Simultaneous Plan Recognition and Monitoring (SPRAM) for Robot Assistants

Michael Karg
Institute for Advanced Study
Technische Universitt München
Lichtenbergstrasse 2a
D-85748 Garching, Germany
Email: kargm@in.tum.de

Alexandra Kirsch
Department of Computer Science
University of Tübingen
Sand 14
D-72076 Tübingen, Germany
EMail: alexandra.kirsch@uni-tuebingen.de

***Abstract*—Personal robots that are useful helpers for humans need to know about common tasks and habits of their human partners to be able to react adequately to human behavior. Thus, knowledge about human task performance becomes an inevitable part of a robotic system that is aimed to work together with humans in human centered environments like a household. But high uncertainties within the robots sensors as well as unpredictable behavior of humans and partial occlusions usually make it hard to achieve certainty about human task execution. We propose a probabilistic module that performs Simultaneous Plan Recognition and Monitoring by maintaining a belief state about human task execution and simultaneously monitoring plans that the human is likely to execute. This way the robot is able to infer possible reactions even if it may not exactly know which plan its human partner is executing. In the application example of a household robot observing a typical morning routine of a person, we show that our module is able recognize and monitor a set of activities of daily living in real time and predict places that the human is likely to visit in near future.**

I. Introduction

Service robots that work in human environments should carry out tasks that their human supervisors are not able to do or do not want to do and assist the human with tasks that cannot be done alone. While a useful service robot will probably very often receive explicit instructions by the human, an efficient robot will also sometimes have to make decisions on its own while being aware of the current state of the human. Consider for example a cleaning robot that is supposed to clean the apartment. Having detected that the human had breakfast and went to work but probably didn't clean the table, the robot should start to clean the table and maybe even start vacuuming the flat. On the other hand, vacuuming the living room while the human is sitting on the sofa watching TV or sleeping may not be a very good idea. For a robot to be this situation-aware, it must have knowledge about the tasks and activities of its human partner. Thus a model about currently and also previously executed activities of the human is essential for service robots that are efficient helpers in human homes.

But unless your apartment is equipped with a lot of sensors, it will not be possible for a robot to observe all actions of the human at every time. Therefore, it makes sense to maintain a probabilistic model of human activities that relies on only partial observations of the human position and actions. The robot will have to deal with high uncertainties regarding the estimation of the current activity of the human partner and in many cases it may not be possible for it to exactly know what the human is doing. This is why we propose the idea of Simultaneous Plan Recognition and Monitoring (SPRAM) that enables us to use (partial) observations of the human to maintain a probabilistic model about several human activities and their state of execution. This has the advantage that even if we are not sure about which activity we are observing, we can still draw conclusions about the human intentions and react adequately. Imagine, for example, a household robot that is observing a human in a kitchen during his daily morning routine. Even if the robot is not sure if the human is preparing cereals, curd-cheese or bread for breakfast, it could still infer that the human is preparing a meal, which has the consequence that the table has to be cleaned afterwards. Thus, we would even be able to generate predictions about e.g. future locations and or objects needed by the human even if we don't exactly know what the human is doing

II. Related Work

A large number of approaches for activity recognition heavily rely on different variations of the Hidden Markov Model (HMM) [3, 10], Hierarchical Conditional Random Fields (CRFs) [13] or Hierarchical Maximum Entropy Markov Models (MEMM) [16].They often assume that the observation and transition probabilities only depend on the current state of the human and the observations are independent of each other. Some approaches introduce extensions that try to overcome the limitations by e.g. the inclusion of durations [6] or hierarchies [11, 4]. Many systems for the recognition of activities of daily living (ADLs) heavily rely on object- and motion detections and have to equip the environment with many sensors. Perkowitz et al. [12] equip a wide variety of objects in a human household with RFID tags to detect sequences of objects that are used by a human while performing an everyday activity. Buettner et al. [3] use WISPs, a combination of RFID tags and accelerometers, to generate sequences of objects that are moved by a human performing an everyday activity.

