BEHAVE - Behavioral analysis of visual events for assisted living scenarios

Carlos Fernando Crispim-Junior
INRIA Sophia Antipolis
2004 Route de Lucioles, BP 93
06902 Sophia Antipolis, France
carlos-fernando.crispim_junior@inria.fr

Jonas Vlasselaer
KU Leuven
Celestijnenlaan 200A-bus 2402
Leuven, Belgium
jonas.vlasselaer@cs.kuleuven.be

Anton Dries
KU Leuven
Celestijnenlaan 200A-bus 2402
Leuven, Belgium
anton.dries@cs.kuleuven.be

François Bremond
INRIA Sophia Antipolis
2004 Route de Lucioles, BP 93
06902 Sophia Antipolis, France
francois.bremond@inria.fr

Abstract

This paper proposes BEHAVE, a person-centered pipeline for probabilistic event recognition. The proposed pipeline firstly detects the set of people in a video frame, then it searches for correspondences between people in the current and previous frames (i.e., people tracking). Finally, event recognition is carried for each person using probabilistic logic models (PLMs, ProbLog2 language). PLMs represent interactions among people, home appliances and semantic regions. They also enable one to assess the probability of an event given noisy observations of the real world. BEHAVE was evaluated on the task of online (non-clipped videos) and open-set event recognition (e.g., target events plus none class) on video recordings of seniors carrying out daily tasks. Results have shown that BEHAVE improves event recognition accuracy by handling missed and partially satisfied logic models. Future work will investigate how to extend PLMs to represent temporal relations among events.

1. Introduction

World population is aging and in a few years from now the number of seniors should surpass the number of people able to provide care. As a consequence, there is a demand for information and communication technologies devoted to fill this gap of human resources, ranging from tools to support the timely diagnosis of aging-related diseases up to systems to promote the well-being of seniors. Medical studies [28] have shown that disabilities during the performance of Instrumental Activities of Daily living (IADL) are a relevant cue to demonstrate how much at risk people are of developing dementia. Nonetheless, very little work has been made in the sense of automatically diagnosing dementia, like Alzheimer's disease (AD), by the alterations it provokes into the visual patterns of human behaviors, such as IADLs [20][23].

This paper introduces BEHAVE, BEHavioral Analysis of Visual Events, a probabilistic framework for the recognition of daily activities at assisted living scenarios, such as nursing homes and people's residences. BEHAVE follows a person-centered, event recognition pipeline to recognize the instances of activities performed by different people in a continuous flow of visual data. The main contribution of the proposed method is its capability of handling uncertainty at event level (probabilistic logic models) associated with the robustness of a color-depth sensing pipeline for people detection and tracking. Prior work on the literature relies on the strengths of color-depth sensing to overcome the noisy observations of the scene and performs event recognition on a deterministic fashion. BEHAVE computes relevant relations between instances of real-world physical objects on a deterministic fashion, but reasons about event realization on a probabilistic fashion. Hence, BEHAVE is more tailored to handle the uncertainties of real-world, assisted living settings where different sources of uncertainty intervene in event recognition task.

Next subsections describe the related work (Section 2), the proposed framework (Section 3), the experiments carried out (Section 4), the results and discussion (Section 5) and the conclusions of this study (Section 6).

Figure 1. Visual event recognition. Event models may be defined by the spatio-temporal relations between semantic regions of the scene (*e.g.*, table, medical cabinet, landline telephone - red polygons) and dynamics of the monitored person, like silhouette's shape (in purple) and spatial relations (*e.g.*, closeness to the telephone table).

2. Related Work

Visual action recognition (short activities) has been extensively studied in the past decades [22][31] and has greatly benefited from latest advances on deep learning engineered features [15]. But, even though methods based on Deep Neural Networks [15], Bag-of-visual-words [31] and Fisher vectors [12] cannot be matched at the task of video classification, their performance at action detection, *i.e.*, online recognition of human activities, is still far from desirable. In this latter task, a method must localize the spatio-temporal boundaries of potential activities in a continuous video stream and then infer whether an event is taking place within those boundaries. The accurate assessment of the duration of an event instance is quintessential for applications in medicine, where the time taken by a person to carry out an activity is an indicator of his/her autonomy.

