International Journal of Social Robotics (2019) 11:437–462
https://doi.org/10.1007/s12369-019-00513-2

Understanding of Human Behavior with a Robotic Agent Through Daily Activity Analysis

Ioannis Kostavelis[1] · Manolis Vasileiadis[2] · Evangelos Skartados[1] · Andreas Kargakos[1] · Dimitrios Giakoumis[1] · Christos-Savvas Bouganis[2] · Dimitrios Tzovaras[1]

Accepted: 3 January 2019 / Published online: 18 January 2019

Abstract

Personal assistive robots to be realized in the near future should have the ability to seamlessly coexist with humans in unconstrained environments, with the robot's capability to understand and interpret the human behavior during human–robot cohabitation significantly contributing towards this end. Still, the understanding of human behavior through a robot is a challenging task as it necessitates a comprehensive representation of the high-level structure of the human's behavior from the robot's low-level sensory input. The paper at hand tackles this problem by demonstrating a robotic agent capable of apprehending human daily activities through a method, the Interaction Unit analysis, that enables activities' decomposition into a sequence of units, each one associated with a behavioral factor. The modelling of human behavior is addressed with a Dynamic Bayesian Network that operates on top of the Interaction Unit, offering quantification of the behavioral factors and the formulation of the human's behavioral model. In addition, light-weight human action and object manipulation monitoring strategies have been developed, based on RGB-D and laser sensors, tailored for onboard robot operation. As a proof of concept, we used our robot to evaluate the ability of the method to differentiate among the examined human activities, as well as to assess the capability of behavior modeling of people with Mild Cognitive Impairment. Moreover, we deployed our robot in 12 real house environments with real users, showcasing the behavior understanding ability of our method in unconstrained realistic environments. The evaluation process revealed promising performance and demonstrated that human behavior can be automatically modeled through Interaction Unit analysis, directly from robotic agents.

Keywords Human behavior understanding · Daily activities interpretation · Interaction Unit analysis · Bayesian networks · Mobile robots

1 Introduction

The field of social robotics has garnered a significant amount of attention from the research community, in order to allow artificial agents' entry at homes and everyday life, aiming to assist humans at social, physical and cognitive level [25]. The necessity for artificial agents to operate in a large spectrum of diverse applications raises significant challenges, the most outstanding of which involve the robot's mobility in human populated environments [8], the demand for context related robotic operational behavior in crowded environments [7]

✉ Ioannis Kostavelis
gkostave@iti.gr
http://www.iti.gr

Manolis Vasileiadis
m.vasileiadis16@imperial.ac.uk

Evangelos Skartados
eskartad@iti.gr

Andreas Kargakos
akargakos@iti.gr

Dimitrios Giakoumis
dgiakoum@iti.gr

Christos-Savvas Bouganis
christos-savvas.bouganis@imperial.ac.uk

Dimitrios Tzovaras
tzovaras@iti.gr

[1] Information Technologies Institute, Centre for Research and Technology Hellas, 6th Km Charilaou-Thermi Road, 57001 Thermi, Thessaloniki, Greece

[2] Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK

and the prerequisite for human behavior understanding from an artificial agent [38]. Several methods in the field of social robotics offer solutions that sufficiently tackle many of the aforementioned barriers [13,21]. However, the field of human behavioral understanding with artificial agents has witnessed more superficial solutions [2,38].

A number of definitions about human behavior exist in the literature, where despite the fact that the majority of them argue regarding the actual factors that influence and form a human behavior (e.g. genetics or social factors, etc.), they all converge on the aspect that the human behavior can be analyzed by the way humans act and interact, given an external or internal stimulus, with other humans or objects of the environment [4,47,53,61]. From a psychological point of view, understanding of human behavior involves the analysis of the psychological signals, the empathy observation and other social observations that frequently require the expertise of psychologists to formulate a behavioral model [40]. However, from an engineering point of view, such psychological signals are difficult to be modeled through an artificial agent and, thus, more quantitative observations should be extracted. This is typically performed by studying human daily activities and extracting patterns of reaction upon humans' interaction with their environment during these activities. Moreover, through observation and modeling of daily activities, it is also feasible to understand normal and abnormal behaviors revealing pathological situations [53]. Thus, it can be reasonably deduced that an artificial agent can extrapolate the observations obtained during daily activities monitoring to formulate human behavior patterns. Furthermore, the idea of analysing patterns of daily activities for the determination of the human behavior relies on the fact that artificial agents should be able to assess human behavior by monitoring the subordinate steps of an ongoing activity and decipher whether or not a human needs assistance [9]. This attribute comprises an interesting pre-condition for a higher level of autonomy and proactivity in human–robot interaction, as it constitutes a fundamental block in the formulation of robot behavioral models [14].

Towards addressing this challenge, the current work is focused on the human behavior understanding and it is grounded on the belief that robots need to understand the behavior of humans at various levels of abstraction in order to operate in a human compatible and socially acceptable manner. A starting point for the apprehension of human behavior can be obtained by systematically analyzing the way daily activities are performed. Therefore, we adopted the Interaction Unit analysis which is a methodology that allows an in-depth activity decomposition, while analytically outlining the way actions are synthesized into complex daily activities, and simultaneously determining the type of objects manipulated by the human in each action. Each Interaction Unit is associated with a specific behavioral factor that allows human behavior specificities identification. The modeling of human behavior is addressed with a specifically designed Dynamic Bayesian Network that operates on top of the Interaction Unit framework, offering quantification of the behavioral factors and activity recognition. Since the developed method targets operation onboard a robotic platform, custom-tailored light-weight solutions for human actions and manipulated objects detection have been developed. Specifically, for human action recognition a skeleton based algorithm has been designed, while for the differentiation among large and small manipulated objects, cluster-based tracking solutions have been implemented, both utilizing depth data from the robot's sensors. To allow efficient robot monitoring of an activity in progress, an automated procedure for selection of robot parking positions within the environment has also been developed, respecting human's personal space and maximizing the view of the area of interest related to the ongoing activity. A representation of the robot during the activity "having a meal" is graphically illustrated in Fig. 1.

At this point, it should be stressed that the objective of this work is not the establishment of a classical human activity recognition method by assigning a discrete activity label. It aims to go one step deeper than this and further scrutinize the recognized activity through the IU analysis that produces meaningful deductions on *how* the activity has been performed in order to understand the normal and abnormal behavior of the user based on the way that s/he executes the daily activities. Thus, this will offer an in-depth understanding of human behavior through the monitoring of well-defined daily activities modeled by the IUs.

To the best of our knowledge, this is the first complete work in the field of social robotics where human behavior understanding is performed through a step-wise IU analysis of human activities and has been explicitly integrated on an artificial agent providing a bottom-up analysis of human actions and objects monitoring, all coupled with robot social navigation strategies. Albeit the fact that the IU analysis is not a new method, its formulation through a DBN and its integration with an active robotic agent for the understanding of human behavior through daily activities constitutes a unique asset of our method. The analytical contributions of this work are summarized as follows:

- The enhancement of the theoretical work of Interaction Unit analysis with features that will enable human behavior understanding in realistic environments.
- The modeling of the Interaction Unit analysis with a machine interpretable method suitable to operate onboard a robot.
- The integration of the collateral subordinate modules required for activity monitoring with robot's mobility.

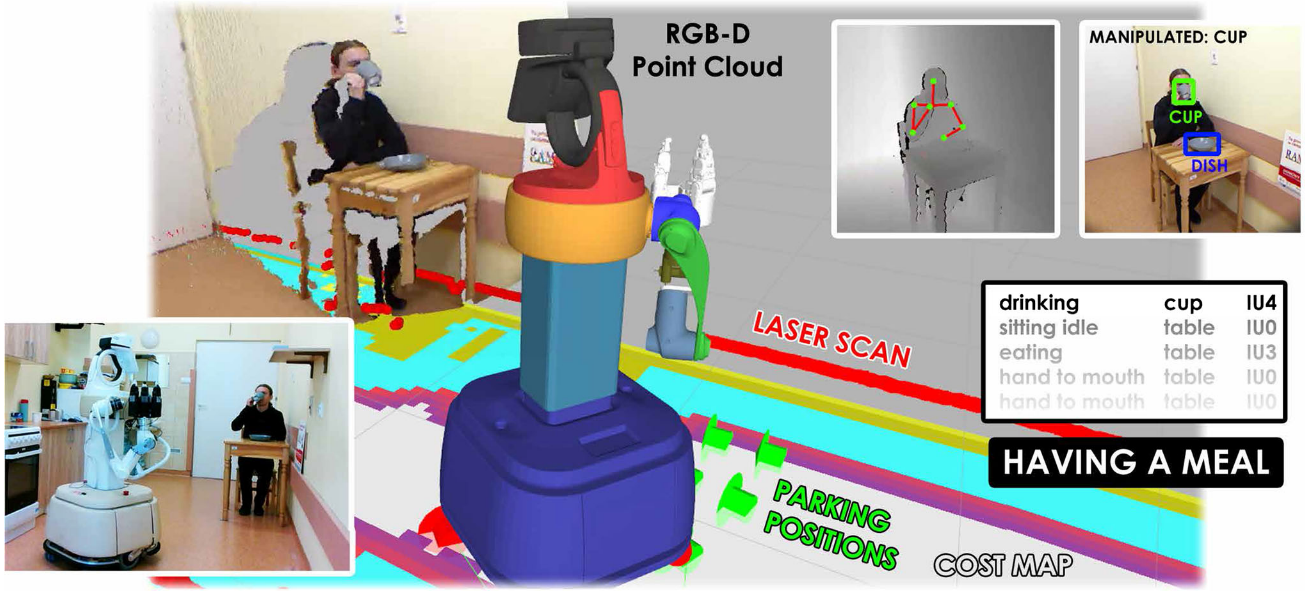


Fig. 1 An illustrative example of the developed framework. On the left, the reference frame of the environment setup along with the robot and the human performing the "having a meal" activity is exhibited. In the middle, the robot visualization (rviz) provides details about the environment apprehension of the robot perspective, where the RGB-D point cloud and the laser scans are presented. Moreover, in the rviz the auto-generated robot parking positions along with the metric cost map are visualized. On top right, the reference frames from the on-board RGB-D sensor demonstrated the skeleton tracking and the detected small objects in the scene are proposed. On the bottom right, the inferred IU steps related to the current robot observations are appended

- To realization of a unified framework for human behavior monitoring capable of operating with limited sensorial input of a robot.

The rest of the paper is organized as follows. In Sect. 2 we present the state of the art on daily activity analysis and human behavior understanding. Section 3 provides insights about the selected activities and their decomposition with the Interaction Unit analysis. Section 4 presents the modelling of the human behavior with a Dynamic Bayesian Network, while Sect. 5 analyzes the peripheral components required for the human action recognition, the detection of manipulated objects, the autonomous selection robot's parking position, and provides implementation details. Section 6 concerns the experimental evaluation of the presented methodology and conclusions are drawn in Sect. 7.

2 The State of the Art

Human activity analysis and recognition has been widely studied during the last years, as it is being leveraged in a variety of computer vision applications including surveillance and patient monitoring systems, as well as systems that involve interactions between people and machines [1]. Moreover, activity recognition and analysis plays a key role for supporting people in their activities of daily life. Consequently, studying the way humans perform their daily activities comprises a reliable starting point for the extraction of required contextual information for the formulation of human behavior models [18]. The current section firstly analyzes the existing works in the field of human activity recognition and afterwards, discusses the existing applications for the establishment of human behavioral models either with robotic agents or with a grid of sensors installed in the environment.

2.1 Human Activity Recognition and Analysis

Following Ziaeefard's recent literature survey [60], the majority of activity recognition methodologies concern the study of the sequential execution of atomic actions. A common separation line between these methodologies comprises the way atomic actions are recognized. Martinez et al. [26] presented a method for silhouette based human action modeling and recognition. The motion templates matching of human actions was performed using motion history of depth images (MHI) and, then, these templates were projected into a new subspace using the Kohonen Self Organizing features. On top of it, Hidden Markov Models (HMM) were used to track the mapped behavior on the temporal sequences of MHI. This approach achieved significant performance, yet has not been evaluated on recent datasets recorded in realistic environment and has never been integrated into a mobile robotic agent. This parameter necessitates among others the positioning of a robot in such a place that will sufficiently observe the scene. Efficient space-time representation of humans based on 3D skeleton data is of great importance in the field of human activity modeling, with a comprehensive survey of the developed methods having been conducted by Han et al. and can be found in [17]. In the work described in [45], the authors presented a method suitable for recognizing activities from 3D human skeleton coordinates. After a key poses extraction step, which constituted the atomic action template, a classification scheme was applied. The key poses frame was set by minimizing the kinetic energy, which was calculated over the joints movements and combined with a Support Vector Machine (SVM) classifier. The method achieved over than 92% and 84% accuracy on the Cornell Activity and MSR Action3D datasets, respectively. Santos et al. [41] introduced a novel approach that improved the segmentation accuracy in human action sequences. The method addressed the temporal segmentation problem of body part trajectories in Cartesian Space, utilizing Discrete Fast Fourier Transform and Power Spectrum as features. The trained classifier was a Dynamic Bayesian Network (DBN) and the entropy of its inference used to adjust continuously the parameters in a sliding window. In this work the adaptation of the sliding window parameters has been evaluated on UC-3D dataset and shown precise results, with an overall ratio of 95%. In a similar method which also relied on 3D skeleton features, Piyathilaka and Kodagoda [30] presented an algorithm that incorporated the importance of weights for skeleton 3D joints used, to train a DBN. It was assumed that each human activity is a collection of different poses that evolve over time and its evaluation on CAD-60 dataset exhibited on average 75% classification accuracy on activities such as “medication intake” and “cooking”. However, the evaluation data included mostly frontal and not occluded recordings that do not resemble closely to realistic human daily activities recordings. Faria et al. [12] employed 3D skeleton features extracted from RGB-D images to train multiple classifiers through Dynamic Bayesian Mixture Models (DBMM). The method was evaluated on CAD-60 dataset with 12 different human activities and exhibited remarkable performance i.e. more than 90% classification accuracy in all cases overcoming other state-of-the-art methods ranked at the CAD-60 website. The authors in [50], computed heterogeneous features to decompose an activity into sub-activities, where for each activity an HMM was trained utilizing those features. The model relied on a maximum-entropy Markov model, where a two-layered hierarchical structure was used. The evaluation of the method was performed on more realistic data and achieved poor performance in activities such as “cooking” (43%) and “pill box manipulation” 56%, further underlining that activity recognition under realistic environments is a very challenging task, due to partial observability of the environment and the human and noisy measurements. Rahmani et al. [32] combined the discriminative information from depth images and 3D joint positions. Their method used depth gradient histograms, local joint displacement histograms and joint movement occupancy volume features. Random Decision Forests (RDF) were exploited for feature pruning and classification, leading to increased performance evaluation on MSR Daily Activity Dataset (more than 92%) which is more realistic than the CAD-120 one, however a per class performance is not available.

