Project no.:	027657
Project full title:	Perception, Action & Cognition through Learning of Object-Action Complexes
Project Acronym:	PACO-PLUS
Deliverable no.:	D8.3

Title of the deliverable: **Demonstration of the PACO-PLUS final scenario and the cognitive control architecture as defined in WP1 on ARMAR and/or parts of it at the SDU and/or JSI platform.**

Contractual Date of Delivery to the CEC:	31. May 2010
Actual Date of Delivery to the CEC	06. July 2010
Organisation name of lead contractor for this deliverable:	KIT/SDU
Authors:	All
Participants:	All
Work package contributing to the deliverable	WP8, WP1
Nature:	R & D
Version	1.0
Total number of pages:	8
Start date of project:	1st Feb. 2006
Duration:	52 months

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract:

The deliverable consists of demonstrations of several components developed in the project, for which we provide the videos. The different components have been integrated into the different robot platforms in the project at KIT, SDU and JSI. The needed components for the final demonstration have been integrated and tested on ARMAR at KIT.

Keyword list: Demonstration of the PACO-PLUS final scenario

Table of Contents

0. Executive Summary

This deliverable shows the demonstrations realized on the PACO-PLUS robotic systems in WP8 toward the implementation of the final project scenario. The deliverable consists of components that are by the time of writing still in the phase of final integration into a complete humanoid robot operation in human environments. Final demonstration will be shown at the review meeting both as video and live.

The deliverable consists of the following demonstrations

- Grasping and grasp learning in a cognitive architecture using the Early Cognitive Vision system (CoVis), which processes different kinds of visual descriptors and their relations. The scene interpretation is done based on the object memory that has been built up in an exploratory process (Birth of the object). Grasping experience is then represented by experiments generated by means of an object specific grasping OAC which are then stored in the episodic memory to be used at a later stage for the establishment or refinement of grasp densities which form the object specific grasp memory.
- Box-based generation of grasping hypotheses for known objects in the final kitchen scenario. Given an object detection and pose estimation skill, grasp generation methods can be applied to known objects in order to assign a list of grasp hypotheses to each model.
- Methodology that enables the generalization of the available sensorimotor knowledge to new situations. New actions are synthesized by applying statistical methods, where the goal and/or other characteristics of an action are utilized as queries to create an optimal control policy with respect to the current state of the world. The developed methods combine prior sensorimotor knowledge about objects and movements to synthesize the optimal action in any situation, thus providing a methodology for learning early object-action complexes by imitation and coaching.
- Symbolic planning for the kitchen scenario: This work demonstrates the integration of the PKS planning system (and associated modules) on the ARMAR robot platform, through a series of object retrieval and manipulation tasks in the KIT kitchen environment. High-level plan generation, execution, and re-planning in the event of failure are demonstrated.
- Markerless, active observation of human actions. In this demonstration, the developed markerless stereo-based upper body motion capture system is applied for the active observation of actions. For this, the system had to be optimized to run at full frame rate, so that active head movements are smooth and do not deteriorate the motion tracking.
- Grasp recognition system in which the robot imitates the grasp performed by the human in front of him. A human performs a grasp without any markers on their arms or hands. The robot observes the arm movement and the grasp type, maps them to his embodiment and performs the grasp on a similar object situated in front of him.
- Active visual search on a humanoid active head, where multi-view object representations are acquired by rotating the object held in the robot’s hand in the field of view of the head. For object search, gaze sequences are generated in order to identify one or multiple instances of a searched object in the scene.
- Haptic exploration of unknown objects on a humanoid robot with a five-fingered hand. Exploration strategies based on dynamic potential fields are presented to both acquire models of unknown objects as well as to implement exploration-driven grasping strategies.

- Integration of the rule learning system (RSYS) and the PKS planner on ARMAR for the on-line learning of PKS-rules
- Action selection for active vision. Implementation of a system that builds models of objects through active perception and manipulation.

1. Grasp learning in a cognitive architecture (SDU, ULg)

The video "**Grasp-Learning.m2v**" shows grasping and grasp learning in a cognitive architecture. The Early Cognitive Vision system first processes different kinds of visual descriptors and their relations. This is done at a high frame rate by the use of a hybrid computer architecture consisting of multi-core PCs and GPUs. The actual scene is interpreted based on the existing object memory (pose estimation) that has been built up in an exploratory process (birth of the object). Grasping experience is then represented by experiments generated by means of an object specific grasping OAC, which are then stored in the episodic memory to be used at a later stage for the establishment or refinement of grasp densities which form the object specific grasp memory.