Furthermore, in the context of assisted living applications, adopted methods must be also capable of discriminating between the activities of different people in the scene. State-of-the-art methods on visual action classification [31][12][15] still lack this feature, since they describe video clips in a holistic fashion. In summary, to address the above described requirements, a method must firstly detect and track the different instances of people in the scene and only then perform event recognition focusing on the spatio-temporal patterns of each person. These methods are called person-centered approaches.

2.1. Person-centered Event Recognition

Existing work in person-centered event recognition may be categorized according to the visual information they rely on: color- and color-depth based approaches. Color-based approaches have been extensively studied in the past years with the advent of more robust people detection algorithms, like, Poselets [5], Deformable Part Models [16], and most recently, Faster R-CNN [26]. However, these methods still require a large number of training samples to generalize their performance between training and test scenarios. In addition, their recognition performance tends to degrade when they are applied on recordings with low image resolution or when people's postures deviates from pedestrian detection scenarios. These restrictions pose difficulties at the direct application of these methods to the recognition of daily activities in assisted living scenarios, where the described limitations are common.

Event recognition based on color-depth sensing (*e.g.*, Microsoft Kinect) tend to be more resilient to illumination changes than color-based methods, since it relies on off-the-shelf 3D (depth) measurements of the scene. Prior work on action recognition and color-depth sensors have explored the spatio-temporal patterns defined by human body joints (skeletal representation [27]) to describe actions of interest [19][14][4] [3]. But, despite the fact that skeletal representations can distinguish the activities of different individuals, the original method for skeleton detection has a limited range of operation (3-5m for Kinect, version 1 and 6-8m Kinect for version 2), and expects the person to be facing the color-depth sensor. These assumptions generally do not hold in assisted living settings, since the scene is commonly cluttered by home appliances and furniture, and the monitoring range easily exceeds 5 meters.

To overcome the limitations of skeleton-based methods, recent studies have extended algorithms for people detection from background subtraction method [25] for the depth signal domain. These advances have enabled the design of event recognition systems more resilient to illumination changes, but less constrained by the distance between people and the sensor [3][9]. However, most systems follow a knowledge-driven approach and do not handle the uncertainty of real-world observations, like missed and ambiguous observations, limiting event recognition to reasonably controlled scenarios at real-world settings.

2.2. Probabilistic Event Recognition

Knowledge-driven methods (Kd) provide a flexibe framework to represent events given relations between heterogeneous types of data, like sensor readings, common sense knowledge and domain semantics [24][18][7][1]. But they tend to underperform in the presence of uncertain data. This behavior is due to their deterministic reasoning mechanism, which lacks the means to handle missed, incomplete and unreliable observations. To overcome this weakness, researchers have studied principled ways to combine knowledge-driven formalism (First-order logic, description logic, constraint programming) with probabilistic inference.

Probabilistic Logic Programming and Statistical Relational Learning (SRL) are among the most prominent approaches [17][13][11].

PLP methods, like ProbLog2, extend a subset of first-order logic (FOL) formulas with probabilistic facts [17][11]. Probabilistic facts weight the likelihood of answers to user queries by the available evidence. Prior work on PLP and event recognition has mostly focused on color-based methods and on recognizing short-length activities, like meeting, moving and leaving objects at the scene [29]. Alternatively, SRL methods have focused on the idea of weighted FOL formulas [13]. Weighted formulas allow one to quantify how likely an event model is given the partial satisfaction of the set of logic formulas that describe it [30]. PLP methods are still less explored in the task of visual event recognition compared to weighted FOL methods.

This paper explores the notion of probabilistic facts to handle uncertainty of missed and ambiguous observations at assisted living scenarios and longer daily activities (*e.g.*, "prepare drink","talk on the telephone","make a payment"). To address the problem of recognizing daily activities, the proposed method couples the probabilistic inference of ProbLog2 with a color-depth sensing pipeline for people detection and tracking (more resilient to scene characteristics) and an ontological representation to define abstractions of real-world objects and the possible relations among them.

3. BEHAVE Framework

The proposed framework can be decomposed into three main modules (Fig.2): people detection, people tracking and probabilistic event recognition. Next sub-sections describe in more details each of these modules, with a particular attention to probabilistic event recognition, the main contribution of this work.