By considering the objects in the environment, the authors in [35] employed salient proto-objects for unsupervised discovery of object and object-part candidates and used them as a contextual cue for activity recognition. Since object knowledge alone is not adequate to discriminate activities, they used the Space Time Interest Points (STIP) motion descriptors, which were then passed into a multi-class SVM. This method achieved sufficient performance for classes such as “having a meal” (66.7%), and “medication intake” (83.3%), however the classification accuracy was degrading with the environment complexity increase. Similarly, the work in [22] introduced a complete framework that observed a scene with a human and objects for a specific time frame. The ultimate goal was to anticipate the future activities for the given time frame. Each activity was modeled with a Conditional Random Field, where a given graph structure had two types of nodes, namely sub-activity and object. This was a complete framework that connected activities with objects, while partially dealt with the temporal segmentation issue. Another STIP based method proposed in [59], relied on features for depth-based action recognition. Different interest point detectors and descriptors were combined to form various STIP features, while the bag-of-words representation and SVM classifiers were used for memorization. The main drawback of this method is that the appropriate combination of detector/descriptor needs to be found in each testing example to achieve adequate performance. However, the right combination ranks the papers performance within the state of the art and, regarding the CAD-60 dataset, the method achieved increased performance in classes of interest i,e. “medication intake” and “cooking”, 67% and 100% classification accuracy respectively.

The authors of [56] introduced an extension of local binary feature descriptors suitable for activity recognition tasks. The descriptor is suitable for real-time applications due to the computational advantage of computing a bag-of-words representation with the Hamming distance. This method retained real time performance, however the reported results cannot be considered as promising, since it has not been evaluated on realistic datasets i.e. with partial human observability and variation on recording distances. The work in [28] presented a descriptor for activity recognition from videos using a depth sensor. It described the depth sequence using histograms, capturing the distribution of the surface normal orientation in the 4D space of time, depth, and spatial coordinates. This descriptor has been evaluated on MSR action dataset and a new acquired dataset and yielded more than 88% overall classification accuracy. Compared to the work referred in [59], it had the advantage that there is no need to search for the appropriate combination of features. A different approach to the problem was introduced by Wang et al. [55], where a new representation of human actions, i.e. actionlets, proposed, which deals with the errors of the skeleton tracking and characterizes better the intra-class variations. A data mining algorithm was proposed to discover the discriminative actionlets, offering the attribute to be highly representative of one action and highly discriminative compared to the other actions. A multiple kernel learning approach was also employed to learn an actionlet ensemble structure, yielding remarkable performance results, yet the computational cost is considered to be rather expensive. An additional scalable approach for activity recognition based on objects is described in [57], however, the objects in this work were tagged with RFID labels, which means that such an approach would not be feasible in real life applications.

2.2 Activity Recognition with Robots

Although the above mentioned works reflect the current state-of-the-art in the field of activity recognition, in their majority they are designed to operate on video sequences for different types of applications, rather than for actual human robot interaction. For the realization of social robots capable of operating in human populated environments, the understanding of the human actions and the respective robot reaction is imperative [7], while monitoring should be established in these occasions typically through the robot's onboard sensors. Thus, a more relevant work is the one described in [9], where the authors modeled pose trajectories using directions traversed by human joints over the duration of an activity and represented the action as a histogram of direction vectors. The descriptor benefited computational efficiency as well as scale and speed invariance, however the experiments were performed in segmented sequences. In addition, in [30] the authors developed a human activity recognition method by applying 3D skeleton features directly on a DBN model. Although this method is similar to the one proposed herein, the authors did not consider the objects that the person interacts with and, therefore, important information about modeling of human behavior was omitted. In a more recent work, Coppola et al. [10] introduced a method for automatic detection of human interaction from RGB-D data that enabled social activities classification. In this work the authors defined a new set of descriptors, suitable to operate in realistic data, while developed a computational model to segment temporal intervals with social interaction or individual behavior and tested the method on their own publicly available dataset. In addition, the authors in [31] demonstrated the effectiveness of DBN in time-dependent classification problems, where in their work reported experimental results regarding semantic place recognition and daily-activity classification.

The authors in [33] developed an activity recognition framework suitable for industrial applications. In this work, feature vectors appropriate for recognition have been proposed, however, the learning framework was trained only with human gestures and simple actions related to the manipulation of an object without actual object detection and tracking feedback. The main drawback of this approach is that the performance significantly degrades when a human interacts with objects due to occlusions that stem from the presence of objects. In another work introduced in [22], the authors incorporated object affordances into human activity recognition. Specifically, a joint model of the human activities and object affordances as a Markov random field was designed, where the nodes represented objects and sub-activities, and the edges represented the relationships between object affordances, their relations with sub-activities, and their evolution over time. Although this method achieved promising results, the evaluation was constrained to sequences acquired from the frontal view of the user. Tackling the issue differently, the authors in [29] used a laser range finder to build stochastic models of the observed movement patterns thus, formulating predictive models of the human activities in terms of their in-between interaction. Yet, this method proved more suitable for high-level activity inferences and less appropriate for human behavior understanding.

2.3 Behavior Understanding from Daily Activities Monitoring

Understanding of user behavior through daily activities monitoring is a demanding task given that simultaneous modeling of human actions and their respective interactions with objects should be determined [39]. When it comes to robotic applications the task becomes even more challenging since the amount of onboard sensing devices is limited, while the understanding of human activities mostly boils down to finding good representations of the sensed primitives [38]. The partial observability of the scene primitives and the uncertainty of human actions gave thrust to the development of strategies that express the perception uncertainties, the stochastic human behavior and the typical mission objectives with explicit Partially Observable Markov Decision Process (POMDP) models [44]. Authors in [18] proposed a knowledge driven method for automatically generating activity analysis and recognition based on POMDP models and context sensitive prompting systems. The approach starts with a description of a task and the environment in which this task will be carried out, using a method that is relatively easy to generalize in various environments.

Contrary to the earlier works, the present one aims at the introduction of an autonomous human behavior understanding approach with a mobile robotic agent, capable to infer normal and abnormal behavior on the sequence of the ongoing set of itemized actions and objects that synthesize the expected activity. It takes advantage of a robot's mobility, to position itself in a spot that will enable light-weight human and environment perception solutions to operate in realistic conditions, compensating occlusions and sensors' limitations. It takes into account the human personal space, ensuring the cohabitant's comfort, while the robot observes the daily actions from automatically generated parking positions. Moreover, this method integrates the Interaction Unit analysis with a robotic vision system, instead of utilizing different sensors scattered in the environment, to sense the human movements and the environment changes. It interprets the Interaction Unit analysis with DBN models enabling human behavior monitoring instead of simply using it as a prompting system.

3 Problem Formulation with Interaction Unit Analysis on Daily Activities

3.1 Introduction to Interaction Unit Analysis

IU analysis was initially introduced by Ryu and Monk [37] and comprises a psychologically motivated approach for transcoding interactions relevant for fulfilling a certain task. It has been mainly inspired by the problems encountered by system developers when designing the interactive behavior of novel hand-held devices to be applied in human computer interaction applications. In such applications, IU specifies the visible system state that leads the user to take some action. In addition, the IU makes explicit the state of the goal stack at the start and end of the unit, and the mental processes (recall, recognition, or action) required. In our case the system state is considered to be the environment state (e.g. objects) and the actions correspond to the user's actions within daily activities. The mental processes correspond to the behavioral factors that accompany each human action. Through IU analysis, the intimate connection between human actions and the environment can be described in a format suitable for user-machine interaction. Thus, the IU analysis is used to formalize a machine interpretable task description. Specifically, each task is decomposed into a sequence of IUs that describe the task that needs to be fulfilled by the human in order to execute it.

After the introduction of the IU methodology from Ryu and Monk, Grześ et al. [15], developed a representative work, according to which IU was utilized to analyze specific sequences of actions which were later encoded into a POMDP framework. This POMDP acted as a prompting system to assist people with dementia and developmental disabilities. However, the proposed system utilized an ad-hoc method for transcoding the IU analysis into the POMDP model. The main drawback was that while each of the factors was well defined, fairly detailed and manual specification was required to enable the translation. The idea behind IU analysis was to guarantee the logical adequacy of users' task sequences and to model those sequences in a meaningful way, suitable to be interpreted by a machine. In the paper at hand, the IU methodology is extended to analyze the activities of daily living in a machine interpretable manner, by gathering observations regarding human's actions and the environment state, using the robot vision system in order to understand the human behavior.

The first set of "activities of daily living" (ADL) has been been identified by gerontologists to allow the assessment of the level of autonomy of elderly persons [20]. In accordance to [18], activities of daily living can be decomposed into simple atomic actions which if they are associated with specific behavioral factors, they can formulate a nominal human behavior within an activity. The first step of the work proposed herein, comprises the identification of the type of activities that will be monitored by the artificial agent. Based on the work presented in [6,48], it is typical for older adults to face difficulties with the execution of complex tasks from activities of daily living, e.g. cooking, hygienic procedures, medication intake, dressing up, preparing meals, shopping, eating etc. Thus, the criteria utilized for the activities selection involve their daily repeatability (participants should be familiar with the examined activities), the natural environment in which they take place (typically house environment) and their ability to be decomposed into a sequence of nominal atomic actions. In this work we refer to atomic actions (or simply "actions") as the elementary process of a human doing something, mostly relevant to physical motion e.g. the action "hand to mouth". Activities are considered as a sequence of N executed atomic actions the time continuity of which can result in the completion of a task, e.g. several "hand to mouth" actions indicate the activity "eating". A syndetic notation among the identified actions and the behavioral factors is described in [36,37], where through the IU analysis a notation that allowed the conjunction of atomic actions with behavioral factors is provided. The sequential expression of such behavioral factors constitutes a human behavior through a systematically defined daily activity.

3.2 Association of Interaction Units with Behavioral Factors

The method developed herein utilizes an adapted variation of the IU analysis to build a model of the user activity by identifying and associating atomic actions with the simultaneously manipulated objects, facilitating feasibility of the identification of environmental pre- and post-conditions with a robotic agent. The behavior analysis builds upon an ideal sequence of behaviors i.e., the simplest sequence of IU steps labeled with specific behavioral factors that would complete an activity. Thus, this list of behaviors is divided into the logical sub-parts, viz. IU steps, that structure an activity. The division into logical subtasks is a standard strategy and people with cognitive impairment commonly stall at the boundary between subtasks [3,27]. The behavioral factors consist of three different elements, namely the "recognition factor, the "recall factor and the "action factor". More precisely:

- The **Recognition factor** (**Rn**) refers to the users perception and understanding with respect to the context of the task given an external stimulus, e.g. the fact that the user recognizes that there is a cup of water on the kitchen table. This can be contrasted with the operation "recall".
- The **Recall factor** (**Rc**) refers to the users ability to remember something without being able to see the required information directly. An example of this is that the user recalls that the pill box is in the drawer and needs to reach it, in order to get the pill box. Recall can be considered in general as a problem for people with dementia and we expect most errors to occur for IUs associated with recall operations.
- The **Action factor** (**Ac**) refers to the users interaction with objects and the environment state and is a shorthand for "recognizing the affordance of acting along with an object", for example opening the cupboard.

3.3 Interpretation of Daily Activities with Interaction Units

Considering the selection of the studied activities in this work, four different types have been found to be most appropriate namely, "cooking", "meal preparation", "having a meal" and "medication intake". We tackle herein each IU step as an adjunct execution of an atomic action over an object, which is coupled with a behavioral factor, that determines the normal and abnormal behavior of the user during the activity. It should be stressed that the validity of each behavioral factor that justifies the expected user's response at each IU step has been reinforced with the expertise of phycologists and neuroscientists[1]. Thus, the incorporation of a new activity in the model can be done manually, since it requires the specification of each action with respect to IU analysis that has to be performed explicitly by psychologists and neuroscientists in order to associate precisely the expected user's response with a behavioral factor. The level of detail that each activity can be analysed in terms of IU steps, depends on the various actions and objects that the robot will be capable of apprehending, given its limited sensorial input, i.e. RGB-D sensor and laser scanners. Additionally, it should be noted that in some occasions, due to visual occlusions (e.g. the human bends towards the open fridge and the robot is parked behind him/her) further information to the duration of this IU step is not feasible to be extracted, therefore it is expected that some activities modelled herein will not include all the behavioral factor labels. Indicatively, Table 1 discusses the IU analysis for the activity "medication intake". Specifically, in this table the entity **IU** is referred to the ID of each interaction unit, the sequence of which form, the pattern of the activity; **Object** is the type of the large or small object involved in the respective activity (see Sect. 5.4). The entity **Action** corresponds to the atomic action expected to be performed by the user (see Sect. 5.3). Considering the **Behavioral Factors**, it encapsulates the actual label of the respective IU step in order to be encoded later in the DBN model, the label of which has been provided to the sequence after the experts' analysis. Moreover, **Normal response** and **Abnormal response** describe the user's typical or abnormal behavior respectively, that should be observed by the end of the specific IU. The **Priority** component comprises a strict identifier (with the label mandatory (M) / not mandatory (NM)) that immediately triggers the event of abnormal behavior detection when such an IU step has not been observed at all within a sequence.