When it comes to activity monitoring, one popular application area is elder care. Pollack et al. [14] present Autominder, a cognitive orthoic system based on Quantitative Temporal Bayesian Networks that uses activity monitoring for elderly with memory impairments to issue reminders if e.g. a person forgot to take his medicine. Also Cesta et al. [5] propose a proof-of-concept intelligent environment for elderly people. They perform proactive monitoring based on constraint-based temporal knowledge to detect abnormalities in the behavior of people using predefined behavioral patterns that were defined by the caregivers according to the user's medical needs. But human activity models don't have to be predefined. Beetz et al. [1] learn partially ordered models of a human table setting task from full body motion tracking data.

Humans generally tend to represent spatial regions not only geometrically but also according to their functional use. So for a robot interacting in a human-populated environment, it must understand its environment in terms of human spatial concepts [19]. One step towards this understanding for machines is done by Liao et al. [9]. They use hierarchical conditional random fields to learn patterns of human behavior from GPS traces, recognize significant places that the human visits during everyday activities and label them according to their function (office, home, ...). Stulp et al. [15] propose a representation of the utility of positions in the context of action-related mobile manipulation. They define so-called *ARPlaces* as probability distributions in reference to the pose of objects to model the probability for a successful grasp. Klenk et al. [8] find that "the ability to understand and reason about spatial regions is essential for cognitive systems performing tasks for humans in everyday environments". In their work, they define context dependent spatial regions for cognitive systems that are learned by qualitative spatial representations and semantic labels. Townsend et al. [17] found that humans tend to pattern daily actions into sequences which they repeat at particular times in particular places. That means a the majority of activities of daily living (ADLs) are based on habits and thus are mostly carried out in the same way at the same locations.

III. Modeling Human Activities Using Spatio-Temporal Plan Representations

To capture habit-based ADLs in our experiments, we decided to use spatio-temporal plan representations that can be acquired from motion tracking data and a semantic environment map we showed in [7]. The spatio-temporal plan representations (STPRs) cover the locations that a human visits while performing an activity in reference to objects (e.g. furniture) in the environment and include the durations that a human spends at those semantically annotated locations. Although this way of modeling human activities seems perfect to model ADLs that the human performs out of habit, some of those activities might look similar with regard to the locations and durations the human visits. To overcome these shortcomings and improve activity recognition results, we extended the STPRs with object detections where available. The object detections can either be obtained directly from sensor observations (RFID, vision, ...) or can even be partly inferred using partial order models like in [1].

We define a spatio-temporal plan description $stpr_n$ as a sequence of n tuples that have a location l_i, a duration t_i and a vector of object detections $o_i^{m_t}$ as elements where m_t denotes the number of detected objects at the current location at time t.

$$sptr_n = ((l_1, t_1, o_1^{m_t}), (l_2, t_2, o_2^{m_t})..., (l_n, t_n, o_n^{m_t})) \qquad (1)$$

The locations l_i are stored in the spatial model ψ which is generated from observations of human 2D positions and a semantically annotated map of the environment that can in wide parts be acquired autonomously [2]. For simplicity reasons, we use a set of Gaussians P_i linked to semantically annotated instances of furniture objects o_i in our semantic map:

$$\psi = \{l_1, l_2, ..., l_n\}, l_i = (P_i, o_i) \qquad (2)$$

In [7], we learned a general model of the locations of humans relative to storage locations of objects for pick- and place actions out of three categories of containers: cupboards drawers and general surfaces. As we show, locations for pick- and place actions can be transferred and we use those learned relative locations combined with a semantic environment map to generate a spatial model ψ that represents locations where we expect the human to be when performing pick- and place actions.

IV. Simultaneous Plan Recognition and Monitoring

Our proposed approach of Plan Recognition and Monitoring (SPRAM) consists of two components that are executed simultaneously, the activity recognition module and the monitoring module. The activity recognition module constantly estimates probabilities for the different activities that the human is likely to perform while the monitoring module tracks the progress of every plan.

