effective when the interaction of the inhabitant with household items is considered. Analyzing how objects are manipulated can be particularly useful, in combination with other sensor data, to detect anomalies in performing ADLs, and hence to support early diagnosis of cognitive impairments for elderly people. Recent improvements in sensing technologies can overcome several limitations of the existing techniques to detect object manipulations, often based on RFID, wearable sensors and/or computer vision methods. In this work we propose an unobtrusive solution which shifts all the monitoring burden at the objects side. In particular, we investigate the effectiveness of using tiny BLE beacons equipped with accelerometer and temperature sensors attached to everyday objects. We adopt statistical methods to analyze in real-time the accelerometer data coming from the objects, with the purpose of detecting specific manipulations performed by seniors in their homes. We describe our technique and we present the preliminary results obtained by evaluating the method on a real dataset. The results indicate the potential utility of the method in enriching ADLs and abnormal behaviors recognition systems, by providing detailed information about object manipulations.

ACM Classification Keywords
I.5.1 [Pattern Recognition: Models: Statistical]

Author Keywords
Activity recognition; smart homes; sensing

Introduction
The recent improvements in sensor technologies are having a deep impact on a long lasting challenge in ubiquitous computing and ambient intelligence, namely the recognition of human activities [4] and in particular activities of daily living (ADL).

The advantages of having sensors on everyday artifacts for ADL recognition have been identified long ago, exploring solutions mainly based on accelerometers and RFID [2, 11], however the technology has not been sufficiently reliable and cost-effective for a wide scale deployment. A common argument against using sensor-augmented objects as opposed to wearables for ADL recognition, in addition to technological issues, has been the difficulty in identifying the subject that is performing the activity in case of multiple inhabitants of the same space [1]. On this respect, there has been some progress on this issue both on the technological side (miniaturization of identifying beacons) and on wearable-free solutions based on data analysis [7]. An other approach to recognize specific object manipulations without neither sensors on objects nor wearables takes advantage of audio and/or video recording [13], but this solution is often perceived as too obtrusive.

Our investigation is driven by a specific application domain: the recognition of fine grained anomalies in performing instrumented activities of daily living by elders at risk of cognitive impairment [12]. Clinicians need to identify manipulations of specific objects in a home environment including omissions, substitutions and improper manipulations. For example, these include reaching and opening a wrong medicine box, using the wrong tool to perform an action or unnecessarily repeating a given manipulation. The system described in this paper is not intended by itself to support early diagnosis based on improper object manipulations. However, reliable object manipulation monitoring is an essential subsystem of a more complex monitoring environment. In particular, what we describe is intended to substitute the RFID-based subsystem used in [12] to monitor the use of items in preparing and consuming meals as well as taking medicines. As shown in Figure 1, in order to recognise anomalies in performing these high level activities other sensor subsystems are used, including sensors revealing presence, pressure, temperature, power consumption and more.

In our experience on deployments in the real homes of the elderly for continuous monitoring, solutions based on wearables are critical: there is no guarantee that wristbands or pendants are constantly worn, not to mention smartphone or RFID readers that have been proposed for the advantage of identifying the specific manipulated object. There are also indications of a general adversity or disaffection of users to wearables targeted to healthcare related applications [5]. Similarly, cameras and microphones are sometimes tolerated in retirement residences, but much less in private homes.

Our major contribution are experimental results on the effectiveness of unobtrusive object manipulation recognition, using current commercial low cost and low energy consumption multi-sensor devices that can be attached to everyday objects. A closely related work is [10], which uses acceleration data acquired from sensors on items to evaluate surgeons' skill in manipulating precision tools. With respect to that work we monitor manipulations relevant to our application domain, which are more coarse grained and of a different nature. We collected a dataset of more than two thousands labeled manipulations, and we report encouraging preliminary results on their recognition through machine learning techniques applied on accelerometer data collected from the objects. We believe that our study contributes to the design of a sensing subsystem that could be effectively integrated in the smarthome environments used in several previous works on monitoring complex activities at home [4, 9, 6], independently from the algorithmic method being used, since object manipulations may be considered as simple events.