3.1. People Detection

The recognition of people in the scene is carried out with the depth-based framework of [25]. The framework extends the standard detection range of color-depth sensors from 3-4 meters (Microsoft and PrimeSense) to 7-9 meters away. It works as follows: first, it performs background subtraction in the depth image to identify foreground regions that contains both moving objects and potential noise. These foreground pixels are then clustered into objects based on their depth values and neighborhood information. Among these objects, people are detected using a head and shoulder detector and tracking information about previously detected people.

3.2. People Tracking

People tracking step [6] takes as input the video stream and the list of objects detected in the current and previous frames using a sliding time window. First, a link score is computed between any two detected objects in this time window using a weighted combination of six object descriptors: 2D and 3D positions, 2D object area, 2D object shape ratio, color histogram and dominant color. Then, successive links are formed to represent the several paths an object can follow within the temporal window. Each possible path of an object is associated with a score given by all the scores of the links it contains. The object trajectory is determined by maximizing the path score using Hungarian algorithm [21].

3.3. Probabilistic Event Recognition

This step consists in using a probabilistic inference engine to infer whether the available evidence satisfies the probabilistic logic models of target activities. BEHAVE adopts the ProbLog2 language and its inference engine to model and recognize the relations that intervene between target activities and its subparts. Probabilistic Logic models are logic programs that formulas (or facts) are annotated with probabilities [17]. ProbLog2 uses a subset of first-order logic where only Horn clauses are allowed, a universally quantified clause that has at most one positive literal. A crucial difference between the semantics of ProbLog2 and first-order logic (FOL), however, is that the former makes use of the closed world assumption. For more details we refer to [17]. A ProbLog formula (Equation 1) reads as follows: the head h is true with probability p if and only if all its body clauses (b_i) are true. Predicates are generally expressed as boolean variables.

$$p :: h : -b_1, .., b_n \qquad (1)$$

where,

- p: probability of head clause
- h: head clause,
- b_i: body clause i.

To model a target event, we have defined a probabilistic logic model that contains the relations (predicates) between physical objects (terms) that are relevant to this event. We have extended the previously presented ProbLog2 formula (Equation 1) with disjunctive annotations. Disjunctive annotations are a ProbLog2 feature that define mutually exclusive outcomes of a set of body clauses [11] (see Full-model, Example 1). To handle missed and noisy observations of the scene, we have defined a complete and partial model for each event. The latter model quantifies the likelihood of an event given that only a subset of its relations are observed (Example 1, Partial model). For example, a person is in front of the medical cabinet but s/he is not bending towards the cabinet to interact with it, as it would be expected.
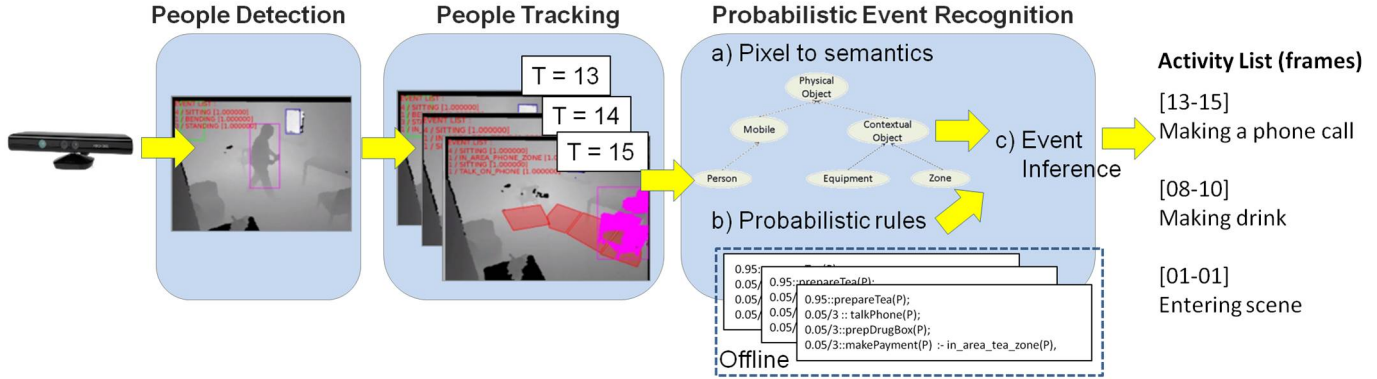
Example 1 *"Probabilistic logic event models"*

