M18-meeting
16 September 2009
Department of Engineering and Computer Science
Universitat Jaume I de Castellon

This document collects the minutes of the M18 meeting Sep 16th 2009 enriched by the notes of the complementary discussions in the preceding and following days during the IURS'09 Summer School in Benicassìm (13-18 September 2009).

Participants: Patrizia Fattori, Claudio Galletti, Konstantinos Chatzidimitrakis, Giacomo Placenti (UBO); Fred Hamker, Frederik Beuth, Mark-André Voss (WWU[1]); Markus Lappe, Katharina Havermann, Robert Volcic (WWU); Marc van Hulle, Nikolay Chumerin, Karl Pauwels (K.U.Leuven); Angel del Pobil[2], Eris Chinellato (UJI), Silvio Sabatini, Giorgio Cannata, Andrea Canessa, Manuela Chessa, Agostino Gibaldi (UG).

Start h 10:20am

1. **Welcome and communications** [Silvio]
   Underlined the main focus of the meeting: (1) Decisions following the 1st-period review report, (2) integration of the different components.

2. **Discussion** [All]
   The reviewers' recommendations are recalled:
   - Identification of an experimental task
   - Increased integration among partners
     - Closer synch between WP2 and WP3
     - Integration of UJI with the rest of the project
   - Explicit use of a non-linear (i.e., space variant) mapping.

   Identification of an experimental task
   General agreement on the task suggested in the review report: *"fixating on and pointing to some visual stimulus (e.g., a new object in the environment or any object of a particular characteristic, e.g., the green object)"* provided that *"this visuo-motor behavior be implemented through mechanisms that are grounded in physiologically-plausible mechanisms, in particular, that they capture key processes regarding the bilateral interaction between motor and perceptual processes"*.
   However, several important points have been raised:
   - Valorization of the "fragment" concept, which represents a key research issue of the project.
   - Enrichment of the saliency concept: not only based on low-level features, but rather related to (stereo-based) *object identity*. WP3 activities should

[1] Chemnitz University of Technology
[2] E-mail discussion

address this issue (currently the progress along this line is suffering a delay, though the development of V2 and V4 visual descriptors is being to start).
- In (re-)defining the task it is important to chose one that could be *shared* between the human and the robot (cf. the planned UJI-WWU joint activity).
- Highlights on the advantages of the peripersonal wokspace: proprioceptive information (related to the knowledge of the position of the hand in space) provides significant cues that add to the ones coming from the visual and oculomotor systems.

Concerning the "sticky hand" metaphor, there is a general consensus on the fact that the task is somehow 'un-natural' (while manipulating an object, we continuously look at it, but when we pull or push an object, it is not necessary to keep fixation on it) *but* it can be considered a drastic simplification of a manipulation process.
Furthermore, it has been observed that fine eye movements
- have no direct relation with reaching
- would imply grasping and manipulation, which evade the current scope of the Eyeshots' project.

In conclusion, the "sticky hand" will be considered as a simulacrum of manipulation, which might be only instrumental to show/prove the active capabilities of the vision system, without explicitly addressing grasping and manipulation problems.

Explicit use of a non-linear (i.e., space variant) mapping

A space variant sensing of the visual signal is an intrinsic feature of the active vision paradigm (cf. foveations). In the first phase of the project, space variant processing is assumed to be potential, and stays in the background, without affecting the general approach of the proposed models and architectures. The importance of an *explicit* use of the space-variant sensing is stressed (since any active vision system must rely on it to be convincing).

Manuela presents the first results obtained by UG-Dibe for disparity estimation in a log-polar geometry and the approach followed. The space-variant disparity processing retains the maximum resolution (and thus disparity acuity) in the fovea, while being capable of detecting large disparity values in the periphery.

Disparity detection takes place in the cortical domain. Yet, a direct measurement of the mapped disparity in the cortical plane can be thought as instrumental to simplify the space-variant processing, and its biological plausibility can be discussed. Basically, the space-variant transformation implies: (1) space-variant low-pass filtering (to remove high spatial frequencies) and (2) space-variant sub-sampling (to reduce spatial resolution).

Open problems:
- Space-variant sub-sampling and space-variant convolutions in the convolutive networks (cf. WP2)
- Possible diversification of the mappings to be adopted in the different levels of a general architectural model of cortical processing (cf. different cortical areas, in WP3 and WP4)

[ex-post editor's note: retinocortical transformations might have implications both for space-variant sensing and for cortical processing, e.g., by simplifying the geometry of the feature space in the mapped domain. For what concerns the space-variant sensing, only (which is the key feature for active foveations), space-variant image sensing (and thus the corresponding receptive fields) can be well approximated by a patch-wise approach: overlapping image patches whose size increases with eccentricity, and, relatively to the patch size, uniform spatial low-pass filtering and sub-sampling].

