Project no.: IST-FP6-IP-027657

Project full title: Perception, Action & Cognition through Learning of Object-Action Complexes

Project Acronym: PACO-PLUS

Deliverable no.: D3.1.2

Title of the deliverable: Test Data "Action Representation I"

Contractual Date of Delivery to the CEC:	**31st January 2007**	
Actual Date of Delivery to the CEC:	**31st January 2007**	
Organisation name of lead contractor for this deliverable:	**Aalborg University (AAU)**	
Author(s): V. Krüger, D. Grest, D. Kragic, J.-O. Eklundh, T. Asfour, P. Azad		
Participants(s): AAU, KTH, UniKarl, BCCN, JSI		
Work package contributing to the deliverable:	**WP3**	
Nature:	**R**	
Version:	**1.0**	
Total number of pages:	**9**	
Start date of project:	1st Feb. 2006	**Duration:** 48 month

Projectco-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:

This report describes the dataset "Action Representation I". This dataset contains mainly three datasets, one dataset recorded at AAU, one recorded at KTH and a third one recorded at UniKarl. The first two datasets are focussed on movements of one arm and hand with plain movements and movements manipulating objects, the third dataset contains full-body movements.

The AAU dataset contains different actions performed by 7 different volunteers. The performances were recorded with 4 synchronized color video cameras. In order to acquire ground-truth data, the individuals were equipped with electro-magnetic sensors at their joints. In total, 7 sensors were used. A total of 500 GB of data was recorded and prepared for use. The sensor data was used for evaluations of HMM-based action recognition approaches. The full set will be used to evaluate our vision-based motion capture systems.

The second dataset recorded at KTH is sensor based only. It contains recordings of 20 different individuals and a large number of different manipulation tasks. The UniKarl dataset contains 400 full-body movement sequences of 20 different individuals. The UniKarl data is represented in the Master Motor Map format (MMM).

Keyword list: marker data, video data, motion capture

Contents

1 Introduction

This report describes the dataset "Action Representation I". This overall dataset contains three datasets: one dataset was recorded at AAU in November 2006, the other one was recorded at KTH in October 2006 and a third dataset recorded at UniKarl. All datasets are the results of experiences from previously recorded dataset. The AAU and the KTH datasets focus on movements of one arm and hand with and without involving objects. The UniKarl dataset focusses on full-body movements. The aim of the dataset recorded at KTH is to allow testing and verification of the action representation and recognition approaches as developed in WP 3.1 and WP3.2. The aim of the dataset recorded at AAU is to allow testing of vision-based motion capture systems for action recognition. While the KTH dataset has a large variety of different actions and activities, the AAU dataset contains a subset of these actions and activities while providing also visual data useful for visual tracking from 4 synchronized digital video cameras. We have started our research by using electro-magnetic sensor data for training and testing. Excluding the vision problem allows to focus on the problems of action representation and recognition. It is, however, the aim during the project period to replace the sensor data with the data coming from a 3D visual tracker, present developements within PACO-Plus are [3; 6]. We intend to use the AAU data to test our 3D trackers [3; 6]. The sensor data will be used as ground-truth information for tracking verification.

We will discuss the AAU dataset in detail in Sec. 2 The KTH dataset shows a large variety of different actions involving a large number of different objects. This dataset will be discussed in detail in Sec. 3. Within the consortium not only one human motion capture system and one action recognition system exist, but several approaches are available, each of those feasible for another case, having its own strengths and drawbacks. Moreover, although the main demonstrator is ARMAR III at the University of Karlsruhe, other partners have to be able to use data and run experiments on their own demonstrator. To overcome the data compatibility problems arising form this circumstance, UniKarl has specified the so-called Master Motor Map (MMM) [2], including a set of 400 actions performed by 20 different individuals recorded as point-light displays using the VICON optical motion capture system. The UniKarl dataset will be discussed in Sec. 4.

The datasets are available at `http://cvmi.aau.dk/~pradeep/motion/`[1]. The webpage contains

1. Access to the sensor and video data of the AAU dataset,
2. Access to the sensor data of the KTH dataset,
3. Access to the sensor data of the UniKarl dataset,
4. A set of Matlab routines for accessing the AAU dataset,
5. full technical description of the datasets.

2 Dataset at AAU

The dataset recorded at AAU contains different actions performed by 10 different volunteers. The aim of the AAU dataset is to provide a basis for vision-based tracking and action recognition approaches [3; 6]. The focus is to provide visual data with a small set of actions and activities.

We have recorded 7 different subjects. The persons involved in the study were not trained in any special way - each action started by having an arm in a relaxed, vertical position.

