Recognition and Understanding Situations and Activities with Description Logics for Safe Human-Robot Cooperation

Jürgen Graf, Stephan Puls and Heinz Wörn
Institute of Process Control and Robotics (IPR)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
{graf, puls, woern}@ira.uka.de

***Abstract*—Recognition of human activities and situation awareness is an important basis for safe human-robot-cooperation. In this paper, a recognition module is presented and discussed. The usage of Description Logics allows for knowledge based representation of activities and situations. Furthermore, reasoning about context dependent actions enables conclusions about expectations for robot behavior. This approach represents a significant step towards a full-fledged cognitive industrial robotic framework.**

Keywords – cognitive robotics, Description Logics, ambient intelligence, situation and action recognition, human-robot cooperation.

I. Introduction

Industrial robotics is a challenging domain for cognitive systems, especially, when human intelligence meets solid machinery like most of today's industrial robots.

Hence, guaranteeing safety for human workers, safety fences are installed to separate humans and robots. As a consequence no real interaction or cooperation sharing time and space can be found in industrial robotics.

Some progress has gained in the past so that some modern working cells are equipped with laser scanners performing foreground detection. But with these systems one is not able to know what is going on in the in the scene and therefore could not contribute something meaningful for challenging tasks like safe human-robot cooperation.

We are conducting research on recognition of and reasoning about actions and situations in a human centered production environment, in order to enable interactive and cooperative scenarios.

This paper focuses on using Description Logics (DLs) [8] as means for representation of knowledge and as reasoning facilities for inference about activities and situations. Furthermore, conclusions about user expectations about robotic behavior can be drawn.

In Section II, some related research work on reasoning about scenes and situations will be presented. In Section III, the framework will be introduced, which enables the sensor data processing and subsequent knowledge based reasoning. In Section IV, DLs will be briefly introduced and the module realizing the communication with a Description Logics reasoner, knowledge base management and reasoner result management will be presented in detail. Also the modeled situations and activities are explained. Section V discusses experimental results which have been carried out for both, predetermined test cases and under real-life conditions. In Section VI, a summary is given. Finally, some hints for future work are also mentioned.

II. Related Works

There are a lot of approaches for action recognition systems based on probabilistic methods, e.g., hidden Markov Models (HMMs) [16, 17, 18], as their theoretic foundation is well understood and applications in speech recognition have shown their capabilities.

Based on arguments, that HMMs are not suitable for recognition of parallel activities, instead propagation networks [19] have been introduced. The propagation network approach associates each node of the network with an action primitive, which incorporates a probabilistic duration model. Also conditional joint probabilities are used to enforce temporal and logic constraints. In analogy to HMMs, many propagation networks are evaluated, in order to approximate the observation probability.

In [20], arguments are put forward, that recognition of prolonged activities is not feasible based on purely probabilistic methods. Thus, an approach is presented which uses parameterized stochastic grammars.

The application of knowledge based methods for action recognition tasks is scarce, but work on scene interpretation using DLs has been conducted.

In [9], DLs are used for reasoning about traffic situations and understanding of intersections. Deductive inference services are used to reduce the intersection hypotheses space and to retrieve useful information for the driver.

In [10], scene interpretation was established using DLs. Table cover scenes are analyzed and interpreted based on temporal and spatial relations of visual aggregate concepts. The interpretation uses visual evidence and contextual information in order to guide the stepwise process. Additionally probabilistic information is integrated within the knowledge based framework in order to generate preferred interpretations. This work is widened to cope with general multimedia data in [11], in which a general interpretation framework based on DLs is presented.

In [12], a comprehensive approach for situation-awareness is introduced, which incorporates context capturing, context abstraction and decision making into a generic framework. This framework manages sensing devices and reasoning components which allows for using

 ISBN: 978-1-61208-108-3

different reasoning facilities. Thus, DLs can be used for high level decision making.

These last examples show that the usage of DLs bears great potential. Hence its adoption in the situation and action recognition task incorporated into the MAROCO framework.