Modeling manipulations

We define as *object manipulation* the interaction of an individual with an object of interest with the objective of achieving some task within the execution of a particular ADL. More formally, we define a manipulation instance as $m = \langle o, M, t_s, t_e \rangle$, where o is the object manipulated, M is the manipulation type, t_s and t_e are respectively the start and end time of the manipulation. Given i_A an instance of an ADL A, we say that a manipulation $m \in i_A$ if m is performed during i_A.

Example 1 *Considering as object of interest a glass, some possible types of manipulation of that object could be: using the glass to drink while eating a meal, moving the glass on the table while preparing the table, emptying the glass in the sink, inserting the glass in the dishwasher, and so on. A manipulation instance could be:* $m = \langle$glass*,* drinking*, 12:45:32, 12:45:45*$\rangle$ *where* $m \in i_{eating}$ *(a manipulation which consists in using the glass to drink during the consumption of a meal).*

We point out that not every type of object manipulation is interesting for monitoring ADL execution, hence we divide the manipulation types in two categories: *relevant* ad *irrelevant*. We consider a manipulation *relevant* if the task that is achieved by performing the manipulation is crucial to monitor a particular ADL; *irrelevant* otherwise. Of course, classification of manipulation types in *relevant* and *irrelevant* has to be decided accurately by domain experts. Manipulations considered as *relevant* are further classified in specific sub-classes.

Example 2 *Suppose that we're interested in monitoring the activity of taking medicines. In this scenario, we consider a manipulation relevant if a medicine package is extracted from its repository, or if it is opened; while it is considered irrelevant if a medicine package is just displaced inside the repository while searching.*

The technique

In this section, we illustrate our technique to analyze the data coming from accelerometer positioned on objects, in order to recognize specific manipulations.

Recognition Framework

The system is considered as part of a smart home environment instrumented with several environmental sensors. The general architecture is shown in Figure 1. Each object of interest has attached a wireless device which incorporates a 3-axis accelerometer sensor. Each device communicates periodically the raw sensor data to the *Smart-object data processing* module, along with the device's unique identifier. This module is in charge of: a) segmenting the accelerometer data in order to identify the manipulation occurrences, b) extracting several features and c) applying machine learning techniques in order to recognize the specific manipulation performed. Since each type of object has an associated set of specific manipulations (e.g., a

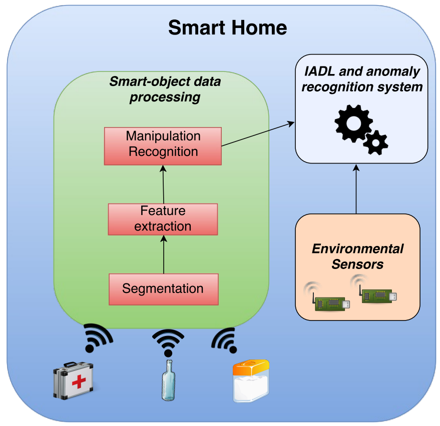

Figure 1: General architecture

bottle of water is used to pour/drink water, differently from a medicine box that is used to extract pills), we built a specialised classifier for each object type. Of course, this does not mean using a different classifier for each object: for instance, a bottle of water and a milk box can be manipulated similarly and a single classifier is in charge of recognizing the manipulations of both objects. Detected manipulations, along with measurements acquired from smart-home environmental sensors, are transmitted to a system which is in charge of recognizing ADLs performed by the monitored subject and the possible abnormal behaviors.

Segmentation and feature extraction
We pre-process data transmitted from the objects in order to identify the manipulation occurrences. To do this, we analyze 3-axis accelerometer data in order to detect whether an object is in motion. This is done by using a straightforward threshold based method on accelerometer data which detects when the object starts and stop moving. Each manipulation occurrence $occ_i = \langle o, t_s, t_e, \vec{x}, \vec{y}, \vec{z} \rangle$ is represented by: the object o manipulated, the start time t_s (i.e. the time instant where the object started moving), the end time t_e (i.e. the time instant where the object stopped moving) and the accelerometer data on the three axis. The output of segmentation module is a set of n manipulation occurrences $O = \{occ_1, occ_2, \ldots, occ_n\}$.

From each manipulation occurrence, we build a feature vector which comprises more than 40 different features regarding statistics on accelerometer data and the duration of the manipulation.

