Acquiring Models of Everyday Activities for Robotic Control

Radu Bogdan Rusu
Intelligent Autonomous Systems
Computer Science Department IX
Technische Universität München
Boltzmannstr. 3
85748, Garching, Munich
+49 (89) 289-17780
rusu@cs.tum.edu
http://www9.cs.tum.edu/people/rusu

PhD-Supervision: **Prof. Michael Beetz, PhD**

Doctoral Seminar, Linz, 2006

Abstract
Intelligent sensor equipped environments can be of much greater help if they are capable of recognizing the actions and activities of their users, and inferring their intentions. An intelligent kitchen that recognizes what a person is looking for can highlight the target object. An oven noticing that the cook is on the phone can reduce the heating temperature, in order to avoid the meal getting burnt. In my dissertation research, I investigate the representation of models of everyday activities and study how such models can be learned from sensory data.

Keywords
Context-aware computing, intentional activity recognition, probabilistic activity models, sensor data fusion, artificial intelligence, cognitive robotic systems

Problem Statement and Research Question
Understanding human activities and characterizing them into expressive and detailed activity models is one of they key issues of today's current pervasive computing systems. If such a system could recognize and understand automatically its user's behavior, it could interact in a more efficient and friendly manner. Unfortunately, the current model construction techniques are based on supervised learning and require specifications from their human counterparts, such as

labeling the acquired sensor data. Our vision is to build technical cognitive systems that create and use models in a straightforward manner, by combining already existing online information with the system's context history.

Recent advances in the development of pervasive computing with an emphasis on wireless sensor networks, enable us to gather rich datasets of sensor streams. Understanding these data streams up to the level of inferring informative and detailed models of human activity from them is a challenging task.

Modern industrial robots are capable of performing very accurate and fast motions and manipulations. Their control programs are very concise and repetitive, but they can never be used for activities such as setting up the breakfast for a person, since they cannot deal with the uncertainties and probabilities associated with daily complex human activities.

The main motivation of our research is to build cognitive systems that can recognize and interpret accurately intentional activities of daily living, thus understanding and learning how to provide their users with a better overall experience.

By being able for instance to determine what activity a person is currently performing, and more importantly, how is (s)he performing that activity, we can build statistical representations of those daily activity experiences.

Our research focus concentrates on obtaining detailed models of daily activities and using them in both better understanding human actions and creating intelligent robotic household assistants. Before reaching our goal, we believe that we will encounter on the way, a series of questions that should be answered:

- What is the most appropriate way to represent these probabilistic models?
- How accurate can the models be, and how do we define a good model?
- Once built, how can the model be used by the cognitive system?
- What types of sensor streams and how much context history does a system need in order to build rich models?

A possible scenario could be the following one: after having learned for some time models of basic activities by observing the actions of people in a context-aware kitchen, the system starts its normal operation. A person selects a recipe from the World Wide Web, on an embedded touchscreen computer, and starts cooking. The system supports its user by highlighting where the tools are, what is the status of the ingredients, where should the actions be performed, etc.

The motivation of looking at complex everyday activities is that they do not require repetitive, constant actions such as the industrial ones, their execution is probabilistic and interleaved, they deal with a lot of uncertainties and need detailed action parametrization.

Approach and Methodology

So far, research in this field has been restricted on either building models of activities for ADLs ([1], [3], [8]) from RFID sensor data, or learning and classifying a predefined set of isolated activities, mostly from accelerometer sensor data. Very few, if none, have combined the two, or thought about the possibility of combining that information with already unsupervised learned models that a context-aware environment can give.

Also, in the context of building motion blueprints (described as movement primitives in [4],[5],[6] and [9]), the research was mostly focused on proving that a complex motion can be decomposed into a set of predefined, handcrafted, behavioral primitives, from either video data ([2]) or other types of complex motion capture systems. These data streams were rather hard to acquire and quite cumbersome to wear or act in front of, for the respective subjects.

We start building on these previously published results, but we envision technical cognitive systems that can record, analyze, interpret, and build models of human activities in an unsupervised straightforward fashion. These models will be later on used to build motion blueprints as a starting point for the developmental learning process of household robotic assistants, as well as for plan-based controllers.

During the last few months, we managed to create an excellent software infrastructure, building on top of the internationally renowned Player/Stage platform[10] (of which we are active developers), as well as on other opensource projects such as TinyOS[1]. This infrastructure is used both for sensor network deployment and data acquisition and serves as a bridge for data interpretation using machine learning algorithms projects such as Weka[2]. We conduct simulation experiments of household robotic assistants operated by plan-based controllers, as well as motion blueprints generation, clusterization and usage on our Powercube 6-DOF manipulators.