To the best of our knowledge this is the first paper which incorporates description logics in the domain of cognitive robotics. For reasons of this, it was not possible to compare the runtime analysis results to concurrent research groups.

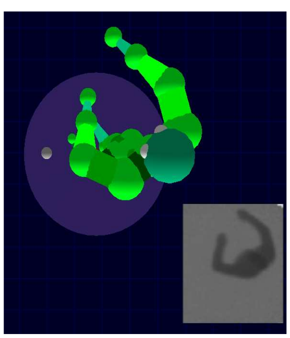
There are some investigations concerning runtime analysis of descriptions logic reasoners (see [21], e.g.) but they are far away from the robotics community and finally they show that the pellet system which was used in this publication is one of the best with respect to the given constraints of the software architecture of MAROCO.

The main motivation writing this paper is introducing the description logics approach into the domain of cognitive robotics. There are just a few other research groups which are dealing with description logics in a similar research domain and the most related ones were referenced in this paper. Most attention was spent on extending the cognitive robotic system MAROCO with description logics and building a knowledge base for action and gesture recognition.

The markerless tracking of a human body in real time is not at the core this paper. But this paper is the first which brings together markerless real time tracking of a human body together with a safe robot path-planning module and the description logic approach. Thus, this paper intends to present interesting results that are gathered from experimental investigations using description logics.

III. The MAROCO Framework

The MAROCO (human robot cooperation) framework [2,3] is an implemented architecture that enables human centered computing realizing a safe human-robot interaction and cooperation due to advanced sensor technologies and fancy algorithms [6,7].



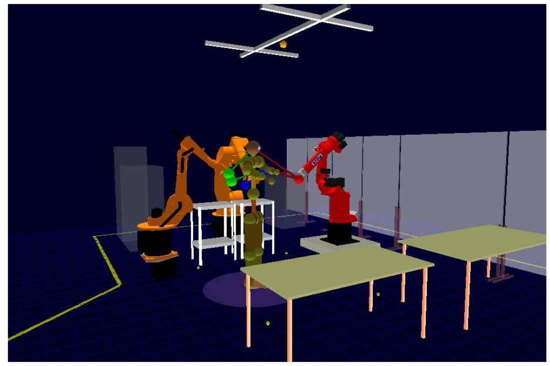


Figure 1. (Up) Reconstructed human model from depth images. (Down) Environmental scene model consisting of several kinematical chains. Three different industrial robots and a human model. All agents and robots have been reconstructed by MAROCO and are integrated into the virtual model in real-time including safety features extraction, risk estimation and path planning.

Every system implementing machine intelligence needs sensors. The MAROCO system analyzes image sequences that are gathered from a 3D vision system [1] based on time-of-flight principle which is mounted to the top of the ceiling of the working cell (see Fig. 1). Modules dedicated to image sequence analysis make it possible to estimate more than a dozen of kinematical parameters, e.g., head orientation, upper body orientation, arm configuration, etc., of a human model without using any markers (Fig. 1). The technical details of the methods realizing the real-time reconstruction of the kinematical model are not in the focus of this paper. Details can be found in [3,6,7].

As safety is one of the most demanding features when industrial robots get in contact with human workers, MAROCO is focused on estimating the risk for the human worker depending on the scene configuration. A variety of methods are integrated into the framework like pure functional evaluation, machine learning tools, e.g., support vector machines, and a two-threaded adaptive fuzzy logic approach, which at the moment makes the race [7].

Having estimated the risk, one is interested in finding a procedure minimizing the risk for both, the worker and machinery. Re-planning is an efficient tool minimizing the risk. A method for re-planning the path of the robot with respect to safety and real-time capability is presented in [4].

The kinematical model also allows for recognition of human activities and situations inside the robot working area. Using Description Logic (DL) reasoning facilities, conclusions about occurring situations, actions, their temporal relations and expectations about robot behavior can be drawn. This is what will be shown in the next sections.

IV. The Recognition Module

This section is dedicated to discuss the recognition module including its components and modeled knowledge base after a very brief introduction to DLs.