Figure 2. BEHAVE framework: visual event recognition may be decomposed into three steps: People Detection, People Tracking and Probabilistic Event (Activity) Recognition. These steps are carried out frame-wisely.

```
#P = Instance identifier of a physical object of type
# ''Person''
#Full−model
0.95 :: prepareTea(P);
0.05/3 :: talkPhone(P);
0.05/3::prepDrugBox(P);
0.05/3::makePayment(P) :− in_area_tea_zone(P),
                        bending(P).
#Partial−model
0.85 :: prepareTea(P);
0.15/3 :: talkPhone(P);
0.15/3::prepDrugBox(P);
0.15/3::makePayment(P) :− in_area_tea_zone(P),
                        \+bending(P).
```


Figure 3. Physical object sub-tree of the ontology language

of physical objects of type "Person" (What is the individual likelihood of each person at "preparing a drink" now?, Example 2B).

Example 2 *"Querying the most likely event model"*

```
query(Phone(5)). %A
query(Drink(_)). %B
```

4. Experiments

The goal of BEHAVE is to serve as an objective framework for the automatic recognition of human daily activities in daily tasks. Hence, we have chosen to evaluate it with color-depth recordings of seniors carrying out unscripted, real-world activities during a clinical protocol for the study of Alzheimer's disease. Senior participants aged 65 years and above (N=19) were recruited by Institut Claude Pompidou (ICP) in Nice, France. The clinical protocol asks participants to undertake a set of physical tasks, single and dual gait tasks, and IADLs in an observation room furnished with home appliances. The set of IADLs consists in the activities: prepare drink (Drink), talk on the telephone (Phone, *e.g.*, calling/answering the telephone), prepare pill box (DrugBox) and make payment using a tablet (Payment). Semantic spatial zones were defined *a priori* to incorporate semantic knowledge about the scene geometry into logic event models. Recording sessions have taken ~ 15 minutes

To help domain experts at modeling the relations that intervene among the real world objects participating in the activities of interest, BEHAVE framework has adopted a declarative constraint-based ontology language to define the possible physical objects and relations among them [10]. Physical Objects (P.O.) are abstractions for real-world objects involved in the realization of an event (*e.g.*, a person, a kettle, the kitchen). Figure 3 describes the P.O. branch of the employed ontology. A P.O. instance may be dynamically detected by the methods underlying the probabilistic event recognition, like people detection and tracking steps, or be provided as prior knowledge about the geometry of the scene.

Probabilistic event recognition is performed by running a temporal sliding window over the video recording, querying the ProbLog2 engine at each frame about the probability of target activities given available evidence (Example 2). The evidence of each frame consists of the observed relations between detected instances of physical objects. An activity query may be indexed by the tracking identifier of a person (What is the likelihood that person 5 is "preparing a drink"?, Example 2A) or be open to all available instances
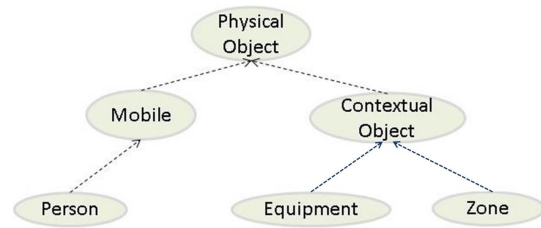
for each participant. The reported performance consists in the frame-wise accuracy between ground-truth annotations and the output of evaluated methods.

To compare the proposed method to state-of-the-art methods, the present evaluation has adopted the knowledge-driven framework (KER) described at [10] as a baseline. This baseline has showed a accuracy higher than holistic approaches based on dense trajectory features [2] for assisted living settings similar to ours [8][10].

