A Probabilistic Ontological Framework for the Recognition of Multilevel Human Activities

Rim Helaoui
Research Group
Data & Web Science
University of Mannheim
rim@informatik.uni-mannheim.de

Daniele Riboni
EveryWare Lab
Dept. of Computer Science
University of Milano
daniele.riboni@unimi.it

Heiner Stuckenschmidt
Research Group
Data & Web Science
University of Mannheim
heiner@informatik.uni-mannheim.de

ABSTRACT
A major challenge of ubiquitous computing resides in the acquisition and modelling of rich and heterogeneous context data, among which, ongoing human activities at different degrees of granularity. In a previous work, we advocated the use of probabilistic description logics (DLs) in a multilevel activity recognition framework. In this paper, we present an in-depth study of activity modeling and reasoning within that framework, as well as an experimental evaluation with a large real-world dataset. Our solution allows us to cope with the uncertain nature of ontological descriptions of activities, while exploiting the expressive power and inference tools of the OWL 2 language. Targeting a large dataset of real human activities, we developed a probabilistic ontology modeling nearly 150 activities and actions of daily living. Experiments with a prototype implementation of our framework confirm the viability of our solution.

Author Keywords
Activity recognition, Hybrid, Multilevel, Situation, Modeling, Representation, Reasoning, Recognition, Probabilistic modeling, Ontology.

ACM Classification Keywords
H.1.2 User/Machine Systems: I.2.1Applications and Expert Systems; I.2.4 Knowledge Representation Formalisms and Methods;

INTRODUCTION
The recognition of ongoing human activities is a decisive aspect of context-awareness, needed to realize the vision of ubiquitous computing. Hence, many research efforts have been recently concerned with automatically recognizing human activities from light-weight wearable and environmental sensors. The main approaches can be generally classified into data-driven, knowledge based, and hybrid approaches.

UbiComp'13, September 8–12, 2013, Zurich, Switzerland.

http://dx.doi.org/10.1145/2493432.2493501

A multitude of works tried to apply diverse data-driven methods to recognize human activities. The majority proposed supervised learning approaches such as Hidden Markov Models [21], Conditional Random Fields [4] and Dynamic Bayesian Networks [27]. Despite being well suited to simple activities, data-driven techniques have a number of shortcomings when applied to the recognition of complex high-level activities, especially in terms of portability, extensibility, and support for common-sense knowledge. As an alternative to data-driven approaches, some researchers adopted different logical modelling and reasoning algorithms. This includes expressive description logics (DLs) [5], possibly coupled with rule-based languages [23]. Despite their ability to overcome many of the limitations of data-driven approaches, these knowledge-based systems suffer from other restrictions. Most importantly, the lack of support for probabilistic reasoning makes them inadequate to cope with the variability of complex activity execution, and with the inherent uncertainty of sensed and inferred context data.
Combining both paradigms has been the motivation of some recent works, including [7] and [22]. However, the loosely-coupled nature of most hybrid activity recognition systems proposed so far provides only a partial solution to the limitations of data-driven and knowledge based approaches.

As a step towards tightly coupling data-driven and knowledge-based approaches, in a previous work [8] we introduced a novel multilevel activity recognition framework, based on the use of *log-linear DL* [19], a description logics that supports both modelling and reasoning with uncertainty by combining log-linear models [13] and description logics [2].

This paper extends that preliminary work [8] with a detailed description of activity modeling, the reasoning framework and algorithms, and extensive experiments with a real-world dataset. For the modeling part, we hierarchically decompose complex activities in simpler ones, up to their simplest atomic components, and describe their semantic structure. For instance, the activity "prepare dishwasher" is a component of the more complex activity "cleaning up", and it is composed by a sequence of simpler gestures such as "open dishwasher" and "put kitchenware in dishwasher". For the reasoning part, we employ probabilistic ontological reasoning to progressively infer higher level activities from their simpler components. As input, we use atomic actions (e.g., "grab dishwasher door", "pull dishwasher door") recognized through the application of standard machine learning techniques on sensor data.

Complex human activities are inherently uncertain, since *i)* the same complex activity can be carried out in many different ways, and *ii)* the same sensor events may indicate the execution of multiple (and possibly incompatible) activities. Non-probabilistic approaches fall short in the recognition of such complex activities. For instance, suppose that a person is performing some actions in the kitchen and, based on the sequence of sensors observations, two complex activities are feasible: *preparing dinner* and *dishwashing*. Since those activities cannot be performed at the same time, with a cautious approach, a non-probabilistic reasoner should discard both activities. This approach determines high precision but low recall. On the contrary, if the reasoner takes a less cautious approach, and ignores the incompatibilities, it should infer both activities, resulting in higher recall but low precision. With our probabilistic DL approach, the reasoner takes into account incompatibilities among activities, and infers the most probable one, considering the (possibly incomplete) sensors observations, and the confidence values of the different activity definitions. With respect to data-driven techniques, the expressiveness of the adopted DL, which is the same of OWL 2 [6], allows us to support modelling and recognition of very specific actions, which are not likely to occur in training datasets of activities used by supervised learning approaches.

Compared to other approaches combining description logics [2] with probabilistic models such as Probabilistic Description Logics [16] and DISPONTO [3], Log-linear description logics allow to integrate heterogeneous features facilitating the parametrization of the models. In particular, unlike probabilities, the values of the weights associated with the uncertain constraints do not affect the satisfiability of the ontology. To the best of our knowledge, the publicly available log-linear DL reasoner Elog [20] is the only reasoner which responds to the need of finding the most probable consistent ontology instead of computing the a-posteriori probability of axioms such as in [26]. A further motivation of our choice is related to the successful application of Log-linear models to the problem of activity recognition in a previous work [7]. There, the activity recognition problem was addressed by employing Markov Logic Networks, a statistical relational paradigm combining log-linear models and first order logic [25].

Starting from a large dataset of multilevel activities acquired in a realistic setting, we have developed a probabilistic ontology, modelling nearly 150 activities of daily living at different granularities. We have developed a prototype implementation of our recognition framework, which includes an interface to populate the ontology with sensor data instances, a probabilistic reasoner to compute the most probable ontological description of the current situation, and an OWL 2 reasoner to infer the performed activities at different granularities. We have performed an in-depth experimental evaluation of our framework with the multilevel activity dataset. The experimental results are promising, and confirm the viability of our solution. For the sake of this work, we concentrated on the probabilistic aspects of our framework. However, as we discuss in the conclusions, this approach has the potential to be improved in several directions; in particular, by applying more sophisticated techniques for reasoning with temporal features, and by fully exploiting available activity datasets to fine-tune the confidence values associated to the different activity definitions.