A. Activity Recognition

For the activity recognition part, we use an Hierarchical Hidden Markov model (HHMM) as in [4] since HHMMs can decently model the hierarchical nature of human activities. In contrast to a Hidden Markov Model (HMM), in a HHMM every state can itself be a (H)HMM. Every time, a state in a HHMM is activated, a so *vertical transition* occurs, i.e. a state of the underlying HMM is activated. The process is repeated until a *production* state is activated. Production-states are the only states that emit observations like in a common HMM and states, that do not emit observations are called *internal states*. After a production state is activated, horizontal transitions occur until a *terminal production state* is reached. Then state control is given back to the corresponding internal state. A (H)HMM consists of a set of states, a set of observations, state transition probabilities, and observation probabilities as well as an initial distribution about the belief over all states. In our case, we are using STPRs to generate the transition probabilities between the states of our HHMM using a Maximum Likelihood estimation where the states of our HHMM correspond to the semantically annotated locations of a STPR. The Maximum Likelihood Estimation results in a

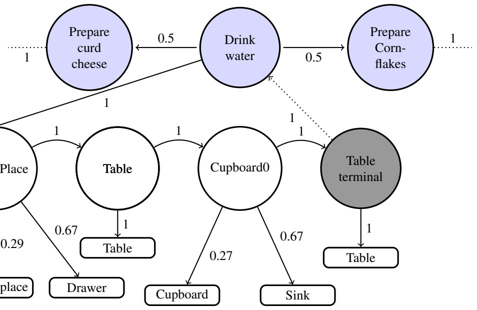

Fig. 1. A hierarchical HMM for activity recognition generated from a STPR. The light gray nodes of the graph represent the plan-states which are HMMs themselves. For visualization, only the "Drink Water" HMM is shown in detail.White nodes are the production states which in our case correspond to locations where the observed human is standing. The dark gray node is a terminal (production) state which, when reached, leads to a horizontal transition in the HHMM. The rectangles represent the observations that are expected by the production states.

HHMM such as the one shown in figure 1. For better visualization, the picture assumes transition probabilities between states that do not appear in the STPRs to be zero. In the current implementation, we applied Laplacian Smoothing with $k = 1$ to prevent model over-fitting. This will prove to be important due to wrong or missing detections of locations resulting from sensor noise and undetected locations in the Kinect data, as we will show in the experiments in section V-B. The observations correspond to the semantically annotated locations where our system observes the human to be standing for a short time.

Every time a new observation is added to the HHMM, the Forward-Backward Algorithm can be used to efficiently calculate the belief probability distribution $P(X_t|O_{1:t})$, where X_t represents the a state and $O_{1:t}$ are the observations so far. In our current implementation, we use the "flattening" method [18] and convert the HHMM to an equivalent HMM and use the standard Forward-Backward Algorithm to calculate the probability for each state and thus the probability of the internal states, i.e. the probabilities for each plan being executed.

B. Activity monitoring

Being a special case of a Dynamical Bayesian Network, the Hidden Markov Model makes use of the Markov assumption for the sake of computational and mathematical tractability. This means that every state only depends on its predecessor, which in our application of states representing locations in a household, means that the locations that a human will visit next is only dependent on the location where he is standing now. This is a rather strong assumption, since, if you imagine a typical table-setting task, your location will also depend on how often you already have visited certain places. So to predict future locations of a person, we want to keep track of the current state of the task execution. As Sung et al. state, activity recognition [16] mostly is not 100 % sure what plan the human is executing. They found that uncertainties are highest among very similar activities, even when using sensor data without occlusions. Therefore, it makes sense to monitor those that have high probabilities of being executed and find similarities among them. In our case, where a typical morning routine of a human that only consists of a limited set of activities, it is even tractable to monitor all of the plans in real time. Using STPRs as models of our human tasks and information about observed visited locations as well as probabilities for each plan as input, we propose a simple monitoring routine that keeps track of the state of execution of every plan and predicts locations that are likely to be visited in near future. Therefore, it maintains a *locations cache* for every plan, that keeps track of a list of locations that have been visited while the probability for the specific plan was high. If the probability for a plan falls below a certain threshold, i.e. the plan is unlikely to be executed at the moment, the locations cache is reset and the list is empty. Using this simple routine, we are able to distinguish between observations that are likely to belong to a certain plan and others that do not. This is important, since, for example, we do not want to mark the place "table" as visited in the Prepare for work" plan when we are quite sure that the human is currently not preparing for work. In this case we rather want to keep all places of the "Prepare for work" plan as possible next locations but give them a low weight. To obtain a probability $P^{next}_{mon}\left(l_i|a\right)$ about which plan-dependent locations we expect to be visited by the human in near future, we test if all locations of the STPR l_i^{STPR} have been visited the expected number of times using the information of the locations cache. This gives us information about if we expect a location to be visited again in near future or not and we simply define the probability $p^{next}_{mon}\left(l_i|a\right)$ for a location l_i to be visited in near future to 1 if the total number of visits according to the locations cache $l_i^{loc-cache}$ is smaller than the expected total number of visits and set it to 0 otherwise (*Note:* This is an extremely simple way of monitoring that we use as proof-of-concept. A more elaborate approach is part of our future work.).

