EUROPEAN COMMISSION
Community Research

COGNITIVE SYSTEMS
ROBOTICS

SEVENTH FRAMEWORK PROGRAMME

Grant agreement for a small or medium-scale focused research project (STREP)

Annex I – Description of Work

Project acronym: *EYESHOTS*
Project full title: *Heterogeneous 3-D Perception across Visual Fragments*
Grant agreement No: *217077*

List of Beneficiaries

Beneficiary			Country	Date of entry	Date of exit
No.	Name	Short name			
1 (coordinator)	University of Genoa	UG	Italy	Month 1	Month 36
2	Westfälische Wilhems-University Münster	WWU	Germany	Month 1	Month 36
3	University of Bologna	UNIBO	Italy	Month 1	Month 36
4	University Jaume I, Castellòn	UJI	Spain	Month 1	Month 36
5	Katholieke Universiteit Leuven	K.U.Leuven	Belgium	Month 1	Month 36

Table of Contents

PART A

A1. Budget breakdown and project summary

A1.1 Overall budget breakdown for the project

Grant agreement Preparation Forms

EUROPEAN COMMISSION
7th Framework Programme on Research, Technological Development and Demonstration

Collaborative Project

A3.2: What it costs

Proposal number (1)	217077	Proposal acronym (2)	EYESHOTS

ONE FORM PER PROJECT

Participant number in this project	Organisation short name	Estimated eligible costs (whole duration of the project)				TOTAL A+B+C+D	Total receipts	Requested EC contribution
		RTD / Innovation (A)	Demonstration (B)	Management (C)	Other (D)			
1	UG	750,538.48	0.00	89,323.38	0.00	839,861.86	0.00	652,227.24
2	WWU, Muenster	835,840.00	0.00	2,000.00	0.00	837,840.00	0.00	628,880.00
3	UNIBO	456,256.00	0.00	0.00	0.00	456,256.00	0.00	342,192.00
4	UJI	495,361.28	0.00	2,979.00	0.00	498,340.28	0.00	374,499.96
5	K.U.Leuven	534,934.40	0.00	1,000.00	0.00	535,934.40	0.00	402,200.80
TOTAL		**3,072,930.16**	**0.00**	**95,302.38**	**0.00**	**3,168,232.54**	**0.00**	**2,400,000.00**

A1.2 Project summary

The research intends to investigate the interplay existing between vision and motion control, and to study how to exploit this interaction to achieve a knowledge of the surrounding environment that allows a robot to act properly. Robot perception can be flexibly integrated with its own actions and the understanding of planned actions of humans in a shared workspace. The research relies upon the assumption that a complete and operative cognition of visual space can be achieved only through active exploration of it: the natural effectors of this cognition are the eyes and the arms.

Crucial but yet unsolved issues we address are object recognition, dynamic shifts of attention, 3D space perception including eye and arm movements including action selection in unstructured environments. We propose a flexible solution based on the concept of visual fragments, which avoids a central representation of the environment and rather uses specialized components that interact with each other and tune themselves on the task at hand.

In addition to a high standard in engineering solutions the development and application of novel learning rules enables our system to acquire the necessary information directly from the environment.

The study and models of human/primate behaviour, based on specific experiments, guide many of our envisaged solutions.

Three main objectives will be addressed:

- A robotic system for interactive visual stereopsis [composed of: an anthropomorphic mechatronic binocular system; and software vision modules based on cortical-like population, to be used as an experimental platform].
- A model of a multisensory egocentric representation of the 3D space [constructed on binocular visual cues, signals from the oculomotor systems, signals about reaching movements performed by the arm].
- A model of human-robot cooperative actions in a shared workspace [relaying on the concept of shared attention to understand the intention or goal of the communicating partner].

PART B

B1. Concept and objectives, progress beyond state-of-the-art, S/T methodology and work plan

B1.1 Concept and project objectives

- Problem and main goal -

Humans act in space. Sometimes, interactions in space are explicit, as we point, reach or grasp the things around us. Other interactions are implicit, an awareness of where we are and what things surround us. In general, to interact effectively with the environment, it can be argued that humans might use complex motion strategies at ocular level (but possibly extended to other body parts, e.g. head and arms, so possibly using multimodal feedback), to extract information useful to build representations of the 3D space which are coherent and stable with respect to time.

Such representations rely on a multisensory active exploration of the environment. In particular, purposive (active) vision is an important source of information and provides a number of cues about the 3D layout of objects in a scene that could be used for planning and controlling goal-directed behaviors.

Although computer and robot vision are today making technological progress, the current active vision solutions are increasingly driven by applications like robotic manipulation or surveillance and are still far from reaching a real purposive behavior. More precisely, research in robot vision is more geared towards active "looking at" scenes and objects rather that "seeing", gathering the visual data and having to act in a limited time and space.

"Seeing" is something we do rather than a sequence of hierarchical interpretative processes. From this perspective, the experience of "seeing" in not necessarily generated, but it express itself in the behavior.

Following these premises, we will study the topic of seeing using an embodied artificial intelligence by considering the cognitive valence of an *early* perception-action embodiment in the visual system.

The necessity of an embodiment to achieve intelligent behavior is not a novel issue in itself (e.g., the "enactivism[1]" paradigm). Yet, in general, in active vision systems, the perception-action loop closes at a "system level" (by decoupling de facto the vision modules from those dedicated for motor control and motor planning), and the exploitation of the computational (feedback) effects of voluntary explorative eye movements on the visual processes are very rare in artificial artifacts. On the contrary, a wide number of neurophysiological experiments report of modulatory effects of motor and premotor signals on the visual receptive fields across several cortical areas (e.g. the gain fields), postulating their role in gaining a perceptual visuospatial awareness in head-centred coordinates for visually guided actions in the peripersonal space. Such motor components, actively contribute to stabilize/improve the 3D perception of space, but also allow us to achieve a global awareness by enabling/establishing loose links among visual fragments of the observed scene (global spatial reference). Such links evolve in a dynamic way, are influenced by attentive processes and memory, and are task-contingent (action/goal-oriented).

From this perspective, the concept of an active fragmented vision represents a dynamic cognitive interpretation of the scene, which does not imply a real metrical 3-D reconstruction of the observed space, but instead a loose representation of objects that are actively bound on time for the task at hand (in terms of affordance, salience, and planning of actions).

The goal of the project is thus to develop a perceptual agent capable of achieving a full 3D awareness for interaction control/planning in the peripersonal space. The sophistication of perceptual capabilities will ultimately be measured in terms of their value to the agent in executing its tasks.

[1] Enactivism claims that sensory-motor activity and embodiment is crucial in perceiving the environment and that machine vision could be a much simpler task if considered in this context (e.g., see Dario Floreano and writings by Merleau-Ponty, O'Regan, Hurley, Noe).

The final awareness will derive by building a knowledge of the sensorimotor laws that drive the relation between possible actions and the resulting changes in incoming visual information [O'Regan and Noe, 2001], and by the integration of this knowledge into planning behavior.

Two important intertwined questions will be raised, which actually define a next big step in the development of artificial perceptual agents (autonomous robotic agents):

1) How to design a robot vision system that, through intentional (i.e., voluntary) eye movements, is able to "see", not only of being able "to look at" saliencies?
2) How can the effect of active eye movements and of arm reaching actions be expressed as joint visuo-motor features, patterns and relationships for a perceptual awareness of space?