In the originally specified IU analysis [18], the effect on each action is coded on the goal stack and the environment implicitly, through the new goal stack and critical environment in the subsequent IU. Each group ends in a dummy IU that shows the goal stack and environment after the last action. In this work, we simplified the model by directly associating each IU behavioral factor with an expected combination of human action and manipulated object instead of having groups of dummy IU steps, allowing the direct modeling with the DBN, the inference of which has physical meaning and is easily interpretable.

4 Probabilistic Modeling of Human Behavior

Having discussed the interpretation of ADLs with IU analysis, a model that allows inference over the observed human and the environment space should be constructed. The majority of the methodologies mentioned in Sect. 2 focus mainly on the identification of the human activities and on the formulation of prompting systems for decision making rather than on the extraction of inference about the human behavior through daily activities. The proposed methodology overcomes the latter issues by exploiting IU analysis along with a DBN. A Bayesian network (BN) is a directed acyclic graph

[1] Full reference of the experts that provided the annotation of the behavioral factors can be found in the "Acknowledgments" section

Table 1 IU analysis and behavioral factors of "medication intake" activity

IU	Human action	Manipulated object	Environment state	Behavioral factor recognition/recall/action	Abnormal response	Priority
1	Reach	Pill box	Pill box closed on table	The pill box is on the table	Forgets where the objects are, looking in the different area. Picks up different than desired object. Easily distracted	M
2	Alter	Pill box	Pill box in hand, opened	The pill box is closed	Opens the box and does not remember the reason	NM
3	Hand to mouth	Cup	Cup with water on table	Take the pill	Gets easily distracted	M
4	Alter	Pill box	Pill box closed	The pill box is opened	Gets easily distracted	M
5	Reach	Pill box	Pill box on table	Place the pill box on table	Gets easily distracted, misplaces the objects	NM
6	Reach	Cup	Cup with water on table	The cup is on the table	Forgets where the objects are, looking in a different area Takes the objects and later does not use them. Pick up different than the desired object	M
7	Hand to mouth	Cup	Cup with water on table	Drink water from the cup	Gets easily distracted	M
8	Reach	Cup	Cup empty on table	Place the cup on the table	Gets easily distracted	NM

(DAG) in which the nodes represent random variables and the edges encode the conditional dependencies between the variables and specify the joint probability distribution over them. A DBN is a BN that represents temporal sequential data and the term "dynamic" refers to the fact that the network is used to model time sequences, where in the examined case, it corresponds to the interpretation of a human behavior through the time interval of a daily activity [5].

4.1 Formulation of IU Related Bayesian Network

The developed DBN model exploits information from the performed human atomic actions and the participation of objects in those actions in order to produce distinct IU steps that related to a behavioral factor. The derived information from both is encoded as two distinct nodes, i.e. random variables in the DBN architecture. It is worth noting that both nodes are observed, not hidden. Specifically, the action recognition module infers about the actions performed by the human given a time interval. Thus, it is possible to have the occurring actions with respect to a global clock as well as their durations. The actions are modeled as a discrete observed variable attaining values equal to the number of actions, $o_1 \in [1, 2, \ldots, N_a]$, where o_1 is the discrete observed variable, N_a is the number of actions and each value represents a specific action, e.g. $1 \rightarrow reach$, $2 \rightarrow open$,..., $N_a \rightarrow close$ etc. More details about the real time action recognition module are provided in Sect. 5.3.

The manipulated objects are modeled as a discrete observed variable o_2 and the dimensionality is equal to the number of objects types, N_{ob} , i.e. $o_2 \in N_{trackingstates}^{N_{ob}}$. Each dimension in this vector is "registered" to a specific object type, for example the first dimension refers to the "cup", the second to the "pill box" and so on, and tracks the existence or absence of a small object on the identified workspace above the segmented supporting surface (Sect. 5.4). It should be noted that modeling of manipulated objects in the DBN could be performed either as a number of unidimensional, multi-state variables or with only one multidimensional, multi-state variable. Both approaches have been examined and none of them exhibited notable superiority against the other. Therefore, we kept the second approach, so as the physical presentation of our method to match with its implementation approach, i.e. two observable variables are modeled through two entities, for the sake of clarity. The global clock is also used to place in the same time frame the manipulated objects along with the performed actions. This combination indicates the manipulated object by a human in conjunction with the respective action at a particular time. Object types that did not appear yet in the examined scene as well as the ones that are present in the workspace at that particular time but are not associated with the concurrently executed action, are not taken into consideration. For example, we consider the case where the user performs the "medication intake" activity and three known objects are initially present in the scene, a "cup", a "pill box" as well as a "remote-control". In the course of a successful execution of the considered activity the states of all the known objects will be affected as follows: all irrelevant, non-visible known objects, such as "pot", "spoon" etc, will never appear in the examined workspace, while the irrelevant "remote-control" object, which is already within the examined workspace, will remain still. For the corresponding dimensions of these object labels the o_2 vector will remain at the "not-detected" and "in-workspace" values respectively, in all time instances. Thus, none of them will be associated with any of the executed actions. Dimensions corresponding to the "cup" and "pill box" labels will receive the "manipulated" state at specific time instances, when movement of these objects is detected. Implementation details about real time object tracking and event triggering can be found in Sect. 5.4.

Concerning the hidden node h_1 of the DBN architecture, it is a discrete variable representing the IU steps for each activity which are associated with a behavioral factor. Thus, the number of possible values that this variable can obtain depends on the respective activity. For example, activity "cooking" has 14 IU behavioral factors and so the obtained values will be in the range of this variable, while activity "medication intake " has 8 behavioral factors, affecting the range of the respective hidden variable accordingly. Each activity is modeled in a separate DBN, where the number of activities we are interested in is four (4). As a result, the methodology trains four different DBNs where the observed variables o_1, o_2 are identical for all DBNs, while the h_1 is a discrete activity specific one, ranging with respect to the modeled activity.

An example DBN structure is illustrated in Fig. 2, for an activity A_i, having N_s number of IU behavioral factors, where the node dependencies are represented by edges between the consecutive BNs. A DBN's structure can be completed by specifying all the nodes and edges that belong to two consecutive slices. Hence, a DBN is defined by two components. The first one, comprises the set of nodes that represent all random variables for a given position (time slice) in the sequence, and the edges between those nodes (the intra edges). This set of nodes and edges forms the BN that will be duplicated along the length of the sequence. The second one, is a set of edges that connect nodes in two consecutive slices in the sequence (the inter edges). The network in time slice t has two observable nodes, the observed action o_1 and the relative translation of objects o_2. Objects and actions are separately recognized using specifically designed custom-tailored solutions to allow operation on a robotic platform, but they are coupled via the h_1 node. Only valid combinations of actions o_i and objects o_i can produce nominally defined IU behavioral factors in each activity. That is, the

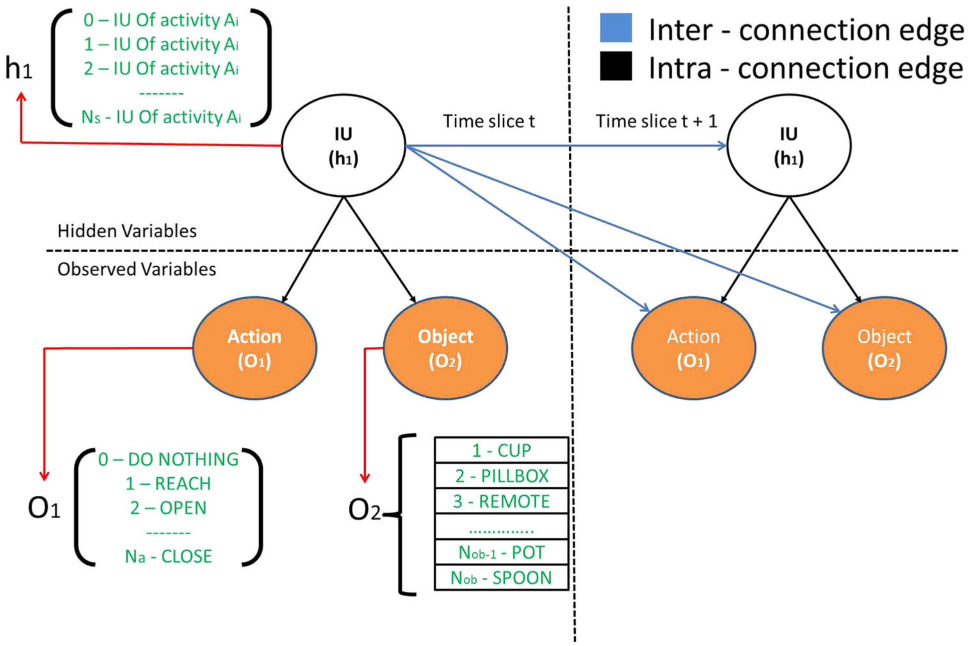


Fig. 2 Illustrative example of the DBN structure. Red arrows are explanatory and describe the variable type and content. Black and blue are edges indicating intra and inter connectivity among variables, respectively

two observable nodes are used as a pair to recognize the current IU behavioral factor, while the appearance of a not valid combination is associated with a control behavioral IU step according to which the user does not perform a relevant action to the examined activity. The interconnecting edges between successive time slices shows the dependence among the IU behavioral factors and the dependence of future actions from the current IU step (behavioral factor). Moreover, they aim at capturing the temporal relations between those steps within each activity separately.

4.2 Training and Inference of DBN Models

In the presented architecture, there are four different activities and, thus, an equal number of DBNs. Each DBN is trained with the respective samples, corresponding to each activity. It is worth noting that input samples retain the same format, i.e. observed nodes are common for all DBNs. Of course, the temporal length might be different in each sample or activity; however it is invariant in the case of DBNs (Fig. 3). This fact will be exploited later, in the inference procedure. The hidden node differs in each DBN model, both in terms of range as well as in its semantic meaning. The latter implies that even when two activities have equal number of IU steps, the ith IU step in activity A_k is different from the ith IU step in activity A_j and correspond to different behavioral factors. To sufficiently train a DBN we need temporally-aligned sequences of the observed and the hidden nodes, which will allow estimating of the Conditional Probability Density (CPD) tables among the variables which are connected through edges. For example, in the case of the action variable o_1 which is discrete, the CPD is a matrix that holds the probabilities of each combination of o_1 and parent values (h_1). The discrete variable o_1 has size N_a and a parent h_1 of size N_s, which will lead to a $N_s \times N_a$ matrix as its CPD.

During the inference phase, query samples that correspond to observed variables o_1, o_2 are provided to the trained DBN models, aiming at the computation of the log-likelihood for all the DBNs. This step leads into probabilities regarding the label of the respective query sequence. The DBN with the highest probability labels the query sample. The most probable DBN on which the query sample belongs to, is further processed by computing its Viterbi path [19,42]. The latter, is a sequence of values of a hidden variable that best explains the observed data. Specifically, in this architecture, given the observed values for T time instances, the Viterbi path will be a sequence of T values of the hidden variable h_1, which corresponds to the T successive IU steps performed within the respective sample. The sequence of the performed IU steps is then exploited to model the normal or abnormal behavior of the user considering specific criterions.