$$p^{next}_{mon}\left(l_i|a\right) = \begin{cases} 1 & \text{if } \left|l_i^{STPR}\right| - \left|l_i^{loc-cache}\right| > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Using the Forward Step of the Forward-Backward in our HHMM algorithm gives us the probability distribution $P^{next}_{FW}\left(l_i^{t+1}|a_j^{t+1}\right)$ over the production states (the locations) l_i^{t+1} to be visited in the next step $t+1$ conditioned on the activity a_j^{t+1}. We now use the probabilities of our monitoring step $P^{next}_{mon}\left(l_i|a\right)$ to penalize plan-dependent locations in $P^{next}_{FW}\left(l_i^{t+1}|a_j^{t+1}\right)$ that are not expected to be visited again according to the monitoring by giving them a low weight. Thus we obtain a merged probability distribution over plan-dependent locations that we expect to be visited in the next step: $P^{next}_{merged}\left(l_i^{t+1}|a_j^{t+1}\right)$. Marginalizing out a_j^{t+1}, we get a probability distribution over all locations to be visited in the next step taking into account all activities:

$$P(l_i^{t+1}) = \int\limits_{a_i} P\left(l_i^{t+1}|a_j^{t+1}\right) * P\left(a_j^{t+1}\right) da. \tag{4}$$

Since we model locations as two dimensional gaussians, $P(l_i^{t+1})$ can be seen as a spatial model comprised of a set of weighted, two dimensional gaussians that represent locations that are likely to be visited in the next time and are updated with every belief-update of the activity recognition. This rather simple approach already allows a robot to have a guess about where the human is likely to be or which space is likely not visited by the human.

V. Application: SPRAM for Household Robots

Continually recognizing and monitoring human task performance enables a robot to maintain a probabilistic spatial model about likely locations a human will visit in near future even without being 100 percent sure about what activity the human is performing. Thus it can react adequately to the behavior of its user.

A. An Activity Dataset for a Human Morning Routine

We investigated typical morning routines of the annotated MIT PlaceLab PLCouple dataset [1] and found that in a common morning routine of a human, the set of different plans that a human performs is very limited (11 activities over 10 weeks performed in a kitchen). Although the annotations of the PLCouple dataset are publicly available only one of the two partly cooperating persons was tracked using a RFID reader due to financial restrictions. Also the full audio and video data are not available due to privacy issues, so many tasks that are part of a typical human routine were not available. With the aim of creating a dataset that captures a typical morning routine of a person, we investigated a period of 14 workdays of a voluntary test person that did not know about our system. We told the test person to write down the activities that he performed before going to work and the locations where he stood still while performing those activities over 3 weeks. We decided to limit our experiment to actions that happen in the kitchen, which consist of preparing a drink, drinking a glass of water, preparing breakfast, having breakfast, cleaning the table, packing a bottle of water into the backpack and leaving the room with the backpack. To obtain motion tracking data and object detections of those morning routines, we told the participant to reenact his morning routines in an experimental kitchen equipped with two Kinects (one for motion-tracking and one for visual marker detections on objects). The experimental kitchen is shown in Figure 2.

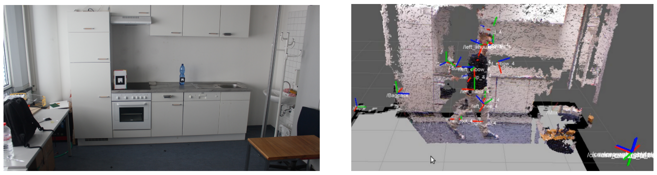

Fig. 2. The experimental kitchen environment is shown on the left picture, while the right picture shows the sensor data of the two Kinect sensors. The motion tracking returns coordinate transformations for each joint of the human while the visual marker detection returns a coordinate transform for each detected object. The map and the Pointcloud data are only shown for visualization.