RELATED WORK

There are only few works that keep both semantic description of the activities and their recognition process tightly-coupled. Such an approach has been adopted in [5] and in [28]. Similarly to our work, the authors load the current contextual information to populate their generated ontology then employ inference reasoner to obtain the most specific equivalent activity to an *unknown* class modelling the current one. Hence, activity recognition can be mapped to equivalence and subsumption reasoning. Springer *et al.* [28] rather focus on inferring high-level situations to trigger the correct reactions in smart houses. They test their system with simple cases such as recognizing whether a ringing person is authorized to enter or not.

Much closer to our approach is the work of Chen *et al.* [5]. There, the authors proceed to an incrementally specific recognition of the activities through the progressive activation of the sensors. Unlike our work, this top down approach fails to recognize fine-grained activities unless the higher one is correctly recognized. Besides, the evaluation data used is not naturalistic. It was collected in a partially predefined and strictly sequential manner including fixed time interval separating the complex activities. Like [28], their approach does not address the inherent uncertainty aspect in human activities. Particularly, their model implicitly assume a deterministic mapping from the context data to the activities' descriptions.

Bridging the gap between such a purely symbolic approach and supporting uncertainty, was the concern of several works recently. Following a lazy instance based approach, Knox *et al.* [11] use a vector of the sensors' values to define their cases. A semantically extended case base is created through extracting ontological relationships between sensors, locations and activities. This allows them to reduce the resulting number of cases. Further efforts to exploit semantic information to improve the recognition system are detected in [30] and [29]. Relying on the subsumption hierarchy, the former involves ontology to handle unlearned objects and map them into learned classes. At the recognition step, parametric mixture models are applied. In the latter, the subsumption hierarchy helps automatically inferring probability distributions over the current actions given the object in use. Thus, the integrated common-sense knowledge is used to learn a Dynamic Bayesian Network-based activity classifier. Other attempts to cope with uncertainty involve applying a hierarchy Bayesian networks based on the ontology's instances such as in [15]. All these works dissociate the inference step from the semantic model. This aspect limits the ability of incorporating rich background and common sense knowledge. It also strips the system from other advantages of symbolic reasoning such as consistency check.

To the best of our knowledge, [9] and [10] are the only ontology-based tightly-coupled human activity recognition approach from dense-sensing.

The first [9] is an extension of the work presented by Kurz *et al*. in [14] focusing on the autonomous selection of the best set of available sensors to recognize a given goal. Using an ontological description of domain knowledge, the authors propose to use the semantic information between the different goals. This allows to reason with sub and super concepts to refine the recognition goal in case of missing sensing capabilities and exploit available data of related sensors. Thanks to the introduced "context predicates" and "Degree of Fulfilment (DoF)" notions, recognition goals can be modelled and inferred while enabling a weight distribution to sensor-goal mappings. Despite the promising aspect of this top-down approach, it was only evaluated with a simplistic low-level scenario involving one single recognition goal: "Locomotion". Moreover, the subsumption axioms do not support uncertainty, which might limit the applicability of the framework under realistic settings. In the second work [10], the authors model the interrelationships between sensors, contexts and activities. They use the resulting hierarchical network of ontologies to generate evidential networks. Following Dempster-Shafer theory of evidence, they calculate and define the heuristic relationships between the network's nodes in form of evidential mappings. These mappings are used through seven steps of evidential operations as inference process. Obviously, their evidential network discloses limited expressiveness compared to our DLs language.

DESCRIPTION LOGICS AND LOG-LINEAR DESCRIPTION LOGICS

Description logics are a commonly used representation formalism for knowledge bases. There are numerous tools and standards for knowledge representation and reasoning using DLs. The DLs framework allows the representation of facts about individuals (class and property assertions), as well as axioms expressing schema information. Log-linear DLs integrate description logics with probabilistic log-linear models [19]. In particular, Log-linear DLs allow to model both probabilistic and deterministic dependencies between DL axioms, extending uncertain axioms with weights. Regarding the expressiveness of the language, Log-linear DL supports the same operators of the well-known OWL 2 language.

Formally, a log-linear knowledge base $C = (C^{\mathsf{D}}, C^{\mathsf{U}})$ is a pair consisting of a *deterministic* knowledge base CBox C^{D} and an *uncertain* knowledge base CBox $C^{\mathsf{U}} = \{(c, w_c)\}$ with each c being an axiom and w_c a real-valued weight assigned to c. The *deterministic* CBox contains axioms that are known to hold; for instance, axiom c_1 = "*Meeting* is a subclass of the activities having at least two actors". The *uncertain* CBox contains axioms with weights: the greater the weight of an axiom, the more confidence there is for it to hold. In the following we illustrate an example of the weighting mechanism, applied to activity axioms.

Example 1. *Suppose that we want to model the fact that most persons (but not everyone) wear rubber gloves when washing dishes. With log-linear DL, we can model this knowledge by adding the following axiom c_2 to the uncertain CBox, and assigning a strong weight to it: c_2 = "*DishWashing *is a subclass of the activities whose actors wear rubber gloves".*

The deterministic CBox is assumed to be coherent and consistent; i.e., whatever the instantiation of classes and properties, no inconsistencies can be derived. For instance, given the deterministic axiom c_1, it is impossible to derive that a person is performing a meeting by herself. On the contrary, the uncertain CBox may be inconsistent, as shown below.

Example 2. *Continuing Example 1, despite the uncertain axiom c_2, it is feasible to have in the knowledge base both the following (uncertain) assertions:* "p is a person that is not wearing gloves"*, and* "p is an actor of DishWashing".

Those inconsistencies can be resolved by exploiting the semantics of log-linear DLs, which is based on probability distributions over *coherent* and *consistent* knowledge bases. The weights of the axioms determine the log-linear probability distribution. For a log-linear knowledge base $C = (C^{\mathsf{D}}, C^{\mathsf{U}})$ and a CBox C' with $C^{\mathsf{D}} \subseteq C' \subseteq C^{\mathsf{D}} \cup \{c : (c, w_c) \in C^{\mathsf{U}}\}$, we have that

$$\Pr_C(C') = \begin{cases} \frac{1}{Z}\exp\left(\sum_{\{c \in C' \setminus C^{\mathsf{D}}\}} w_c\right) & \text{if } C' \text{ is \textit{coherent}} \\ & \text{ and \textit{consistent};} \\ 0 & \text{otherwise} \end{cases}$$

where Z is the normalization constant of the log-linear distribution $\Pr_C$.