The subjects were recorded with

[1] Login: `pacoplus`, Password: `pacoaccess`.

Figure 1: The left figure shows the control unit of the Motion Start system, the right figure shows the Extended - Range Transmitter.

1. 4 synchronized digital color cameras.

2. 7 electro-magnetic sensor devices that provide ground-truth data.

The cameras have a resolution of 640×480 pixels and the recording was done with 20 Hz. The data is compressed with a lossless compression to prevent loss of image quality. The cameras are Bayer cameras. The available image data is raw data from the CCD chip, thus no white-balancing was done. Preprocessing was done by us as follows:

1. The raw Bayer data was converted to RGB 24 bit image data.

2. Images were rectified using the internal and external camera calibration data.

For recording the ground-truth data we used the **Motion Star** of **Ascension** [1], see Fig. 1.

The system captures a $6D$ data vector for each sensor containing its $3D$ position and $3D$ orientation. Capturing speed was set to 25 Hz. The video camera system and the Motion Star system were not synchronized.

As we are mainly interested in one-arm actions, the 7 sensors from the Motion Star system were connected to the hand (4 sensors), to the arm (2 sensors) and to the torso (1 sensor), see Fig. 2 for a precise decription of the sensor locations. In Fig. 3, one can see a snapshot of the entire setup. Visible are the four blue video cameras on the left, the cables for the electro-magnetic sensors on the bottom-right and the transmitter of the motion star system on the bottom left.

We have four different categories of actions:

1. Pointing

2. Grasping: reaching out and grasping an object. Sequences end with a touching of the object, no lifting or other motion

3. Displacement: grasp and displace object

4. Rotating: grasp and rotate the object

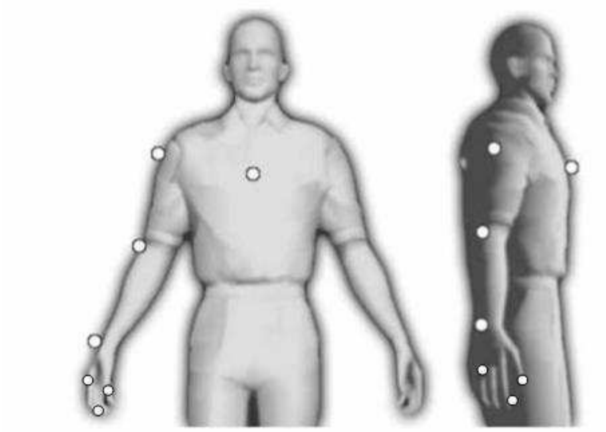

Figure 2: The dots mark the position of the electro-magnetic sensors for the AAU dataset.

Each of these movements were carried out at three different scales (reaching distances) and three different directions and with 5 repetitions.

A total of 0.5 TB of data was recorded and prepared for use. In addition, we have provided matlab code to simplify access of the data. The sensor data was used for evaluations of HMM-based action recognition approaches. The full set will be used to evaluate our vision-based motion capture systems.

3 Dataset from KTH

The second dataset recorded at KTH is sensor based. It contains recordings of 20 different individuals and a large number of different manipulation tasks. The aim of this database is to aid investigation and experiments concerning modeling of actions with action primitives (see also WP 3.1 and WP3.2).

The dataset was also described thoroughly in [7] where it was used for evaluation.

To generate the measurements for the training data, a Nest of Birds sensor was used, see Fig. 4 (right). The Nest of Birds simultaneously tracks the position and orientation of four sensors, referred to transmitter emitting pulsed DC magnetic field. The placement of the sensors is shown Fig. 5: thumb, hand, lower arm and upper arm. The persons involved in the study were not trained in any special way - each action started by having an arm in a relaxed, vertical position. Apart from the variation in their height and velocity with which an action was performed, the following variations were introduced to the training data:

- The objects were put on two different heights
- The person was standing at three different angles with respect to the table: 0, 30 and 60 degrees

Each action was performed three times for all combinations of the above heights and orientations resulting in total 18 training sequences per person and action thus 360 training sequences for each action.

The sensors were positioned as follows (see Fig 4):

1. On the chest for reference
2. On the middle of the hand
3. On the thumb

Figure 3: The image shows the setup of the recording session of the AAU dataset. One can see the four blue video cameras and the cables for the motion star system (bottom right). At the bottom left, the transmitter of the motion star system is visible.