A. Description Logics

In this paper, DLs [8] are used to formalize knowledge about situations, actions and expectations. DL is a 2-variable

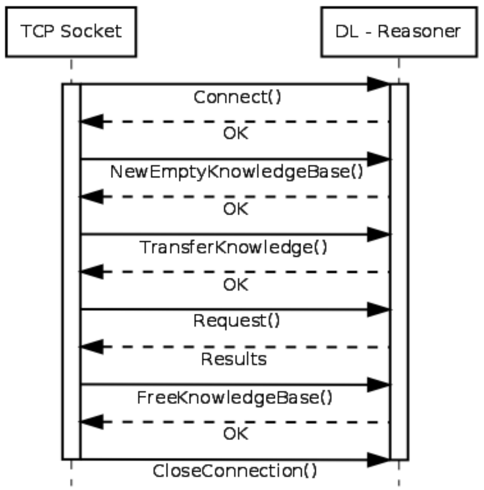


Figure 3. Communication between interface component and DL reasoner.

fragment of First Order Logic and most DLs are decidable. Thus, sound, complete and terminating reasoning algorithms exist.

A DL knowledge base is divided distinctly into general knowledge and knowledge about the individuals in the domain. The former defines the terminology of the domain and its axioms are declared in the terminology box, hence TBox. The latter defines assertions about individuals and, therefore, is declared in the assertion box, hence ABox. This allows for modular and reusable knowledge base and thus for more efficient coding of knowledge [9].

Due to DL's open world assumption, it can deal naturally with incomplete information, which is essential in reasoning about sensor data.

B. *The Module Design*

The recognition module needs to fulfill at least the tasks of establishing a communication interface with the Description Logics reasoner, managing the knowledge base and managing the reasoner results.

In Figure 2, components of the module are presented. The communication via TCP and the XML parsing are done by the components marked as DIG-interface. The DIG-interface is a W3C standard developed by the Description Logic Implementation Group for communication with Description Logics reasoners in the realm of the semantic web and is introduced in [5]. Many reasoners [13,14,15] support this interface definition, which allows the separation of application and reasoner by the means of programming language and execution place.

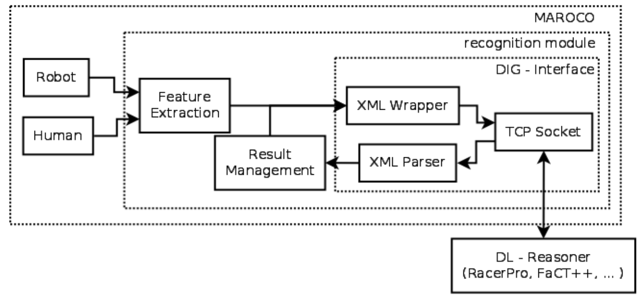


Figure 2. Components of the recognition module.

The DIG-interface follows a functional approach called *Tell&Ask* [8]. After defining a knowledge base – the *tell* operation – reasoner results and information can be retrieved – the *ask* operation. The modification of an existing knowledge base after using an *ask* operation is not defined by the DIG-interface. Therefore in each run time cycle the recognition module creates a complete knowledge base, which will be released in the end (see Fig. 3).

As a consequence the recognition module needs to manage an up-to-date model of the knowledge base, which consists of domain specific knowledge and assertions dependent on the current kinematical human model and robot specific parameters. This distinction corresponds in Description Logics with TBoxes and ABoxes even though the DIG-interface does not distinguish between them. The domain specific knowledge is modeled a priori, the assertional knowledge is created in each runtime cycle afresh. The modeled knowledge base will be explained in more detail in Section IV C.

As the assertional knowledge depends on kinematical parameters a feature extraction component is applied in order to fill the attribute values of the assertions. The following features are important w.r.t. the component *Human*: Angles of both elbows, Angles of both arms to shoulder respective to the up-axis, Angle difference between head orientation and robot, Walking velocity and used tool.