Manipulation recognition
The next step is to infer, for each feature vector, the specific manipulation performed with the related object. As previously described, for each type of object we're interested in distinguishing between *irrelevant* manipulations and a set of specific *relevant* manipulations. Since we're not interested in detecting the fine-grained types of *irrelevant* manipulations, they're grouped together into a single class called *Irrelevant*. We experienced that tree-based discriminative learning models proved to be effective to solve this class of problems. In particular, we adopt a supervised approach, using state-of-the-art classifiers like *Random Forest* [3] and *AdaBoost* [8] depending on the specific object.

We adopted two different classification approaches:

- Direct classification
- Multi-layer classification

In the following we describe these techniques.

Direct classification
Our first straightforward approach consists in directly distinguishing the fine-grained manipulations using a multi-class

classifier. This method is shown in Figure 2. Hence, de-

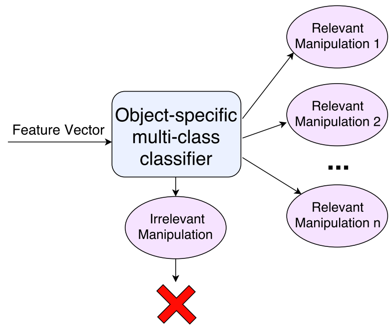

Figure 2: Direct classification schema for a specific object

pending on the type of object from which the manipulation comes, a specific classifier is used. Every single classifier is trained with a set of relevant manipulations and irrelevant manipulations of the specific object type.

Multi-layer classification
With the objective of improving the above mentioned method, we also propose a different approach, which is represented in Figure 3. Instead of directly detecting the manipulation type from the feature vector, we use two layers. In the first layer a binary classifier is in charge of distinguishing, for a specific object, *relevant* manipulations from the *irrelevant* ones. This classifier is trained with relevant manipulations (all grouped together in the same class) and irrelevant manipulations of the specific object. Only the manipulations which are classified as *relevant* are forwarded to the second layer, while the others are discarded. In the second layer, a multi-class classifier is in charge of recognizing the specific relevant manipulation performed on the object. Hence, this classifier is trained only with the corresponding relevant manipulations.

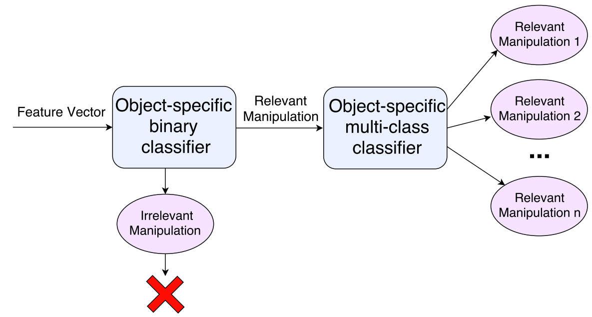

Figure 3: Multi-layer classification schema for a specific object

Experimental evaluation
In this section we describe our experimental setup, how we acquired a dataset of manipulations, and finally we present our preliminary results.

Experimental setup
Driven by requirements from clinicians, we currently focus on fine-grain monitoring of three complex activities: *preparing a meal*, *consuming a meal* and *taking medicines*. While the final deployment of stable versions of the system is in real homes, our experimental activity is conducted in a smart-room lab. Activity recognition is performed by processing data coming from a wide variety of environmental sensors, including pressure pads, temperature sensors, power meters, magnetic switches, presence sensors and more. The experiment reported in this paper is intended to verify the viability of substituting our RFID based solution for recognizing manipulations of specific items. For this purpose we selected specific objects: a) medicine boxes as they have a key role in monitoring adherence to prescription and their improper manipulation can also be a useful indicator, b) a liquid bottle as it is an example of an item used in meal consumption, may have to be refrigerated,

(a) Liquid bottle

(b) Medicine boxes

(c) Knife

Figure 4: The monitored objects

and may also play a role in monitoring water consumption, and c) a kitchen tool, a knife in particular, as a tool being used both in meal preparation, and in meal consumption. These objects are shown in Figure 4 with their sensing device attached.

