Learning in PACO-PLUS

Justus Piater, Mark Steedman, Florentin Wörgötter

April 16, 2009

Abstract

This report gives an overview of the use of machine learning in the PACO-PLUS project. It lists the major learning problems that occur, the machine learning techniques employed to solve them, and describes their roles and interrelations within the PACO-PLUS architecture. Learning occurs throughout the project, mostly for solving individually formulated problems. In addition, the project has produced novel, interrelated learning problems and methods, and has contributed new theoretical insights on existing methods.

1 Introduction

One of the insights that motivated the PACO-PLUS project was the realization that the real world is much too complex and unpredictable for sophisticated interaction to rely exclusively on fixed, pre-programmed behaviors. Therefore, an important principle of PACO-PLUS is to allow a robot to *discover* information about its environment, and to exploit the discoveries in its interaction with the environment. Thus, learning is an overarching and ubiquitous aspect of the project.

Any learning involves the optimization of an objective function $f(\mathbf{x})$ over a parameter vector $\mathbf{x}$. From the viewpoint of an autonomous agent in control of the $\mathbf{x}$, this requires the closure of a feedback loop that enables the agent to observe information useful to optimize $f(\mathbf{x})$. Machine learning paradigms differ in the way this feedback is provided. Two types of feedback mechanisms are explicitly put forward in PACO-PLUS:

- the observation of the consquences of the agent's own actions, useful for exploratory (trial-and-error) learning, e.g. by reinforcement learning,
- the observation of demonstrations of desirable behavior by an external teacher, useful for imitation learning, e.g. by supervised learning.

These two paradigms are complementary and can be fruitfully combined to speed up the more general but potentially slow exploratory learning by using imitation learning to bias exploration towards promising regions of parameter space.

Exploratory learning requires closed perception-action loops. The PACO-PLUS architecture (Fig. 1) permits the closure of perception-action loops through any of its three levels, thus providing ample opportunities for exploratory learning (augmented by imitation learning, if desired) almost everywhere in the system. In general, the degree of adaptivity of such cycles increases with the number of levels involved.

Figure2 shows how the major learning problems that occur in PACO-PLUS are situated within this architecture. Their roles and interplay are described in the following sections.