The project addresses these problems at different levels, integrating the contributions coming from different disciplines (engineering, neuro-physiology, and psychology). At the lowest level investigation is focused to understand the mechanisms of depth vision based on / intertwined with eye motor control. At higher level, the project aims at allowing an artificial intelligent system to master the peripersonal space in which it is embedded and the objects present in it.
Specifically,

- We address the problem of arm-reaching in the three-dimensional space around the observer as a purposive visuomotor task in which synergic processes between visual perception modules (depth vision) and effectors (eye and arm movements) strongly contribute to the organization of the adapted behavior.
- We will develop interactive vision modules for the perception of 3D spatial relations and relative 3D motion for controlling spatially directed actions (e.g., reaching), and, in general, visually-guided goal-directed movements in the peripersonal space.
- We will experiment, study, model, and implement structural mechanisms and adaptive integration strategies to enrich the front-end vision modules with an ("active") motor component.
- The resulting representation of the visual space will embed a strong motor component and will be specifically suited for a dynamic interaction of a robot with the environment. On this basis, cooperation between a robot and a human partner, which share the same peripersonal space will be considered.

- Specific scientific and technological objectives -

The EYESHOTS project will address three Principal Objectives:

Objective 1: Development of a robotic system for interactive visual stereopsis.
A first objective concerns the development of a robotic system composed of: (1) an anthropomorphic mechatronic binocular system, and (2) software vision modules based on cortical-like population of disparity detectors with different characteristics for foveal and parafoeveal representations of the visual field, to be used as an experimental platform. The function of the system is to interactively explore the 3-D space by active foveations. Benefits of the motor side of depth vision are expected to be bidirectional by learning optimal sensorimotor interactions: on one hand the system learns to see in 3-D through eye movements, and, on the other the system learns coordinated binocular eye movements in 3-D through vision. Binocular alignment and stereo matching is favored by the structural paradigms of the binocular eye coordination. Fast (real-time) binocular fusion around the fixation point is achieved by dynamically adjusting the response of disparity detectors.

Objective 2: Development of a model of a multisensory egocentric representation of the 3D space.
The representation is constructed on (1) binocular visual cues, (2) signals from the oculomotor systems (position of the eyes), (3) signals about reaching movements performed by the arm.
Egocentric representations require regular updating as the robot changes its fixation point. Rather than continuously updating based on motor cues or a visual mechanism (i.e. optic flow), the model updates only the egocentric relationship and object-to-object relationships of those objects currently in the field of view. During motion, the model covertly and overtly shifts attention to objects in the environment to maintain the

model's current awareness of the environment. The updating of the internal representation of spatial relations requires binding processes across the different visual fragments. Spatial awareness of the environment provides the model with the capability to interact with 3D environments. The model can maintain awareness of objects and visual features as the robot moves its eyes in the 3D space. The model can encode and update the 3D spatial location of objects and if the model needs to view an object outside of the current field of view, the model can request a saccade to a remembered spatial location. The model is also able to request arm movements to reach spatial locations in the peripersonal space to interact with objects in the 3D environment.

Objective 3: Development of a model of human-robot cooperative actions in a shared workspace.
By the mechanism of shared attention the robot will be able to track a human partner's overt attention and predict and react to the partner's actions. This will be extremely helpful in cooperative interactions between the robot and a human.

- Expected Results -

By the end of the three years the following results will be achieved (to answer a variety of fundamental questions about to what extent, and how a more "physical" coupling between perception and action can enrich 3D robot vision capabilities):

1. Implementing strong "dynamic" and "pro-active" components in which the effect of eye movements and of arm reaching actions will express as *joint* visuo-motor features, patterns and relationships for a perceptual awareness of space;
2. Building a contingent knowledge of the sensorimotor laws that govern the relation between possible actions and the resulting changes in incoming visual information.
3. Binding of objects into a global workspace for cognitive task control.

To measure the success of such achievements, three different subsystems will be considered:

1. the oculomotor system [an anthropomorphic robotic vision head (mobile eyes, fixed neck)]
2. the visually-based reaching system [a robotic arm and a stereo vision system]
3. the human/robot system [eye-tracking and hand tracking].

From the technological and engineering point of view, the project will follow a bottom-up strategy aiming at the development of an integrated head-eye-arm platform. The bio-inspired stereo head-eye system and the experimental set-up consisting of a standard robot arm and head-eye system will be developed in parallel. The integration of the two systems will be pursued during the last part of the project. However, priority will be given to the tests described below, performed separately on the two systems and before the final integration.
Tests to be performed on the different subsystems include:

- Coordination of binocular eye movement to minimize the motion of the epipolar lines.
- Anticipatory control of vergence.
- Learning receptive fields for interactive stereo processing.
- Sensorimotor representation of the peripersonal space (stereovision cues + oculomotor signals + arm-reaching signals).
- Shared attention in the peripersonal space.

The expected capabilities of the final system(s) are illustrated by the following three examples that relate to the corresponding three objectives of the project:

1. The mechatronic binocular eye system will exhibit stable and fast eye movement behaviours, switching between reflex-like vergence movements and voluntary 3D exploration.
   <u>Expected result: active exploration.</u>
2. The robot will reach different momentarily assigned targets (objects located in the peripersonal space) without fixating them continuously and while keeping visual exploration of the scene (through eye movements).
   <u>Expected result: situation awareness,</u> i.e.:

The peripersonal space is fully "measured" including the awareness of limb position and movement, which is essential for limb coordination

- without building a 3D reconstruction of the space
- by using motor representation (motor maps for reaching and eye movements) to loosely couple visual fragments in headcentric coordinates.

3. Same scenario of example 2, with the robot engaged in a cooperative behaviour in the workspace shared with a human (see Task5.4 for details). E.g., the robot will reach targets that a human observer is fixating.
Expected result: situation awareness of the shared workspace.

- Evaluation activities and "benchmarking" -

- The computational advantages of the solutions adopted in the head-eye system (aimed to fully exploit the "motor side" of depth vision) will be assessed with respect to conventional approaches to binocular robot vision. Quantitative evaluation of the differences with standard platforms and their role in active vision will be carried out.
- The most desirable characteristics will be (i) flexibility, allowing the system to explore its surrounding environment in the most efficient way, (ii) good dynamical performances, for fast reaction to environment changes and high speed scrutiny, e.g. minimal number of saccades needed for fixating a target, (iii) compactness, for easy integration in space limited systems such as mobile robots, and (iv) high accuracy and robustness.
- The voluntary explorative behaviour of the vision system (e.g., over an object's surface) will be a distinctive feature of the perceptual performance of our system. The vision system will perform an attentive process of "seeing" ruled by short-range saccades. The geometric properties of the object under observation will drive visual exploration rather than mere saliency data.
- The final system should be able to localize a particular target object in a visual scene. This will be evaluated by a benchmark involving a number of different scenes and different target objects. The reference model to which our model will be compared with is a standard, bottom-up saliency model.
- The performances of the EYESHOTS binocular robotic platform will be systematically compared with those obtained with more conventional platforms (pan/tilt/vergence heads). The added value of its integration with the arm system will be assessed with respect to representation accuracy and speed.
- Comparisons with similar platforms developed in other projects (such as *POP*¸ *Decisions-in-motion*, *RobotCub*) will be considered. From the geometric and kinematic point of view, we compare how structural features affect the system model, the control procedures and performance and possibly the visual processes as well (e.g., (i) negligible influences of the viewing distance on retinal motion information due to eye movements; (ii) reduced computation time of stereo matching at reasonable mechanical costs).
- Such comparisons will be made publicly available on the website and discussed during joint events with consortia of related projects.
- The functional testing of the system's components (see Workplan description and the List of Milestones for details) will lead to a statistical characterization of their performance (e.g., in terms of accuracy, robustness [e.g., low lighting conditions], velocity, and the associated computational cost).