The sensing devices

In order to monitor objects manipulation, we take advantage of current off-the-shelf devices: Estimote's Stickers. A sticker is a packaged PCB with a battery-powered ARM CPU equipped with 3-axis accelerometer, temperature sensor, and a Bluetooth Smart radio able to periodically broadcast its sensed data in a short range (a few meters). Their tiny packaging makes it easy to attach them on objects as shown in Figure 4. Each sticker can be easily distinguished by a unique identifier which is particularly useful to improve manipulation detection by exactly knowing which kind of object is manipulated. Estimote Stickers adopt a proprietary communication protocol called Nearables; Table 1 reports the data frame of this protocol. In our setup, each sticker broadcasts a packet every 100 milliseconds while it is moving; every 200 milliseconds otherwise. A BLE scanner is in charge of collecting the data coming from each sticker.

Sensor data analysis

We perform data acquisition by scanning the BLE signal through a mobile device. In order to perform segmentation, we exploit the value of the *Motion* field transmitted in every packet by the stickers. This field is set to *true* when the sticker is in motion. Our experiments revealed that this value provides sufficient accuracy for determining begin and end of our manipulations. Hence, for a specific sticker we consider all the consecutive data packets with the *Motion* field set to *true* as part of the same manipulation occurrence.

A labeled dataset is used to construct the predictive model.

Field	Description
Identifier	Unique identifier
Motion	Whether the sticker is moving (boolean)
xAcceleration	X-Axis acceleration
yAcceleration	Y-Axis acceleration
zAcceleration	Z-Axis acceleration
Temperature	Stickers temperature value (in Celsius)
Orientation	Physical orientation of the sticker
RSSI	Signal strength
Power	Signal strength at 0 meters
Battery Level	Sticker's battery level

Table 1: Nearables data frame

We performed experiments with different type of models for each of our considered objects, and we selected Random Forest for the liquid bottle and the medicine boxes, and AdaBoost for the knife manipulations.

Segmentation, feature extraction and classification are performed in real time on the mobile platform. The output serialized in JSON format is sent to a REST server for integration with events detected by processing data coming from other sensors as illustrated in Figure 1.

Dataset collection

Since our recognition technique is based on supervised machine learning, a critical task is the acquisition of a sufficiently large and significant dataset of object manipulations. The dataset must also be annotated with the ground truth related to each manipulation. In order to facilitate this task, we developed a mobile application. The application starts with a simple screen consisting in only one button. When that button is clicked, the bluetooth scanner starts acquiring Nearables data packets, which are internally stored. After a

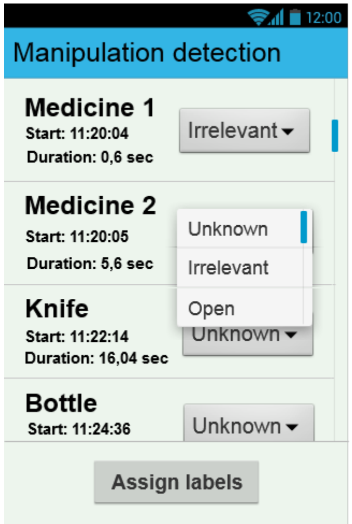

Figure 5: Application layout

few manipulations we conclude the experiment, and the app performs segmentation, it creates a set of three axis acceleration data for each manipulation and then allows the user to label each one with the ground truth (Figure 5).

As already mentioned, in this experiment, we focus on manipulations of a liquid bottle, a medicine box, and a knife. For the purpose of this first assessment of our system we collected manipulations performed by six different adults without physical impairments. They executed those manipulations spontaneously within realistic scenarios of activities of daily living executed in a smart room lab (e.g. cooking, taking medicines, . . .). The total number of manipulations is 2058, with 887 manipulations involving the liquid bottle, 656 the medicine boxes and 515 the knife. Out of the total, 1365 manipulations are considered relevant, while the rest are considered irrelevant. This distinction is clearly application dependent, and in our case it has been driven by the scenarios of our e-health domain and by the interest in specific manipulations by the clinicians.

It is important to consider that the specific way in which we perform segmentation can lead to group more than one manipulation into a single one; for example, if a subject extracts the water bottle from the fridge and pours the water in a glass as a single action without interruption, the motion value of Nearable remains true and the whole action will be segmented as a single manipulation. On the contrary, if the bottle is moved from the fridge to the table, and then used to fill a glass, the system will identify two manipulations. We considered alternative segmentation methods, but actually observed that the presence of these ’composed’ manipulations is sometimes a benefit for our specific application, considering the final recognition accuracy.