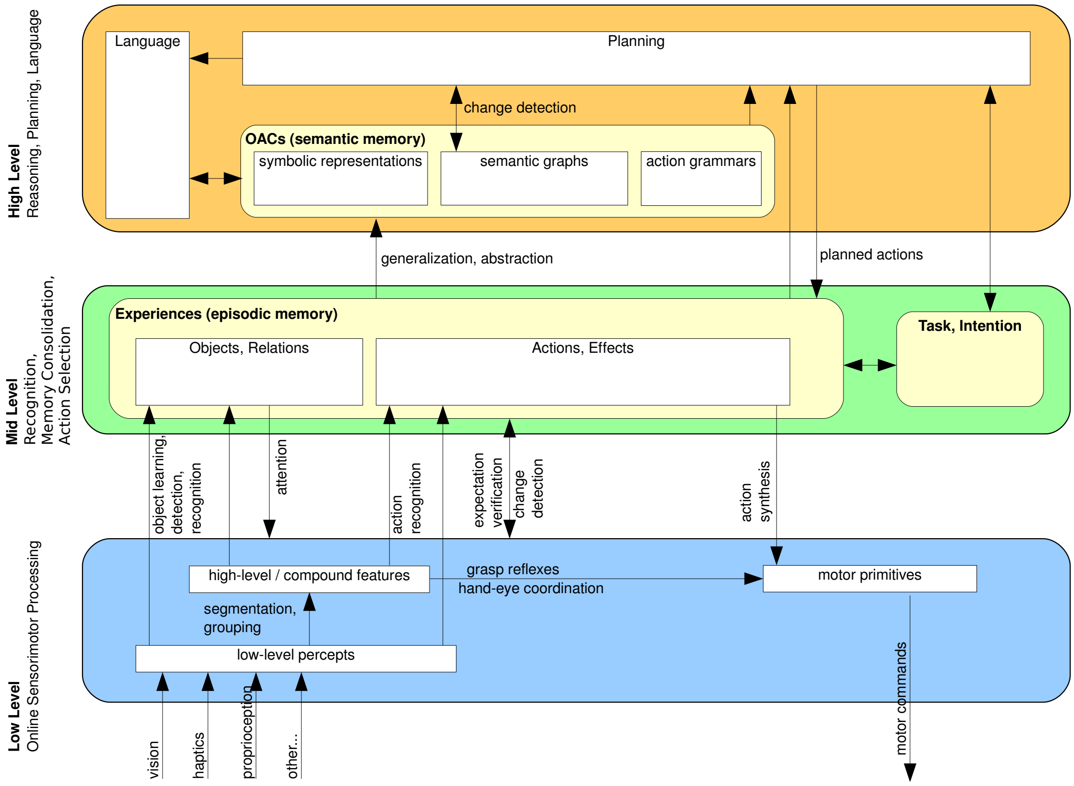

Figure 1: The psychologically-inspired PACO-PLUS cognitive architecture. The agent's functionalities are organized into three levels that are distinguished by their level of abstraction of their view of the world, and that each occupy a dedicated role within the system. All levels are concerned with both perception and action. Processing generally flows clockwise: Raw sensory data are received on the bottom left and are increasingly abstracted on their way up. The high level generates plans based on sensory information, which are turned into concrete motor commands on their way down on the right. Executed motor commands have effects on the environment, which trigger new sensory input, closing a perception-action cycle. Each level can close perception-action cycles without going through levels above.