Increased integration among partners

- Operative proposals for WP2-WP3 interactions already circulated. Subgroup discussions required [Fred-Marc] (see details below).
- Actions for a better integration of UJI with the rest of the project. Subgroup discussion required, specifically on what can be developed and implemented from the theoretical framework proposed by UJI [Eris-Markus-Patrizia-Claudio-Fred-Silvio] (see details below).

Mechatronic anthropomorphic vision system

Following the reviewers' comments (raised at the end of the review meeting, and not included in the written report), Giorgio reported that UG-Dist is developing a simulator of the bio-mechanical system. The simulator (that will integrate the UG-Dibe active vision module developed by Manuela) should be ready in the near future. More specifically, a library of ocular models and cameras (dynamic/kinematic) will be completed by the end of October 2009.

Several groups are interested in this integrated simulator. Emphasis is put on the fact that, in addition to the "tracking a surface" mode (useful for smooth control of vergence on a visible surface) the simulator should include saccadic mechanisms, too, which play a key role for the project.

Giorgio further observed that, concerning the physical realization of the mechatronic binocular vision system, the engineering phase is ready to start, depending on the decision of the Consortium. No contrary positions emerged. Accordingly, the realization of the system can proceed as planned.

2a. Subgroup meetings and discussions

WP2-WP3 interactions

[Fred, Marc, Nick, Mark-Andrè, Frederik, Manuela, Silvio]

- Fred, Marc, Nick, Mark-Andrè, Frederik:

K.U.Leuven will consider Chemnitz' disparity-tuned feature-selective cells, that were trained on the statistics of natural scenes, as inputs to our convolutional network and train the network to solve the vergence task. For this, a study visit of Nick Chumerin (K.U.Leuven) to Chemnitz is planned.

Since the population of feature-selective cell responses is expected to lead to sparse, asymmetric population responses, the vertical disparity contribution in the image can not be cancelled out by pooling over the population. The presence of object discontinuities also will pose this problem. We wil compare the performance of the convolutional network with the one we have trained for regular (Gaussian) disparity-tuned cell populations.

If the performance is not satisfactory, we will need to consider mixed population configurations (feature-selective + regular cells), perhaps using a pruning strategy.

- Fred, Manuela, Mark-Andrè, Frederik:

The simulator developed by UG-Dibe will be used by WWU-Chemnitz in order to obtain snapshots of different scenes acquired by the laser scanner with different fixation points. To this aim the software module will accept as inputs the scene to be observed (VRML model) and the fixation point and will provide as outputs the stereo pair (left and right images). The software module will be ready and provided to the partners by mid November.

- Fred, Silvio: considerations on the quadrature properties of the simple cell receptive fields learned by the model. Quadrature pairs are necessary for building binocular energy units. As an alternative (to avoid "wiring-by-hand" upon learned resources), the model could/should directly learn binocular complex cells with proper orientation and disparity tuning.

WP2, interactive stereopsis under limited accuracy of the gaze direction

[Karl, Manuela, Silvio]

K.U.Leuven's approach for disparity estimation under limited accuracy of the gaze direction will provide gaze information (corrected fundamental matrix) to WP1 (UG-Dist) as required for the interactive exploration of the fragment. This will be the subject of Deliverable 2.2a.

In close cooperation with UG-Dibe, the gaze correction procedure will be adjusted to operate on the population-based representation. This will enable the required coordinate transformations by means of gain modulation or radial basis function network techniques. This will be the subject of Deliverable 2.2b.

The computer vision approach will serve as a performance baseline to compare the distributed model, which eventually include space-variant mapping (see work already developed by UG-Dibe). Such a mapping will enable a reduction of the computational load and a widening of the range of disparity values that can be computed for a given scale.

WP2, functional characterization of the learned disparity-vergence cells: comparisons

[Marc, Nick, Agostino]

We discussed the ways to produce a more unified approach to characterize the vergence approaches proposed in D2.1. We agreed that this characterization should be done in the format vergence signal vs. disparity (see Fig. 10 in deliverable D2.1).

Nick: Right now, this kind of plots is not possible to produce for the actual convolutional vergence-control network (CVCN), because the output of the CVCN is a vergence angle, but not a vergence signal. But this problem might be solved by modification of the vergence control database and proposed CVNC. In this case we should add the desired vergence signal value to each sample in vergence control dataset and retrain the network. It seems to be not difficult, but the question is how to produce the desired vergence signal value from given data (disparity, scene structure, desired fixation point, actual fixation point and so on)? One of the possible solutions is in approximation of the vergence signal by the weighted difference of actual and desired vergences.