Liquid bottle’s manipulations
The total number of manipulations of the bottle that we acquired is 887. We consider 500 of them as relevant because their detection is useful to monitor the activity of meal consumption or even just *drinking* (e.g., "extract from the fridge" or "pour water"). Table 2 shows how we classify manipulations of the bottle.

Class	**Include**	**Description**
Irrelevant	Minor displacement	Bottle is displaced in the same place
	Irrelevant	Bottle is moved, but not by a person (e.g., movements of the fridge)
	Displacing in the fridge	Bottle is displaced inside the fridge
	Opening/closing fridge door	When the bottle is in the fridge door and it is opened, bottle moves
Relevant displacement	Displaced	Bottle is displaced from a place to another which is not a fridge
	Inserted	Bottle is displaced from a place to the fridge
	Extracted	Bottle is displaced from the fridge to a place
Drinking/Pouring	Drink	Bottle is taken from a place and is brought to lips and tilted
	Pour	Bottle is taken from a place and liquid is poured in a glass

Table 2: Liquid bottle’s manipulations

Medicine box's manipulations
The total number of these manipulations is 656. We consider 474 of them as relevant, because their detection is useful to monitor the activities "taking medicine" (e.g. "extract from the repository" or "open medicine box"). Table 3 shows how we classify manipulations of medicine boxes. Note that distinguishing manipulations like "displacing the medicine M box" and "accessing the content of medicine M box" is very important in our domain, since the first if not followed by the second may be an indication that the patient prepared the medicine but in the end forgot to take it.

Knife's manipulations
The total number of manipulations involving the knife is 515. We consider 391 of them as relevant because their detection is useful to monitor the activities "preparing meal" (e.g. "extract from the drawer" or "cut something"). Table 4 shows how we classify manipulations involving the knife.

Results
Table 5 summarizes our results on the recognition of object manipulations. We use a 10-folds cross-validation method. The table shows both the results using the direct classification approach and the ones using the layered approach. Despite several extensions will be required, we considered these results encouraging since they show that direct classification with a simple segmentation strategy and state-of-the-art machine learning already provides quite adequate accuracy for our application requirements. We expected more from the layered approach that shows improvements only on specific object manipulations.

Class	**Include**	**Description**
Irrelevant	Handle	Medicine box is taken and handled
	Irrelevant	Medicine box is moved, but not by a person (e.g., hit the repository)
	Displacing in the repository	Medicine box's manipulations when someone searches for the correct one
	Opening/closing repository drawer	When the medicine box is in the repository and it is opened, medicine box moves
Relevant displacement	Displaced	Medicine box is displaced from a place to another which is not the medicine repository
	Inserted	Medicine box is displaced from a place to the correct repository
	Extracted	Medicine box is displaced from the repository to a place
Accessing content	Opened	Medicine box is taken from a place and a blister pack is extracted in the same place or in another

Table 3: Medicine box's manipulations

Conclusions and discussion
In this paper, we proposed an unobtrusive and effective approach to monitor specific manipulations of objects as a necessary component for the recognition of normally and abnormally performed ADLs in smart homes. Extensive experiments with a dataset consisting of more than two thousands manipulations show encouraging results.

Technological limitations
The use of BLE accelerometers attached to objects addresses important drawbacks of different technological

Class	Include	Description
Irrelevant	Irrelevant	Knife is moved but not by a person (e.g., the repository is shaked)
	Displacing in the repository	Knife's manipulations when someone searches for a tool or silverware
	Opening/closing repository drawer	When the knife is in the repository and it is opened, the knife is moved
Relevant displacement	Displaced	Knife is displaced from a place to another which is not the knife repository
	Inserted	Knife is displaced from a place to the correct repository
	Extracted	Knife is displaced from the repository to a place
Cutting	Cut	Knife is taken from a place and something is cut in the same place or in another

Table 4: Bread knife's manipulations

solutions proposed in the literature. However, it currently has several limitations: First of all, energy consumption, since we observed that the high transmission rate we used reduced the battery life to levels not acceptable for a real home deployment. Energy consumption was also affected by the need to increase the standard transmission power in order to cover at least the whole room. A second problem we found is interference when these devices are close to metal objects. Other problems arise when the monitored objects are dipped in water or exposed to high temperatures, since the devices would be damaged. However, we are confident that technological evolution will soon solve these limitations, while the ones affecting other approaches are not only technological, but involve user acceptance and privacy issues that may be more difficult to overcome.