2. Box-based grasp generation (KTH)

The video "**Box-based-Grasping-of-Unknown-Objects.mp4**" shows the evaluation of the capabilities of the framework for grasping based on box decomposition to deal with unknown objects.

3. Learning early object-action complexes by imitation and coaching (JSI, KIT)

Acquiring new sensorimotor knowledge by imitation and coaching is a promising paradigm for robot learning. To be effective, action learning should not be limited to direct replication of movements obtained during training, but must enable also the generation of actions in situations that a robot has never encountered before. These videos demonstrate a methodology that integrates the available motor knowledge with perceived object information, thus enabling the generalization of the available sensorimotor knowledge. New actions are synthesized by applying statistical methods, where the goal and other characteristics of an action related to objects that the robot acts upon are utilized as queries to create an optimal control policy with respect to the current state of the world. Nonlinear dynamic systems are employed as an underlying motor representation. The proposed approach enables the generation of a wide range of policies without requiring an expert to modify the underlying representations to account for different task-specific features and perceptual feedback.

The generalization of reaching actions that takes the current object position acquired by active vision into account is shown in the videos **discrete-reach-open.mov**, **discrete-reach-closed.mov**, and **hoap-grasp-bear.mov**. The same methodology has also been applied to the generation of throwing actions (**discrete-throwing.mov** and **discrete-throwing-sensing.mov**). We demonstrate the coaching and generalization of periodic movements in the videos **periodic-examples.mov**, **periodic-generalized1.mov**, **periodic-generalized2.mov**. Finally, we show the adaptation to the external force signal in the video **periodic-force-learning.mov**

4. Symbolic Planning for the kitchen scenario (UEDIN, KIT)

In this work we demonstrate the integration of the PKS planning system (and associated modules) on the ARMAR robot platform, by considering a set of object retrieval and manipulation tasks in the KIT kitchen environment. The planner is responsible for constructing sequences of actions that direct the robot in achieving particular user-defined goals. These sequences include actions for moving between different work areas in the kitchen, hand selection for grasping, and destination-dependant placement operations, involving objects that might be toppled or occluded. Plans are communicated from the planner to the robot using a new ICE-based communication architecture, developed at KIT as part of WP1. Using this framework, the robot requests actions and reports the status of action execution back to the planner, while the planner services robot requests and relays high-level actions and sensing requests to the robot. All high-level planning activities are directed by a plan execution monitor, which acts as a wrapper around the planner and message communication framework, developed as part of WP4 and WP5. In particular, the execution monitor inspects discrepancies between states of the world predicated by the planner, and those actually observed by the robot. In the event of action or plan failures, the execution monitor is responsible for controlling all re-planning and high-level re-sensing activities in the system.

This demonstration highlights the latest versions of the robot action planner, plan execution monitor, and ICE-based communication architecture integrated on the ARMAR robot platform, and is reported in the deliverables D5.1.3 and D1.2.3.

5. Markerless active observation of human actions (KIT, JSI)

In this work, we deal with the problem of active observation of human actions, in particular with markerless upper body motion capture using an active robot head for the purpose of imitation learning.

Experiments from the previous reporting period showed that for a meaningful active observation of actions, i.e. active arm and object motion tracking, the up to then achieved processing rate of 15-20 Hz was too low, deteriorating the accuracy of the acquired trajectories significantly. The reason is that active tracking results in jittery head movements when operating on the basis of low frame rates, and jitter can lead to severe motion blur in the images, which in particular makes the computation of the edge-based cue for arm motion tracking impossible.

As active tracking is naturally a problem that cannot be solved by offline post-processing (the robot head motion for active tracking must be computed online), the upper body motion tracking system was optimized to run at maximum frame rate. For this purpose, two modules were optimized: image pre-processing and the edge cue as part of the evaluation function of the particle filter for arm tracking.