Given that the label of the activity has been inferred by maximizing the log-likelihood, the method associates the IU behavioral factors with the respective IU reference table that describes the ongoing activity. Missing expected IU behavioral factors indicate some abnormal behavior of the ongoing activity; the incident is registered accordingly. The appearance of prolonged repeated patterns of an IU step that is associated to a specific behavioral factor (Rn, Rc, Ac) are considered to be also an abnormality indicating that the user stalled in the boundary between two IU steps. Another parameter that is considered for the modeling of the human behavior is the user specific duration required for each activity to be completed. Since the time steps are constant, i.e. the tempo-

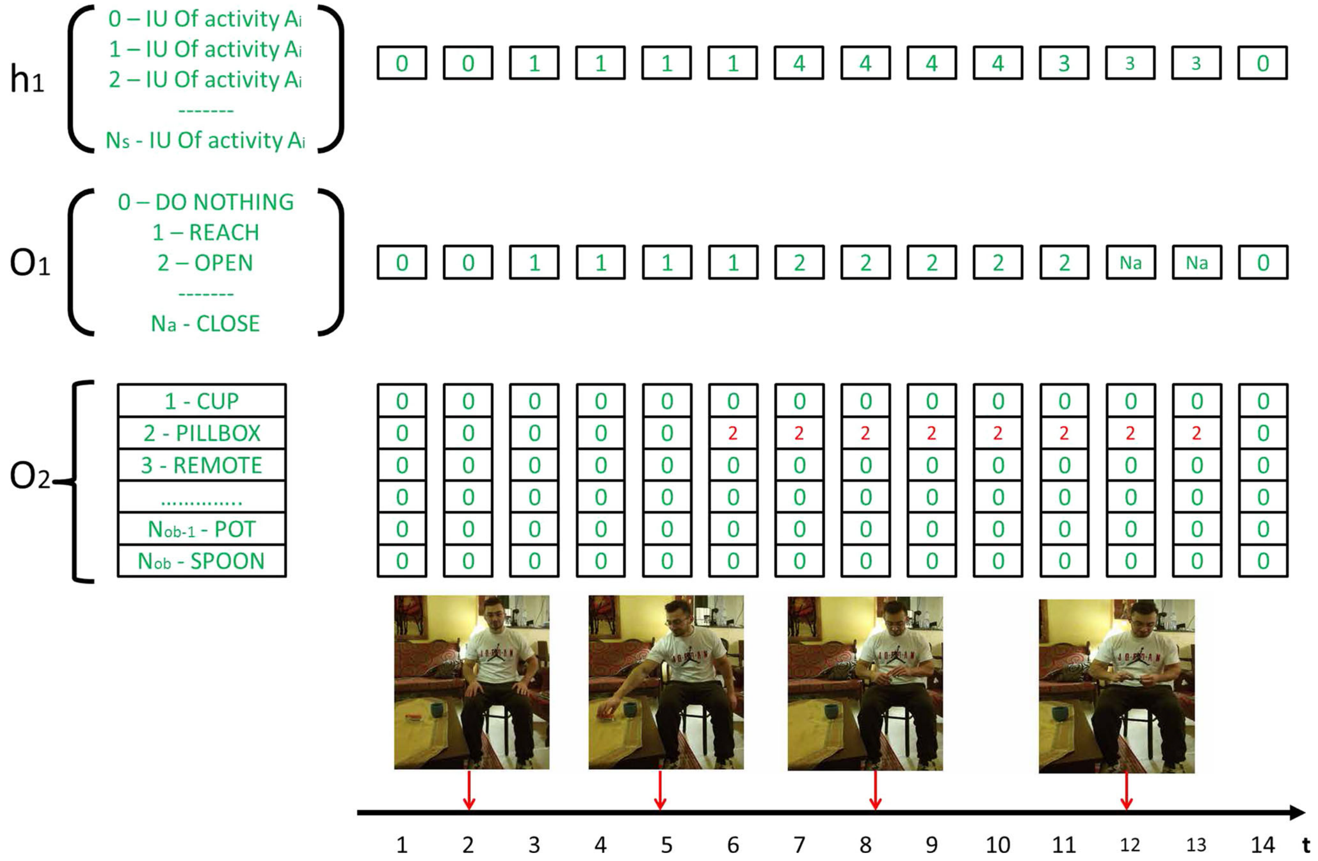


Fig. 3 Indicative description of a training sample. Both hidden and observed variable values are used for training, the dimensionality of each slice is always $N_{ob} + 2$ while the temporal length of the sample may vary, in this particular example it is equal to 14, which means that each time instance t is associated with an examined frame i.e. coupled action and manipulated object

ral difference between t and $t+1$ is equal to the difference in $t+5$ and $t+6$, the Viterbi path is utilized to provide statistics regarding the duration of each IU, the duration of the entire activity as well as the most probable sequence of executing the IUs.

5 Autonomous Robot Monitoring

5.1 Robot Parking Positions Selection

Considering that monitoring of human behavior through daily activities should be performed by a mobile robot equipped with limited vision sensors, the selection of the robot parking position that allows continuous human observation is essential. Our robot is equipped with a single RGB-D camera and laser scanners of approximate 360° field of view and should be able to autonomously select parking poses that will enable it to sufficiently track the human actions and the manipulated objects in the scene, something that is further limited by the on board RGB-D sensor. In addition, a balance between the short and long distances from the camera and the user is required since long distances allow full body view of the human and, hence, better tracking while close distance views are beneficiary for the detection of small objects. Another constraint that should be considered during the calculation of the robot's parking pose is the discreet presence of robot during activity monitoring, respecting thus the human's comfort during human–robot coexistence.

These constraints are tackled herein by developing a custom solution, tailored to the human presence and robot dimensions. Firstly, the human pose in the scene is identified (see Sect. 5.3) and the average human state vector $(X_h, Y_h, Z_h, \theta_h)$ is computed. Then, the human's "personal space", inspired by Hall's [16] proxemics theory, is modeled by centering a Gaussian Kernel around (X_h, Y_h), imposing thus soft constrains regarding the human approaching borders instead of crisp ones, in order to handle occasions with limited free space for robot maneuvering. The robot apprehends the human"personal zone" as a circular area around the human of outer radius greater than 1.2 m [51] and searches for an optimal observation range outside this perimeter for

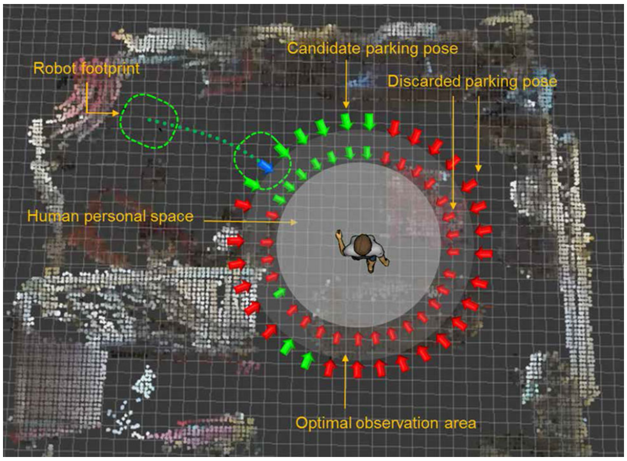


Fig. 4 The autonomous selection of robot parking position scheme. The human personal space and the search space i.e. observation area, are superimposed on a 2D metric map . The generated robot parking positions are presented with arrows among which, the red ones are discarded due to the fact that the robot footprint does not fit among the existing obstacles in the static 2D metric map, those marked with green are the candidate ones and the blue is the selected robot parking position

the identification of a suitable parking pose that satisfies the requirements for human frontal facing and the robot's footprint fitting among the static obstacles of the global metric map. The robot's footprint affordance to the static metric map is controlled with a spatial decomposition technique, i.e. kdTree, by searching with neighborhood radius of size equal to the robot's footprint radius, ignoring thus those poses that intersect among the static obstacles and the footprint's points [34]. The selection of the most appropriate parking pose is performed by applying an Euclidean distance minimization criterion among the robot current pose and the calculated ones, see Fig. 4.

5.2 Global Human Observation in the Scene

The first step towards human action recognition is the robust detection and tracking of the human in the observed scene. However, relying only on the robot's RGB-D sensor for human tracking introduces partial observability of the human actions since the camera's FoV is restricted and human tracking is not feasible during robot manoeuvring; e.g. the robot traverses from its monitoring parking pose towards an activity related parking pose and in the meantime the human is in the middle of an ongoing activity. Therefore, in order to ensure constant situation awareness about the human presence, we fused two specific human tracking algorithms by exploiting data form robot's laser scanners and RGB-D sensor:

RGB-D based human tracking A human detection and tracking framework suitable to operate with low-cost depth sensors at real-life situations addressing limitations such as body part occlusions, partial-body views, sensor noise and interaction with objects [54] has been adopted. In particular, a human template is initialized in the first frame of a tracking sequence, through a two-step initialization process, using as input the human pose estimation provided by the Microsoft Kinect v1 built-in skeleton tracker [46]. The human pose tracker then employs the articulated SDF model [43] which utilizes the articulated skinned human template to track the human pose in sequences of depth images, extracting thus the 3D positions of the skeletal joints required for the recognition of atomic actions. The attributes of this method that facilitate operation on real life robotic applications are provided through a series of complementary tracking features. In particular, the **Free space violation** criterion reduces the possibility of a part of the human template, to not correspond to any part of the input data; The **Body part visibility** is a factor which ensures that only the visible body parts are used in the optimization process of the SDF model, since occlusions of body parts due to obstacles and constraints of the camera's FoV are commonly encountered in realistic environments. Last, the **Leg intersection** criterion counteracts the occasional mix-up of the lower limbs, using a body part representation similar to [52].

Laser based human tracking The laser based human tracking is mainly inspired by the work presented in [24], in accordance to which, the laser scans are clustered according to the distance and a feature vector is extracted for each cluster using specific geometrical features. A random forest classifier is trained with these features to model human presence or absence. During the inference procedure, all the detected pairs of legs that produce a high probability score are considered as potential humans. However, the method produces many false positive observations, yet fusion with skeleton tracking reduced the false positives through a blacklisting procedure [23]. Specifically, when laser-based human tracks are matched with skeleton based tracks in terms of global coordinate system, the human position is locked. The rest of the noisy information of the laser-based tracks are blacklisted, and their confidence degrades with time. In occasions where the human is out of the RGB-D camera FoV, while has been previously matched with the leg-based tracker, the observation is passed to the white list and the monitoring of the human position is resumed.

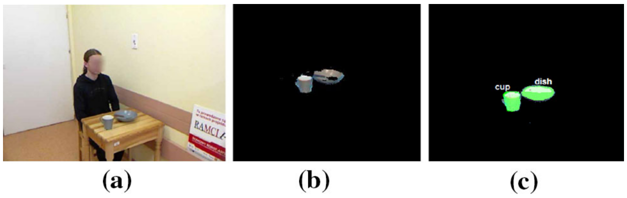
5.3 Human Action Recognition

The human action recognition module is the one developed in [49], which is specifically designed to operate in realistic conditions, with robotic platforms. It employs the tracked humans skeleton joints and by extending the classic EigenJoints [58] method, it improves recognition robustness for a series of actions involved in common daily activities. The method utilizes novel features to take into account action specificities such as joints travelled distance and their evolution trend in subsequent frames in relation to the reference one. In addition, it associates specific actions with information related to the users manipulated objects, taking into account that several actions may be similar, yet performed with different objects e.g. “eating” can be analysed as a “hand to mouth” atomic action with object “spoon” and drinking can be analysed also as a “hand to mouth” action with object “cup”. Moreover, this method operates with continuous input video streams without the demand of pre-segmented action sequences which is a prerequisite on real time gathered data by the on-board camera. Considering also the observed variables of the DBN, the detected actions correspond to the observed variables o_1 fed to the DBN during the training and inference phase.

5.4 Detection and Tracking of Environment Objects

During daily activities, humans interact with a variety of large objects e.g. “fridge”, “cupboard”, as well as with small objects e.g. “cup”, “pot”, “pill box”, which share great heterogeneity between each other. This renders the task of human manipulated objects tracking—by a single detection/tracking method—a challenging task that, necessitates the adoption of holistic solution for differentiation among tracking of large and small objects. Towards this direction, the environment has been organized in a hierarchical fashion where on top of the metric map, a semantic model is constructed, which is a tree structure that retains the relationships among objects-places and describes explicitly the domestic environment in terms of human concepts (Fig. 5).

In this schema small objects are organized in terms of their attributes and their relations to large objects. The position of the large objects is expressed in the same coordinate system as the metric map. The design principle of this component is based on the assumption that the human continuously interacts with objects in the course of the activities of interest; i.e. “cup” when the human drinks water, “table” when the human is sitting at the kitchen table without performing a specific simple activity. Therefore, the monitoring of the large objects operates continuously, considering as the currently large manipulated object, the one that has minimum distance with human.



(a) (b) (c)

Fig. 5 Initialization of the small object monitoring: **a** the RGB reference image, **b** the isolated workspace area and **c** the outcome of the object detection algorithm

The initiation of the tracking component of small manipulated objects relies on the assumption that the human interacts with small objects placed on proximal to him/her large objects. Each large object defines a *workspace* and the small objects in the scene are expected to be present in this workspace. Upon identification of an ongoing activity and given that the robot has reached its parking position, the monitoring of small objects is initialized as follows; firstly, the closest to the user large object is determined e.g. the table, and this information is utilized as a triggering event for the small object detection and tracking. Secondly, RANSAC plane segmentation is applied to extract the dominant plane defined by the large object which acts as a supporting surface for the small objects expected to be in the scene. The area above the supporting surface is defined herein as *workspace*. The convex hull of the large object’s points that satisfy the plane equation is extracted and, thus, a geometrical description of the workspace with respect to the robot’s viewpoint is obtained. The points above the plane are projected on it and the properties of the convex description of the surface are utilized to efficiently crop the scene. On the formulated workspace, a depth-based object detection algorithm [11] is applied and, thus, each pixel is labeled either as one of the known N_ob objects or as unknown. It is revealed that small objects are not statically associated with large objects, however, information from large objects is extracted to determine the type of small objects within its workspace and to deduce whether these objects are manipulated by the user.

To infer which of the detected objects is currently manipulated by the user a depth-based tracking rule is applied, in accordance to which, a voxel grid filtering step merges the points labeled as a specific object in a structure that enables change detection during human manipulations. Comparing all subsequent frames with a reference frame, all labeled clusters are being checked in terms of their occupancy and, hence, all detected objects are being tracked. If a new cluster enters the monitored workspace, object detection is executed specifically on the area defined by this cluster. If any change, removal or addition of a small object takes place, the reference frame is updated accordingly.