B1.2 Progress beyond the state of the art

- Scenario (state of the art) -

ROBOTICS (purposive vision). Current models of vision generally assume action and perception are still sequential processes: it is the effect of action that influences the perception (e.g., eye movements serve to select a scene for perception) and not the acting itself, even less its planning (but see, Hamker 2005a, Hamker 2005b). In this project we plan to consider visual perception and action selection operate in parallel, thus really closing the loops and taking full advantage of a concurrent/anticipatory perception/action processing.

- **PSYCHOLOGY (eye movements and visual fragments)**. Until recently it was believed that the brain constructs a composite image of the external world by combining images from consecutive fixations (Jonides et al., 1982), but little experimental evidence supported this idea. Every fixation provides a new piece of local information, called a visual fragment, about a certain part of the scene, and these fragments stay to some degree separate across saccades. While general scene changes are often unnoticed during a saccade, changes at the saccade target are noticed and pre-view experiments suggest that some aspects of the current view of the scene (the current fragment) influence the perception during the subsequent fixation. Most fragile during a saccade is location information. Intra-saccadic displacements of visual stimuli, even of the saccade target, are usually not noticed. When the saccade target is repeatedly displaced, however, humans unconsciously adapt the gain of the saccade and compensate for the visual expectation that the postsaccadic saccade target should appear in the center of the visual field at the new point of fixation (McLaughlin, 1967). This adaptation of oculomotor parameters recursively modulates perceived location of objects in the scene (Bahcall and Kowler, 1999; Awater et al., 2005), which suggests the existence of a closely knit integration of action and cognition in scene perception.

- **NEUROSCIENCE (visuospatial awareness and planning behaviour)**. The research in neurobiology provides large evidence that the brain uses representations of the three-dimensional (3D) space, which are both perception and action oriented. The most famous examples are so-called "Mirror Neurons". These neurons fire in the case of e.g. a monkey performing an action or watching somebody else performing the same action (for review see (Rizzolatti and Craighero, 2004)). Less known in the robotics community are the neurons, which change their visual receptive fields according to motor actions, such as the position of the eyes and of the limbs.

Even though the specialization of different brain pathways towards action-oriented and perception-oriented vision is well recognized, their interrelation seems to be tighter than previously expected (Himmelbach and Karnath, 2005). Recent research in which we have participated (Singhal et al. 2007) has shed new light on the theory of the dorsal and ventral visual streams (Goodale and Milner, 1992). Mutual projections are very likely connecting advanced areas of the two streams, and allow, on the one hand, to potentiate the effectiveness of action planning and selection through the use of conceptual knowledge about the world and, on the other hand, to use visuomotor insights (i.e. affordances) for improving the capabilities of classifying and recognizing objects.

Focusing on the action-oriented dorsal stream, neuroscientific research, including findings from UNIBO, show that the sensorimotor transformations regarding reaching and vergence eye movements are likely coupled in the posterior parietal cortex of primates. An artificial agent endowed with an implicit coupling between its ocular and limb motor systems can more easily take advantage of both proprioceptive and exteroceptive signals in order to interact with its environment and construct an awareness of it.

- Scientific and technological baseline of the Consortium -

The ideas expressed in the project and the proposed approach move from the following scientific ground. The project aims to develop and integrate synergically these concepts towards a significant advance of the state of the art.

- **MECHATRONIC EYE SYSTEM**

Baseline Many eye-head robots have been designed in the past few years but little attention has been paid to emulate eye kinematics and actuation similar to those of the human eye. The emulation of the actual ocular mechanics adopting a spherical eye-ball, actuated by tendons pulled by DC motors, has been first introduced in (Cannata et al., 2007a; Biamino et al., 2005; Cannata et al., 2006a; Cannata and Maggiali, 2006; Cannata et al., 2006b Cannata et al., 2007b). The implementation of saccades and smooth pursuit obeying to the geometric principle of Listing's Law (Tweed and Vilis, 1987, 1988, 1990; Furman and Schor, 2003; Straumann et al., 1996), has been investigated. We have formally proven (Cannata et al., 2006a; Cannata and Maggiali, 2006; Cannata et al., 2006b; Cannata et al., 2007b) that the mechanical design of the tendon driven spherical robot eye, ensures the correct approximation of Listing's Law. This confirms the hypothesis supported by severaly physiologists that these types of ocular motions are mainly due to the mechanical structure of the ocular plant (Koornneef, 1974; Miller, 1989; Demer et al., 1995; Clark et al., 2000; Demer, 2000).

Progress The arguments discussed above open new interesting questions.

(1) The first one refers to the problem of the coordinated eye movements where Listing's Law must be extended to keep into account the effect observed by physiologists of the tilting of Listing's Planes, (Furman and Schor, 2003). From a mechanical point of view this effect could be intuitively reconducted to an extension of the model proposed in (Cannata et al., 2006b). In fact mechanical implementation of Listing's Law is strongly affected by the path of the four recti extra ocular muscles which are constrained by soft connective tissue known in the literature as soft pulleys, (Demer et al., 1995). Recent physiological experimental observations of innervation of the soft-pulleys, (Demer et al., 2000), suggest that they could not act as a passive mechanism but could contribute to define the orientation of Listing's Planes. The investigation of the control strategies required to ensure the proper placement of the soft-pulleys as well as the study of the mechanics required to their implementation in a robot eye will be one of the objectives of the project.

(2) The second important point is the study of the interplay between the evolution of the ocular mechanics and the visual processing capabilities of the brain. Assuming that ocular mechanics and neural image processing and control mechanisms have followed a common evolutionary path, the current equilibrium (i.e. the way humans coordinate vision and ocular movements), will be investigated trying to highlight the interactions among neural processes and mechanical structure of the eye plant.

Main Partner : UG (G. Cannata)
Related WP(s) : WP1

- **DISTRIBUTED ARCHITECTURES FOR STEREOMOTION**

Baseline Considering jointly the binocular spatiotemporal constraints posed by moving objects in 3-D space, psychophysical and physiological data evidenced that motion detection and stereoscopic depth perception are processed together in the visual system by a population of cells tuned to both motion and disparity (DeAngelis and Newsome, 1999; Pack et al, 2003), but still the question arises whether stereomotion is processed on the basis of information that can be gathered directly from the luminance modulation or requires a specific feature tracking motion system [Lu and Sperling, 2001]. Our claim is that these positions can reconcile themselves if one assumes a proper spatial organization of short-range joint mechanisms for detecting motion and binocular disparity. Recent experimental observations [Anzai et al, 2001] on a joint encoding of motion and depth by visual cortical neurons in relation to the Pulfrich effect, provide evidence supporting our view. We have demonstrated that dense stereomotion information can be obtained, without tracking, from binocular local measurements, provided that such measures are characterized by high significance and robustness [Sabatini and Solari, 2004; Sabatini et al., 2002]. From this perspective, early vision is no longer associated to a mere pre-processing of the image signal, but relates to complex behaviours, to perceptual decisions and, eventually, to active perception.