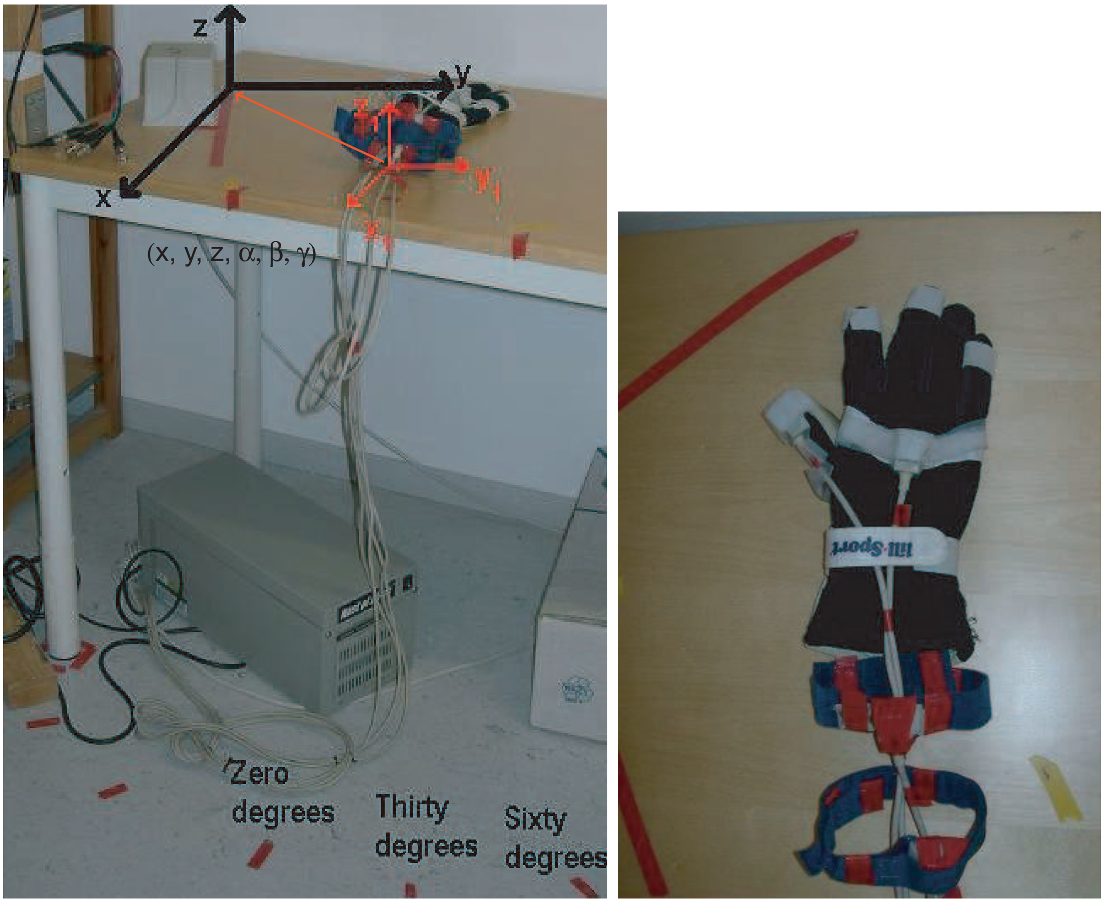

Figure 4: Nest of Birds sensor (left), and Glove with the four sensors (right).

4. On the index finger

The following activities were considered:

1. Push an object forward, short distance
2. Push an object forward, longer distance
3. Push an object to the side, close distance
4. Push an object to the side, far away
5. Rotate an object, close
6. Rotate an object, far away
7. Pick up an object, close
8. Pick up an object, far away
9. Move an object, (pick it up and put it down), close
10. Move an object, (pick it up and put it down), far away

Fig. 5 shows two example images stored during a push activity training - the activity is performed with the object being placed at two different heights. To motivate the choice of these activities, the reader may consider a robot being a part of a coffee drinking scenario. A *pick up* activity could be representing the fact of picking up the cup to take a swig of coffee; *put down* an object could represent leaving the cup of coffee after taking a swig, *rotate* an object would be similar to fold a napkin placed on the table, and finally, let us suppose that the person who sat down in front of you taking a coffee asks for the sugar bowl close to you and you *push* the bowl sliding over the table to bring it closer to him/her. The activities considered in this work are major building blocks of any similar task.