The small object tracking system operates in three (3) states for each considered object in the scene, i.e. “not-detected”, “in-workspace”, “manipulated”. When the component launches, all small objects can be either in the “not-detected” or in the “in-workspace” state. While the human executes an activity it is possible to add to the observed workspace a previously unseen but known object. This object will transit from the “not-detected” to the “in-workspace” state. When the human picks up an “in-workspace” object, its state changes to “manipulated”. Tracking states for all

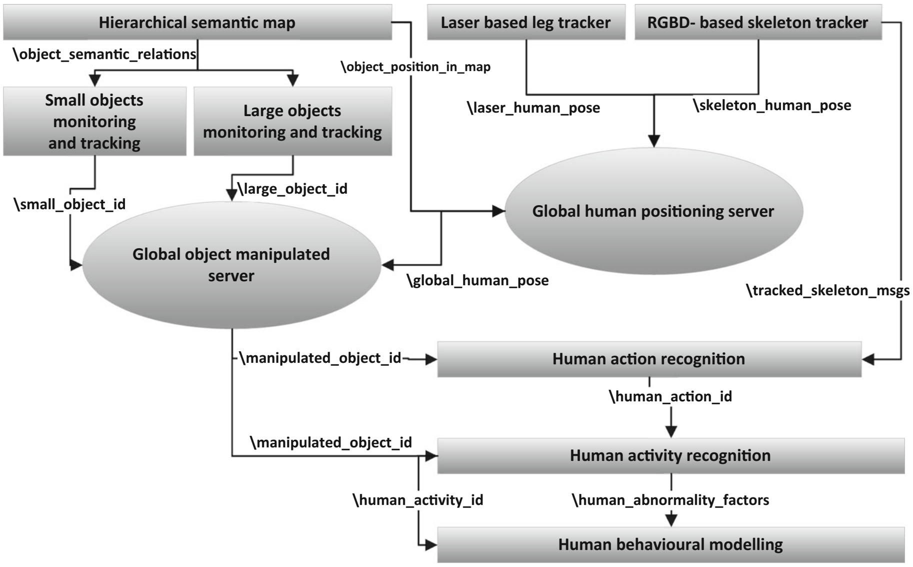

Hierarchical semantic map
\object_semantic_relations
Small objects monitoring and tracking
Large objects monitoring and tracking
\object_position_in_map
Laser based leg tracker
RGBD- based skeleton tracker
\laser_human_pose
\skeleton_human_pose
Global human positioning server
\large_object_id
\small_object_id
Global object manipulated server
\global_human_pose
\tracked_skeleton_msgs
\manipulated_object_id
Human action recognition
\human_action_id
\manipulated_object_id
Human activity recognition
\human_activity_id
\human_abnormality_factors
Human behavioural modelling

Fig. 6 Block diagram of the software components involved in human behavior understanding and their interconnection with specific input output

object labels are reported as an o_2 vector to the DBN at each execution cycle.

5.5 Robot-Wise Implementation Details

The above described subordinate modules have been implemented within the Robot Operating System (ROS) framework and their operation along with the developed software interfaces is exhibited in the architecture diagram illustrated in Fig. 6. An in depth observation of this diagram proves that the simultaneous monitoring of the human actions and the respective human manipulated objects is a challenging task when the observations stem from solely vision input, i.e. RGB-D camera and Laser scanners. The *global human positioning server* is responsible to continuously provide data regarding human's location during the monitoring phase and compensates the situations where the human is out of the FoV of the RGB-D sensor. The *global manipulated-object server* continuously provides observations about the objects manipulated by the human during activity monitoring. Thus, monitoring of the large objects continuously operates as a background process considering the minimum distance among the human position (*global human positioning server*) and the registered large objects in the hierarchical semantic map. Consequently, labels of large objects observations are reported to the DBN retina, yet when changes on the position of a small object is observed, the *global manipulated-object server* infers the label of the respective small object. It is apparent that the *global manipulated - object server* provides high priority to the small objects tracking which is activated after the robot has reached a parking position suitable for monitoring a specific activity.

5.5.1 Runtime Analysis

The run-time performance of the overall autonomous monitoring framework depends on the individual performance of its core modules, described above, as they sequentially provide the information necessary for human behavior inference. The vision algorithms (human tracking, action recognition and object detection and tracking) present the highest computational burden introducing a performance latency in the system.

The overall framework requires approximately 150 ms to capture at least one sample from each module, with the workload being shared over a 2-PC configuration (Intel i7-5930K, NVIDIA GTX970) installed in the robot, i.e. PC1 and PC2. More specifically, human tracking, human action recognition and parking position selection processes run on PC1, while object detection and tracking run on PC2, effectively utilizing the available CPU and GPU resources. The final activity inference is performed on PC1, after all the above modules had concluded their execution. The runtime requirements of each module are described in detail in Table 2.

6 Experimental Evaluation

The evaluation of the proposed human behavior understanding method has been performed on two distinct test sites. The first one concerned the assessment of the method on a simulated room located in the premises of the Medical University of Lublin (LUM) at Poland. This simulated room was utilized for the acquisition of a realistic dataset with daily activities and served for the training and tuning of the DBN models. Additionally, the autonomous operation capacity of our method for human behavior understanding has been ini-

Table 2 Run-time analysis of the overall autonomous monitoring framework: time required to capture at least one sample from each individual module

Robot parking positions selection	< 0.01 s
Human tracking	0.06 s
Human action recognition	0.08 s
Object initial detection	1.0 s (executed once)
Object tracking	< 0.01 s
Activity inference	< 0.02 s (depends on number of samples)

tially evaluated on this simulated environment. The second test site on which the proposed method has been evaluated consists of 12 different real houses with real users located in Barcelona. The robot, with the trained DBN models, had been installed in each house and the user had the opportunity to interact with it for seven days. In this period the selected daily activities were repeated from the user frequently, i.e. once per day and, thus, the capability of the proposed method to perform human behavioral understanding has been evaluated with large scale experiments in diverse and uncontrolled environments.

6.1 Evaluation in a Simulated Room (LUM Test Site)

6.1.1 Dataset Acquisition

The subjects that participated at the LUM test site for the dataset capture were elderly patients, selected by medical personnel based on their mental and physical state and their ability to perform the required activities. During the recording process, each subject was asked to perform a series of everyday activity scenarios for at least three repetitions, while being monitored by the robot's on-board camera.

Specifically, four (4) activities were selected, namely *meal preparation, cooking, having a meal, medication intake*. The amount and type of the activities that the dataset contains was selected by group of the neuroscientists and psychologists associated to this work. The main selection criteria can be found in Sect. 3.1 and was deemed adequate to perform behavior understanding. It should be stressed, that even if the amount of the selected activities is disparate, which renders the issue of activity recognition a relatively easy task that can be addressed by several state-of-the-art methods, their subsequent execution steps (IUs) are essential for the extraction of meaningful deductions for human behavior understanding. For each activity sequence, a specific recording protocol was followed as described in Table 3 the human pose and actions as well as the manipulated objects were detected using the pose tracking, action recognition and environment tracking methods described in Sect. 5, respectively. The resulting captured data provided a realistic depiction of the monitoring conditions that an assistive robotic platform may encounter when deployed in real home environments, making it suitable for the evaluation of the proposed framework. In total 123 discrete activity sequences, performed by 18 different subjects, distributed in the four activities as outlined in Table 4 were captured and used for evaluation purposes. In addition, Fig. 7 exhibits representative sample frames for each scenario.

6.1.2 Evaluation of the Activity Recognition Performance (LUM Test Site)

The first step towards efficient behavioral modeling comprises the classification of the ongoing activity, so as for the robot to be able to understand the type of the ongoing activity, select the most appropriate parking position for close distance observation, recall the respective IU model and infer correctly about the normal or abnormal behavior of the user. The recorded data were used for the training and testing of the four different activities, following a leave one out cross validation procedure. Table 5 presents the detailed results of the classification process, in accordance with which it is revealed that the developed activity recognition method presents high activity classification potential, as it achieved overall precision and recall rates of over 98% respectively.

During the inference phase of the activity recognition module, the observations of human actions and manipulated objects presented sequentially to the DBN's retina and the DBN models were queried iteratively after gathering N amount of observations. Specifically, the system was evaluated progressively in a slice-oriented manner, where the observation window of each slice was set to $N = 30$, i.e. approximately 3 s, while the activity label for the overall sequence was inferred through majority voting; the activity label that appeared in the majority of the N-size slices was used to describe the whole sequence. The selection of the parameter N was a decent compromise among the inference frequency and the situation awareness for the autonomous recognition of a new ongoing activity. According to this parameter, the system expected to gather N pairs of synchronized observations (human simple actions and manipulated objects), and then the DBN models were queried and inferred about the class that the aggregated sample (from the begin of the sequence) belonged to. Therefore, it was observed that during the experimental evaluation, for the first 2–5 slices the trained DBN models could misclassify the activity type of the query sequence. However, when the system gathered more data, the overall class discrimination capability of it

Table 3 The activities recording protocol regarding the environment setup

1. *Meal preparation*	
a.	The user is in the kitchen area
b.	The user opens the fridge
c.	The user gets an object from the fridge
d.	The user closes the fridge
e.	The user moves the object towards the kitchen bench
2. *Cooking*	
a.	The user is in the kitchen area
b.	The ingredient of interest is in the drawer, the pot is in the cupboard
c.	The user opens the drawer and gets the ingredient of interest
d.	The user places the ingredient on the kitchen bench
e.	The user closes the drawer
f.	The user opens the cupboard and gets the pot
g.	The user places the pot on the bench
h.	The user closes the cupboard
i.	The user gets the pot and places it under the sink
j.	The user opens the tap of the faucet and then closes it
k.	The user places the pot on the burner grate
l.	The user turns on the burner grate.
m.	The user waits
n.	The user turns off the burner grate
3. *Having a meal*	
a.	The user moves towards the table
b.	The user sits down
c.	The user starts eating
d.	The user starts drinking
e.	The user stands up from the chair
4. *Medication intake*	
a.	The user sits on the chair near the table
b.	The pill box and a cup of water are nearby, on the table
c.	The user takes a pill and drinks water

Note that the order of the participant’s atomic actions was not binding

Table 4 Distribution of 18 different subjects to the four activities in LUM dataset

	Meal preparation	Cooking	Having a meal	Medication intake
Participants	13	8	10	14
Repetitions	39	20	32	32

was revealed, and the total sequence was more accurately classified. This phenomenon is typically observed in activities that are executed by humans in the same place and are rather difficult to be identified even from a human observer. Specifically, “Cooking” often is misclassified with the “Meal preparation” activity during the beginning of the observation (first slices), since they both take place in the kitchen area, while “Having a meal” can be misclassified with the “Medication intake activity, as both take place in the kitchen table. However, as we propagate deeper in an activity, the discrimination ability of the DBN models increases since different types of objects are involved in each activity.

6.1.3 Evaluation of IU Analysis for Behavior Understanding (LUM Test Site)

For the evaluation of the IU steps detection process, a subset of the overall LUM dataset was selected, including only the activity sequences where the monitored subjects executed all the IU steps. Next, every activity sequence was passed through the respective activity IU step detection module. For

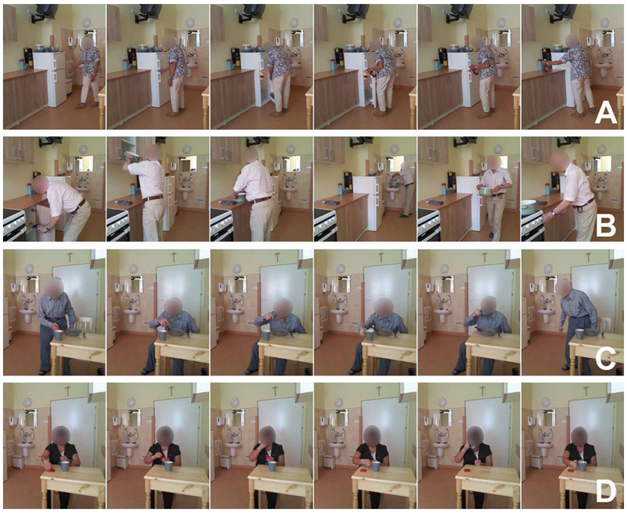


Fig. 7 Sample sequences from the LUM dataset of the four activity scenarios: **a** "meal preparation", **b** "cooking", **c** "having a meal", **d** "medication intake"

each IU step of an activity, the IU step detection rate was estimated, defined as the rate of the activity sequences where the IU step was detected, to the total number of the selected activity sequences. Table 6 summarizes the IU detection rates for the four complex activities. It should be pointed out that the IU analysis is not utilized to boost the performance of the DBN based activity recognition, but it is used as a mean to interpret the inference of the DBN model by analyzing the behavioral factors of each element of the queried Viterbi path and, thus, understand the human behavior.

In the "meal preparation" activity, a 73.71% average IU step detection rate was achieved. IU steps 1 and 2, which correspond to "reaching and opening the fridge", achieved very high detection rates, over 89%. IU 3 "close the fridge" presented relatively lower detection rate, and the reason is the challenging viewpoint and corresponding occlusions during human tracking, as the user would usually overlap with the fridge door after opening it, thus making it harder to track the "close the door" action. In the "cooking" activity, all of the IU steps achieved a detection rate of over 80%, leading to an overall IU step detection rate of 87.50%. In the "having a meal activity", an 95.62% overall IU detection rate was achieved, with the two mandatory IU steps, namely IU 3 "start eating" and IU 4 "start drinking", achieving almost perfect detection rates. IU steps 4 and 5, i.e. "sitting down to the kitchen table" and "standing up from the kitchen table", achieved more than 90% detection rate while the erroneous estimations in these cases stemmed mainly from human occlusions from the kitchen table. Finally, in the "medication intake activity" 82.03% overall IU detection rate was achieved, with all of the IU steps being detected at least 75% of the times. The reason for the relatively low recognition rate of IU 1 and IU 4 steps during the "medication intake" activity is the fact that some users opened/closed the pill box very close to their body and behind the monitored surface i.e. kitchen table, where occlusions from the kitchen table took place.

From the above, it is concluded that the IU analysis achieved sufficient detection results, proving the ability of the method to infer over the examined samples. The effectiveness of the developed IU-step based method has been validated, both in terms of activity execution evaluation and selection of suitable IUs to detect and track for each activity. Additionally, it should be mentioned that the human activity recognition module is built and assessed on top of the object monitoring and human action recognition software components; as such, it is expected that errors on object detection and action recognition can be propagated to the hierarchy and, thus, can be inherited to the human activity recognition and IU analysis method.