	Accuracy (%)		
	Multi-layer classification	Direct classification	Total occurrences
Bottle			
Total	90,41	**91,54**	887
Irrelevant	91,73	**94,83**	387
Rel. displacement	85,28	**85,71**	231
Drink/Pour	**92**	91,82	269
Medicine box			
Total	92,07	**92,98**	656
Irrelevant	88,46	**91,75**	182
Rel. displacement	84,67	**85,40**	137
Accessing content	**97,03**	96,73	337
Knife			
Total	**96,88**	96,11	515
Irrelevant	**97,58**	97,56	124
Rel. displacement	**95,51**	93,58	156
Cut	**97,44**	97,02	235
Total			
Total	92,56	**93,14**	2058
Irrelevant	91,91	**94,51**	693
Relevant	**93,51**	93,06	1356

Table 5: Results

Future work

We intend to extend our work in several directions. We plan to extend the number of objects and manipulations, and compare different recognition techniques. Acceleration data can also be usefully combined with fine grained indoor positioning data, as well as other sensor data to refine manipulation detection. We want to do more experiments acquiring data from senior subjects while performing manipulations as part of complex activities.

REFERENCES

1. Oliver Amft and Gerhard Tröster. 2008. Recognition of dietary activity events using on-body sensors. *Artificial intelligence in medicine* 42, 2 (mar 2008), 121–36.
2. Michael Beigl, Hans-w Gellersen, and Albrecht Schmidt. 2001. MediaCups : Experience with Design and Use of Computer- Augmented Everyday Artefacts. *Computer Networks* (2001).
3. Leo Breiman. 2001. Random Forests. *Machine Learning* (2001), 5–32.
4. Liming Chen, Chris D. Nugent, and Hui Wang. 2012. A Knowledge-Driven Approach to Activity Recognition in Smart Homes. *IEEE Trans. Knowl. Data Eng.* (2012).
5. James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No Longer Wearing: Investigating the Abandonment of Personal Health-tracking Technologies on Craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*.
6. Barnan Das, Diane Cook, Narayanan Krishnan, and Maureen Schmitter-Edgecombe. 2016. One-Class Classification-Based Real-Time Activity Error Detection in Smart Homes. *IEEE Journal of Selected Topics in Signal Processing* (2016).
7. Ifat Afrin Emi and John A Stankovic. 2015. SARRIMA : Smart ADL Recognizer and Resident Identifier in Multi-resident Accommodations. In *Wireless Health '15*.
8. Yoav Freund and Robert E Schapire. 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. System Sci.* 55, 1 (1997), 119 – 139.
9. Enamul Hoque, Robert F. Dickerson, Sarah M. Preum, Mark Hanson, Adam Barth, and John A. Stankovic. Holmes: A Comprehensive Anomaly Detection System for Daily In-home Activities. In *2015 International Conference on Distributed Computing in Sensor Systems*. IEEE, 40–51.
10. Aftab Khan, Sebastian Mellor, Eugen Berlin, Robin Thompson, Roisin McNaney, Patrick Olivier, and Thomas Plötz. 2015. Beyond Activity Recognition: Skill Assessment from Accelerometer Data. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM.
11. Donald J. Patterson, Dieter Fox, Henry Kautz, and Matthai Philipose. 2005. Fine-Grained Activity Recognition by Aggregating Abstract Object Usage. In *ISWC '05: Proceedings of the Ninth IEEE International Symposium on Wearable Computers*. IEEE Computer Society.
12. Daniele Riboni, Claudio Bettini, Gabriele Civitarese, Zaffar Haider Janjua, and Rim Helaoui. 2016. SmartFABER: Recognizing fine-grained abnormal behaviors for early detection of mild cognitive impairment. *Artificial Intelligence in Medicine* (2016).
13. Jianxin Wu, Adebola Osuntogun, Tanzeem Choudhury, Matthai Philipose, and James M. Rehg. 2007. A Scalable Approach to Activity Recognition based on Object Use. In *Proc. IEEE International Conference on Computer Vision (ICCV 07)*. IEEE Press, 1–8.