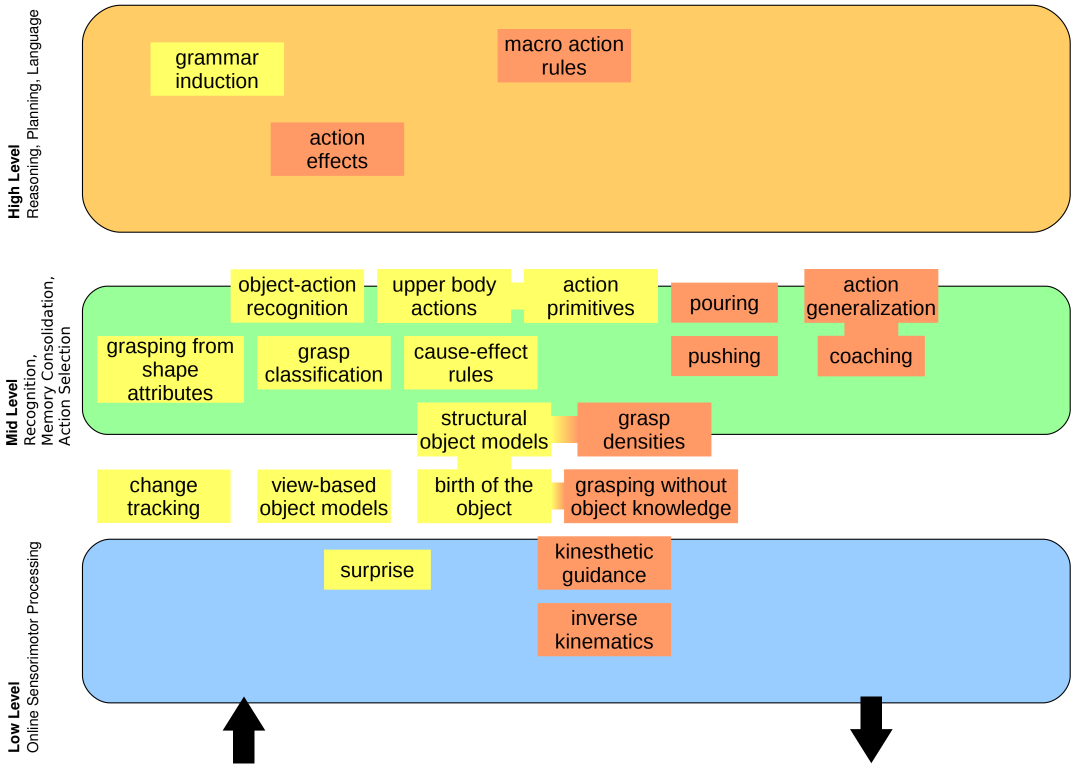

Figure 2: Overview of the key learning problems within the PACO-PLUS architecture, and the major links between them. Learning problems related to objects and recognition are shown in yellow; learning of actions and action parameters is shown in pale red. The labels correspond to the bold-face keywords in the following tables.

Table 1: Learning at the Low Level

Objective	*Methods*	*Pubs.*	*Partn.*	*WP*
Static Recognition				
Rule learning system: mapping raw sensor readings to symbolic conditions; trigger **surprise** on mismatches	Online supervised Bayesian classification assuming normally-distributed sensor data	[5, 4]	CSIC, BCCN	WP6
Action Learning				
Inverse kinematics from as few robot movements as possible	Parametrized self-organizing maps; gradient-based and exact methods using rational Bézier surfaces and prior knowledge	[28, 39]	CSIC, UniKarl	WP8
Learning actions via **kinesthetic guidance**	Locally Weighted Projection Regression, local Gaussian processes; incremental training	[3]	CSIC	WP7

2 Learning at the Low Level

The low level constitutes the interface of the robot's computational resources with the physical world via sensors and effectors. It is responsible for translating raw sensor readings into meaningful percepts, and for transforming motor behaviors into executable motor commands. The low level does not contain a symbolic, long-term memory; its operations and learning remain close to the signal level. The three low-level learning problems are summarized in Table 1.

The **surprise** learning system is trained online to perform Bayesian classification of raw sensor readings into semantic categories, assuming normally-distributed sensor data. Unexpected contingencies yield misclassifications (surprises) and give rise to learning.

Inverse kinematics of redundant manipulators is a long-standing, low-level problem in robotics for which no generally-accepted solution exists. It is often addressed by learning approaches; this is also the case here with the objective of minimizing the number of physical training movements.

Complex low-level manipulations are tedious to pre-program and hard to learn by exploration. Here, this is addressed under the programming-by-demonstration paradigm where actions are learned from **kinesthetic guidance** given by a human trainer.

3 Learning at the Mid Level

Figure 2 reveals that most learning takes place at the mid level and its interfaces to the low and high levels. This concentration arises by construction: The primary objective of the mid level is to provide an infrastructure for learning from experience. Its central component is a long-term, episodic memory where concrete sensorimotor experiences are stored. This allows the construction of abstract concepts via statistical analysis of experiences, which is the basic principle of OAC formation in PACO-PLUS.

The wide variety of learning problems situated at the mid level fall into two broad categories. Problems of the first category (Table 2) are concerned with recognition and form part of the bottom-up pathway in Fig. 1; they are shown as yellow boxes in Fig. 2. Problems of the second category (Table 3) learn action parameters and participate in closing perception-action loops in Fig. 1; these are shown as pale red boxes in Fig. 2.

Most of the learning problems listed in Tables 2 and 3 serve specific, well-delineated purposes, and address them using well-chosen, well-established methods of machine learning. We will not give an in-depth description of each and every learning problem here; we direct the reader to the tables and the references to the publications therein. Rather, we will highlight some of the core components, connections and complementarities.