6.1.4 Activity Recognition with Autonomous Robot Monitoring (LUM Test Site)

The next step of the experimental evaluation process comprises the assessment of the robot's ability to recognize the type of an ongoing activity and autonomously select the most appropriate parking position for close distance observation. In this context, the trained activity recognition framework was exposed to a continuous sequence of observations in accordance to which the users where asked to perform sequentially the activities, while the robot had to select and navigate towards a parking pose convenient for the monitoring of this activity. Initially, the robot was parked on its charging station which is normally located in a discreet spot

Table 5 Confusion matrix of the activity classification results on the LUM activities dataset

Activities	Classified AS			
	Meal preparation	Cooking	Having a meal	Medication intake
Meal preparation	38	1	0	0
Cooking	0	20	0	0
Having a meal	0	0	32	0
Medication intake	1	0	0	31
Total	39	21	32	31

Table 6 IU step detection results and behavioral modelling assessment for the four activities at the LUM dataset

IU 1	IU 2	IU 3[a]	IU 4[a]		Overall
Meal preparation					
100%	89.74%	64.10%	79.41%		83.31%
IU 1	IU 2[a]	IU 3[a]	IU 4[a]		Overall
Cooking					
85%	100%	85%	80%		87.50%
IU 1	IU 2	IU 3[a]	IU 4[a]	IU 5	Overall
Having a meal					
93.75%	93.75%	100%	96.87%	93.75%	95.62%
IU 1[a]	IU 2[a]	IU 3[a]	IU 4		Overall
Medication intake					
75%	81.25%	96.87%	75%		82.03%

[a] Mandatory IU step

in the home environment where the robot observes the human and large objects. During this experiment, each user performed progressively all the target activities, while there were intermediate segments where the user was doing nothing relevant to the target classes. In this occasion, the human activity recognition module should have the ability to continuously infer about the ongoing activity based on the log-likelihood, while deductions about the IU steps should take place only when the system is confident about the ongoing activity. To model this confidence factor, the target samples undergo a thresholding procedure on the log-likelihood level, where each sample is examined among all the DBN models and the resulting log-likelihood value is assessed. In cases that it is below a predefined threshold T_{ll}, experimentally estimated during the activity classification experiments described in Sect. 6.1.2, this segment of the sequence is considered as unknown and during this phase, the user does not perform any of the target classes or is in the begging of an activity, where the system has not yet gathered enough observations. In order to reduce the impact on the overall activity recognition, of individual slices generating momentarily false activity proposals, the inference of the ongoing activity is performed taking into consideration not only the latest slice, but also all the previous slices that have generated the same activity label, in the form of a history buffer. The history buffer is reset and cleared only when the current slice, along with the slices in the buffer, produces an activity label different than the last slice. While this approach does indeed increase the activity recognition robustness, it also introduces a bias in favor of longer activities, which produce larger history buffers, thus requiring more non-activity slices to ”forget” the activity after its actual conclusion.

Once the robot was confident about a certain activity the autonomous parking positions generation module was activated and upon selection of the most appropriate one, the robot traversed towards that pose. For the evaluation of the autonomous recognition capability, the sequences of all the activities performed 20 times in a specific order i.e. “meal preparation”, “cooking”, “having a meal” and “medication intake”. The ordering of this sequence is conceptually justified by considering the rationale where the user firstly prepares the ingredients for the meal, then proceeds with the actual cooking and, afterwards, the eating activity follows. The “medication intake” activity was retained last, mainly due to the fact that typically, medication is received after a meal while also this activity is contextually different with the rest of the activities and is expected to take place separately within the day.

The results obtained from this experiment revealed the ability of the system to recognize in the correct order all of the examined sequences. During this experiment, the DBN models are queried periodically in slices of N acquired observations (i.e. every 3 s), stored in a buffer -where observations are concatenated- and the log-likelihood of all DBN networks is assessed after N observations. Whenever the DBN models obtained continuous log-likelihood values above the predefined threshold T_{ll}, the recognized class index was assigned to the respective class and the autonomous robot position selection was activated. Upon selection of the most appropriate parking pose, the robot navigated towards this direction. During the robot’s locomotion, the software components for human action and object manipulation where deactivated and activated again upon reaching the target pose. Figure 8 summarizes the outcome of the aforementioned evaluation procedure and demonstrates the capacity of the

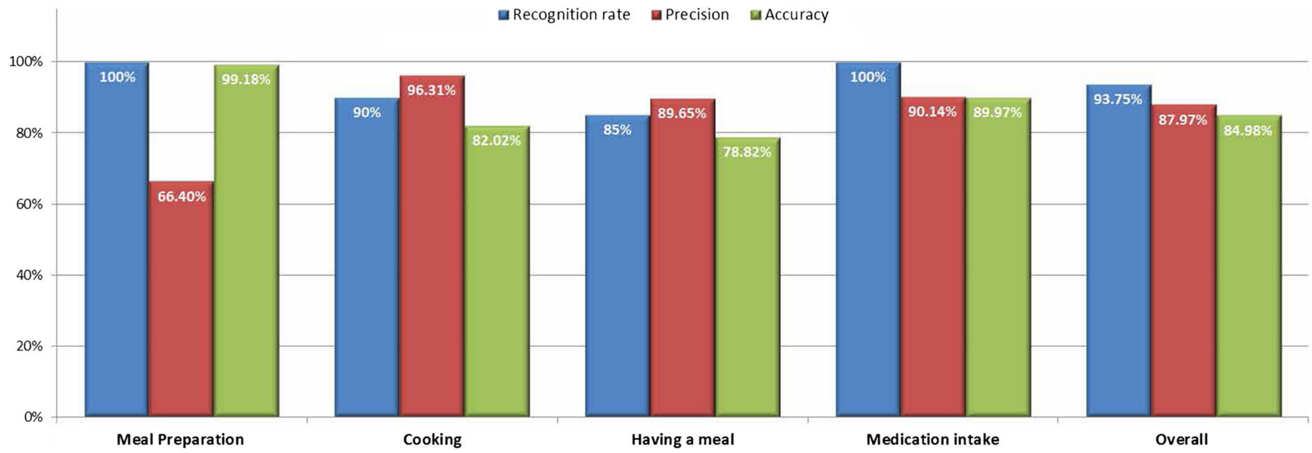


Fig. 8 Experimental results for continuous activity recognition at the LUM test site

developed methodology to recognize correctly the ongoing activities within the long testing sequences, as the detection rate achieved was over 93%.

Further experimental assessment of the autonomous recognition mode comprises a more detailed view, by considering also the temporal accuracy during the evaluation of the long sequences. From this experiment, it is revealed that the performance decreases mainly due to the fact that the evaluation is performed online and the algorithm needs some time to "forget" the last activity, reset the history buffer and accurately infer about the class label of the new one. Specifically, in intermediate time durations between successive activities, the observations are getting noisy since information of the previous and the current ongoing activities are retained, while also the robot is traveling from one position to another. Consequently, the log-likelihood for all the trained classes obtains values under the threshold T_{ll} and as a result, the query sample is classified as unknown or misclassified to the previous activity class. The outcome of this experiment is summarized in ("Precision Column") Fig. 8.

More specifically, by observing the results in Fig. 8 it is revealed that for the "meal preparation" activity we have low precision mainly due to the fact that at its last moments it overlaps with the first moments of the "cooking" activity. Yet, this can be justified if we consider that both activities have similarities in the involved IU steps (objects and simple activities) e.g. reaching the kitchen bench or the cupboard, and the observations need to propagate deeper in the "cooking" activity in order for the log-likelihood to obtain a higher value for this class while gradually "forgetting" the previous one. Concerning the "Cooking" activity high precision results were obtained, revealing that the recognition of this activity is not heavily affected from the "meal preparation" activity, something that is reasonable since the "Cooking" activity has a relatively long average duration. The "having a meal" activity retained high precision scores, while the average recall results were caused mainly from the preceding "cooking" activity, which due to its long average duration needed a longer time period to be "forgotten" and as a result, would overlap with the initial moments of the "Having a meal" activity. Finally, the proposed framework achieved high precision and recall values for the "Medication intake" activity indicating that the recognition of this activity was affected less from the other activities.

6.2 Evaluation in Real Environments (Barcelona Test Site)

The subjects that participated at the Barcelona test site experiments were elderly patients selected by the medical personnel of Fundacio ACE Barcelona Alzheimer Institute & Research Centre and fulfilled the same inclusion criteria with the ones that participated at the LUM test site. In total, 12 subjects were selected, with the experiments taking place in their personal real home environment, representative examples of which are presented in Fig. 9. The subjects were asked to execute the same four activities, with at least one repetition each day, for seven days in total, while the robot was observing them from a selected parking position. Apart from two home environments were the robot was not able to enter the kitchen area due to space limitations, and, thus, the "meal preparation" and "cooking" activities were not examined, in the rest of the apartments all four activities were executed. During this experimental procedure 308 activity instances were examined in total, corresponding to all the participants for the seven days of interaction. The distribution of these activities is presented in Table 7.

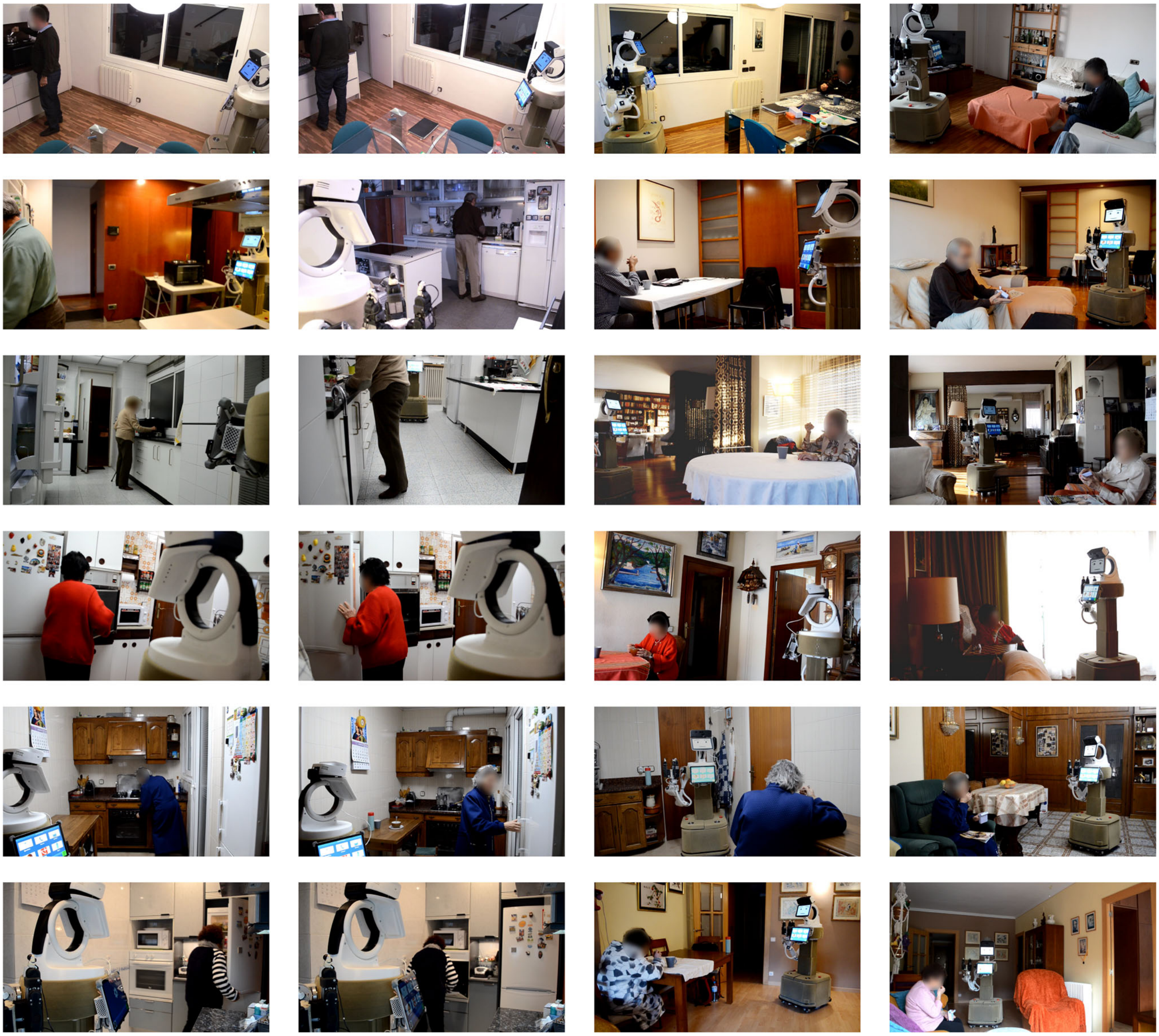


Fig. 9 Representative instances of the robot observing human behavior through daily activities in six houses at the Barcelona test site. Each row corresponds to a different house while from left to right the robot observes the “meal preparation”. “cooking”, “having a meal” and “medication intake” activities

Table 7 Distribution of the 12 different subjects for the four activities at the Barcelona test site

	Meal preparation	Cooking	Having a meal	Medication intake
Participants	10[a]	10[a]	12	12
Repetitions	7	7	7	7
Total executions	70	70	84	84

[a] 2 participants did not performed the indicated activity due to space limitations

6.2.1 Autonomous Behavior Understanding (Barcelona Test Site)

This experimental procedure comprises the evaluation of the proposed method to perform autonomous behavior understanding through the daily activities recognition, the selection of the most appropriate parking position for close distance observation and the assessment of the IU analysis. For these experiments the DBN models and the log-likelihood threshold T_{ll} that distinguishes the ongoing activities, were retained from the LUM test site. Considering the robot's charging station, this was selected with the criterion to ensure maximum visibility of the user and the large objects of interest, where the examined actions took place in each house. Each participant performed the target activities sequentially within the day while there were also intermediate segments were the user was doing nothing relevant to the target classes. Thus, when the robot gathered such observations for a significant amount of time, the history buffer was cleared and reset. In situations where the DBN models obtained continuous log-likelihood values above T_{ll}, the system was activated and the robot calculated the most appropriate pose and moved towards the selected pose to resume the activity recognition and behavior understanding.