The feature *used tool* is not supported by existing sensors at the moment and is therefore simulated. It can have one of the following values: *none*, *measurement tool* or *working tool*. The simulation of this parameter can be influenced directly by user input using standard human machine interfaces. As a result complex working scenarios can be modeled and analyzed.

The component *Robot* provides the parameters for: gripper status and movement status.

During feature vector creation, extracted values are mapped onto sharp sets. The knowledge base is then populated with corresponding set strings which can be used for comparative operations during reasoning.

One major aspect of understanding human activity is modeling temporal relations between different actions. In this work, these relations are introduced by defining an *after*-role. Hence a certain action can only be recognized if certain other actions occurred prior. This *after*-role can be regarded as defining preconditions onto actions. Previously recognized actions need to be included in the knowledge base in order to allow for correct recognition of current actions. All recognized actions are stored by the reasoner result management component and are retrieved during recreation of the knowledge base.

C. *The Knowledge Base*

In Figure 4, the ontology about situations and activities which are modeled by the knowledge base are presented. The concept *Situation* has the attribute *Number Humans* to distinguish between the concepts *Robot alone* and *Human present*.

In the situations of *Human present*, or its sub-concepts, *Activities* can *take place*, which are *done by* a *Human*. This defines the corresponding concepts and relating roles.

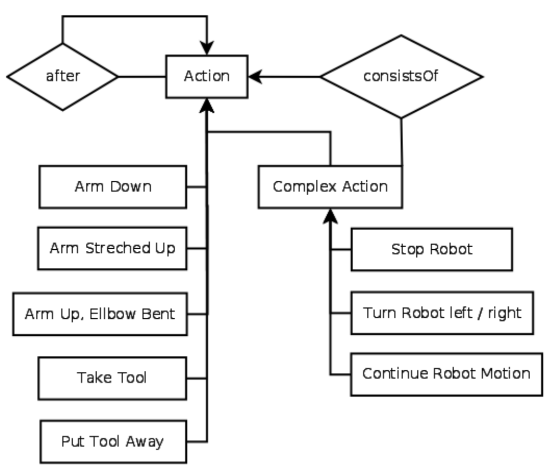


Figure 5. ER model of the action ontology.

In Figure 5, the ontology concerning *Actions* and *complex Actions* is shown. As pointed out above, actions can have a temporal relation expressed as *after*-role. The action *Put Tool Away* can only occur after the action *Take Tool*. This role is also exploited in complex actions, e.g., *Continue Robot Motion* can only be signaled after *Stop Robot*.

Actions can be regarded as atomic concepts, whereas complex actions consist of other actions, regardless of atomicity. The concepts *Take Tool* and *Put Tool Away* are considered atomic, because they are defined by and based on the single attribute *Used Tool*. This attribute is directly altered by user input, thus does not result from sensor data analysis. The role *doneBy* which is defined for activities is also modeled for actions. For reasons of readability this relation is not depicted.

The occurrence of the situation *Cooperation* implies that there are *expectations* towards the robot behavior. Moreover, an expectation can be *triggered by* an action (see Fig. 6). This allows for reasoning about expectations without necessarily recognizing a triggering action. This implicit relation is also exploited between the activities *Monitor*, *Hold Tool* and *Actions*.

V. EXPERIMENTAL RESULTS

For reasons of experimental analysis of the implemented activity and situation recognition different courses of action were executed and the recognition results were recorded.

In order to analyze different scenarios efficiently means of automated feature value presetting were implemented. The overall analysis is based on these presets and on actual sensor data processing. Hence natural movements and transitions between actions can be tested and special use cases can be investigated.