Progress We will design a general cortical architecture based on local operators (e.g., binocular motion energy units) capable of providing a full description of 3D motion event originated by eye movements, by projecting it into a subspace of elemental features (motion, disparity and orientation). Such description will rely upon space and time phase information gathered from a band-pass spatio-temporal transformation of the binocular visual signal. Coherent stereo-motion correspondence constraints will be directly embedded in the structure of such visual operators, rather than being considered at higher semantic level of data fusion. By mapping the architectural principles of the cortical model in a population-based algorithms for motion and stereo processing, we expect to obtain a high degree of flexibility and potential adaptability of the front-end vision modules: each cell can be modulated by attentional signals fed back from higher visual processing stages or by non-visual signals from the effectors.

Main Partner : UG (S.P. Sabatini)
Related WP(s) : WP2, WP1

- ☐ <u>LEARNING BRAIN-STYLE ARCHITECTURES</u>

Baseline Several visual learning strategies in hierarchical neural network models have been suggested for developing receptive fields that extract feature descriptors of increasing complexity (Fukushima, 1980; Linsker, 1986; LeCun et al., 1989,1998; Mozer, 1991; Riesenhuber and Poggio, 1999; Wersing and Körner 2003). Most of these networks consist of interleaved S- and C-layers, as in the primary visual cortex, and has been introduced by Fukushima in his Neocognitron (1980). Since 1989, Yan LeCun and co-workers have introduced (LeCun et al., 2005) a series of similar network architectures based on developing local feature descriptors. Recent versions of this network, such as LeNet5 (LeCun et al., 1998), which has also been applied to a wider range of image classes, including binocular ones, has been shown to act as an efficient framework for nonlinear-dimensionality reduction of image-sets (Hadsel et al., 2006). The architecture is similar to Fukushima's but the network does not learn autonomously since it relies on back-propagation. But the big advantage is that the whole network is, from first till the last layer, optimized for the given task, making this approach useable for real-world applications. Riesenhuber and Poggio (1999) presented a biologically plausible architecture where only the last stage is a classifier (or a regressor) that is learnt in a supervised manner. Serre, Wolf and Poggio recently extended the model and showed how it can learn a vocabulary of visual features from natural images. Learning in their model is constrained to the tuning of simple cells to random snapshots of local input activity generated by presentations of objects of interest. The architecture has been applied to modeling of V4 and IT neuron responses and also as an input stage to a classifier for object and face recognition. One should note that these architectures start from the pixels in the input image, leading to a rich variety of detectors, but that often cannot be traced back to biologically plausible cues.

Progress We introduce a new application of hierarchical (convolutional) network architectures: binocular eye coordination based on near and far disparity detectors (vergence control). We also introduce a new learning paradigm for those networks. In particular, the first layer of the network will consist of different disparity detectors (far- and near-tuned detectors) that have been trained on image statistics. In this way, we are directly using biologically-plausible binocular neurons, rather than starting from binocular imagery and letting the development of binocularity depend on the outcome of learning (which does not necessary lead to biologically plausible detectors). Furthermore, we introduce an additional learning constraint in that only a limited set of detectors can be selected by the next layer, thus forcing the learning procedure to come up with its own optimal, yet sparse set of near- and far detectors. Sparse learning has not yet been fully applied to convolutional network architectures.

Main Partner : K.U.Leuven (M. Van Hulle)
Related WP(s) : WP2

- ☐ **STEREOSCOPIC OBJECT RECOGNITION (object identity)**

Baseline Purposive visual perception relies on a fragment-based representation. With respect to an object identity several cues are simultaneously represented and used. In Computer Vision the information of depth is extracted from the disparity and typically represented in a separate depth map (Büker et al., 2001) used for action in space or used for a 3D reconstruction (Sumi et al., 2002) but rarely to determine object identity. Humans fuse the two stereo images into a percept that is qualitatively very different from the two monocular components (Read, 2004). Inputs from left and right eyes are combined in primary visual cortex (V1), where many cells are tuned for binocular disparity. Complex cells in V1 show a tuning to a one preferred disparity. However, V1 is not the source of stereoscopic depth perception. V1 rather provides local estimates of absolute disparity. V2 neurons begin to show sensitivity to relative disparity and seem to be selective for disparity-defined edges (von der Heydt et al., 2000; Qiu & von der Heydt, 2005). Neurons in V4 and area MT are sensitive to the orientation of disparity defined planes (Hinkle & Connor, 2005; DeAngelis, 2000). The observed disparity selectivity appears to be relevant for object recognition (Yoshiyama et al., 2004). The Parietal Reach Region (PRR) is involved in planning reaches, and integrates head and eye position signals to encode the next intended reach in eye centered (or retinal) coordinates. Details how this is done are unknown but depth vision appears to modulate PRR neurons. Concluding, disparity is a factor that influences perception and contributes to our rich appearance of a 3D world in a number of different ways.

Progress We will take an approach to learn these distributed representations from the statistics of the stereo camera images. Typically, a combination of linear filtering and redundancy reduction based on second or higher order statistics is used for learning. The principal component analysis (PCA), independent component analysis (ICA) and independent factor analysis or sparse coding (IFA) have been successfully used to learn a set of basis functions (Hancock et al., 1992, Harpur & Prager, 1996, Olshausen & Field, 1996, Bell & Sejnowski, 1997). Especially ICA and IFA let the receptive fields converge to edge-filters, which exhibit similar properties as V1 cells. Despite this success there remain several open questions. From our point of view, one outstanding issue is the generalization of present approaches to higher levels of visual processing. Yet, there are only a few attempts to extend learning to model higher areas of visual processing. We have recently developed a learning algorithm to learn increasingly complex features (Hamker & Wiltschut, submitted to Network). This algorithm will be extended to operate also under the control of visual attention such that learning of objects can take place in cluttered scenes. The ability to learn such high-level representations in 3D scenes would lead to a major breakthrough in computer vision. In this project object-identity will be an important link to goal-directed action selection.

Main Partner : WWU (F. Hamker)
Related WP(s) : WP3

☐ **SELECTING VISUAL FRAGMENTS (visual attention)**

Baseline Loosely bound distributed visual fragments are required to be bound together on demand for the task at hand. Such binding processes can be well described by concepts of visual attention. Most computer vision approaches to object recognition and attention reduce the concept of attention to the idea of selecting a part of the image to speed up processing. Other approaches follow the classical approach and assume that an object is selected by an attentional focus, processed preferably and then recognized. Typically a saliency map is used to indicate a location of interest (Milanese, 1993, Itti & Koch, 2000). Walther & Koch (2006) combined the saliency map approach with a hierarchical model of object recognition by performing an additional segmentation around the location of interest as determined by the saliency and then used this segmentation to generate of focus of attention to gate the content in the object recognition module. While this improves the recognition performance in the presence of more than one object, the approach is severely limited, since the performance critically depends on the goodness of segmentation. We will argue that this is a chicken-egg problem, while the segmentation needs object information, but object information cannot be provided until the attentional focus is determined.