Under the given syntax and semantics, the central inference task is the maximum a-posteriori (MAP) query: *"Given a log-linear ontology, what is a most probable coherent ontology over the same class and property names?"*

Example 3. *The application of the MAP query to the simple ontology shown in Examples 1 and 2 would return a consistent ontology, in which one of the two cases may happen:*

- *Case* 1*: axiom c_2 belongs to the most probable consistent ontology. In this case, either assertion "p is not wearing gloves" or "p is washing dishes" is discarded, based on the computed log-linear probability distribution.*
- *Case* 2*: axiom c_2 does not belong to the most probable consistent ontology. In this case, the two assertions are coherent.*

As we illustrate in the following section, this inference mechanism is the core of our activity recognition framework.

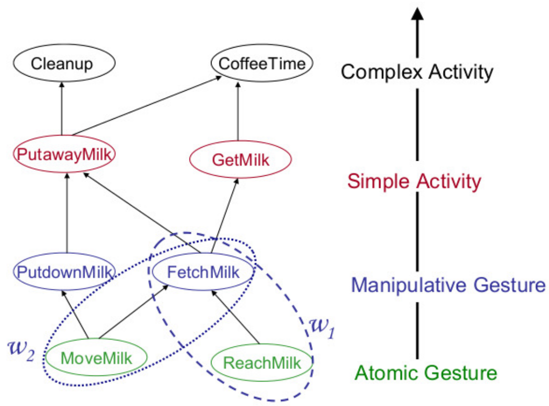
ONTOLOGY-BASED MULTILEVEL ACTIVITY MODELLING AND RECOGNITION FRAMEWORK

In this section, we describe our ontology-based technique for multilevel activity recognition.

Modelling and representation

In order to illustrate the multilevel activity structure, we adopt the frame proposed in [17] for the Opportunity dataset.

- *Atomic Gestures* (Level 4) are those actions that cannot be decomposed in simpler ones, as "*release dishwasher*



w_1: $\mathrm{FetchMilk} \sqsupseteq \mathrm{ManipuilativeGesture} \sqcap \exists\ \mathrm{hasActor.(Person} \sqcap \exists\ \mathrm{hasAtomicGesture.ReachMilk)}$

w_2: $\mathrm{FetchMilk} \sqsupseteq \mathrm{ManipuilativeGesture} \sqcap \exists\ \mathrm{hasActor.(Person} \sqcap \exists\ \mathrm{hasAtomicGesture.MoveMilk)}$

Figure 1. An example of the Multilevel activities and the weighted axioms defining the *Manipulative Gesture* "Fetch Milk"

door'" or "*reach glass*". Typically, they have a very short duration; hence, for simplicity we consider them as instantaneous events.

- *Manipulative Gestures* (Level 3) are characterized by the execution of simpler *Atomic Gestures*. For instance, the *Atomic Gestures* "*reach dishwasher door*", "*push dishwasher door*", and "*release dishwasher door*" characterize the *Manipulative Gesture* "*close dishwasher*". The duration of *Manipulative Gestures* is limited to a very few seconds. In some cases, they can be executed concurrently. For instance, an individual can put a glass on the table with the left hand while she is opening the fridge door with the other hand.

- *Simple Activities* (Level 2) are characterized by temporal sequences of *Manipulative Gestures* and modes of locomotion. For example, the *Simple Activity* "*put tableware in dishwasher*" is characterized by first "*fetch an object*", then "*open the dishwasher*", "*put down the object*", and finally "*close the dishwasher*". Of course, the same *Simple Activity* may be characterized by different sequences of *Manipulative Gestures*. Its typical duration is limited to a few seconds.

- *Complex Activities* (Level 1) are characterized by the concurrent execution of *Simple Activities*. For instance, the *Complex Activity* "*clean up*" is characterized by the concurrent execution of *Simple Activities* as "*put tableware in dishwasher*" and "*clean table*". Typically, *Complex Activities* have a sensible duration (from a few minutes to a few hours).

As anticipated, we use log-linear DLs [19] to define a probabilistic ontology that formally represents the multilevel structure of the considered activities. The uncertainty support provided by that formalism allows us to associate different values of confidence to the different axioms describing the activity structure. The weighted definitions of activities compose our *uncertain* CBox C^{U}, while we express the incompatibility of certain activities (i.e., activities that cannot be executed at the same time) as disjointness axioms in the *deterministic* CBox C^{D}. The axioms weights can be manually defined based on background knowledge, or automatically learned from a training set of performed activities. Figure 1 shows the global multilevel structure and gives an example of two different ways to describe the *Manipulative Gesture* "*fetch milk*". Once in terms of the *Atomic Gestures* "*reach milk*" with some weight w_1 and once in terms of the "*move milk*" with some weight w_2.

Multilevel activity recognition

We propose the following technique to recognize increasingly complex activities based on simpler ones. As imposed by real-life scenarios, the method is not limited to sequential performance of activities but also covers concurrent ones.

Recognizing atomic gestures

At the bottom-line of our technique, *Atomic Gestures* (Level 4) are recognized through the application of supervised learning methods, based on data acquired from different sensors:

- Body-worn sensors (accelerometers, gyroscopes) are used to capture movements, simple actions, and body posture. For instance, through data captured from several on-body accelerometer and a custom-made motion jacket, it is possible to infer whether an individual is *pushing* or *pulling* an artifact.
- RFID tags and readers to detect objects usage.

By matching the used artifact (e.g., a door) with the action inferred from body worn sensors, it is easy to infer the performed *Atomic Gesture* (e.g., *pushing a door*).

Recognizing manipulative gestures

Manipulative gestures (Level 3) are recognized based on the inferred *Atomic Gestures*, and on the semantic structure of activities.

In particular, in order to recognize the *Manipulative Gestures* performed during a time window τ_3, we perform four operations as stated below.

1. We collect all the *Atomic Gestures* that were inferred during τ_3;
2. We represent them as ontological classes and assertions, and add them to the probabilistic ontology;
3. We instantiate properties that relate each *Atomic Gesture* with its actor;
4. We exploit both probabilistic and standard ontological reasoning tasks to infer the most probable *manipulative gesture(s)* performed by the actor, as explained below.

After adding the instances, the resulting ontology may be inconsistent. For example, according to the multilevel activity structure, the set of *Atomic Gestures* recognized during τ_3 may lead to the derivation of two (or more) incompatible *Manipulative Gestures*. These inconsistencies may be due to sensor failures, or simply to ambiguity of interpretation of the current set of *Atomic Gestures*. Inconsistencies are resolved by computing the *most probable consistent ontology* C^*, as explained in the previous section. The *Manipulative Gestures* during the time window τ_3 are then inferred through standard subsumption and equivalence reasoning on C^*.