One core component of the mid level is a block of four interrelated learning systems that address the entire chain from organizing low-level percepts to learning about grasp affordances of objects. It begins by identifying a sparse set of three-dimensional visual features as belonging to a rigid object by grasping features and moving the grasped object around (**birth of the object** and **grasping without object knowledge**). On the basis of these object features, 3D, probabilistic, **structural object models** are learned that are useful for object detection, pose estimation, and learning of grasp affordances (**grasp densities**) via exploration and demonstration.

These highly-structured, computationally-expensive 3D object representations are complemented by a second object representation system that uses active learning to construct **view-based object models** for rapid search, detection and coarse pose estimation.

While grasp densities are designed for learning to grasp *familiar* objects, a complementary method addresses the problem of grasping *unfamiliar* objects via generalization from **shape attributes**.

Other methods address the recognition of grasps, actions and scene changes, and form essential components of scene interpretation and learning from demonstration.

Learning action parameters is a difficult problem. The PACO-PLUS project explores complementary methods such as exploration for **pouring** fluids and **pushing** objects, and **coaching** via demonstration and qualitative instructions given by a human teacher.

4 Learning at the High Level

The high level is concerned with reasoning, planning and language. The representations manipulated at the high level are almost exclusively symbolic. Thus, the wide variety of classical learning problems, which mostly concerns continuous-to-categorical or continuous-to-continuous mappings, is absent here; their domain is the mid level. Learning at the high level produces categorical-to-categorical mappings (Table 4).

One learning objective is the generation of **macro action rules** to coalesce sequences of actions into a single meta-action and thus simplify the planning problem with growing experience.

Another problem concerns the learning the **effect of actions** in the state space considered for planning. This produces empirically-validated action-effect rules that can be used for planning.

The third problem concerns **grammar induction**. Learning grammars and parsing models is quite unlike other learning modules in PACO-PLUS. Most learning problems in the project are standard classifier and associative learning problems, or involve unstructured reinforcement learning. But parsing is not a classification problem. Almost all trees in the 1M word Penn Treebank constitute unique labels.

A parsing model has to provide a measure of how similar each of a large number (routinely, thousands, sometimes millions) of possible parse trees for a novel sentence are to the trees it

Table 2: Learning at the Mid Level: Objects and Recognition

Objective	*Methods*	*Pubs.*	*Partn.*	*WP*
Object Representations				
3D object reconstructions (**birth of the object**)	Bayesian filtering to accumulate 3D information over multiple views	[17, 26]	SDU, KTH	WP4
Probabilistic, **structural object models** for recognition and pose estimation	Unsupervised feature clustering, spatial co-occurrence statistics, Markov networks	[11, 24]	ULg, SDU	WP4
View-based object models; manipulation for figure-ground segmentation and snapshot acquisition	Gaussian processes for background modeling, Bayesian estimation of Gaussian mixtures for object appearance modeling	[37, 40, 21]	JSI, UniKarl	WP2.1
Static Detection, Recognition, Estimation				
Visual **grasp classification** and mapping	k-nearest-neighbor classification and regression	[15]	KTH	WP3.2
Object models; object recognition and scene **change tracking**	Graph representations; Group Method of Data Handling	[30]	BCCN, SDU	WP4.2
Grasping unfamiliar objects by generalization **from shape attributes**	Neural-network regression to connect shape features (from box decomposition) to grasp quality measures (using GraspIt simulator as a trainer)	[14, 12]	KTH	WP4.1
Cause-effect rules: learn preconditions that yield expected postcondition under given action	Online constructive induction; best-first search among candidate preconditions	[6, 7]	CSIC, BCCN	WP6
Dynamic Action Recognition				
Human **upper body actions**	Hidden Markov Models, Expectation-Maximization	[8, 13]	AAU, UniKarl, JSI	WP3
Action primitives	Unsupervised statistical clustering	[29]	AAU	WP3
Visual **object-action recognition**	Conditional Random Fields	[16]	KTH	WP3.2