Progress We have developed a model where an oculomotor feedback signal tunes the receptive field structure (Hamker & Zirnsak, 2006). The tuning of the receptive fields, in addition to the overall gain enhancement, increases the processing resources around the location of the next fixation. Furthermore, a shrinkage of the receptive fields reduces the influence of clutter on the population response. Thus, tuning of the receptive field structure improves object recognition in a number of ways that are less error prone than a spotlight model which only gates visual processing. The dynamics of receptive field changes have only been demonstrated in terms of visual space. We will extend our work and will demonstrate that object recognition in cluttered 3D scenes will benefit from the tuning of receptive fields. Moreover, as with our previous models we will use feedback to implement a gain enhancement for intention matching coherent representations (Hamker, 2005), but here we will demonstrate this principle for the visual search of real world objects in 3D cluttered scenes.

Main Partner : WWU (F. Hamker)
Related WP(s) : WP3

☐ **CONSTRUCTION OF PERIPERSONAL SPACE ACROSS EYE MOVEMENTS**

Baseline When multiple visual fragments are fixated sequentially the visual system needs to track not only their content (see above considerations) but also their position in order to establish awareness of the location of fragments in peripersonal space. Spatial localization across eye movements can either be achieved by taking into account the relative location of a fragment with respect to visual reference objects (e.g. other fragments) or by using motor information about the direction and amplitude of the intervening saccade (Bridgeman, 1994). The visual references information from the saccade target is particularly important and is recruited immediately after the saccade is finished (Deubel et al., 1998; Lappe, 2000; Awater & Lappe, 2006). Motor information is particulary important when visual information is missing. We have recently reported that changes in motor parameters induced by a process known as saccadic adaptation (McLaughlin, 1967) influence the perceptual location of stimuli presented before saccade execution (Awater et al., 2005). This suggests a direct involvement of motor signals in the construction of perceptual space.

Progress We hypothesize that motor signals used for saccade execution are also used for the awareness of the spatial locations of the sequence of visual fragments and that, because these motor signals are plastic, perceptual awareness of peripersonal space is dynamic as well, in the sense that it is adpated to the motor parameters. Using saccadic adaption as an experimental paradigm we will study the shaping of perceptual space by sensorimotor contingencies. A motor-contingency of peripersonal sensory space allows a direct mapping between a fragment's location and the motor command to reach the fragment by an eye movement. Such a direct mapping eases motor planning and allows quicker goal-directed movements.

Main Partner : WWU (M. Lappe)
Related WP(s) : WP5, WP1

- **EVIDENCES OF JOINT VISUO-MOTOR FEATURES IN THE PARIETAL CORTEX**

Baseline The medial parieto-occipital cortex is located in the caudal part of the superior parietal lobule. It contains neurons responsive to visual stimuli (Galletti et al., 1996; Galletti et al., 1999), as well as cells modulated by somatosensory inputs, mainly from the upper limbs (Breveglieri et al., 2002), and arm movement-related neurons (Galletti et al., 1997; Fattori et al., 2001). Recently, it has been reported the existence in the medial parieto-occipital cortex of reach-related neurons able to code the direction of arm movement (Fattori et al., 2005). Gaze position effects in areas V6 and V6A of the medial parieto-occipital cortex have been extensively studied both on spontaneous activity (gaze fixation in darkness) and on visual responses (visual stimulation of the receptive field) (Galletti et al., 1995). About 48% of visual and 32% of non-visual neurons showed eye-position related activity in total darkness, while in about 60% of visual neurons the visual response was modulated by the direction of gaze. Another study from our group (Kutz et al., 2003) investigated the saccade-related activity in area V6A, showing that about 10% of tested neurons presented responses correlated with saccades (with fixation targets projected on a tangent screen). We also studied the effect of gaze direction on reach-related neural responses (Marzocchi et al., 2005). We found that eye position modulated 40% of the cells during fixation period, about 60% during forward arm movement, and 60% during holding phase. In most cases the modulation observed during arm movement was not correlated with the one during fixation period. In all the above described tasks the fixation point was varied on a tangent plane. We plan to extend previous studies, using fixation points in the 3D space.

Progress In conclusion, previous work demonstrated that the monkey medial parieto-occipital cortex is involved in elaborating eye movement signals both fixation and saccadic (Galletti et al., 1995; Kutz et al., 2003), that its neurons deal with the arm-movements performed in depth in the peripersonal space and that their activity is sensitive to the direction of reaching (Fattori et al., 2001; Fattori et al., 2005). Other very recent work demonstrates an involvement of the human medial parieto-occipital cortex in vergence eye movements and in reaching movements in different depths (Culham et al, 2007). There are all the premises to hypothesise the involvement of the monkey medial parieto-occipital cortex in vergence eye movements and in addressing arm movements in depth according to vergence signals. Considering the involvement of this region of brain in coding eye- and arm- movements, the medial parieto-occipital cortex could be a crucial node adressing ocular and arm motility in the 3D space. An open question in this field is the sensorimotor link in 3D space perception and awareness, specifically explored through active reaching movements. There are some data on the frame of reference of reaching during the planning of these movements. However, what is the frame of reference of these arm movements during and after action executions is still largely unexplored. We intend to tackle this problem with the proposed research, and we intend to do this in a very crucial node of the dorsal visual stream, at the boundary between passive sensory modalities and active eye- and arm-movements.

Main Partner : UNIBO (P. Fattori)
Related WP(s) : WP5, WP4

☐ **SHARED ATTENTION AND EYE MOVEMENTS**

Baseline Although attention can in principle be spatially directed without overt eye movements, most attention shifts in normal behavior are accompanied by eye movements (Henderson, 2003), and, in fact, saccadic eye movements mandate the allocation of attention to the saccade goal before the onset of the saccade (Deubel & Schneider, 1996). In goal-directed action, gaze shifts precede arm movements and object manipulation in a highly predictive manner (Johansson et al. 2001). The same coupling between eye movements and action goals occurs in human observers that monitor such action sequences (Flanagan & Johansson, 2003), demonstrating that the sequence of eye movements in goal directed actions is shared between actor and observer. Furthermore, perception of the gaze direction of others is known to automatically trigger the allocation of attention towards the gaze point of the other person (Driver et al. 1999, Ristic & Kingstone, 2005), demonstrating that overt gaze shifts induce shared spatial attention. These principles of human shared attention can be used by the model in order to monitor the attention of a human cooperation partner and predict his actions in a well-defined interaction scenario. The overt allocation of attention by the human partner can be monitored by eye tracking and used via knowledge of the sequence of eye- and hand-actions to predict upcoming hand movements. This knowledge can be used to (a) avoid interference of the robot arm with the arm of the human partner (a very important safety concern!) and (b) to begin the planning of the subsequent action of the robot. Indications for shared processing of information in a combined workspace are also seen in the selectivity of arm motion-related neurons in the parietal cortex: neurons that are active during arm movements of a monkey become sensitive also to view of the arm movements of a second monkey when both monkeys were able to reach for the same food object in the shared workspace (Fujii et al. 2007).