As part of image pre-processing, image undistortion, low-pass and edge filtering as well as color segmentation were replaced by calling the highly optimized implementations from the Keyetech Performance Primitives (KPP, www.keyetech.de), which achieve speedups of up to factor 10, depending on the type of image processing function. Furthermore, memory consumption was reduced with regard to cache usage. The edge cue was optimized by applying integer arithmetics only for the implementation of the Bresenham algorithm for line traversal. With these optimizations, the system achieves processing rates of 40-50 Hz, running on conventional hardware (Core 2 Duo 3.0 GHz).

The video "**ActiveObservationARMAR.avi**" shows the active observation of an action, involving upper body motion and object manipulation.

6. Grasp recognition (KTH, KIT)

This demo (see videos **"GraspRecognition_Part-1.avi"** and **"GraspRecognition_Part-2.avi"**) exemplifies a situation in which a person wants to teach a robot how to grasp a particular object. For this purpose the robot observes how the human performs the action, paying special attention to how the arm is moved, which grasp type is performed and which object is grasped.

The arm movement is observed with the wide-angle stereo pair in ARMAR, and processed with an upper body tracker (Azad et al. 2008). The grasp type is observed with one foveal camera in ARMAR, compared with a database of hand poses and classified as a particular grasp type (Romero et al. 2010).

For this demo, we restricted the grasp types to be just one of the following three: power grasp from top, power grasp from the side, and pinch grasp. Finally, the object recognition is based on comparison with different views generated from an object database (Azad et al. 2009). All this information is used to generate an approach movement with the Master Motor Map interface, and apply a grasp (defined as the correspondent grasp to the one performed by the human) to the recognized object (Do et al. 2009). The system runs in real time, and the human does not need to wear any markers or special devices.

7. Active visual search (KIT, JSI)

In this work we consider the whole cycle of active acquisition of appearance-based object representations to the visual search of the acquired objects using the ARMAR-III active head (Welke et al. 2009, 2010). For this purpose, objects were put in the five-fingered hand of the robot and multi-view representations are acquired by rotating the object in the field of view of the head. The developed system allows for (1) identification and ascertaining of object candidates using active peripheral and foveal vision, (2) generation of gaze sequences in order to identify one or multiple instances of a searched object in the scene and (3) memorization of the processing results from foveal and peripheral processing in an ego-centric scene memory (see "**full_acquisition.avi**" and "**VisualSearch.avi**").

8. Haptic Exploration (KIT)

We present the results of our work on haptic exploration of unknown objects with a five-fingered humanoid hand. Exploration strategies based on dynamic potential fields, which have been developed in the project (see Bierbaum et. al 2009, 2009a), have been transferred to the five-fingered hand and successfully evaluated using several daily objects such as a transparent bottle, door handles, etc. The developed methods have been integrated into a software framework for haptic exploration.

The videos "**Haptic_Exploration_Doorhandle.wmv**" and "**Haptic_Exploration_Bottle.wmv**" show simulation results of the haptic exploration of a bottle and of a door handle with ARMAR-III. During exploration, the robot's hand-arm system is guided by a potential field based exploration control approach (see WP2.2). An attractive potential field (red dots) is initiated from a bounding box estimation of the object. The potential field is updated dependent on the progress of the exploration. Once a contact between the object and a tactile sensor is detected, a repelling potential is added to the potential field and the object model is extended by the new contact point. The simulation uses a detailed geometric model of the robot's tactile sensor system (light red regions at fingers and palm). Only contacts at actual sensor locations are registered to provide a realistic estimation of the contact information that can be acquired by this sensor system. For introducing known obstacles to the robot, an additional repelling potential field (orange dots) was used to limit collisions with the table or the door surfaces respectively. See also WP1 and WP2.2.

9. Integration of the RSYS rule learning system and the PKS planner on ARMAR for the on-line learning of PKS-rules (CSIC, UEDIN, KIT)

In the demo (see "**RSYS-on-ARMAR.wmv**") we present two highlights of the RSYS learning mechanisms integrated with the PKS planner on ARMAR. The first one presents a detailed description of the mechanism for refining an incomplete PKS-rule that produces an inconsistency (surprise) in the outcome when it is executed. The second demonstration is referred to as the macro rule generation method that combines multiple one-step rules in a macro rule with various actions, so as to reduce the deliberation efforts of the PKS planner. To make the exposition clear for the demo we use the simple task of placing cups on a sideboard using a grid description of the world. Each cup is assumed to be placed inside one cell, and ARMAR is restricted to a set of actions providing horizontal and vertical movements using grasping.