Recognizing simple activities

Since *Simple Activities* are defined as temporal sequences of *Manipulative Gestures*, we need to express temporal relations about the occurrence of the specific gestures executed by the user.

EXAMPLE 4. *Consider, the* Simple Activity *"put in dishwasher". A possible sequence describing that activity is:* "the individual takes an object, then opens the dishwasher door, puts the object inside, and closes the dishwasher door".

However, as explained in [24], the description logic underlying OWL 2 [6] does not natively support temporal reasoning; this limitation is shared by log-linear DLs. In order to address this limitation, we adopt an ad-hoc method to perform a simple form of temporal reasoning with sequences of *Manipulative Gestures*. To recognize the *Simple Activities* executed during a time window τ_2, we keep a list of the *Manipulative Gestures* that have been executed within that time interval. The list elements are represented as instances of *triadic* properties, each relating the *Manipulative Gesture*, its actor, and the order of execution within τ_2. However, in the DLs we use, only *binary* properties can be expressed. Hence, we rely on an ontology pattern[1], and represent the triadic property by means of a class T-MANIPULATIVEGESTURE, having three properties to represent (i) the actor (e.g., ALICE), (ii) the performed *Manipulative Gesture* (e.g., OPENDISHWASHER), and (iii) its order of execution within τ_2. This is illustrated in Example 5.

EXAMPLE 5. *Continuing Example 4, we define the* Simple Activity *"put in dishwasher" in terms of the sequential occurrences of* Manipulative Gestures *"fetch object","open dishwasher", "put down object", and "close dishwasher":*

$$\begin{aligned}
&\textsc{PutInDishwasher} \sqsubseteq \textsc{SimpleActivity} \sqcap \\
&\forall\, \textsc{hasActor}.\big(\textsc{Person} \sqcap \exists\, \textsc{hasT-ManipGesture}. \\
&\quad(\textsc{T-ManipulativeGesture} \sqcap \\
&\quad\exists \textsc{hasManipGesture}.\textsc{FetchObject} \\
&\quad\sqcap \exists \textsc{hasOrder} = 1) \sqcap \\
&\quad\exists\, \textsc{hasT-ManipGesture}. \\
&\quad(\textsc{T-ManipulativeGesture} \sqcap \\
&\quad\exists \textsc{hasManipGesture}.\textsc{OpenDishwasher} \\
&\quad\sqcap \exists \textsc{hasOrder} = 2) \sqcap \\
&\exists\, \textsc{hasT-ManipGesture}.(\textit{[other gestures \dots]})\big).
\end{aligned} \tag{1}$$

[1] http://www.w3.org/TR/2004/WD-swbp-n-aryRelations-20040721/

In order to recognize all the *Simple Activities* performed during τ_2, we go through the following steps:

1. We collect all the *Manipulative Gestures* that were inferred during τ_2;
2. We represent them as ontological classes and assertions, and add them to the probabilistic ontology;
3. We instantiate the properties that relate each *Manipulative Gesture* with its actor;
4. We update the order of execution of the collected *Manipulative Gestures*, and add the list to the probabilistic ontology, using the ontology pattern explained above;
5. We use the ELOG reasoner[2] [20] for log-linear DLs to compute the *most probable consistent* ontology from the probabilistic ones;
6. We perform standard ontological reasoning on the obtained *most probable consistent* ontology to derive the set of *Simple Activities* executed during τ_2.

Recognizing complex activities

In order to recognize *Complex Activities* in terms of *Simple Activities*, we proceed similarly to recognizing *Manipulative Gestures* in term of *Atomic Gestures*. Here, the *Simple Activities* are collected over a longer time window τ_1. During such a window, several *Simple Activities* might be performed which helps discriminate between the *Complex Activities* and allows the omission the intra-activity temporal relationships, as shown in the following example.

EXAMPLE 6. *Suppose that we want to represent the* Complex Activity *"clean up" in terms of its component simple activities "put in dishwasher" and "clean table". Hence, we can add the following axioms to the knowledge base:*

$$\begin{aligned}
&\textsc{CleanUp} \sqsubseteq \textsc{ComplexActivity} \sqcap \\
&\quad\forall\, \textsc{hasActor}.\big(\textsc{Person} \sqcap \\
&\exists\, \textsc{hasSimpleActivity}.\textsc{PutInDishwasher}\big), 1.8
\end{aligned} \tag{2}$$

$$\begin{aligned}
&\textsc{CleanUp} \sqsubseteq \textsc{ComplexActivity} \sqcap \\
&\quad\forall\, \textsc{hasActor}.\big(\textsc{Person} \sqcap \\
&\exists\, \textsc{hasSimpleActivity}.\textsc{CleanTable}\big), 1.5
\end{aligned} \tag{3}$$

*Axiom (*1*) has weight* 1.8*, and essentially states that if a person is currently putting things in the dishwasher, she is probably performing the complex activity "clean up". Axiom (*2*) is similar, apart that it has a slightly lower weight, and considers activity "clean table" instead of "put in dishwasher".*

EXPERIMENTAL EVALUATION

In this section we first describe the dataset we used for our experiments as well as our multilevel activity ontology. Then we outline the prototype implementation of our system. Finally, we present our experiments and the achieved results.

[2] https://code.google.com/p/elog-reasoner/

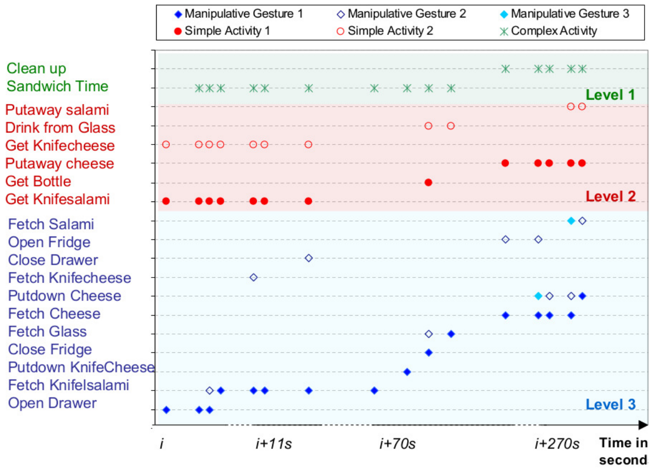


Figure 2. Illustration of a typical portion from the multilevel structured activities. This shows a high concurrency degree. As it can be depicted, the user can be simultaneously involved in "fetch cheese", "putdown cheese", "fetch salami" at Level 3, and "putaway cheese", "putaway salami" at Level 2

Dataset

To evaluate our approach we use real-life data collected in highly rich networked sensor environments. The dataset is a part of the EU research project "Activity and Context Recognition with Opportunistic Sensor Configuration"[3]. In a smart room simulating a studio flat, a total of 72 sensors with 10 modalities are deployed in 15 wireless and wired networked sensor systems in the environment [17]. Several users participated in a naturalistic collection process of a *morning routine* with the focus on maximizing the activity primitives.