Figure 10 summarizes the detailed results on the classification process of the real environments at the Barcelona test site. The overall detection rate is above 92%, however it is decreased when compared to the operation in the simulated environment. This is related to the fact that in the complex realistic environments there are more occlusions from peripheral objects and the performance of the small and large object detection and the human tracking degrades. The precision capability of the methods ("Precision" column Fig. 10) is also decreased since the method needs more time to "forget" the last activity and reset the history buffer, especially in situations where the two activities do not start immediately one after the other and the robot gathers observations irrelevant with the trained models and, thus, the log-likelihood value for all the trained classes obtained falls below the threshold T_{ll}. In general the overall analysis regarding the accuracy, precision and recall follows the same pattern with the similar experiment conducted in the simulated environment. However, the performance is degraded, mainly due to noisy measurements from the action and object detection components, introduced by the challenges in the realistic environments.

Table 8 IU step detection results and behavioral modelling assessment for the four activities at the Barcelona test site

IU 1	IU 2	IU 3[a]	IU 4[a]		Overall
Meal preparation					
94.29%	87.14%	84.29%	87.14%		88.21%
IU 1	IU 2[a]	IU 3[a]	IU 4[a]		Overall
Cooking					
88.57%	91.43%	84.29%	87.14%		87.86%
IU 1	IU 2	IU 3[a]	IU 4[a]	IU 5	Overall
Having a meal					
90.48%	88.10%	91.67%	72.62.87%	82.14%	85.00%
IU 1[a]	IU 2[a]	IU 3[a]	IU 4		Overall
Medication intake					
86.90%	83.33%	86.90%	83.33%		85.12%

[a] Mandatory IU step

Meanwhile, another part of the experimental assessment concerns the methods' ability to correctly infer about the behavior factors that allow understanding of the user's behavior through the IU analysis. Thus, each recognized activity sequence was passed through the respective IU step detection module. The inference regarding the IU steps was performed after the robot selected and moved towards the parking position. As stated above, each participant performed each activity at least once per day; as a result the maximum rep-

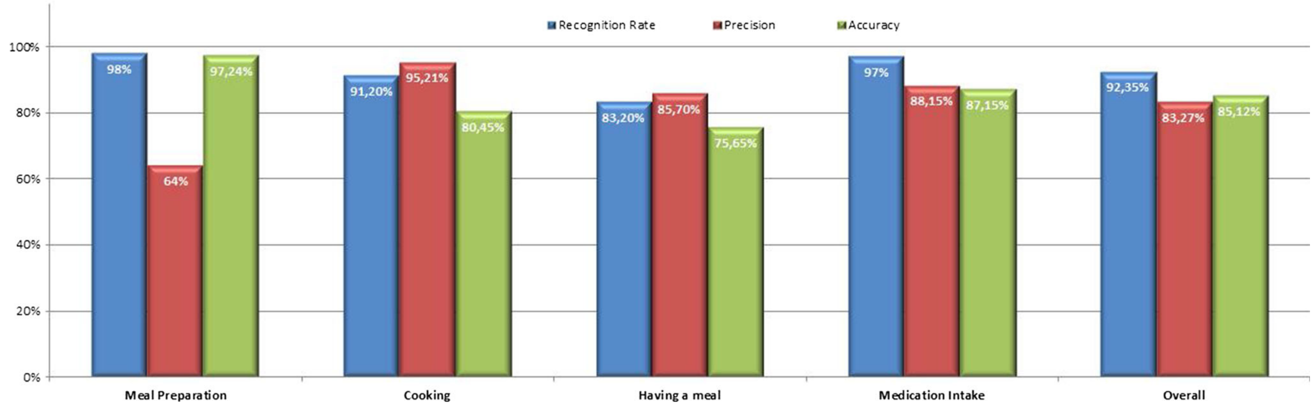
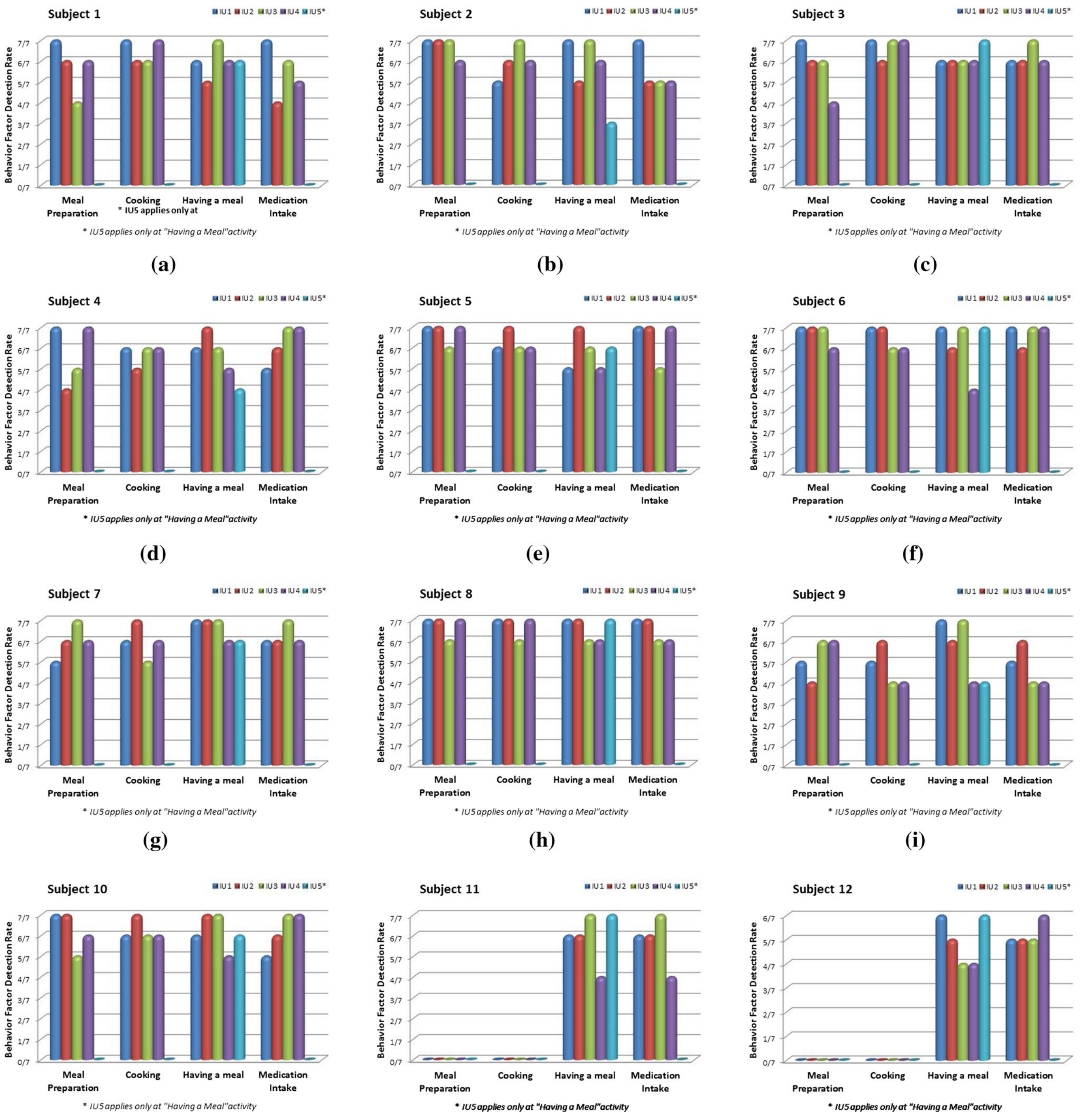


Fig. 10 Experimental results for continuous activity recognition at the Barcelona test site



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Fig. 11 Per subject behavior modelling evaluation at the Barcelona test site. For *subject*11 and *subject*12 the evaluation on "meal preparation" and "cooking" activities was not applicable due to space limitations

etitions of each activity per participant was seven (7), apart from two house environments where the "cooking" and "meal preparation" activities were omitted. The users were asked to perform the activities naturally, while the robot was observing them. The robot inferred continuously regarding the IU analysis and decided over the normal and abnormal behavior after the completion of the activity, i.e the log-likelihood value was below the threshold T_{ll}. The evaluation of the robot's capability to correctly assess the user's behavior was performed offline where the sequences of the obtained data were examined by the experts to annotate and compare the robot's inference with the actual human behavior, i.e. to check for missing mandatory IU steps. In Table 8, the average cross-subject performance for the detection of each IU step and the overall IU detection rate per activity is summarized. The method achieved over 85.00% detection rate on the real environments and when compared with the respective results in Table 6, it is revealed that the performance remains adequate. Specifically, for the "meal preparation" activity the performance was slightly increased and this was due to the fact that in many cases the robot selected a parking position with good visibility that allowed successful observation of the large objects and the human actions. Regarding the "cooking" activity the behavior understanding capability of the method was approximately the same, while the behavior understanding during the "having a meal" activity exhibited lower performance when compared to the simulated environment. This was due to the fact that the robot typically parked in large distances from the user while s/he was having a meal considering the proxemics to respect the personal space and, thus, the performance of the object detection component degraded, especially in situations where the table was cluttered with several objects. The behavior understanding during "medication intake" was approximately the same with slightly better performance in $IU1$, when the user was sitting in an open space and the robot was capable to approach in close distance to observe the activity. The per subject behavior modelling analysis is summarized in Fig. 11 where for each subject the IU detection rate for the seven days of interaction is exhibited. Note, that for $subject11$ and $subject12$ only two of the activities were examined.

Last, it should be noted that the selection of appropriate parking positions for the monitoring of each activity ensures the necessitated controlled nature of the observed environment, where human and object detection tracking algorithms operate accurately, an attribute that positively impacts on the action recognition mode and, hence in the behavioral understanding mode, since the RGB-D camera is closer at the user and the depth data becomes more reliable. In addition, the manipulated objects involved in each scenario hold major role in the discriminative ability of the DBN models and the selection of optimal parking pose for monitoring allows precise detection of the small objects. Concerning the execution rate, it strongly depends on the gathering rate of the data exposed to the DBNs retina and is regulated from the slowest software component; in our specific case, this is the global manipulated object server, which operates at a rate of 10Hz. The inference rate of the human behavior understanding module depends on the amount of the data existing in the buffer, and indicatively, for relatively large sequences of 5min, it takes approximately 1sec for inference, while this execution rate increases linearly with the time that DBN models are exposed to the data.

7 Conclusions

In this work, the ability to understand the human behavior through daily activities, with a robot, has been presented. Each selected activity has been analysed in terms of Interaction Unit analysis, which is a typical psychology-inspired method to model human behavior. We modeled the problem with a specifically designed Dynamic Bayessian Network that allowed the transcription of Interaction Unit analysis into a machine interpretable manner. This approach facilitated both activity recognition and in-depth activity analysis in terms of evaluation of the associated to the IUs behavioral factors. Moreover, it should be stated that until now the implementation of Interaction Unit analysis required explicit measurements of the environmental state and human actions, which was addressed by applying several sensors to obtain the necessary recordings. In the proposed method, it has been demonstrated that behavior understanding through IU analysis can be performed with a mobile robot equipped with a minimum set of sensors i.e. RGB-D camera and laser scanners. The latter has been achieved by taking advantage of the robot's mobility, where by placing the robot towards the region of interest, better view from the human and environment state is obtained. To this end, a component responsible for the automatic selection of robot parking positions has been developed, which also takes into consideration the psychological constraints of the proxemics theory. The added value of IU analysis integration with all of the subordinate monitoring modules into a united framework operated by a robotic agent, increases drastically the automation of human behavior understanding through daily activities observation, since it presents a minimum invasion on the environmental setup. Consistent observations from humans and manipulated objects were obtained through custom-tailored delicately designed software components capable of operating in real time and under realistic environmental conditions, thus compensating limitations such as human side-views, occlusions and interactions with realistic daily-used objects. The evaluation of the proposed method has been performed on two test sites. The first one concerned the evaluation of the method in a small scale controlled simulated environment at the LUM test site, while the second one concerned the evaluation on a large scale experiment in unconstrained environments, with 12 real users in their personal home environments at the Barcelona test site. The obtained results show the capability of the robot to accurately differentiate the ongoing activities as well as to precisely infer about the behavioral factors for each specific activity and thus adequately understand the human behavior. Finally, the findings of the research conducted herein could be also analyzed in terms of the overall impact of the system on the users. This necessitates further research efforts to identify how the users accepted the system, for instance by using for example System Usability Scale and other metrics.

Acknowledgements We would like to thank Fundacio ACE Barcelona Alzheimer Institute & Research Centre as well as the Medical University of Lublin, Poland, for offering their expertise in the annotation of the behavioral factors in the IU analysis as well as for offering their premises for the dataset acquisition with RAMCIP robot.