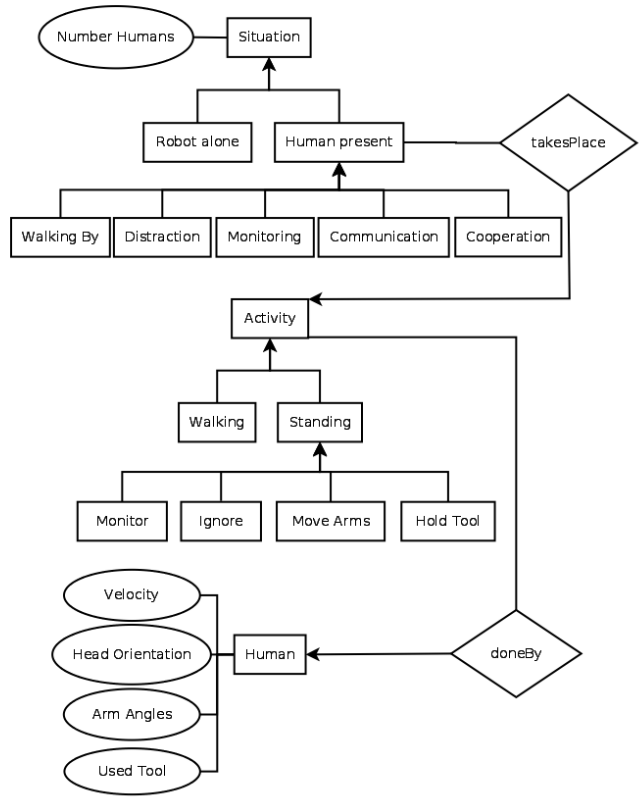


Figure 4. ER model of activity and situation ontology.

In this section, recorded recognition results will be illustrated and discussed.

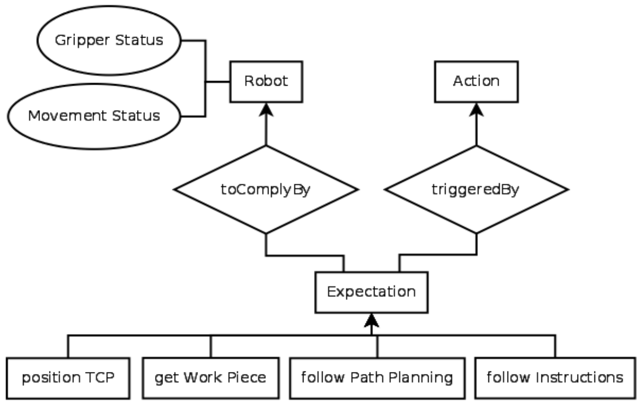


Figure 6. ER model of the expectation ontology.

A. Examplary Result Records

The recorded experimental results contain a timestamp which indicates the starting time of the recognition cycle in milliseconds since program start. This timestamp is then followed by the extracted feature values if there is a human worker in the supervised area. The components of the feature vector are listed in following order: Angle arm left, angle arm right, angle elbow left, angle elbow right, walking velocity, angle difference between head orientation and robot, holding tool, gripper status and robot movement status.

The next number is the timestamp of the final result message from the DL reasoner (Tab. I). Results will be recorded whenever there are new insights. Thus, the last two lines of Table I have no special entries past the last return timestamp.

TABLE I. EXAMPLE RECORD BASED ON SENSOR DATA

29009 29395 RobotAlone
29396 0 0 0 0 1 84 0 0 1 29797 Distraction Ignore
29799 0 0 0 0 1 86 0 0 1 30212
30213 0 15 0 8 1 56 0 0 1 30642

Table II demonstrates the recognition of different situations and activities. Furthermore an additional action and expectation are reasoned and recognized.

During a recognition cycle all recognized concepts are returned from the DL reasoner in a single flush, therefore, the number of lines in the records represents the number of returned responses.

TABLE II. EXAMPLE RECORD BASED ON PRESETS

16160 90 0 0 0 20 0 0 0 1 16965 WalkingBy Walking
. . .
22061 90 0 0 0 20 0 0 0 1 22447
22448 0 0 0 0 0 0 1 0 1 22834 Cooperation HoldTool TakeTool getWorkPiece

B. Results

Tables I and II already indicate that the processing time of a recognition cycle varies around 500 ms. This indication can be shown to hold true by analysis of a large amount of cycles.