B. Experiments

In our experiments, we perform Simultaneous Plan Recognition and Monitoring in the experimental kitchen scenario of the dataset explained in section V-A. We use STPRs as models for our plans and the corresponding HHMMs that we generated as explained in section IV-A. The HHMM is updated as soon as we detect that the human is "standing still", i.e. where his center of mass is not moving more than 25 cm within 0.5 seconds (0.5 m/s). If we detect such a location, we query the spatial model to which semantically annotated location the coordinate of the human corresponds to most likely we did in [7]. While in their case, the locations are not overlapping and unique, our spatial model includes gaussians that overlap which results in locations that can easily be mixed up as can be seen in figure 3. Those overlapping locations are the result of kitchen furniture that are located very close to each other. In our kitchen, for example, the cupboards are located on top of the kitchenette, which results in very similar locations of e.g. the sink and the cupboard that is on top of it. To address this issue, we learn the observation probabilities of the HMM of our activity recognition module using maximum likelihood estimation in a set of training data. This is the most time-consuming step in our approach, but as other experiments showed, the observation probabilities can as well be estimated from the spatial model with slightly less accurate results.

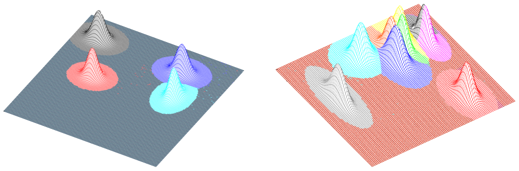

Fig. 3. The left picture shows a spatial model with only a four context dependent spatial regions. In this case, it is easy to distinguish between different locations. The right picture shows the spatial model of a more realistic kitchen environment where some furniture objects are close to each other. Most of the gaussians are located very close together and thus make it hard to reliably detect unique locations due to overlapping.

The fact that some locations are close to each other and cannot clearly be distinguished is not only a drawback. On the one hand, this leads to the activity recognition having a hard time distinguishing different plans, but on the other hand, some of the plans that look quite similar often have similar properties. This can be seen in figure 4 where the plan probabilities (upper picture) and the probabilities for next locations to be visited (lower picture) are plotted over time. While the human is preparing the table for cornflakes, plan recognition returns similar probabilities for the plans "Prepare Cornflakes" and "Prepare Curd Cheese" and the two cleaning plans are also assigned high probabilities during some parts of the human preparing cornflakes. This comes as no surprise since these plans mostly consist of common locations or locations that are close to each other.

By simultaneously monitoring all plans and predicting future locations, we can still create a weighted spatial model as explained in section IV-B and give a robot an idea in which

[1] http://architecture.mit.edu/house_n/data/PlaceLab/PLCouple1.htm

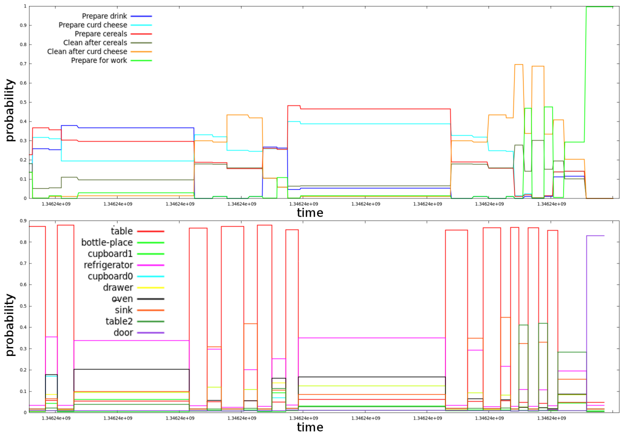

Fig. 4. The upper picture shows plan probabilities of one experiment of the real dataset estimated by the activity recognition. The probabilities in the lower picture correspond to locations in the spatial model that the human is expected to visit next. Although, activity recognition is not able to reliably distinguish between some activities in many cases, some future locations of the human can be predicted with high probabilities due to partial similarities of some activities.

area the human is to be expected soon as can be seen in the lower picture in Figure 4. Although plan recognition is not very sure about which plan the human is currently executing, we can quite reliable detect some of the future locations that the human is about to visit. Even when activity recognition performs bad and is not sure which activity is performed, the location “table” is predicted correctly almost all of the time due to partial similarities in the plans: In almost every case, the human will go to location “table” after he visited another location.