Table 3: Learning at the Mid Level: Actions

Objective	*Methods*	*Pubs.*	*Partn.*	*WP*
Grasping and Other Actions				
Grasping without object knowledge	Regression using Radial Basis Function networks with autonomous labeling of training data	[25]	SDU, KTH	WP4
Grasp densities: continuous representations of object-relative grasp parameters and their success likelihoods	Biased exploration, importance sampling	[9, 10]	ULg, SDU	WP4.2
Pour fluid; improvement of a prior learned (by demonstration) hand position	Kernel-based reinforcement learning methods; biased exploration by "path straightening"	[34, 33, 35]	BCCN, JSI, UniKarl	WP8.1
Coaching: improving robot motor behavior via marker-based or kinesthetic demonstration and qualitative instructions	Iterative trajectory adjustment via transformation functions in Cartesian and configuration spaces	[27]	JSI, UniKarl	WP2.3
Action generalization by interpolating example movements	Locally weighted regression on splines or dynamic movement primitives	[38, 36]	JSI, UniKarl	WP2.3
Goal-directed **pushing** of objects by learning the relation between contact parameters and object response	Neural network regression	[20]	JSI	WP4.1

Table 4: Learning at the High Level

Objective	*Methods*	*Pubs.*	*Partn.*	*WP*
Rule learning system: **macro action rules**	Backpropagation of preconditions and outcomes along sequences of actions	[6, 5, 4, 7, 2]	CSIC, BCCN	WP6
Action effects in terms of before-after state differences for planning (STRIPS, ADL)	Kernel perceptron classification	[22, 19, 23]	UEdin, SDU	WP4.3, WP5.1
Grammar induction	Bayesian statistical learning, Dirichlet processes	[32, 18, 31]	UEdin	WP5.2

was trained on. As in many other machine learning arenas, there are two main varieties of models:

- Grammar-like Bayesian *generative models*, which enumerate the infinite set of possible analyses, assigning probabilities to candidate analyses;
- Perceptron-like *discriminative models* which sum over weighted features to rank the candidate analyses.

WP5.2 uses generative models, because much of the evidence for the nature of the child's grammar (including deviations from the adult grammar) comes from its productions. Generative models can be reversed to make predictions about about such deviations.

The deliverable D5.2.1 [1] outlining the problem shows how the problem of learning a language from paired strings and meaning representations can be viewed as learning a generative parsing model for the entire space of possibilities for universal grammar using an incremental version of the EM algorithm, and shows how correct predictions follow. However, a sound probability model for this problem is quite a complex object theoretically. The probability distributions of all linguistic phenomena are highly skewed, and require discounting for future unseen events, in particular for unseen words. The model used in the paper submitted since the deliverable employs an infinitely expandable Dirichlet process as a prior, in particular the "Chinese Restaurant Process" formulation to assign priors [18]. This model has been run and evaluated on tyhe CHILDES data, including partial comparisons with related work [44, 42, 43].

5 Theoretical Research

In addition to the application of learning methods to learning problems that arise in the PACO-PLUS system, the PACO-PLUS scenario has also motivated fundamental research on learning methods. This research took place under the umbrella of WP6 and mainly addressed the following three topics:

- The relationship between correlation-based (Hebbian) learning and Reinforcement Learning was analyzed. The two were found to be equivalent under certain conditions.
- Methods were developed for learning forward dynamic models; such models are extremely important in both biological and robotic motor control. In robotics, force and velocity controllers require accurate dynamic models, but such are very difficult to obtain. Research on learning forward models is thus a field with a high potential for impact.
- A novel reinforcement learning method with (receptive field based) function approximation and continuous actions has been developed. This method has been compared to the currently best method from the literature (Natural Actor Critic [41]) and performs equally well. As it is based on the generation of a learned vector field it offers different advantages as compared to the Natural Actor Critic.

This RL method has been implemented on a movement recalibration problem (glass filling) on a HOAP robot and will now be transferred to ARMAR.