Progress In conclusion, it is clear that interacting humans can form a shared workspace and a shared control of attention within this workspace by using eye movements and knowledge about the contingencies of the shared task. Social cognition and mirroring networks are important for this. The respective contributions of task-knowledge and internal action planning versus observational information form the partner (i.e. the partners eye movements or other actions) are not clear, however. In human-human experiments we will clarify how shared workspace/shared attention is established and maintained between partners and how it depends on task settings and sensorimotor contingencies. This should allow to build a model of attention tracking based on action planning and eye movement observation to employ a robot with an understanding of human action in a particular task. Such an understanding is fundamental for successful interaction. The robot will become aware of the partner in the scene and of the partner's intentions and planned actions.

Main Partner : WWU (M. Lappe)
Related WP(s) : WP5, WP4

B1.3 S/T Methodology and associated work plan

B1.3.1 Overall strategy and general description

The approach proposed by EYESHOTS project follows the development of four Key Research Actions (KRAs), which have been identified according to the overall project's goal:
(1) Constructing visual perception of space by interactive stereopsis.
(2) Recursive modulation of perception across visual fragments.
(3) Visuospatial awareness and planning behavior
(4) Human behavior replicas by integration/interactive paradigms

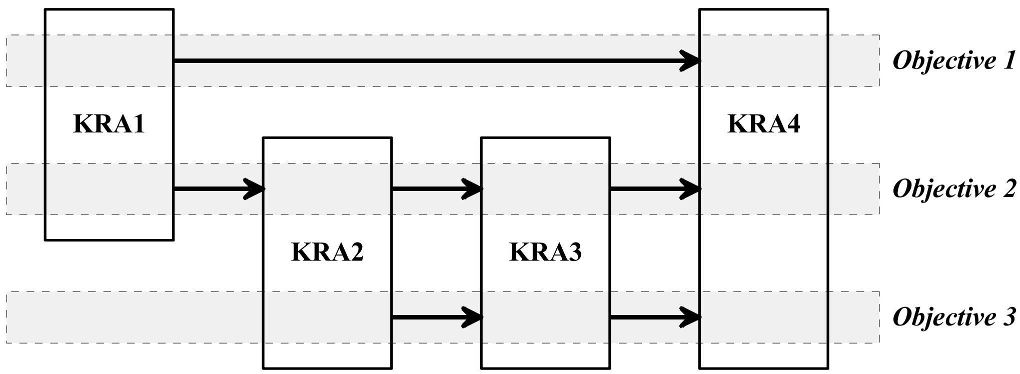


Figure 1: Development of the Key Research Actions

To focus the scientific problem, we introduce the following simplification assumptions: (1) objects in the scene are static (except the "arm", as part of the system, moving in the workspace). (2) The head is fixed. We rarely shift our gaze without moving our heads, nonetheless we introduced this restrictive assumption to isolate a particular function and to have a common setting with the planned neurophysiological experiments. (3) Only saccade-type eye-movements are considered. This will not hamper the behavioral tasks tackled in this research proposal. Indeed, there will rarely be pursuit movements to the motion of one's own arm. Usually, when reaching for a target, gaze is directed to the target and the motion of the arm is controlled from the visual motion of the arm in the peripheral visual field.

KRA 1: Constructing visual perception of space by interactive stereopsis.
Vision is the first source of information about the 3D space. The search for optimal visuomotor coordination to achieve robust and stable percepts does pose a major challenge. KRA1 will focus on:

- Anthropomorphic robotics eye system.
- Static and dynamic disparity detectors.
- Learning disparity detectors.
- Stereoscopic object recognition.

It provides an input to KRA2 contributing to the definition of a visual fragment of the observed scene.

KRA 2: Recursive modulation of perception across visual fragments
Definition of a strategy to achieve a global perception of the 3D spatial relations and relative 3D motion for controlling spatially directed actions (e.g., reaching), and, in general, visually-guided goal-directed movements in the whole peripersonal workspace. KRA2 will focus on:

- An attentional-based section of visual fragments.
- A construction of peripersonal space across eye movements.

The output of this KRA is not a real 3D reconstruction of the scene but a loose coupling among fragments providing awareness of the objects in the scene and information on how the system can interact with them. Only minimal descriptors for performing the required task will be recruited.

KRA 3: Visuospatial awareness and planning behavior
Here we address the problem of constructing an action-minded representation of the 3D space.
This will be achieved through a multisensory description of 3D space obtained through active ocular and arm movements.
KRA3 will contribute to:

1. The definition of joint representation signals of eyes and hand movements in a 3D extrinsic coordinate frame, on which to base the 3D location of a visual target with respect to a point on the body surface. We expect several advantages from such combined representation, respect to computing from signal intrinsic to each system such as version/vergence or joint-angle signals.
2. The definition of shared attention behaviour in common workspaces.

KRA 4: Human behavior replicas by integration/interactive paradigms.
This is a technical KRA concerned in the "translation" of the scientific achievements of KRAs1-3 into operative modules/subsystems (hw/sw robotic systems) characterized by perceptual/cognitive capabilities that emulate the human behaviour.
Two different subsystems will be considered:

1. Binocular eye system [an anthropomorphic robotic vision head (mobile eyes, fixed neck)]
2. Visually-based reaching system [a robotic arm and a stereo vision system]

This KRA will combine the components and the computational paradigms developed in KRAs1-3: the design, testing, and comparative analysis of performances of these systems are crucial for validating the approach.
Specifications for the testing activities will be provided by the development of the other KRAs: the experiments with robots will be a follow-up of the biological experiments. The psychophysical experiments will provide behavioural patterns (I/O specifications), while in-vivo experiments will provide architectural solutions (I/O + internal structural data). In order to meet the concurrent development stages of the research activities of the Workplan (described in Section B1.3.5) we will pursue a dynamic approach adapting the final integration plan in the course of the project to the actual achievements. Initially, tests will be conducted on the different subsystems.

These four Key Research Actions are addressed step by step in the workpackages and their sub-tasks described in Section B 1.3.5 of this document and summarized here below.

The Work Program consists of 8 Work-packages (WPs). There will be five scientific and technological WPs (WP1-5), and three WPs, planned for: training (WP8), dissemination and exploitation of the project's results (WP7), and for general project coordination and management (WP6). WPs 6-8 shall not be described here, please see Section B2.