5. Results & Discussion

Experimental results have showed that the proposed method improves the performance of event recognition task at all target activities (84% versus 76% of baseline method, Fig.5 x Fig.4, respectively). The improvement in performance is due to the capacity of probabilistic models to consider partially satisfied models as an undergoing event for which instantaneous observations have been missed or corrupted by noise. For instance, by failures at steps prior to probabilistic event recognition or by small deviations from the semantic relations that define an activity model. However, the recognition performance of "none" class has been reduced at the expenses of improvements on target activities. The "talk on the telephone" event was the mostly confused with the event "none". The confusion was caused by the occasional partial occlusion of the person by the image border. The occlusion has caused uncertainty about the position of the person in the scene (people detection step), and in consequence about the person being close enough to be using the telephone (probabilistic event recognition step).

On a second step, this evaluation has assessed the similarity between the the duration of events recognized by BEHAVE framework and their true duration computed using ground-truth data (Fig. 6). Results have shown that the automatically assessed duration of activities is statistically correlated with ground-truth data (Pearson R, $p \leq 0.01$; ICP dataset). However, we have observed that the durations of "preparing drink" and "prepare drug box" activities have been underestimated by the proposed framework. The underestimation of the duration of "prepare drug box" event was due to participants carrying out this activity far away from the medication cabinet. In these cases, they would quickly reach out for the desired medication when needed only, returning to an ambiguous upright posture, far away from the related semantic zone. In the case of "prepare drink" event, the underestimation of its duration was mostly due to missed people detection. Further work will analyze how to cope with above described problems, in addition to the high intra-class variance of certain activity classes.
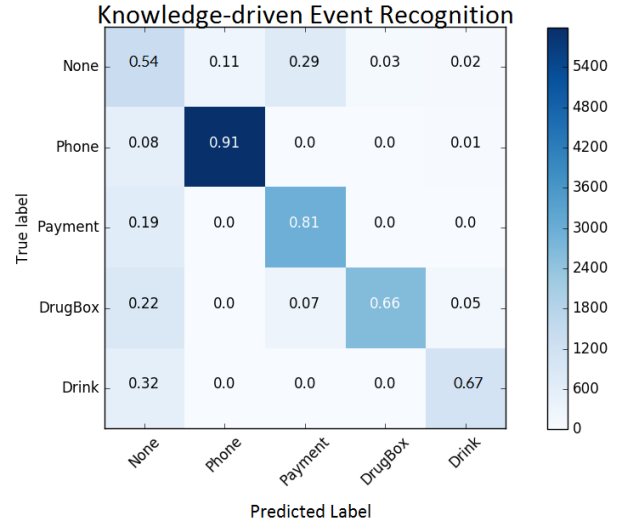

Figure 4. Confusion matrix of the Knowledge-driven event recognition system (baseline). The labels "None", "Phone", "Payment", "DrugBox" and "Drink" refer to, respectively: "no event", "talk on the telephone", "make payment using a tablet", "prepare pill box" and "prepare drink".
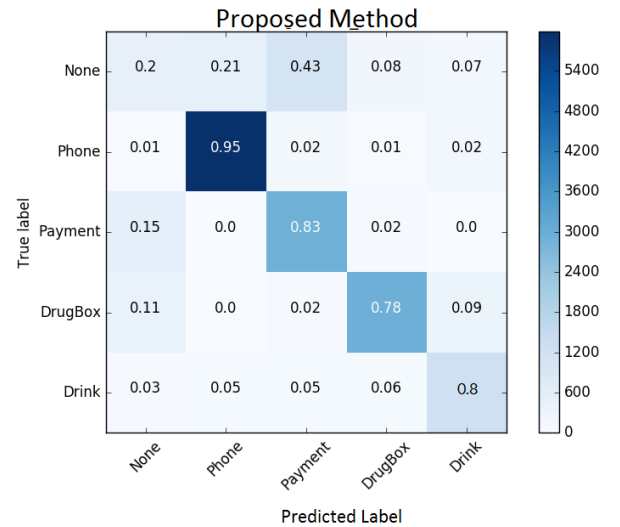

Figure 5. Confusion matrix of the proposed method. The labels "None", "Phone", "Payment", "DrugBox" and "Drink" refer to, respectively: "no event", "talk on the telephone", "make payment using a tablet", "prepare pill box" and "prepare drink".