In summary, the main goal of WP6 is to arrive at scientifically novel learning methods and at a deeper understanding of learning in biological as well as artificial agents leading to the advancement of the field.

6 Conclusions

This report presents the most important occurrences of learning within PACO-PLUS. It shows that learning occurs in all parts of the system, from the low end close to the hardware to the high end concerned with abstract symbol manipulation, and both for perception and for action.

Most learning problems are situated at the mid level. This is not an accident but by design, as the mid level was designed to provide the infrastructure for learning from sensorimotor experience to allow high-level symbols to be grounded in low-level physical interaction. Nevertheless, both the low and the high level address some of their own problems using learning techniques.

The majority of the learning problems present in the system are structured by their designers to map onto well-understood, classical problem formulations. This permits well-motivated choices of state-of-the-art methods for solving them, their rigorous evaluation and theoretical and empirical comparison to related work. This decomposition into independent problems, an unavoidable principle in the construction of any complex system, has led to a wide variety of different learning methods within the system, where almost no two are the same. Many of the chosen methods are based on probabilistic models and statistics (MRF, HMM, EM, CRF, RL), which is an almost inevitable consequence of the project's reliance on exploratory learning based on collecting empirical data. Naturally, this also reflects the current popularity of probabilistic methods in machine learning, computer vision and artificial intelligence.

On the other hand, some problems are unique to the project and emerge as a consequence of the high-level objective of PACO-PLUS to allow an artificial system to construct its own semantics by autonomous interaction with the world. Such problems have led to new learning problem formulations and sets of interrelated learning systems such as the pathway from the birth of the object to grasp densities.

Finally, beyond formulating and solving learning problems, PACO-PLUS has also shed new insight on existing learning methods such as the relationship between Hebbian correlation learning and reinforcement learning, and has contributed entirely new learning methods as described in Section 5.

Publications on Learning in PACO-PLUS

[1] Deliverable D5.2.1, EU Project IST-FP6-IP-027657 PACO-PLUS.

[2] Learning action rules on the basis of observed cause-effects and teacher instruction. Deliverable D7.2.3, EU Project IST-FP6-IP-027657 PACO-PLUS, 2008.

[3] Deliverable D7.1.4, EU Project IST-FP6-IP-027657 PACO-PLUS, 2009. In preparation.

[4] Alejandro Agostini, E. Celaya, Carme Torras, and Florentin Wörgötter. Action Rule Induction From Cause-Effect Pairs Learned Through Robot-Teacher Interaction. In *International Conference on Cognitive Systems*, pages 213–218, 2008.

[5] Alejandro Agostini, E. Celaya, Carme Torras, and Florentin Wörgötter. Learning Rules from Cause-Effects Explanations. Technical Report IRI-TR 04/2008, Institut de Robòtica i Informàtica Industrial, UPC-CSIC, 2008.

[6] Alejandro Agostini, Florentin Wörgötter, E. Celaya, and Carme Torras. On-line Learning of Macro Planning Operators using Probabilistic Estimations of Cause-Effects. De-

liverable IRI-TR 05/2008, Institut de Robòtica i Informàtica Industrial, UPC-CSIC, 2008.

[7] Alejandro Agostini, Florentin Wörgötter, E. Celaya, and Carme Torras. On-line Learning of Macro Planning Operators using Probabilistic Estimations of Cause-Effects. Deliverable D6.6, EU Project IST-FP6-IP-027657 PACO-PLUS, 2008.

[8] Tamim Asfour, Pedram Azad, F. Gyarfas, and Rüdiger Dillmann. Imitation learning of dual arm manipulation tasks in humanoid robots. *International Journal of Humanoid Robotics*, 5(2):183–202, 2008.

[9] Renaud Detry, Emre Başeski, Mila Popović, Younes Touati, Norbert Krüger, Oliver Kroemer, Jan Peters, and Justus Piater. Autonomous Learning of Object-specific Grasp Affordance Densities. Submitted, 2009.