TABLE III. RESULTS FROM EVALUATION

# Recognition cycles	2140		Max [ms]	9705
Ø Response time [ms]	551.78		# > 1000 ms	17 (0.79%)
Min [ms]	216		# > 5000 ms	4 (0.18%)

In Table III, the results of 2140 recognition cycles are summarized. It shows that the average processing time is approximately 550 ms. The lower bound is 216 ms. The casual outliers take up to 10 seconds in worst case scenarios. The number of cycles taking more than 1 second reaches 0.79% of all cycles. The amount of processing cycles consuming more than 5 seconds is 0.18%.

TABLE IV. RECORD FOR ANALYSIS OF LONG RUNTIMES

60260 90 0 0 0 0 0 0 0 1 60740 Comm. MoveArms 480
. . .
64501 90 0 0 0 0 0 0 0 1 64940 439
64940 90 0 0 0 0 0 0 0 1 66475 1535
66475 90 0 0 0 0 0 0 0 1 67017 542
67017 90 0 0 0 0 0 0 0 1 72300 5283
72300 90 0 0 0 0 0 0 0 1 72750 450
72750 90 0 0 0 0 60 0 0 1 73221 Distr. Ignore 471

In Table IV, cycle run times are noted at line's end. These numbers show that long cycle times cannot be related directly to changes in the feature vector. Thus, the recognition process itself might not cause the outliers. This will need further investigation.

By using the kinematical human model, recognition of gestures and human motion can be analyzed (see Fig. 7). Table V shows an example in which a human first watches the robot. This concludes the expectation, that the robot shell follow a planned path. After some time the human moves his arms which results in a communicative situation. Because the arms are moved differently by the human, a *Stop Robot* instruction is recognized in the next recognition cycle. The reasoning results in the expectation that the robot shell comply with the instructions.

Consequently natural movements and actions can be recognized despite the average cycle processing time of approx. 550 ms.

TABLE V. EXAMPLE RECORD FOR NATURAL MOVEMENT

103607 0 0 1 0 3 5 0 1 2 104135 Monitoring Monitor followPathPlanning
. . .
112169 0 0 1 0 1 9 0 0 1 112706
112707 62 9 26 21 3 0 0 0 1 113193 Comm. MoveArms
113194 74 70 21 23 2 6 0 0 1 113823 StopRobot followInstructions
113824 76 88 20 35 5 9 0 0 2 114473

Tables II and V demonstrate that depending on situation and actions expectations are generated. The generation of expectation is also dependent on the robot movement status. Table VI shows that at first a cooperative situation is recognized and a generated expectation *get Work Piece*. At this moment the robot was following a planned path, which is signaled as `1` in the feature vector. In the simulation incorporated in MAROCO, this generated expectation leads to a change of the robot movement status which sets the

 ISBN: 978-1-61208-108-3

corresponding feature value to 2, meaning the robot is obeying instructions. This change allows the reasoning to conclude the new expectation to position the robot's tool center point in order to ease the work that the user is about to do with the work piece.

TABLE VI. EXAMPLE FOR DYNAMIC EXPECTATION REASONING

96795 75 0 21 0 0 3 1 0 1 97287 Coop. HoldTool TakeTool getWorkPiece
97289 75 0 22 0 0 0 1 1 2 97799 positionTCP

This process of interaction between reasoner results and robotic behavior demonstrates the dynamic abilities of the presented approach to recognize and understand situations and actions.