To evaluate the quality of our plan recognition and location prediction, we calculated precision, recall and accuracy for each activity a using ground truth labels of the dataset in the following way:

$$precision = \frac{|t_a \cap t_a^*|}{|t_a^*|}, \; recall = \frac{|t_a \cap t_a^*|}{|t_a|}. \tag{5}$$

t_a represents the time when activity a has been executed by the participant according to the ground truth labels of the dataset and t_a^* stands for the time where the detection estimates activity a to be the most likely one. We also calculate the *accuracy* which is the proportion of true classification results (true positives and true negatives) during the whole observation period t_{obs}:

$$accuracy = \frac{|t_a \cap t_a^*| + |\bar{t_a} \cap \bar{t_a^*}|}{|t_{obs}|}. \tag{6}$$

$\bar{t_a}$ corresponds to the time when activity a has not been performed and $\bar{t_a^*}$ represents time periods when activity a has not been classified as the most likely activity. Table I shows the average of precision, recall and accuracy of activity recognition for 12 experiments as well as the percentage of correctly predicted locations for each activity (meaning the location with the highest probability of our predicted spatial model $P(l_i^{t+1})$ was indeed visited next by the human). Overall activity recognition rates are slightly worse than the ones of e.g. Buettner et al. [3], who attach RFID based sensors to 25 objects and equipped an apartment with antennas. But given the fact that our system is way less intrusive and needs less sensors, the recognition rates are still more than sufficient to be useful for a robot. The locations that were visited by the human could be correctly predicted in 66.3 % of the cases on average. One should note that this evaluation does not reflect our whole predictions since even if the location prediction with the highest probability was wrong or is not significantly unique, the robot still has the spatial model $P(l_i^{t+1})$ which gives it an impression about in which region the human is likely to go next as can be seen in Figure 5 (A more elaborate evaluation of our location prediction is part of our future work).

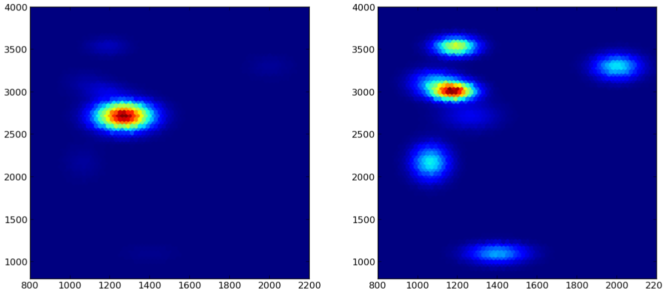

Fig. 5. The spatial model displays probabilities about where the human is expected to go next. In the left picture, one location has a very high probability of being the next one while on the right picture, the predictions are not unique. Nevertheless, the spatial model shows likely next locations with corresponding probabilities.

Activity	Prec.(%)	Recall (%)	Acc. (%)	Loc. (%)
Drink water	35.9	37.0	76.4	61.1
Prepare cereals	51.9	67.5	62.9	52.8
Prepare curd cheese	34.8	25.0	63.0	63.5
Clean after cereals	68.4	23.2	82.3	58.6
Clean curd cheese	85.8	34.1	84.9	70.0
Prepare work	63.4	91.3	92.6	91.7

TABLE I
AVERAGE RESULTS FOR 12 EXPERIMENTS.

Most of our experiments show that the different food-preparing and cleaning-tasks cannot be distinguished reliably using only information about the locations but nevertheless, predictions about future locations can be generated as visualized in figure 4. Some of the locations (in our case the table) can be predicted very reliably while in other cases, our system was not too certain which location would be visited next. However, in this case, we still have a weighted spatial model which gives information about several locations that possibly are visited in near future as shown in Figure 5. In the case of e.g. a household robot, the robot could try to avoid these regions in case he does not want to disturb the human or search those locations if it is looking for the human.

VI. Conclusion

We presented a framework for Simultaneous Plan Recognition and Monitoring that enables a robot to maintain a belief state about human task execution and predict locations where its human partner is likely to go next even if it does not exactly know which plan the human performs. This has the advantage that the robot does not have to heavily rely on accurate plan recognition to generate certain predictions. We evaluated our system in an experimental kitchen using two Kinects and visual markers for object detection. We show that we can generate predictions about future locations of the human during task performance. There are quite some improvements for future work. We are thinking about extension of our HHMMs towards objects and contextual features (e.g. daytime, etc.), more elaborate monitoring techniques, better handling of partial occlusions and the integration of SPRAM into human-aware planning.