[10] Renaud Detry, Mila Popović, Emre Başeski, Norbert Krüger, and Justus Piater. Autonomous Learning of Object-specific Grasp Affordance Densities. In *Approaches to Sensorimotor Learning on Humanoid Robots*, 2009. Workshop at the IEEE International Conference on Robotics and Automation.

[11] Renaud Detry, Nicolas Pugeault, and Justus Piater. A Probabilistic Framework for 3D Visual Object Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009. To appear.

[12] S. Geidenstam, Kai Hübner, and Danica Kragić. Using 2D Grasping Strategies for 3D Object Manipulation. Submitted, 2009.

[13] D. Herzog, Aleš Ude, and Volker Krüger. Motion imitation and recognition using parametric hidden Markov models. In *IEEE-RAS International Conference on Humanoid Robots*, 2008. Daejeon, Korea.

[14] Kai Hübner and Danica Kragić. Selection of Robot Pre-Grasps using Box-Based Shape Approximation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008.

[15] Hedvig Kjellström, Javier Romero, and Danica Kragić. Visual Recognition of Grasps for Human-to-Robot Mapping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008.

[16] Hedvig Kjellström, Javier Romero, David Martinez, and Danica Kragić. Simultaneous Visual Recognition of Manipulation Actions and Manipulated Objects. In *European Conference on Computer Vision*, 2008.

[17] Dirk Kraft, Nicolas Pugeault, Emre Başeski, Mila Popović, Danica Kragić, Sinan Kalkan, Florentin Wörgötter, and Norbert Krüger. Birth of the Object: Detection of Objectness and Extraction of Object Shape through Object Action Complexes. *International Journal of Humanoid Robotics*, 5:247–265, 2008.

[18] Tom Kwiatkowski and Mark Steedman. Computational Grammar Acquisition from CHILDES Data Using a Probabilistic Parsing Model. Submitted, 2009.

[19] Kira Mourão, Ronald Petrick, and Mark Steedman. Using Kernel Perceptrons to Learn Action Effects for Planning. In *International Conference on Cognitive Systems*, pages 45–50, 2008.

[20] Damir Omrčen, Aleš Ude, and Andrej Kos. Learning Primitive Actions through Object Exploration. In *IEEE-RAS International Conference on Humanoid Robots*, 2008.

[21] Damir Omrčen, Aleš Ude, Kai Welke, Tamim Asfour, and Rüdiger Dillmann. Sensorimotor processes for learning object representations. In *IEEE-RAS International Conference on Humanoid Robots*, 2007. Pittsburgh, PA, USA.

[22] Ronald Petrick, Christopher Geib, and Mark Steedman. Extensions to PKS for support of low-level continuous control systems. Deliverable D5.1.2, EU Project IST-FP6-IP-027657 PACO-PLUS, 2008.

[23] Ronald Petrick, Dirk Kraft, Kira Mourão, Christopher Geib, Nicolas Pugeault, Norbert Krüger, and Mark Steedman. Representation and Integration: Combining Robot Control, High-Level Planning, and Action Learning. In *International Cognitive Robotics Workshop*, pages 32–41, 2008.

[24] Justus Piater, Fabien Scalzo, and Renaud Detry. Vision as Inference in a Hierarchical Markov Network. In *Twelfth International Conference on Cognitive and Neural Systems*, 2008.

[25] Mila Popović, Dirk Kraft, Leon Bodenhagen, Emre Başeski, Nicolas Pugeault, Danica Kragić, and Norbert Krüger. An Adaptive Strategy for Grasping Unknown Objects Based on Co-planarity and Colour Information. Submitted.

[26] Nicolas Pugeault. *Early Cognitive Vision: Feedback Mechanisms for the Disambiguation of Early Visual Representation*. Vdm Verlag Dr. Müller, 2008.