Funding This work has been supported by the EU Horizon 2020 funded project namely: "Robotic Assistant for MCI Patients at home (RAMCIP)" under the Grant Agreement with No. 643433.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Aggarwal JK, Ryoo MS (2011) Human activity analysis: a review. ACM Comput Surv (CSUR) 43(3):16
2. Atkeson CG, Hale JG, Pollick F, Riley M, Kotosaka S, Schaul S, Shibata T, Tevatia G, Ude A, Vijayakumar S et al (2000) Using humanoid robots to study human behavior. IEEE Intell Syst Appl 15(4):46–56
3. Baddeley AD, Baddeley H, Bucks R, Wilcock G (2001) Attentional control in Alzheimer's disease. Brain 124(8):1492–1508
4. Berelson B, Steiner GA (1964) Human behavior: an inventory of scientific findings. Trans-action 1:2–2
5. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
6. Bloem BR, Valkenburg VV, Slabbekoorn M, Willemsen MD (2001) The multiple tasks test: development and normal strategies. Gait Posture 14(3):191–202
7. Charalampous K, Kostavelis I, Gasteratos A (2016) Robot navigation in large-scale social maps: an action recognition approach. Expert Syst Appl 66:261–273
8. Charalampous K, Kostavelis I, Gasteratos A (2017) Recent trends in social aware robot navigation: a survey. Robot Auton Syst 93:85–104
9. Chrungoo A, Manimaran S, Ravindran B (2014) Activity recognition for natural human robot interaction. In: International conference on social robotics. Springer, pp 84–94
10. Coppola C, Cosar S, Faria DR, Bellotto N (2017) Automatic detection of human interactions from RGB-D data for social activity classification. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 871–876
11. Doumanoglou A, Kouskouridas R, Malassiotis S, Kim TK (2016) Recovering 6D object pose and predicting next-best-view in the crowd. In: IEEE conference on computer vision and pattern recognition, pp 3583–3592
12. Faria DR, Premebida C, Nunes U (2014) A probabilistic approach for human everyday activities recognition using body motion from RGB-D images. In: The 23rd IEEE international symposium on robot and human interactive communication, pp 732–737
13. Foka AF, Trahanias PE (2010) Probabilistic autonomous robot navigation in dynamic environments with human motion prediction. Int J Soc Robot 2(1):79–94
14. Garrell A, Villamizar M, Moreno-Noguer F, Sanfeliu A (2017) Teaching robots proactive behavior using human assistance. Int J Soc Robot 9(2):231–249
15. Grześ M, Hoey J, Khan SS, Mihailidis A, Czarnuch S, Jackson D, Monk A (2014) Relational approach to knowledge engineering for POMDP-based assistance systems as a translation of a psychological model. Int J Approx Reason 55(1):36–58
16. Hall ET (1966) The hidden dimension. Doubleday, New York
17. Han F, Reily B, Hoff W, Zhang H (2017) Space-time representation of people based on 3D skeletal data: a review. Comput Vis Image Underst 158:85–105
18. Hoey J, Plötz T, Jackson D, Monk A, Pham C, Olivier P (2011) Rapid specification and automated generation of prompting systems to assist people with dementia. Pervasive Mobile Comput 7(3):299–318
19. Karplus K, Brown M, Hughey R, Krogh A, Mian IS, Haussler D (1996) Dirichlet mixtures: a method for improving detection of weak but significant protein sequence homology. Comput Appl Biosci 12:327–345
20. Katz S, Downs TD, Cash HR, Grotz RC (1970) Progress in development of the index of ADL1. Gerontologist 10(1 Part 1):20–30
21. Kim B, Pineau J (2016) Socially adaptive path planning in human environments using inverse reinforcement learning. Int J Soc Robot 8(1):51–66
22. Koppula H, Saxena A (2013) Learning spatio-temporal structure from RGB-D videos for human activity detection and anticipation. In: International conference on machine learning, pp 792–800
23. Kostavelis I, Kargakos A, Giakoumis D, Tzovaras D (2017) Robots workspace enhancement with dynamic human presence for socially-aware navigation. In: International conference on computer vision systems. Springer, pp 279–288
24. Leigh A, Pineau J (2014) Laser-based person tracking for clinical locomotion analysis. In: IROS-rehabilitation and assistive robotics
25. Leite I, Martinho C, Paiva A (2013) Social robots for long-term interaction: a survey. Int J Soc Robot 5(2):291–308
26. Martinez-Contreras F, Orrite-Urunuela C, Herrero-Jaraba E, Ragheb H, Velastin SA (2009) Recognizing human actions using silhouette-based HMM. In: Sixth IEEE international conference on advanced video and signal based surveillance. IEEE, pp 43–48
27. Norman DA, Shallice T (1986) Attention to action. In: Davidson RJ, Schwartz GE, Shapiro D (eds) Consciousness and self-regulation. Springer, New York, pp 1–18
28. Oreifej O, Liu Z (2013) HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences. In: IEEE conference on computer vision and pattern recognition, pp 716–723
29. Panangadan A, Matarić M, Sukhatme GS (2010) Tracking and modeling of human activity using laser rangefinders. Int J Soc Robot 2(1):95–107
30. Piyathilaka L, Kodagoda S (2015) Human activity recognition for domestic robots. In: Mejias L, Corke P, Roberts J (eds) Field and service robotics. Springer, Berlin, pp 395–408
31. Premebida C, Faria D, Souza F (2017) Dynamic Bayesian network for time-dependent classification problems in robotics. In: Prieto Tejedor J (ed) Bayesian inference, chapter 15. InTech, Croatia
32. Rahmani H, Mahmood A, Huynh DQ, Mian A (2014) Real time action recognition using histograms of depth gradients and random decision forests. In: 2014 IEEE winter conference on applications of computer vision. IEEE, pp 626–633
33. Roitberg A, Perzylo A, Somani N, Giuliani M, Rickert M, Knoll A (2014) Human activity recognition in the context of industrial human–robot interaction. In: Asia-Pacific Signal and Information Processing Association. IEEE, pp 1–10
34. Rusu RB, Marton ZC, Blodow N, Dolha M, Beetz M (2008) Towards 3D point cloud based object maps for household environments. Robot Auton Syst 56(11):927–941
35. Rybok L, Schauerte B, Al-Halah Z, Stiefelhagen R (2014) Important stuff, everywhere! activity recognition with salient proto-objects as context. In: IEEE winter conference on applications of computer vision. IEEE, pp 646–651

36. Ryu H, Monk A (2005) Will it be a capital letter: signalling case mode in mobile phones. Interact Comput 17(4):395–418
37. Ryu H, Monk A (2009) Interaction unit analysis: a new interaction design framework. Hum Comput Interact 24(4):367–407
38. Salah A, Ruiz-del Solar J, Mericli C, Oudeyer PY (2012) Human behavior understanding for robotics. In: Salah AA, Ruiz-del-Solar J, Meriçli Ç, Oudeyer P-Y (eds) Human behavior understanding. Springer, Berlin, pp 1–16
39. Salah AA, Gevers T, Sebe N, Vinciarelli A et al (2010) Challenges of human behavior understanding. In: Salah AA, Ruiz-del-Solar J, Meriçli Ç, Oudeyer P-Y (eds) HBU. Springer, Berlin, pp 1–12
40. Salah AA, Lepri B, Pianesi F, Pentland AS (2011) Human behavior understanding for inducing behavioral change: application perspectives. In: International workshop on human behavior understanding. Springer, pp 1–15
41. Santos L, Khoshhal K, Dias J (2015) Trajectory-based human action segmentation. Pattern Recognit 48(2):568–579
42. Schmidler SC, Liu JS, Brutlag DL (2000) Bayesian segmentation of protein secondary structure. J Comput Biol 7(1–2):233–248
43. Schmidt T, Newcombe R, Fox D (2015) DART: dense articulated real-time tracking with consumer depth cameras. Auton Robots 39:239–258
44. Schmidt-Rohr SR, Losch M, Dillmann R (2008) Human and robot behavior modeling for probabilistic cognition of an autonomous service robot. In: IEEE international symposium on robot and human interactive communication. IEEE, pp 635–640
45. Shan J, Akella S (2014) 3D human action segmentation and recognition using pose kinetic energy. In: IEEE workshop on advanced robotics and its social impacts. IEEE, pp 69–75
46. Shotton J, Sharp T, Kipman A, Fitzgibbon A, Finocchio M, Blake A, Cook M, Moore R (2013) Real-time human pose recognition in parts from single depth images. Commun ACM 56(1):116–124
47. Skinner BF (1953) Science and human behavior. Simon and Schuster, New York
48. Sonn U (1996) Longitudinal studies of dependence in daily life activities among elderly persons. Scand J Rehabilit Med Suppl 34:1–35
49. Stavropoulos G, Giakoumis D, Moustakas K, Tzovaras D (2017) Automatic action recognition for assistive robots to support MCI patients at home. In: 10th international conference on pervasive technologies related to assistive environments. ACM, pp 366–371
50. Sung J, Ponce C, Selman B, Saxena A (2012) Unstructured human activity detection from RGBD images. In: IEEE international conference on robotics and automation. IEEE, pp 842–849
51. Takayama L, Pantofaru C (2009) Influences on proxemic behaviors in human–robot interaction. In: IEEE international conference on intelligent robots and systems. IEEE, pp 5495–5502
52. Taylor J, Shotton J, Sharp T, Fitzgibbon A (2012) The Vitruvian manifold: Inferring dense correspondences for one-shot human pose estimation. In: IEEE conference on computer vision and pattern recognition. IEEE, pp 103–110
53. Tsai MJ, Wu CL, Pradhan SK, Xie Y, Li TY, Fu LC, Zeng YC (2016) Context-aware activity prediction using human behavior pattern in real smart home environments. In: 2016 IEEE international conference on automation science and engineering (CASE). IEEE, pp 168–173
54. Vasileiadis M, Malassiotis S, Giakoumis D, Bouganis CS, Tzovaras D (2017) Robust human pose tracking for realistic service robot applications. In: IEEE international conference on computer vision workshops, pp 1363–1372
55. Wang J, Liu Z, Wu Y (2014) Learning actionlet ensemble for 3D human action recognition. In: Wang J (ed) Human action recognition with depth cameras. Springer, Basel, pp 11–40
56. Whiten C, Laganiere R, Bilodeau GA (2013) Efficient action recognition with MoFREAK. In: International conference on computer and robot vision. IEEE, pp 319–325
57. Wu J, Osuntogun A, Choudhury T, Philipose M, Rehg JM (2007) A scalable approach to activity recognition based on object use. In: IEEE international conference on computer vision. IEEE, pp 1–8
58. Yang X, Tian Y (2014) Effective 3D action recognition using eigenjoints. J Vis Commun Image Represent 25(1):2–11
59. Zhu Y, Chen W, Guo G (2014) Evaluating spatiotemporal interest point features for depth-based action recognition. Image Vis Comput 32(8):453–464
60. Ziaeefard M, Bergevin R (2015) Semantic human activity recognition: a literature review. Pattern Recognit 48(8):2329–2345
61. Zipf GK (2016) Human behavior and the principle of least effort: an introduction to human ecology. Ravenio Books, Cambridge

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dr. Ioannis Kostavelis is a Senior Research Associate in the Information Technologies Institute (ITI) of the Centre for Research and Technology Hellas (CERTH), in Thessaloniki, Greece. His main research interests concern robot vision, robot navigation, metric, semantic and social mapping, 3D environment understanding, human activity monitoring and machine learning. He has contributed in more than 60 papers in refereed journals, international conferences and book chapters and involved in several National and European projects related to robotics applications.

Mr. Manolis Vasileiadis received the Diploma degree in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece in 2011 and is currently a Ph.D. candidate at the Department of Electrical and Electronic Engineering, Imperial College London. His doctoral research focuses on human pose estimation from sparse data on embedded systems. Since January 2013, he has been with the Information Technologies Institute at the Centre for Research and Technology Hellas, participating as a research assistant in multiple projects, funded by the EC and the Greek Ministry of Research and Technology. His research interests focus on image processing, computer vision and deep learning optimization for low-power devices.

Mr. Evangelos Skartados is a Research Assistant in Information Technologies Institute, in Centre for Research and Technology, Hellas since 2014. He has participated in three European Union funded projects (CloPeMa, RAMCIP, BADGER) in the field of Robotics. His research interests include Robotic Vision and Machine Learning.

Mr. Andreas Kargakos graduated from Electrical and Computer Engineering Dpt. of Aristotle University of Thessaloniki (A.U.TH.), Greece. He is a research associate in the Information Technologies Institute - Centre for Research and Technology Hellas since 2013. His research interests include fields such as robotics and artificial intelligence.

Dr. Dimitrios Giakoumis is a Senior Research Associate in the Information Technologies Institute (ITI) of the Centre for Research and Technology Hellas (CERTH), in Thessaloniki, Greece. His research interests include affective computing, computer and robot vision, robot navigation, human motion, activity, and behaviour analysis and modelling, signal processing and sensor management, multimodal interfaces and pattern recognition. His involvement in these areas has led to co-authoring of more than 50 papers in refereed journals, international conferences and book chapters.

Dr. Christos-Savvas Bouganis is a Reader in Intelligent Digital Systems in the Department of Electrical and Electronic Engineering, Imperial College London, U.K. He has published over 100 research papers in peer-referred journals and international conferences, and he has contributed three book chapters on digital system design. His current research interests include the theory and practice of reconfigurable computing and design automation, mainly targeting the domains of Machine Learning, Computer Vision, and Robotics.

Dr. Dimitrios Tzovaras is a Senior Researcher of Grade A' in the Information Technologies Institute (ITI) of the Centre for Research and Technology Hellas (CERTH), in Thessaloniki, Greece. His main research interests include network and visual analytics for network security, computer security, data fusion, biometric security, virtual reality, machine learning and artificial intelligence. His research work is summarized in 2 books, 129 publications in International Journals with, 40 book chapters, and 362 presentations in International Conferences. Dr. Tzovaras has participated with his team in more than 80 Research and Development projects of ITI, with a large management record having been the project coordinator, scientific and technical manager in more than 50 EU projects.