The workplan is organized to allow the concurrent development of the S&T activities:

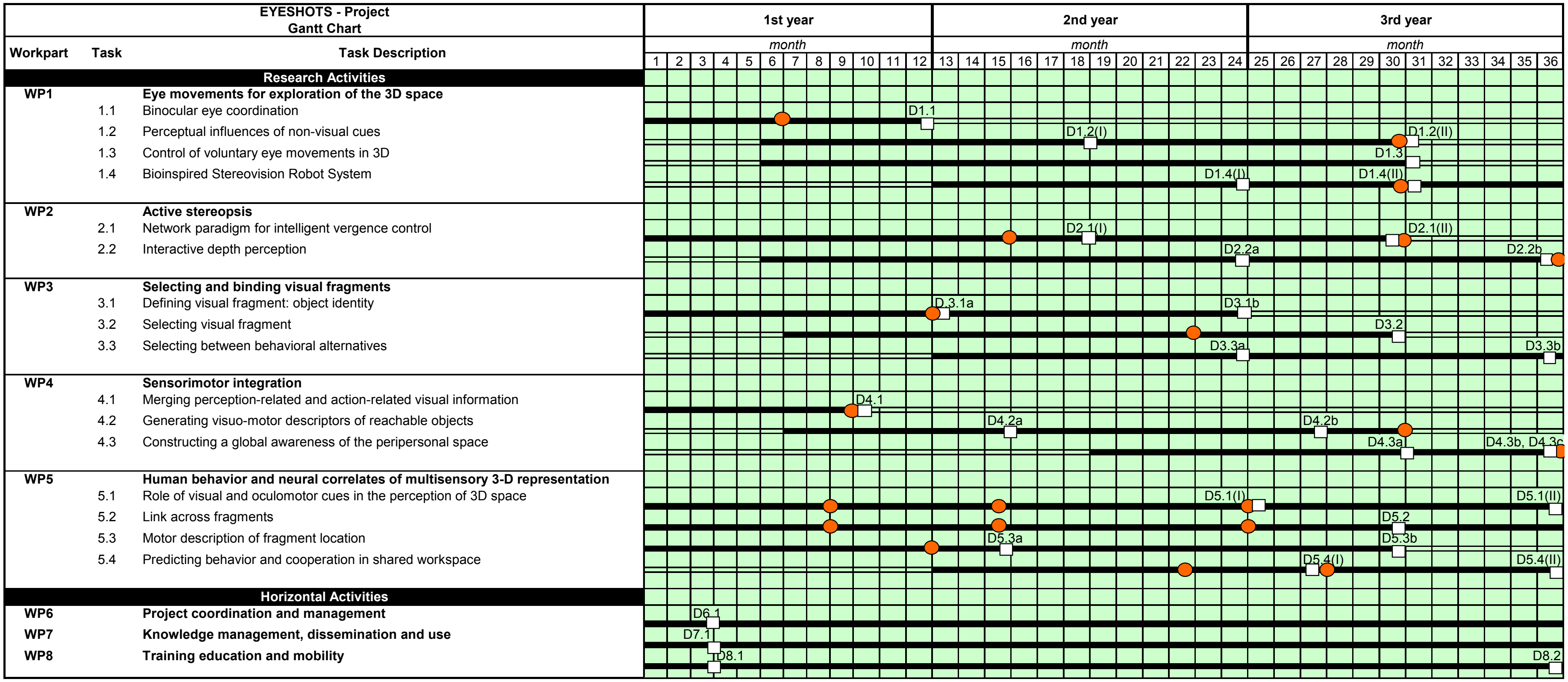
- biomimetic control of eye movements (see WP1)
- interactive stereopsis (see WP2)
- selection and integration of visual information (see WP3)
- visuomotor integration of perception (see WP4).

In order to ensure coordination and maintain the appropriate interaction among the partners various internal workshops are scheduled (see Section B2.1.2 and WP7).
Furthermore, specific experimental activities (see WP5), provide the architectural guidelines for the organization of perceptual behavior.

Specific milestones (see Section B1.3.7) have been set to assess the level of integration of software and hardware, components forming the robotic platforms that will be used for the tests.

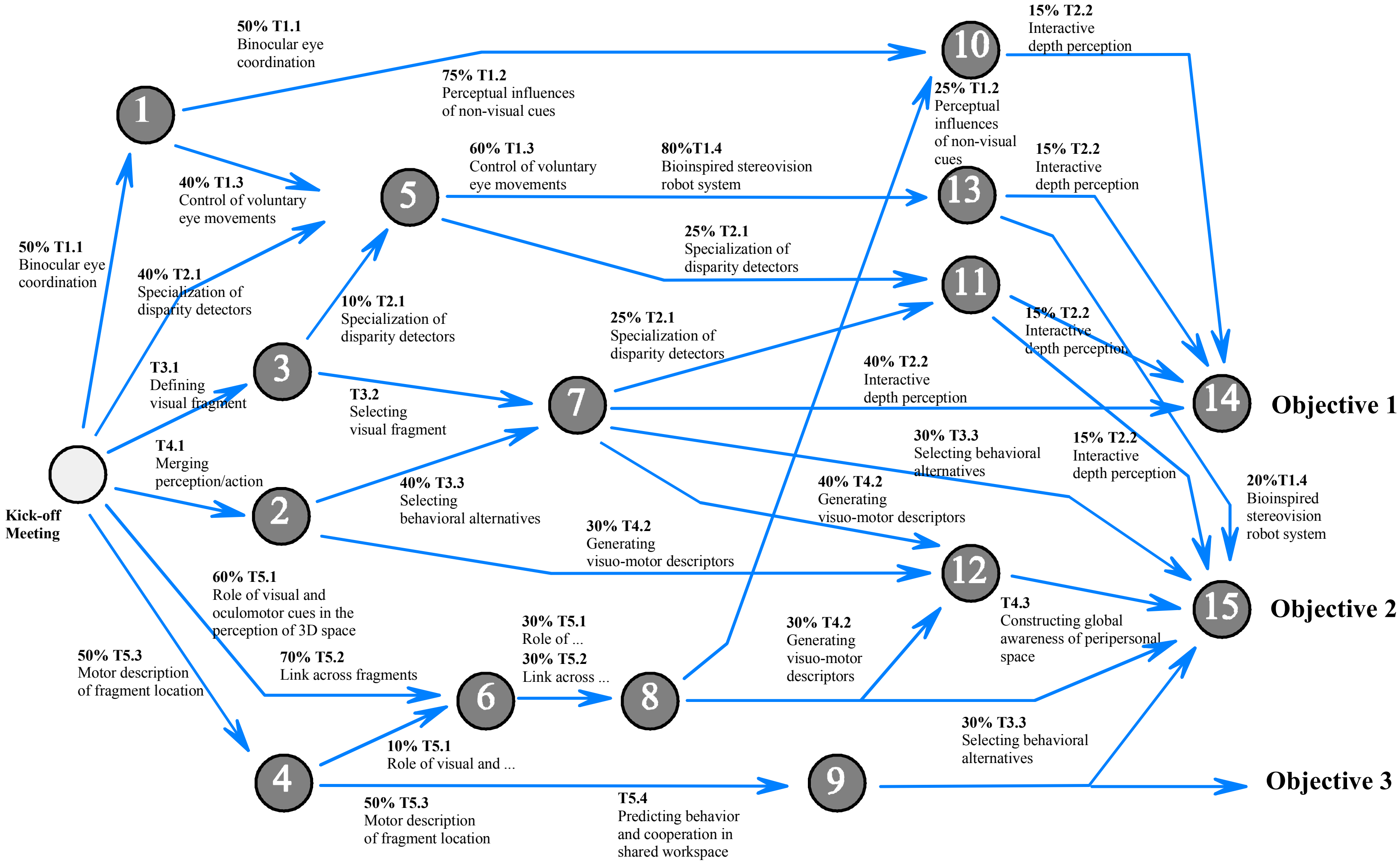
B1.3.2 Timing of work packages and their components:

GANTT chart (circles refer to Milestones, squares refer to Deliverables)