References

[1] Michael Beetz, Jan Bandouch, Dominik Jain, and Moritz Tenorth. Towards Automated Models of Activities of Daily Life. In *First International Symposium on Quality of Life Technology - Intelligent Systems for Better Living*, Pittsburgh, Pennsylvania USA, 2009.

[2] Nico Blodow, Lucian Cosmin Goron, Zoltan-Csaba Marton, Dejan Pangercic, Thomas Rühr, Moritz Tenorth, and Michael Beetz. Autonomous semantic mapping for robots performing everyday manipulation tasks in kitchen environments. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, CA, USA, September, 25–30 2011. Accepted for publication.

[3] M. Buettner, R. Prasad, M. Philipose, and D. Wetherall. Recognizing daily activities with rfid-based sensors. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 51–60. ACM, 2009.

[4] H.H. Bui, D.Q. Phung, and S. Venkatesh. Hierarchical hidden markov models with general state hierarchy. In *Proceedings of the National Conference on Artificial Intelligence*, pages 324–329. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004.

[5] A. Cesta, G. Cortellessa, R. Rasconi, F. Pecora, M. Scopelliti, and L. Tiberio. Monitoring elderly people with the robocare domestic environment: Interaction synthesis and user evaluation. *Computational Intelligence*, 27(1):60–82, 2011.

[6] T.V. Duong, H.H. Bui, D.Q. Phung, and S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-markov model. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 838–845. Ieee, 2005.

[7] Michael Karg and Alexandra Kirsch. Acquisition and Use of Transferable, Spatio-Temporal Plan Representations for Human-Robot Interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012.

[8] Matthew Klenk, Nick Hawes, and Kate Lockwood. Representing and reasoning about spatial regions defined by context. In *AAAI Fall 2011 Symposium on Advances in Cognitive Systems*, 2011.

[9] Lin Liao, Dieter Fox, and Henry Kautz. Extracting places and activities from gps traces using hierarchical conditional random fields. *Int. J. Rob. Res.*, 26(1):119–134, 2007.

[10] N.T. Nguyen, H.H. Bui, S. Venkatsh, and G. West. Recognizing and monitoring high-level behaviors in complex spatial environments. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–620. IEEE, 2003.

[11] N.T. Nguyen, D.Q. Phung, S. Venkatesh, and H. Bui. Learning and detecting activities from movement trajectories using the hierarchical hidden markov model. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 955–960. IEEE, 2005.

[12] Mike Perkowitz, Matthai Philipose, Kenneth Fishkin, and Donald J. Patterson. Mining models of human activities from the web. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 573–582. ACM, 2004.

[13] D.Q. Phung, H.H. Bui, S. Venkatesh, et al. Hierarchical semi-markov conditional random fields for recursive sequential data. *Arxiv preprint arXiv:1009.2009*, 2010.

[14] M.E. Pollack, L. Brown, D. Colbry, C.E. McCarthy, C. Orosz, B. Peintner, S. Ramakrishnan, and I. Tsamardinos. Autominder: An intelligent cognitive orthotic system for people with memory impairment. *Robotics and Autonomous Systems*, 44(3):273–282, 2003.

[15] Freek Stulp, Andreas Fedrizzi, and Michael Beetz. Action-related place-based mobile manipulation. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, 2009.

[16] J. Sung, C. Ponce, B. Selman, and A. Saxena. Unstructured human activity detection from rgbd images. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 842–849. IEEE, 2012.

[17] D.J. Townsend and T.G. Bever. *Sentence comprehension: The integration of habits and rules*. The MIT Press, 2001.

[18] L. Xie, S.F. Chang, A. Divakaran, and H. Sun. Unsupervised discovery of multilevel statistical video structures using hierarchical hidden markov models. In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*, volume 3, pages III–29. IEEE, 2003.

[19] H. Zender, O. Martínez Mozos, P. Jensfelt, G.J.M. Kruijff, and W. Burgard. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 2008.