By analyzing the combined sensor data streams from a cognitive robotic perspective, we hope to be able to later on, build complex arm movements, for our B21 mobile robots, by sequentially combining and interpolating simple motion blueprints.

This work is including research topics from several different areas of science such as: artificial intelligence and learning, common sense reasoning, wireless sensor networks and robotics.

As an example scenario for validating and evaluating our research work, we are currently building a distributed sensor enabled environment (AwareKitchen) in our labs. We are making use of technologies such as B21 mobile robots equipped with Powercube 6-DOF manipulators, small embedded devices such as Gumstix that act as distributed sensor fusion nodes, wireless sensor networks (Motes, Particles), RFID tagged objects and readers, video and IR thermal cameras, as well as other types of sensors (for example, we are working together with people from the

[1] *http://www.tinyos.net*
[2] *http://www.cs.waikato.ac.nz/ml/weka/*

Ludwig-Maximilians University of Munich, in developing new capacity sensors that could be used to detect human motion).

Related Work

While some publications already covered areas of research such as inferring models of activities from sensor data, they mostly referred to the aspects of classifying an activity rather than reasoning about fine-grained details on how the activity was performed. In our research, we would like to take this one step further, since we believe that only very detailed models of human activities can be used in tasks such as building activity blueprints for a robotic household assistant.

Patterson, Fox et al.[1] present several results related to achieving fine-grained activity recognition for context-aware computing applications, but only from RFID data. They show the advantages of adding additional complexity to their previous models ([3], [8]) and conclude with a probabilistic model that can gracefully generalize classes of activities from object instances. While their work has a big impact on recognition and interpretation of RFID data, it does not say however, how is a certain activity performed, in terms of motion and environmental details.

Lee and Kim[2] developed a method for automatic gesture recognition from video data using Hidden Markov Models. They achieved good recognition rates of over 93.14% by introducing the concept of a threshold model (a complex garbage model or filler model) that calculates the likelihood threshold of an input pattern and provides a confirmation mechanism for the provisionally matched gesture patterns. They do not however, research if the same recognition rates can be achieved by using other types of sensor data, such as acceleration (for example), since optimal coverage for building such a model would be difficult to achieve and constrained by using just video data.

Perkowitz et al.[3] describe their experience on using the already existing knowledge from recipe web sites such as ehow.com and combining it with RFID sensor data to recognize more activities than a standard activity recognition system, which can only learn a tiny fraction of the thousands of human activities that are potentially useful to detect. They built an inference engine who accepts models from a mining engine and RFID sensor data as inputs and tries to provide activities with likelihoods as an output. The results are promising, but the resulted system does not make use of the entire knowledge available in a "recipe", thus the models only contain information about the objects involved in an activity and nothing about their properties, states, how are they manipulated, etc. We believe that by combining information from several recipe web sites and building more complex ontology models in knowledge bases such as Cyc, which already have strong commonsense reasoning engines, we will have a deeper understanding of how a certain activity is performed. The next step will be to create plan-based robot controllers that can be used by household robotic assistants to perform daily chores and complex tasks.

Jenkins and Mataric[4] address the problem of creating perceptual-motor primitives for a humanoid robot, from human motion capture data. The data is gathered from a Vicon optical motion capture system, which provides the trajectories of the markers in 3D, and the results are demonstrated in a dynamical humanoid simulator (Adonis). The motion primitives are however, manually crafted, and the motion capture system provides a cumbersome way of acquiring sensor data. We believe that multiple sensor streams, and combination of sensors such as RFID, accelerometers and even small laser units can provide an easier way of acquiring detailed datasets, in a non-intrusive manner. Furthermore, we are evaluating the possibility of building motion blueprints in an unsupervised manner by reasoning about previously acquired knowledge and combining it with the system's incoming fluent sensor streams.

The results presented in the above publications show the necessity of deeper research in the field. There are still plenty gaps to fill, and problems to solve, as we envision robotic assistants and context-aware environments that can automatically learn and build models of human activities, and then use them for interacting in a more efficient and friendly manner with their users.

Preliminary Results

Our efforts are concentrated right now on building the software infrastructure of a such cognitive system. We are currently developing and contributing with a lot of code and experience to the international cognitive robotics community.