The deployed sensors can be classified into wearable sensors such accelerometers and environmental sensors such as RFID tags and readers. The wearable sensors are used to infer 13 body gestures like "reach" and "move" and 4 modes of locomotion like "sit" and "lie". The environmental sensors can detect 17 different objects with which the user can interact. The inferred gestures, modes of locomotion and objects in use are provided in the dataset. The combination of this information results in 86 *Atomic Gestures* such as "*reach fridge*". Apart from the *Complex Activities*, the two levels in between are not annotated. Hence, to be able to evaluate our approach, we completed the annotation task for three subjects $S10$, $S11$, and $S12$ with three different routines each ($ADL1-3$). The duration of the routines varies significantly from a minimum of 13 minutes up to a maximum of 27 minutes. During this time, the sensor inputs are very irregular. As example, in the data collected by user $S10$, the shortest interval between two sensor inputs equals 0.03 seconds while the longest can span up to 240 seconds. The average resides between 1 and 2 seconds.

The annotation task was done by three persons yielding some inevitable differences. The dataset considers a total of 40 *Manipulative Gestures* and 21 *Simple Activities*. However, labels do not cover all the possible activities, leaving some sensor

[3]http://www.opportunity-project.eu

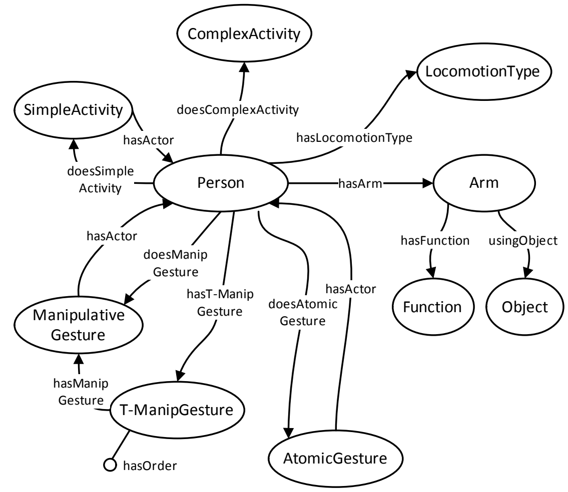


Figure 3. The core classes (represented as nodes) and properties (edges) of our ontology

observations with no annotation at one or more levels. Furthermore, the data does not contain any segmentation scheme. The complete annotation can be downloaded from this link [4]. The resulting multilevel structure exhibits a high degree of concurrency in fine-grained activities. Figure 2 illustrates a portion of this data where the user is first "*having a sandwich*" then "*cleaning up*". In this example, these two *Complex Activities* involve 11 *Manipulative Gestures*, 6 *Simple Activities*. The user can be simultaneously involved in more than one *Manipulative Gestures* and/or *Simple Activities*. A corresponding example in Figure 2 would be: "*close fridge*" and "*fetch glass*" at Level 3, and "*get bottle*", "*drink from glass*" at Level 2.

Multilevel activity ontology

We have developed a probabilistic ontology to fully support the activity recognition scenario of the Opportunity dataset. The ontology includes about 150 actions and activities of daily living at four granularity levels. Only the data generated by the user $S10$ was considered while building our ontology. The data of the other two individual are used to assess the viability of our approach for user-independent scenarios. The ontology has been developed using the Protégé OWL editor [12]. Axioms weights have been manually added to axioms definitions in form of annotations annotations using the *confidence* property. The estimated values of the weights were based on common sense knowledge supported by the observation of the data collected by user $S10$.

Figure 3 shows the core classes and properties of our ontology. Since, in the Opportunity scenario, users wear RFID gloves and accelerometers to detect so called "arm functions" (e.g., push, pull, . . .) and used objects, the ontology includes classes for *Arm*, *Function* and *Object*. The *hasArm* property relates each instance of class *Person* to his/her *Arms*. Each arm is related to its current function and used object by properties *hasFunction* and *usingObject*, respectively. The ontology includes an extensive collection of *Atomic Gestures*, de-

[4]http://webmind.dico.unimi.it/care/annotations.zip

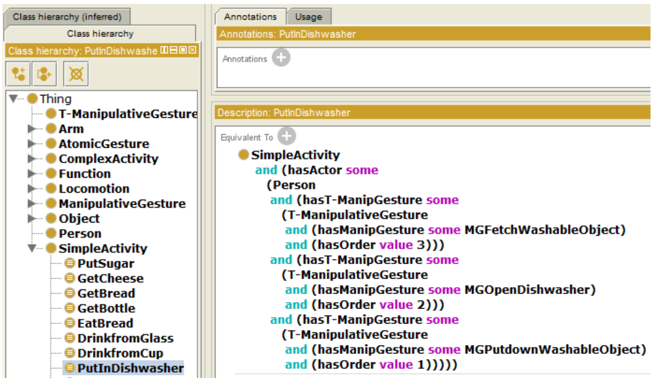


Figure 4. Axiom definition for *PutInDishwasher* in the Protégé editor

fined as specializations of the core *AtomicGesture* class. We have defined 86 *Atomic Gestures* classes, such as *SipCup SpreadCheese*, *StirSpoon*, *UnlockDoor*. The definitions of these gestures is given in terms of the function and used object of the user's arms. For instance, *StirSpoon* is defined as: *"an atomic gesture whose actor has an arm having function Stir and using object Spoon"*.

At a higher level, *Manipulative Gestures* are defined in terms of *Atomic Gestures*. As explained before, multiple axioms are given for the same manipulative gesture, each associated with a confidence value. For instance, *PutDownCup* is defined by the following axioms: *"a manipulative gesture whose actor is performing the atomic gesture ReleaseCup"* with confidence 0.9, and *"a manipulative gesture whose actor is performing the atomic gesture MoveCup"* with confidence 0.15. Our ontology contains classes for 40 classes of *Manipulative Gestures*; most of these classes are defined by 2 or 3 weighted axioms.