6. Conclusions

This paper has presented BEHAVE, a probabilistic framework for the recognition of activities of daily living in assisted living scenarios. It combines robust algorithms for people detection and tracking using color-depth sensing with the probabilistic inference of ProbLog2 language. ProbLog2 models handle uncertainty caused by missed and ambiguous observations of the real world at event-level. Moreover, by relying on semantic information, BEHAVE
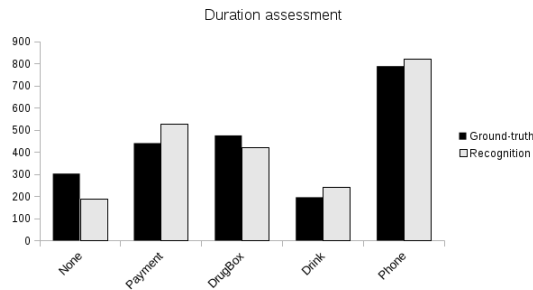
Figure 6. Assessment of event duration - ICP dataset

can also delimit the temporal boundaries of monitored activities, an advantage over most state-of-the-art methods which high-performance is tied to the use of hand-clipped video segments. In summary, BEHAVE is an approach more tailored for real-world settings, such as assisted living scenarios, since it can handle different sources of uncertainty that intervene in the task of event recognition, like partially satisfied event models.

Further work will investigate how to incorporate temporal relations into ProbLog2 models, as well as fine-grained semantic features, like body part relations. Finally, BEHAVE will be also studied as a unbiased and objective tool for the assessment of the autonomy of seniors at daily activities, *e.g.*, at medical consultations or daily at their own homes.

References

[1] A. Artikis, M. Sergot, and G. Paliouras. An event calculus for event recognition. *Knowledge and Data Engineering, IEEE Transactions on*, 27(4):895–908, April 2015. 2

[2] K. Avgerinakis, A. Briassouli, and Y. Kompatsiaris. Activity detection using sequential statistical boundary detection (ssbd). *Comput. Vis. Image Underst.*, 144(C):46–61, Mar. 2016. 5

[3] T. Banerjee, J. M. Keller, M. Popescu, and M. Skubic. Recognizing complex instrumental activities of daily living using scene information and fuzzy logic. *Computer Vision and Image Understanding*, 140:68 – 82, 2015. 2

[4] B. A. Boulbaba, J. Su, and S. Anuj. Action Recognition Using Rate-Invariant Analysis of Skeletal Shape Trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–14, Dec. 2015. 2

[5] L. Bourdev and J. Malik. Poselets: Body part detectors trained using 3d human pose annotations. In *International Conference on Computer Vision (ICCV)*, 2009. 2

[6] D. P. Chau, F. Bremond, and M. Thonnat. A multi-feature tracking algorithm enabling adaptation to context variations. In *The International Conference on Imaging for Crime Detection and Prevention (ICDP)*, London, United Kingdom, Nov. 2011. 3

[7] L. Chen, C. Nugent, and G. Okeyo. An ontology-based hybrid approach to activity modeling for smart homes. *Human-*

Machine Systems, IEEE Transactions on, 44(1):92–105, Feb 2014. 2

[8] C. Crispim-Junior, K. Avgerinakis, V. Buso, G. Meditskos, A. Briassouli, J. Benois-Pineau, Y. Kompatsiaris, and F. Bremond. Semantic event fusion of different visual modality concepts for activity recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2016. 5

[9] C. Crispim-Junior, V. Bathrinarayanan, B. Fosty, A. Konig, R. Romdhane, M. Thonnat, and F. Bremond. Evaluation of a monitoring system for event recognition of older people. In *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*, pages 165–170, Aug 2013. 2

[10] C. F. Crispim-Junior, A. Gmez Ura, C. Strumia, M. Koperski, A. Knig, F. Negin, S. Cosar, A. T. Nghiem, D. P. Chau, G. Charpiat, and F. Bremond. Online recognition of daily activities by color-depth sensing and knowledge models. *Sensors*, 17(7), 2017. 4, 5

[11] L. De Raedt and A. Kimmig. Probabilistic (logic) programming concepts. *Machine Learning*, 100(1):5–47, 2015. 3

[12] C. R. de Souza, A. Gaidon, E. Vig, and A. M. López. *Sympathy for the Details: Dense Trajectories and Hybrid Classification Architectures for Action Recognition*, pages 697–716. Springer International Publishing, Cham, 2016. 2