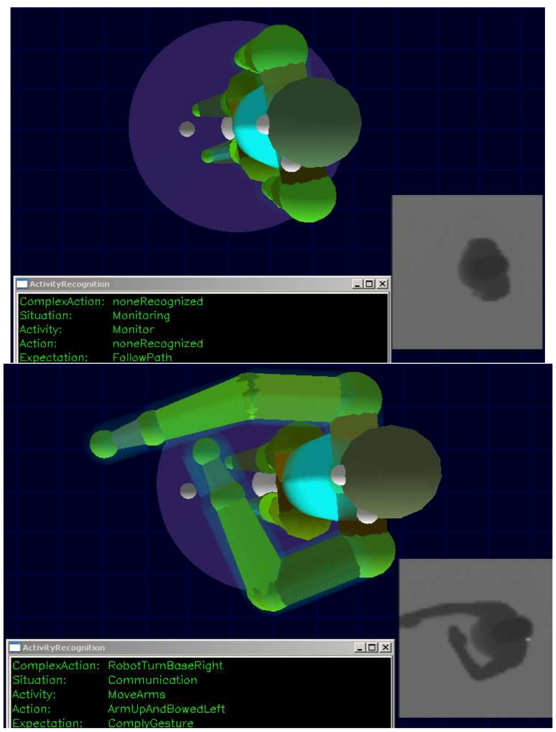


Figure 7. (Top) Human watching the robot. Recognized situation: Monitoring. Recognized activity: Monitor. No specified action recognized. The robot is expected to carry on with its task of following its planned path. (Bottom) Human is communicating with the robot. The complex action to signal a right turning movement is recognized. Recognized situation: Communication. Activity: Move arms. The robot is expected to comply with the users instructions.

C. *Evaluation of Results*

The results demonstrate that the capabilities of the presented approach reach beyond sole activity and situation recognition. By generating expectations towards robot behavior, an understanding of the situation can be achieved. This induction of relations between concepts can hardly be realized by purely probabilistic methods.

The achieved processing cycle time of approx. 550 ms does not allow for safe cooperation based only on the recognition module. Thus, the MAROCO framework uses its implemented techniques and algorithms to enforce safety and real-time capabilities during robot motion. Nevertheless, the measured results will be used to quantify improvements of later developments. To the best of our knowledge, there are no such time related results made available in the field of industrial human-robot cooperation or another related field close to it so far.

VI. SUMMARY AND FUTURE WORK

In this paper, a situation and action recognition module was implemented, which is capable of generating expectations towards robotic behavior.

A knowledge base containing domain and assertional knowledge was modeled. It defines concepts about situations, activities, actions and expectations. These concepts are linked and related by role definitions. Temporal associations of actions are modeled by an *after*-role, which allows preconditioning the recognition of certain actions.

Description Logics are used to define the knowledge base. By implementing the DIG-interface, Description Logics reasoning facilities can be used independently of programming language and execution space.

In order to express value constraints on concept attributes, the feature extraction process maps feature values onto sets, which can be represented as strings in the knowledge base. This allows additionally for support of a wide range of Description Logic reasoners.

During evaluation the effectiveness was shown. Situations, activities and naturally conducted actions are recognized. Expectations are generated and can influence dynamically subsequent processing cycles.

The here presented experimental results are promising for further research in the field of cognitive industrial robotics.

The next steps will be modeling a broader knowledge base in order to incorporate multi-robot setups. Also, the implementation of action plan recognition will deepen the understanding of situations and enable the analysis of complex cooperation scenarios.

It was taken a stand against the probabilistic way of estimating actions from image sequences in the beginning of the related work section. But it is suggested to evaluate different approaches in the near future which also take probabilistic methods into account or maybe fuse different methods bringing together the best of both worlds.

ACKNOWLEDGMENT

This work was supported by PMD Technologies by providing the novel PMD CamCube 3.0 camera.

REFERENCES

[1] http://www.pmdtec.com/products-services/pmdvisionr-cameras/pmdvisionr-camcube-30/ [Last visited on 2010-08-13]

[2] J. Graf and H. Wörn, "An Image Sequence Analysis System with Focus on Human-Robot-Cooperation using PMD-Camera", in VDI Proc. of Robotik 2008, June 2008, pp. 223-226.

[3] J. Graf and H. Wörn, "Safe Human-Robot Interaction using 3D Sensor", in Proc. of VDI Automation 2009, June 2009, pp. 445-456.

 ISBN: 978-1-61208-108-3

[4] J. Graf, S. Puls, and H. Wörn, “Incorporating Novel Path Planning Method into Cognitive Vision System for Safe Human-Robot Interaction”, in Proc. of Computation World, pp. 443-447, 2009.