Simple Activities are defined in terms of the temporal sequences of *Manipulative Gestures* and modes of locomotion of the actor. Our ontology contains classes for 21 classes of *Simple Activities*, such as *DrinkFromCup*, *EatBread*, *LieOnLazyChair*, *PrepareCheeseSandwich*. As for manipulative gestures, we take into account the variability of simple activities execution by assigning multiple weighted axioms to the same simple activities. Figure 4 shows a snapshot of the axiom definition for *PutInDishwasher*, shown in the Protégé editor. The ontology also contains classes for the five complex activities considered by the Opportunity dataset.

Even if the ontology has been developed with the scenario of the Opportunity dataset in mind, it can be reused in large part for many other ADL recognition applications. The ontology is available under this link [5].

Implementation and experimental setup

The proposed approach has been implemented as a prototype system in Java relying on the OWL-API [6]. As depicted in Figure 5, the program automatically parses the sensors observations during a sliding time window; then iteratively outputs the inferred activities. Each level of granularity is recognized using the inferred output from the finer-grained level as follows. The program starts collecting the sensor observations consisting in the individual modes of locomotion, arms function, and used objects. As soon as the elapsed time at starting point exceeds the predefined time window τ_3 for *Manipulative Gesture*, the reasoning process described below is triggered. For our evaluation, we fix τ_3 at 1*second* duration to remain within the average range of the sensor input frequency (see Dataset section).

Each recognized *Atomic Gesture* is added as an instance of the corresponding class in the ontology, and a property assertion of *hasAtomicGesture* is added, to relate it with the user. Each one of these *Atomic Gestures* is used to generate an *unknown Manipulative Gesture* class description. The new axioms are added to the probabilistic ontology and the *Elog* reasoner is initiated to output the most probable consistent one. In particular, ELOG solves MAP queries by transforming the probabilistic ontology into an integer linear program(ILP). It iteratively queries the Pellet[7] reasoner to derive explanations for incoherences or inconsistencies and adds those as constraints to the ILP; resolves them using the Gurobi solver[8]; and repeats these operations until all inconsistencies are resolved.

The resulting consistent ontology serves again as input for Pellet reasoner to infer the equivalent class(es) to the introduced unknown class. These classes represent the predicted *Manipulative Gestures* during the given time window. At the next sensor input, the collected data is deleted and the whole collection-reasoning process is triggered again.

Similarly, in order to recognize activities at the next abstraction level, the predicted classes from the previous one are used to define an unknown class for the current *Simple Activity*. However, instead of deleting the current collection of *Manipulative Gestures*, these are kept as long as no *Simple Activity* has been recognized. From one time window to the next, their order is updated accordingly. Since in our ontology the longest sequence composing a *Simple Activity* consists of *four Manipulative Gestures*, we fix our τ_2 interval to *four* τ_3 intervals. Hence, a *Manipulative Gestures* can have a maximum order of *four* before its deletion.We exploit the explanation features of Pellet to retrieve the axioms that actually fired the recognition of the *Simple Activity*, in order to derive its duration. Finally, the *Simple Activities* predicted during τ_1 are used to generate an unknown *Complex Activity* class modelling the current *Complex Activity*. Complex activities cover longer periods (up to 305 *seconds* for user $S10$). However, to remain faithful to our real-time aspect in our recognition framework, we arbitrary limited τ_1 to 30 *seconds*. The recognition procedure follows the same steps described for the *Manipulative Gestures*.

Evaluation

As described in the previous section, our prototype outputs the prediction results at the end of each time window. In the absence of any inferred Activity, it just outputs "*null*".

[5]http://webmind.dico.unimi.it/care/multilevel_activities.owl
[6]http://owlapi.sourceforge.net/
[7]http://clarkparsia.com/pellet/
[8]http://www.gurobi.com/

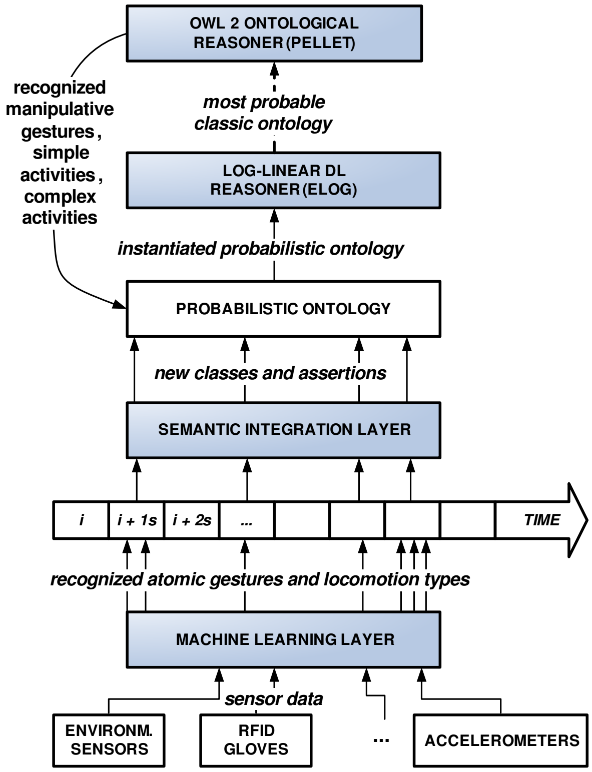


Figure 5. The global architecture of the proposed recognition framework

To evaluate the predictions, we first reproduce the same windowing technique on the ground truth. Then, we apply three well-known evaluation metrics used in pattern recognition: Precision, recall and F1 score to compare it to our predictions as follows: For each output, a correctly predicted activity is counted as *true positive (TP)*. A predicted activity A_1 at time window w_1 that does not match any activity in the reference dataset for the same time window is counted as a *false positive (FP)*. Finally, any activity present in the reference yet absent in the predictions for that same time window counts as *false negative (FN)*.

Results

Table 1 depicts the average *Precision*, *Recall*, and *F1 measure* values over three routines performed by user $S10$. Recall that only the data of this user was considered during the manual definition of the ontology's axioms and their confidence values. *Level 3*, or the *Manipulative Gestures* level, achieves the highest performance with Precision and Recall of 0.87 and 0.82 respectively. These results are positive, considering the large number of manipulative gestures (40).

At Level 2, while we obtain similarly good results for precision, the performance of recall drops significantly. Several aspects might explain this deterioration. On one side, the erroneous and missing predictions from Level 3 will most probably violate the whole *Simple Activity* sequence and result in missing predictions. Since *Simple Activities* usually spread over two, three or four time windows, the consequences of Level 3 errors are amplified at Level 2. These shortcoming can be alleviated by adding axioms for partial sequences of

Table 1. Recognition results for subject $S10$. The values correspond to the average values over three morning routines. The variation (σ) between these routines is also reported.