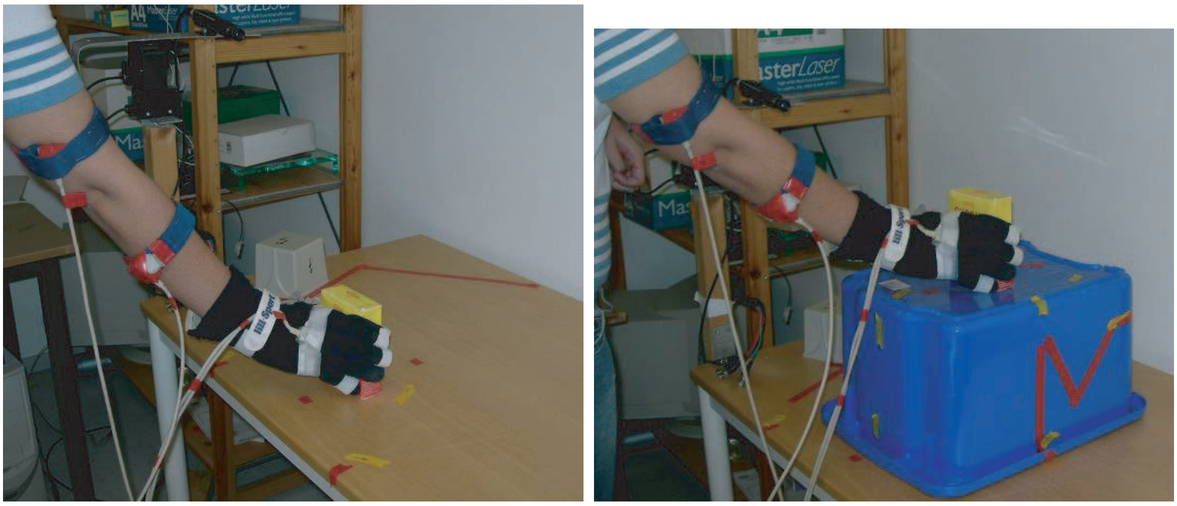

Figure 5: An example of pushing forward an object on the table (left) and an example of pushing forward an object on the box (right).

4 The Master Motor Map from UniKarl

The dataset at UniKarl contains human capture data stored in the MMM Format (Master Motor Map, as described in D8.2.1). The Master Motor Map serves as an exchange format for motion data captured with various sensor systems. Currently, the database consists of movements performed by 20 persons with a total size of 400 movements. The data has been recorded with the VICON optical motion capture system within the German Humanoid Robotics Project (SFB 588) funded by the Deutsche Forschungsgemeinschaft (DFG) [5]. Ten cameras were used with a resolution of 1000×1000 pixels, running at a frame rate of 120 Hz. The marker set was a standard biomechanical marker set consisting of 80 markers [4]. The database is available on: `http://i61www.ira.uka.de/users/asfour/mmm/`.

Within the PACO-Plus project, the marker data has been converted to the MMM format. Currently, a common database is being built which includes the motion capture data acquired at AAU, KTH, and UniKarl, using the Master Motor Map as the common data representation. For this purpose, conversion modules for mapping the marker data of AAU and KTH to the MMM format will be implemented. Furthermore, the markerless human motion capture system developed at UniKarl already produces output in the MMM format. The idea is to have a database consisting of human motion data captured with various sources but in terms of the same representation. This will allow us not only to visualize and reproduce human motion on the robot, but also to develop, train, and evaluate an action recognition system that is independent of the data acquisition techniques.

References

[1] Ascension. *Motion Star Real-Time Motion Capture.* http://www.ascension-tech.com/products/motionstar_10_04.pdf.

[2] P. Azad, T. Asfour, and R. Dillmann. Toward an unified representation for imitation of human motion on humanoids. In *International Conference on Robotics and Automation (ICRA), Roma, Italy*, 2007.

[3] P. Azad, A. Ude, T. Asfour, G. Cheng, and R. Dillmann. Image-based Markerless 3D Human Motion Capture using Multiple Cues. In *International Workshop on Vision Based Human-Robot Interaction*, Palermo, Italy, 2006.

[4] T. Beth, I. Boesnach, M. Haimerl, J. Moldenhauer, K. Boes, and V. Wank. Characteristics in human motion from acquisition to analysis. In *International Conference on Humanoid Robots (Humanoids), Karlsruhe/München, Germany*, 2003.

[5] URL: `http://www.sfb588.uni-karlsruhe.de/`.

[6] D. Grest, V. Krueger, and R. Koch. Single view motion tracking by depth and silhouette information. In *Scandinavian Conference on Image Analysis,*, 2007. submitted.

[7] I. S. Vicente and D. Kragic. Learning and recognition of object manipulation actions using linear and nonlinear dimensionality reduction. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2007*, 2007. submitted.