Demo: rule refinement

The situation consists of two cups placed nearby, one blue and one green. The goal is to place the green cup in the cell on the other side of the blue cup. The robot uses its only rule for plan generation, an incomplete rule for a straight horizontal movement to the goal cell, which does not specify the necessity of empty cells in the trajectory. Since the blue cup is blocking the trajectory, the action cannot be executed and a surprise arises, triggering the rule refinement, and solving the situation.

Demo: Macro rule

The situation selected for the demo involves four cups. The goal is the same as before: place the green cup in the cell on the other side of the blue cup. The red cups are blocking all the possible movements available in the limited set of rules learned so far, which prevents the planner from finding a plan. The teacher provides an instruction to move the blue cup twice without colliding with the other cups, so as to unblock the path. These instructions lead to the generation of a macro rule that permits PKS to generate a plan to reach the goal from the initial complicated situation.

10. Action selection for active vision (CSIC)

The system was tested on a moderate set of motion sequences using both SIFT and CoVis features. The strategy actively chooses the most appropriate viewpoint changes when building object models using visual primitives. A novel parameterization to model uncertainties was explored in which Gaussian probabilities are encoded using a canonical representation. This representation simplifies considerably state augmentation and inference, and as a side effect it also makes it much easier to compute information gain for action selection. This novel representation was pursued further to model the geometric relations between features and camera/sensor viewpoints (Foix et al. ICRA 2010). See video "**Action-Selection.wmv**".

References (all arising from the project)

Azad, P., Asfour, T., and Dillmann, R. (2008). Robust Real-time Stereo-based Markerless Human Motion Capture, Proc. IEEE/RAS International Conference on Humanoid Robots, Daejeon, Korea.

Azad, P., Asfour, T., and Dillmann, R. (2009). Accurate Shape-based 6-DoF Pose Estimation of Single-colored Objects, Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2690-2695, St. Louis, USA.

Bierbaum, A., Rambow, M., Asfour, T., Dillmann, R. (2009) Grasp Affordances from Multi-Fingered Tactile Exploration using Dynamic Potential Fields. In IEEE/RAS International Conference on Humanoid Robots (Humanoids), Paris, France.

Bierbaum, A., Asfour, T., Dillmann, R. (2009a) Dynamic Potential Fields for Dexterous Tactile Exploration. In Workshop on Human-Centered Robotics Systems (HCRS), November, Bielefeld, Germany.

Do, M., Romero, J., Kjellström, H., Azad, P., Asfour, T., D. Kragic and Dillmann, R. (2009). Grasp Recognition and Mapping on Humanoid Robots, In IEEE/RAS International Conference on Humanoid Robots Paris, France.

Foix, S., Alenyà, G., Andrade-Cetto, J. and Torras, C. (2010). Object modeling using a ToF camera under an uncertainty reduction approach. In IEEE International Conference on Robotics and Automation (ICRA), Anchorage, USA.

Gams, A. and Ude, A. (2009) Generalization of example movements with dynamic systems. In Proc. IEEE-RAS Int. Conf. on Humanoid Robots, Paris, France, pages. 28-33.

Petrick, R., Adermann, N., Asfour, T., Steedman, M., Dillmann, R. (2010). Connecting knowledge-level planning and task execution on a humanoid robot using Object-Action Complexes. In 4th International Conference on Cognitive Systems, Zurich, Switzerland.

Romero, J., Kjellström, H., and Kragic, D. (2010) Hands in action: Real-time 3D reconstruction of hands in interaction with objects. In IEEE International Conference on Robotics and Automation.

Ude, A., Gams, A., Asfour, T., and Morimoto, J. (2010) Task-specific generalization of discrete and periodic dynamic movement primitives. IEEE Trans. on Robotics (conditionally accepted).

Welke, K., Asfour, T., and Dillmann, R. (2009) Active multi-view object search on a humanoid head. In IEEE International Conference on Robotics and Automation, pages 417–423.

Welke, K., Issac, J., Schiebener, D., Asfour, T., and Dillmann, R. (2010) Autonomous Acquisition of Visual Multi-View Object Representations for Object Recognition on a Humanoid Robot. In IEEE International Conference on Robotics and Automation (ICRA), Anchorage, USA.