Agostino: In order to go more in depth in the comparison of the convolutional vergence-control network (CVCN) with the linear dual mode (i.e. two "channels") control described in section 4 of deliverable D2.1, we will test the functionality of the latter in a more general case, i.e. not only for a frontoparallel plane and when the gaze direction is straight ahead.
The general case will be characterized by a gaze direction defined by fixed elevation and azimuth angles spanning a defined range, and the plane will be both orthogonal to the gaze direction, and tilted respect to it.
A further comparative test will be done with low-resolution images (41x41), like those used in the CVCN.

A systematic approach to compare the CVCN and the dual mode models will be defined.

WP1-WP2 interactions
[Giorgio, Marc, Karl, Nick]

- Giorgio, Marc, Nick: Coordinated control of ocular movements.

  The problem proposed by K.U.Leuven is that of training a neural-network (NN), based on stereo-vision data, and to use the output of the trained network to drive (using a continuous closed loop control strategy) the gaze direction onto a point of interest in the scene.

  The proposed control problem will be first formulated in a kinematic framework and simulated within the integrated simulation environment (currently under development by UG).

  Actions: Implementation of a first control scheme is expected to start (with support from UG) as long as the first release of the integrated virtual simulator will be available. See also Encl.1.

- Giorgio, Karl: Estimation of relative eye orientation.

  The coordinated control of the robot eyes (as part of WP1) could require the measurement or the estimate of the orientation of the eyes (gaze) with respect to some (head) fixed reference frame. As a matter of fact the present design does not plan the use of any orientation sensor, furthermore, assuming to use similar control strategies on different robotic eyes (e.g. tilt-pan cameras), uncertainties related to camera placement as well as on the kinematics of the structure may require the adoption of image based estimation procedures to compute at any image time sampling time the homogeneous transformation (or at least the rotation matrices) of the left and right eyes with respect to the head.

Actions: Implementation of the K.U.Leuven's visual-based posture estimation algorithms within the simulator and tests with the candidate control algorithms proposed in WP1. See also Encl.1.

WP4, visuomotor representation of reachable objects
[Eris, Markus, Patrizia, Claudio, Fred, Silvio, Katharina, Konstantinos]

In the framework we agreed upon, eyes and arms are treated as separate effectors that receive motor control via different specific sensorimotor representations of the peripersonal space, maintained through basis function networks.

The plan is to combine these two representations to form a unique shared representation of visuo-motor awareness. The integrated representation is obtained by matching the two basis function networks by a Hebbian learning procedure on the robotic setup.

Fundamental points to incorporate to the model are 1) the actual experimental learning framework; 2) the inclusion in the model of the neuroscience data from WP5 regarding area V6A neurons and saccadic adaptation; 3) the final human-robot interaction setup, on which some decisions have already been taken.

See also Encl. 2 [04/10/09].

We agreed upon the following points (also after e-mail discussion with Angel):

- Start working on a core model, *simple* and *flexible* enough to be generalized to account for the experimental outcomes of WP5. A minimal/essential list of grounding papers will be provided.
- Finding a person *fully* (or, at least, mostly) devoted to this task is crucial and of priority importance for the project and for the next review meeting.
- The person (supervised by Eris) should travel in Chemnitz and Münster for some weeks in the next months (e.g., November) to receive further guidance by Fred on the modeling and discuss with Markus and Katharina about how to link the modeling with the experimental data.
- In the first phase, detailed inclusion of experimental data and the integration of the model in the robotic set-up will be postponed, since they evade the objectives of the 2nd reporting period.
- In the background, it is important to prepare the experimental set-up and to smooth-out problems with the stereo head and its integration with an arm. Existing expertise in the UJI lab, and maybe one the two persons to be hired this year by UJI on another project, could help and contribute to this task.

The e-mail exchange (to/from Angel [16-17/09/09] and the "notes on the schedule" from Markus [18/09/09]) is part and parcel of these notes (see Encls. 3 and 4).

3. **Management and other contractual issues** [Silvio]

   Deliverables due by month 24 have been recalled.

   All other issues are postponed to the next meeting.

   Tentative plan for the next M24 meeting:

   > *date*: 1 day and a half in the 2nd or 3rd week of January 2010
   > *place*: TBD (any location easy reachable by partners, to minimize time spent traveling).

<u>End h 18:20pm</u>