[13] P. Domingos and W. A. Webb. A tractable first-order probabilistic logic. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, AAAI'12, pages 1902–1909. AAAI Press, 2012. 3

[14] Y. Du, W. Wang, and L. Wang. Hierarchical recurrent neural network for skeleton based action recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 2

[15] C. Feichtenhofer, A. Pinz, and A. Zisserman. Convolutional two-stream network fusion for video action recognition. *CoRR*, abs/1604.06573, 2016. 2

[16] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010. 2

[17] D. Fierens, G. Van Den Broeck, J. Renkens, D. Shterionov, B. Gutmann, I. Thon, G. Janssens, and L. De Raedt. Inference and learning in probabilistic logic programs using weighted boolean formulas. *Theory and Practice of Logic Programming*, 15(3):358401, 2015. 3

[18] R. Helaoui, D. Riboni, and H. Stuckenschmidt. A probabilistic ontological framework for the recognition of multilevel human activities. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 345–354, New York, NY, USA, 2013. ACM. 2

[19] J. F. Hu, W. S. Zheng, J. Lai, and J. Zhang. Jointly learning heterogeneous features for rgb-d activity recognition. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5344–5352, June 2015. 2

[20] A. Kônig, C. F. Crispim-Junior, A. G. U. Covella, F. Bremond, A. Derreumaux, G. Bensadoun, R. David, F. Verhey,

P. Aalten, and P. Robert. Ecological assessment of autonomy in instrumental activities of daily living in dementia patients by the means of an automatic video monitoring system. *Frontiers in Aging Neuroscience*, 7:98, 2015. 1

[21] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97, 1955. 3

[22] I. Laptev. On space-time interest points. *International Journal of Computer Vision*, 64(2):107–123, 2005. 2

[23] B. E. Lyons, D. Austin, A. Seelye, J. Petersen, J. Yeargers, T. Riley, N. Sharma, N. C. Mattek, H. Dodge, K. Wild, and J. A. Kaye. pervasive computing technologies to continuously assess alzheimers disease progression and intervention efficacy. *frontiers in aging neuroscience*, 7(102), 2015. 1

[24] G. Meditskos, E. Kontopoulos, and I. Kompatsiaris. *The Semantic Web – ISWC 2014: 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part II*, chapter Knowledge-Driven Activity Recognition and Segmentation Using Context Connections, pages 260–275. Springer International Publishing, Cham, 2014. 2

[25] A. T. Nghiem and F. Bremond. Background subtraction in people detection framework for rgb-d cameras. In *Proceedings of 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2014. 2, 3

[26] S. Ren, K. He, R. B. Girshick, and J. Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015. 2

[27] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, pages 1297–1304, Washington, DC, USA, 2011. IEEE Computer Society. 2

[28] S. A. M. Sikkes, P. J. Visser, D. L. Knol, E. S. M. de Lange-de Klerk, M. Tsolaki, G. B. Frisoni, F. Nobili, L. Spiru, A. S. Rigaud, L. Frlich, M. O. Rikkert, H. Soininen, J. Touchon, G. Wilcock, M. Boada, H. Hampel, R. Bullock, B. Vellas, Y. A. Pijnenburg, P. Scheltens, F. R. Verhey, and B. M. Uitdehaag. Do instrumental activities of daily living predict dementia at 1- and 2-year follow-up? findings from the development of screening guidelines and diagnostic criteria for predementia alzheimer's disease study. *Journal of the American Geriatrics Society*, 59(12):2273–2281. 1

[29] A. Skarlatidis, A. Artikis, J. Filippou, and G. Paliouras. A probabilistic logic programming event calculus. *Theory and Practice of Logic Programming*, 15(2):213245, 2015. 3

[30] A. Skarlatidis, G. Paliouras, A. Artikis, and G. A. Vouros. Probabilistic event calculus for event recognition. *ACM Trans. Comput. Logic*, 16(2):11:1–11:37, Feb. 2015. 3

[31] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu. Dense trajectories and motion boundary descriptors for action recognition. *Int. J. Comput. Vis.*, 103(1):60–79, May 2013. 2