The work is currently split into three major components:

- building complex, cognitive models of human activities in the ResearchCyc knowledge database by mining the world wide web, and using them together with sensor data streams by our inference reasoning engine for recognition and characterization of intentional activities;

- analyzing and interpreting streams of sensor data for building motion blueprints automatically, and using them for our B21 robotic assistants as a starting point in the developmental learning process;

- estimate the current activities and their states and provide a reliable source of information through multiple *sensor fusion nodes* in a context-aware environment for generating plan-based controllers for household robotic assistants.

Preliminary results have shown that we can build and cluster object usage in an intentional activity, by mining web sites such as ehow.com and extracting the appropriate instructions for that activity. After processing the information though several filters (such as Wordnet), we managed to create detailed representations of what objects are used and in what way in a certain daily activity.

We are experimenting with fusion of sensor data streams through the usage of wireless sensor networks and Player/Stage[10], for generating motion blueprints which can be used as a starting point in building and learning complex robotic arm movements. Our tests are being conducted on Powercube 6-DOF arms, mounted on a B21 mobile robot.

In the context of plan-based controllers, by modeling activities from sensor data, and fusing that with the existing knowledge about a certain activity, we can create better plans. Detailed recognition of the objects involved in an activity can lead to better image tracker routines which in turn, support a higher accuracy of the plan's execution routines.

Conclusions and Future Steps

To validate the results of our work, we are establishing application scenarios where this type of cognitive systems could prove to be useful. The initial testbed will be our AwareKitchen, a distributed sensing environment which we are currently building.

We will also investigate how we can transfer methods for learning action models that we have developed in the context of automatic football game analysis to this domain ([11]).

Another aspect of future applications of action models of this type, are action-aware control mechanisms for robots. We have proposed methods for the acquisition and usage of action models for the improved control of robot soccer and service robots ([12], [13]).

We are looking forward for cooperating and participating in projects together with other research groups.

References

[1] D. Patterson, D. Fox, H. Kautz, M. Philipose. *Fine-Grained Activity Recognition by Aggregating Abstract Object Usage*. Proceedings of ISWC 2005, Osaka, October 2005.

[2] Hyeon-Kyu Lee , Jin H. Kim. *An HMM-Based Threshold Model Approach for Gesture Recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, v.21 n.10, p.961-973, October 1999.

[3] M. Perkowitz, M. Philipose, K. Fishkin, and D. J. Patterson. *Mining Models of Human Activities from the Web*. Proceedings of the Thirteenth International World Wide Web Conference (WWW 2004).

[4] Odest Chadwicke Jenkins and Maja J Mataric´. *Deriving Action and Behavior Primitives from Human Motion Data.* In the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002), pages 2551-2556, Lausanne, Switzerland, 2002.

[5] Amit R, Mataric MJ. *Parametric primitives for motor representation and control*. Int. Conf. on Robotics and Automation (ICRA) Washington DC, May 11-15 2002.

[6] Jun Nakanishi, Jun Morimoto, Gen Endo, Gordon Cheng, Stefan Schaal, Mitsuo Kawato. *Learning from demonstration and adaptation of biped locomotion*. Robotics and Autonomous Systems 47(2-3): 79-91 (2004).

[7] Ling Bao, Stephen S. Intille. *Activity Recognition from User-Annotated Acceleration Data*. Pervasive 2004: 1-17.

[8] Matthai Philipose, Kenneth P. Fishkin, Mike Perkowitz, Donald J. Patterson, Dieter Fox, Henry Kautz, and Dirk Hähnel. *Inferring Activities from Interactions with Objects*. IEEE Pervasive Computing, October 2004.

[9] Ajo Fod, Odest C. Jenkins, and Maja J. Mataric', (2000). *Automated Derivation of Primitives for Movement Classification*. In Proceedings of the IEEE-RAS International

Conference on Humanoid Robotics (Humanoids), MIT, Cambridge, MA, Sep 2000.

[10] Brian Gerkey, Richard T. Vaughan and Andrew Howard. *The Player/Stage Project: Tools for Multi-Robot and Distributed Sensor Systems*. Proceedings of the 11th International Conference on Advanced Robotics, pages 317-323, Coimbra, Portugal, June 2003 (ICAR'03).

[11] Michael Beetz, Bernhard Kirchlechner, and Martin Lames. *Computerized Real-Time Analysis of Football Games*. IEEE Pervasive Computing, 4(3):33–39, 2005.

[12] Freek Stulp and Michael Beetz. *Action Awareness - Enabling Agents to Optimize, Transform, and Coordinate Plans*. In Accepted for the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2006.

[13] Freek Stulp and Michael Beetz. *Optimized Execution of Action Chains Using Learned Performance Models of Abstract Actions*. In Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005), 2005.