[27] Marcia Riley, Aleš Ude, Christopher Atkeson, and Gordon Cheng. Coaching: An Approach to Efficiently and Intuitively Create Humanoid Robot Behaviors. In *International Conference on Humanoid Robots*, 2006.

[28] V. Ruiz de Angulo and Carme Torras. Learning Inverse Kinematics: Reduced Sampling Through Decomposition Into Virtual Robots. *IEEE Transactions on Systems, Man and Cybernetics – Part B*, 38:1571–1577, 2008.

[29] Sanmohan, Volker Krüger, Danica Kragić, and Hedvig Kjellström. Primitive based action representation and recognition. Submitted.

[30] Nataliya Shylo, Florentin Wörgötter, and Babette Dellen. Ascertaining relevant changes in visual data by interfacing AI reasoning and low-level visual information via temporally stable image segments. In *International Conference on Cognitive Systems*, 2008.

[31] Mark Steedman, Tom Kwiatkowski, and Julia Hockenmaier. The Statistical Problem of Language Acquisition. In preparation, 2008.

[32] Mark Steedman, Ronald Petrick, and Christopher Geib. Grammar Induction. Deliverable D5.2.2, EU Project IST-FP6-IP-027657 PACO-PLUS, 2008.

[33] Minija Tamosiunaite, James Ainge, Tomas Kulvicius, Bernd Porr, Paul Dudchenko, and Florentin Wörgötter. Path-finding in real and simulated rats: On the usefulness of forgetting and frustration for navigation learning. *Journal of Computational Neuroscience*, 2009. In press.

[34] Minija Tamosiunaite, Tamim Asfour, and Florentin Wörgötter. Learning to reach by reinforcement learning using a receptive field based function approximation approach with continuous actions. *Biological Cybernetics*, 2009. In press.

[35] Minija Tamosiunaite, Florentin Wörgötter, and Tamim Asfour. Implementation of nSARSA with function approximation in its two stages as given by tasks 6.8.1 and 6.8.2 first on a simple platform and then on ARMAR. Deliverable D6.5, EU Project IST-FP6-IP-027657 PACO-PLUS, 2008.

[36] Aleš Ude, Andrej Gams, and Tamim Asfour. Generalization of Example Movements with Dynamic Systems. In preparation, 2009.

[37] Aleš Ude, Damir Omrčen, and Gordon Cheng. Making Object Learning and Recognition an Active Process. *International Journal of Humanoid Robotics*, 5(2):267–286, 2008.

[38] Aleš Ude, Marcia Riley, Bojan Nemec, Andrej Kos, Tamim Asfour, and Gordon Cheng. Synthesizing Goal-Directed Actions from a Library of Example Movements. In *International Conference on Humanoid Robots*, 2007.

[39] S. Ulbrich, V. Ruiz de Angulo, Tamim Asfour, Carme Torras, and Rüdiger Dillmann. Learning robot kinematics using rational Bézier forms. In preparation.

[40] Kai Welke, Tamim Asfour, and Rüdiger Dillmann. Object separation using active methods and multi-view representations. In *IEEE International Conference on Robotics and Automation*, pages 949–955, 2008. Pasadena, CA, USA.

References

[41] Jan Peters and Stefan Schaal. Reinforcement learning of motor skills with policy gradients. *Neural Networks*, 21(4):682–697, 2008.

[42] Yuk Wah Wong and Raymond Mooney. Learning Synchronous Grammars for Semantic Parsing with Lambda Calculus. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 960–967. Association for Computational Linguistics, 2007.

[43] Charles Yang. *Knowledge and Learning in Natural Language.* Oxford University Press, 2002.

[44] Luke Zettlemoyer and Michael Collins. Learning to Map Sentences to Logical Form: Structured Classification with Probabilistic Categorial Grammars. In *Proceedings of the 21st Conference on Uncertainty in AI*, pages 658–666, 2005.