User S10	Level 3	Level 2	Level 1
Precision	0.87(σ 0.02)	0.87(σ 0.04)	0.90(σ 0.04)
Recall	0.82(σ 0.01)	0.40(σ 0.05)	0.58(σ 0.1)
F_1	0.85(σ 0.02)	0.55(σ 0.04)	0.70(σ 0.08)

Table 2. Average recognition results over three routines for subjects $S10$, $S11$ and $S12$. Each subject was evaluated using three different routines.Only the data generated $S10$ was considered to build our ontology and define its axioms. The variation (σ) between the results of respective users is also reported.

All Users	Level 3	Level 2	Level 1
Precision	0.85(σ 0.01)	0.86(σ 0.02)	0.91(σ 0.01)
Recall	0.81(σ 0.01)	0.42(σ 0.05)	0.65(σ 0.04)
F_1	0.83(σ 0.01)	0.57(σ 0.05)	0.75(σ 0.03)

manipulative gestures. On the other side, the limited temporal support of our ontological model fails in capturing all the possible *Manipulative Gestures* sequences defining a particular *Simple Activity*. While extensively specifying all the different sequences of *Manipulative Gestures* that may characterize *Simple Activities* is infeasible, this limitation can be solved by extending our modelling formalism with support for expressing temporal relationships.

Despite the low recall at Level 2, the recognition results of *Complex Activities* (*Level 1*) reached a significantly higher $F1$ value of 0.7. This can be explained by two reasons. The first one refers to the small number of *Complex Activities* classes compared to the other levels. The second one consists in the discriminative power of the set of recognized *Simple Activities* belonging to one *Complex Activity*. Indeed, since *Complex Activities* are recognized with relatively low frequency, there is a high chance to correctly recognize enough *Simple Activities* to allow the inference of the correct complex activity. Nonetheless, we acknowledge that obtained results at this level remain unsatisfactory. This is partially due to inadequacy of the used dataset for our multilevel scenario. Basically, the *Complex Activity* provided labels were initially selected to be recognized directly from low-level sensor data. We think that our approach could perform better using a highly contextual data.

Table 2 presents the average recognition performance of our prototype for the three different users. Despite the variety resulting from involving new users, we observe a similar performance level to the $S10$ case in Table 1. We also depict low variation among the three users. Accordingly, the reported results validate the robustness of the approach under user-independent setting.

We have also evaluated the feasibility of our approach for recognizing activities in real-time. We have executed the recognition algorithm on a workstation equipped with an Intel Core i7-2600 3.40 GHz CPU and 8.00 GB RAM, running the 64-bit Windows 7 operating system. Results about the execution

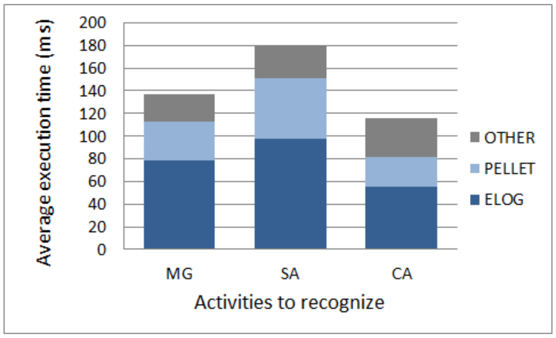


Figure 6. Average execution times of ELOG and Pellet reasoners with *Manipulative Gestures* (MG), *Simple Activities* (SA), and *Complex Activities* (CA)

times of our algorithm are shown in Figure 6. Average execution times for *Manipulative Gestures* (MG) are well below the one-second time window that we use to recognize MGs. *Simple Activities* (SA) are more computationally expensive to recognize than MGs; this is due to the high number of axioms that describe the same SAs. However, even for SAs, average recognition times are well below one second. Recognition of *Complex Activities* (CA) is less expensive, mainly due to the simple axioms we used in their definition. We evaluated in depth the computational cost of the most computationally expensive tasks of our algorithms; i.e., execution of ELOG and Pellet reasoners. In general, ELOG reasoning within our algorithm is more time consuming than Pellet. A relevant part of execution time is taken by other operations; in particular, OWL parsing and dynamic modification of the ontology.

DISCUSSION AND FUTURE WORK

In this paper, we proposed a tightly-coupled hybrid system for human activity recognition. Our framework unites both symbolic and probabilistic reasoning. This is achieved through the adoption of highly expressive log-linear DLs to represent and reason about the current activity at different granularities and complexity levels. The experimental results validate the viability of our approach to address multilevel activities for user independent scenarios.

The benefits of the proposed approach are manifold. Unlike the majority of related works, it supports the inherent uncertain nature of human activities without sacrificing the advantages of ontological modelling and reasoning. These advantages include consistency checking, the ability of integrating rich background knowledge and the simultaneous recognition of coarse and fine-grained activities. The use of a standard description formalism enhances the portability and reusability of the proposed system, and supports the representation of heterogeneous and uncertain context data. Moreover, the declarative nature of DLs reinforces the flexibility and intelligibility of the system.

A major limitation concerns fine-grained activities with relevant and rich temporal intra-relationships. This is due to the simplistic temporal modelling and reasoning employed in our framework. Two alternatives could be considered to address this issue. The first one consists in the use of temporal description logics, in which the temporal and terminological domains are tightly coupled. A relevant instance of these languages was proposed in [1].The second would be to use of a loosely-coupled technique, in which time is treated as a *concrete domain* [18]. We believe that the latter approach is a promising research direction to follow.

Our future work involves new techniques to alleviate the manual ontological engineering through automatically learning activity axioms and weights. We are also considering further modelling techniques to address incomplete data. This can be realized by adding category classes such as "Use Fridge" as a superclass for all the *Atomic Gestures* involving the object "Fridge", to allow the recognition of the composite activities even if the correct "Gesture" is missing from the sensor observations. Future work also includes the validation of our approach with datasets that include richer context sources, and experiments with inaccurate and incomplete sensor data.

Acknowledgments

The authors would like to thank Laura Palmi for her excellent work in ontology design and software implementation. The authors also thank Claudio Bettini and Mathias Niepert for their insightful comments and contribution to a preliminary version of this work. The work of Daniele Riboni has been partially supported by the project "SECURE: Intelligent System for Early Diagnosis and Follow-up at Home", funded by a grant of Lombardy Region and Italian Ministry of Education, University and Research.