[5] S. Bechhofer, “The DIG Description Logic Interface: DIG/1.1.”, in Proc. of the 2003 Description Logic Workshop, 2003.

[6] J. Graf, F. Dittrich, and H. Wörn, „High Performance Optical Flow Serves Bayesian Filtering for SafeHuman-Robot Cooperation”, in Proc. of the Joint 41th Int. Symp. on Robotics and 6th German Conf. on Robotics, pp. 325-332, Munich, 2010.

[7] J. Graf, P. Czapiewski, and H. Wörn, “Evaluating Risk Estimation Methods and Path Planning for Safe Human-Robot Cooperation”, in Proc. of the Joint 41th Int. Symp. on Robotics and 6th German Conf. on Robotics, pp. 579-585, Munich, 2010

[8] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, “The Description Logic Handbook”, 2nd Edition, Cambridge University Press, 2007.

[9] B. Hummel, W. Thiemann, and I. Lulcheva, “Description Logic for Vision-Based Intersection Understanding”, in Proc. of Cognitive Systems with Interactive Sensors (COGIS), Stanford University, CA, 2007.

[10] B. Neumann and R. Möller, “On Scene Interpretation with Description Logics”, in Image and Vision Computing, vol. 26, pp. 81-101, 2008.

[11] R. Möller and B. Neumann, “Ontology-Based Reasoning Techniques for Multimedia Interpretation and Retrieval”, in Semantic Multimedia and Ontologies, part 2, pp. 55-98, Springer London, 2008.

[12] T. Springer, P. Wustmann, I. Braun, W. Dargie, and M. Berger, „A Comprehensive Approach for Situation-Awareness Based on Sensing and Reasoning about Context“, in Lecture Notes in Computer Science, vol. 5061, pp. 143-157, Springer, Berlin, 2010.

[13] V. Haarslev, R. Möller, and M. Wessel, “RacerPro User’s Guide and Reference Manual”, Version 1.9.1, May 2007.

[14] D. Tsarkov and I. Horrocks, “FaCT++ Description Logic Reasoner: System Description”, in Lecture Notes in Computer Science (LNCS), vol. 4273, pp. 654-667, 2006.

[15] E. Sirin, B. Parsia, B. Grau, A. Kalyanpur, and Y. Katz, “Pellet: A practical OWL-DL reasoner”, in Web Semantics: Science, Services and Agents on the World Wide Web, vol.5 (2), pp. 51-53, 2007.

[16] V. Krüger, D. Kragic, A. Ude, and C. Geib, „The Meaning of Action: A Review on action recognition and mapping“, in Proc. of Advanced Robotics, Vol. 21, pp. 1473-1501, 2007.

[17] P. Raamana, D. Grest, and V. Krueger „Human Action Recognition in Table-Top Scenarios : An HMM-Based Analysis to Optimize the Performance”, in Lecture Notes in Computer Science (LNCS), Vol. 4673, pp. 101-108, 2007.

[18] Y. Wu, H. Chen, W. Tsai, S. Lee, and J. Yu, „Human action recognition based on layered-HMM”, in IEEE Inter. Conf. on Multimedia and Expo (ICME), pp.1453-1456, 2008.

[19] Y. Shi, Y. Huang, D. Minnen, A. Bobick, and I. Essa, „Propagation Networks for Recognition of Partially Ordered Sequential Action”, in Proc. of Computer Vision and Pattern Recognition (CVPR), Vol. 2, pp. 862-869, 2004.

[20] D. Minnen, I. Essa, and T. Starner, „Expectation Grammars: Leveraging High-Level Expectations for Activity Recognition”, in Proc. of Computer Vision and Pattern Recognition (CVPR), Vol. 2, pp. 626-632, 2003.

[21] T. Gardiner, I. Horrocks, and D. Tsarkov, “Automated Benchmarking of Description Logic Reasoners”, in Proc. of the 2006 Intern. Workshop on Description Logics (DL2006), Windermere, Lake Districrt, UK, 8 pages, June 2006.