REFERENCES

1. Artale, A., and Franconi, E. A temporal description logic for reasoning about actions and plans. *Journal of Artificial Intelligence Research (JAIR) 9* (1998), 463–506.
2. Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., and Patel-Schneider, P. F., Eds. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.
3. Bellodi, E., Lamma, E., Riguzzi, F., and Albani, S. A distribution semantics for probabilistic ontologies. In *URSW* (2011), 75–86.
4. Buettner, M., Prasad, R., Philipose, M., and Wetherall, D. Recognizing daily activities with rfid-based sensors. In *Ubicomp '09: Proceedings of the 11th international conference on Ubiquitous computing*, ACM (New York, NY, USA, 2009), 51–60.
5. Chen, L., Nugent, C. D., and Wang, H. A knowledge-driven approach to activity recognition in smart homes. *IEEE Trans. Knowl. Data Eng. 24*, 6 (2012), 961–974.
6. Grau, B. C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P. F., and Sattler, U. OWL 2: The next step for OWL. *Journal of Web Semantics 6*, 4 (2008), 309–322.
7. Helaoui, R., Niepert, M., and Stuckenschmidt, H. Recognizing interleaved and concurrent activities using

qualitative and quantitative temporal relationships. *Pervasive and Mobile Computing 7*, 6 (2011), 660 – 670.

8. Helaoui, R., Riboni, D., Niepert, M., Bettini, C., and Stuckenschmidt, H. Towards activity recognition using probabilistic description logics. In *Activity Context Representation: Techniques and Languages*, vol. WS-12-05 of *AAAI Technical Report*, AAAI (2012), 26–31.
9. HÃűlzl, G., Kurz, M., and Ferscha, A. Goal processing and semantic matchmaking in opportunistic activity and context recognition systems. In *The 9th International Conference on Autonomic and Autonomous Systems (ICAS2013)* (March 2013), 7.
10. Hong, X., Nugent, C., Mulvenna, M., McClean, S., Scotney, B., and Devlin, S. Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive Mob. Comput. 5*, 3 (June 2009), 236–252.
11. Knox, S., Coyle, L., and Dobson, S. Using ontologies in case-based activity recognition. In *FLAIRS Conference*, H. W. Guesgen and R. C. Murray, Eds., AAAI Press (2010).
12. Knublauch, H., Fergerson, R. W., Noy, N. F., and Musen, M. A. The protégé owl plugin: An open development environment for semantic web applications. In *Proceedings of the Third International Semantic Web Conference*, vol. 3298 of *Lecture Notes in Computer Science*, Springer (2004), 229–243.
13. Koller, D., and Friedman, N. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009.
14. Kurz, M., HÃűlzl, G., Ferscha, A., Calatroni, A., Roggen, D., Tröster, G., Sagha, H., Chavarriaga, R., del R. Millán, J., Bannach, D., Kunze, K., and Lukowicz, P. The opportunity framework and data processing ecosystem for opportunistic activity and context recognition. *International Journal of Sensors, Wireless Communications and Control, Special Issue on Autonomic and Opportunistic Communications* (2011), 102–125.
15. Latfi, F., Lefebvre, B., and Descheneaux, C. Ontology-based management of the telehealth smart home, dedicated to elderly in loss of cognitive autonomy. In *OWLED*, C. Golbreich, A. Kalyanpur, and B. Parsia, Eds., vol. 258 of *CEUR Workshop Proceedings*, CEUR-WS.org (2007).
16. Lukasiewicz, T. Expressive probabilistic description logics. *Artif. Intell. 172*, 6-7 (2008), 852–883.
17. Lukowicz, P., Pirkl, G., Bannach, D., Wagner, F., Calatroni, A., Förster, K., Holleczek, T., Rossi, M., Roggen, D., Tröster, G., Doppler, J., Holzmann, C., Riener, A., Ferscha, A., and Chavarriaga, R. Recording a complex, multi modal activity data set for context recognition. In *Proceedings of ARCS '10 - 23th International Conference on Architecture of Computing Systems*, VDE Verlag (2010), 161–166.
18. Lutz, C., and Milicic, M. A tableau algorithm for description logics with concrete domains and general tboxes. *J. Autom. Reasoning 38*, 1-3 (2007), 227–259.
19. Niepert, M., Noessner, J., and Stuckenschmidt, H. Log-linear description logics. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)* (2011), 2153–2158.
20. Noessner, J., and Niepert, M. Elog: a probabilistic reasoner for owl el. In *Proceedings of the 5th international conference on Web reasoning and rule systems*, Springer (2011), 281–286.
21. Patterson, D. J., Fox, D., Kautz, H., and Philipose, M. 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 (Washington, DC, USA, 2005).
22. Riboni, D., and Bettini, C. Cosar: hybrid reasoning for context-aware activity recognition. *Personal Ubiquitous Comput. 15*, 3 (Mar. 2011), 271–289.
23. Riboni, D., and Bettini, C. Owl 2 modeling and reasoning with complex human activities. *Pervasive and Mobile Computing 7*, 3 (2011), 379–395.
24. Riboni, D., Pareschi, L., Radaelli, L., and Bettini, C. Is ontology-based activity recognition really effective? In *Proceedings of the Ninth Annual IEEE International Conference on Pervasive Computing and Communications Workshops*, IEEE (2011), 427–431.
25. Richardson, M., and Domingos, P. Markov logic networks. *Machine Learning 62* (2006), 107–136.
26. Riguzzi, F., Lamma, E., Bellodi, E., and Zese, R. Bundle: A reasoner for probabilistic ontologies. *International Conference on Web Reasoning and Rule Systems. LNCS, Springer* (2013).
27. Schneider, M. Plan recognition in instrumented environments. In *Proceedings of the 5th International Conference on Intelligent Environments*, vol. 2 of *Ambient Intelligence and Smart Environments*, IOS Press (2009), 295–302.
28. Springer, T., and Turhan, A.-Y. Employing description logics in ambient intelligence for modeling and reasoning about complex situations. *J. Ambient Intell. Smart Environ. 1*, 3 (Aug. 2009), 235–259.
29. Wang, S., Pentney, W., Popescu, A.-M., Choudhury, T., and Philipose, M. Common sense based joint training of human activity recognizers. In *Proceedings of IJCAI 2007* (2007), 2237–2242.
30. Yamada, N., Sakamoto, K., Kunito, G., Isoda, Y., Yamazaki, K., and Tanaka, S. Applying ontology and probabilistic model to human activity recognition from surrounding things. *IPSJ Digital Courier 3* (2007), 506–517.