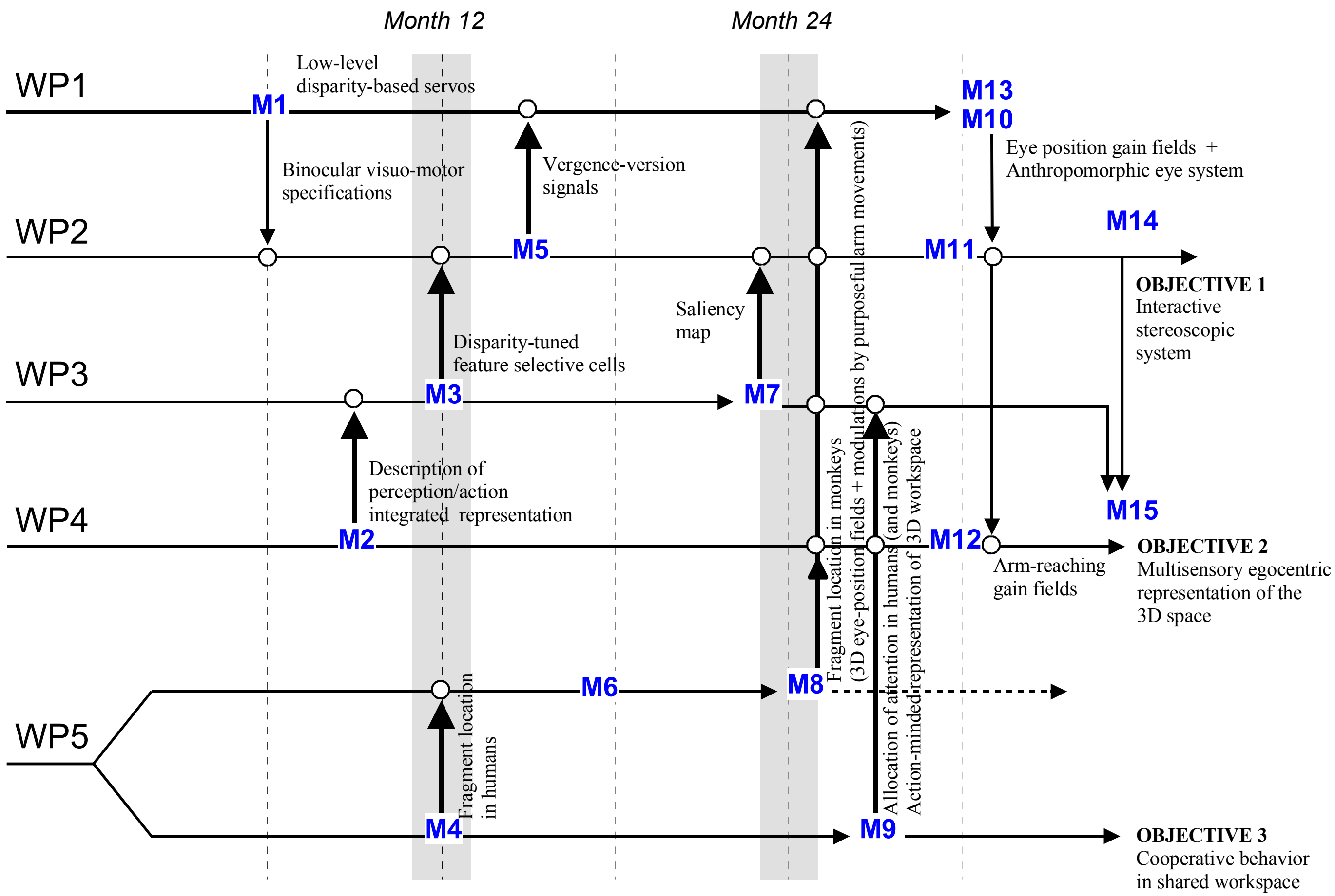
Pert Diagram

Graphical presentation of the project's components, showing their interdependencies. Tasks are represented on the arcs and Milestones on the nodes. The percentage associated to each arc relates to the total *effort* of the indicated Task.



Interdependencies among Work packages

The diagram shows the major transition and decision points. Major results transfer needs to take place with Milestones 2-5 as well as 7-9 (thick arrows). For each Workpart, the diagram indicates the expected achievements in correspondence to the major decision points. Time points 12 and 24 are given at the top.



B1.3.3 Work package list / overview

Work package List

Work package No	Work package title	Type of activity[2]	Lead beneficiary No[3]	Person-months	Start month	End month
WP1	Eye movements for exploration of the 3D space	RTD	1	63.5	1	36
WP2	Active stereopsis	RTD	5	64	1	36
WP3	Selecting and binding visual fragments	RTD	2	66	1	36
WP4	Sensorimotor integration	RTD	4	68	1	36
WP5	Human behavior and neural correlates of multisensory 3D representation	RTD	3	127	1	36
WP6	Project coordination and management	MGT	1	3+12	1	36
WP7	Knowledge Management, Dissemination and Exchanges with similar projects.	RTD	1	6	1	36
WP8	Training, Education and Mobility	RTD	1	6	1	36
	TOTAL			415.5		

Table 1: Workpackage list for the full duration of the project

2 'Types of activities' per WP (only if applicable for the chosen funding scheme – must correspond to the GPF Forms):
RTD = Research and technological development including scientific coordination applicable for collaborative projects and NoEs
DEM = Demonstration - applicable for collaborative projects
OTHER = Other activities (including management) applicable for collaborative projects, NoEs, and CSA
MGT = Management of the consortium - applicable for all funding schemes
COORD = Coordination activities – applicable only for CAs
SUPP = Support activities – applicable only for SAs

3 Number of the beneficiary leading the work in this work package.

B1.3.4 Deliverables list

List of Deliverables – to be submitted for review to EC

Del. no.	Deliverable name	WP no.	Lead beneficiary	Estimated indicative person-months	Nature [4]	Dissemination level[5]	Delivery date (proj. month)
D1.1	Binocular eye coordination and its role in depth vision	1	UG	9	R	PU	12
D1.2	Non-visual depth cues and their influence on perception	1	UG	13.5	R	PU	18,30
D1.3	Control of voluntary transfer of fixations to new depth planes	1	UG	18	O	PU[6]	30
D1.4	Bioinspired Stereovision Robot System	1	UG	23	P	PP	24,30
D2.1	Convolutional network for vergence control	2	K.U.Leuven	33	R	PU	18,30
D2.2a	Algorithm for 3D scene description through interactive visual stereopsis adaptation using a conventional binocular vision platform	2	K.U.Leuven	13	O	PU [6]	24
D2.2b	Algorithm for 3D scene description through interactive visual stereopsis adaptation using the mechatronic system.	2	K.U.Leuven	18	D,R	PU [6]	36
D3.1a	Demonstration of learning disparity-tuned feature selective cells	3	WWU	12	O	PU	12
D3.1b	Demonstration of object selective cells at intermediate complexity showing properties of disparity	3	WWU	12	R	PU	24
D3.2	Object-based top-down selection	3	WWU	22	O	PU [6]	30
D3.3a	Working memory model	3	WWU	10	O	PU [6]	24
D3.3b	Final, fully tested version of the Working Memory Model	3	WWU	10	R	PU [6]	36
D4.1	Description of integrated representation	4	UJI	10	R	PU	9

4 **R** = Report, **P** = Prototype, **D** = Demonstrator, **O** = Sofwtare module

5 **PU** = Public, **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).

6 These deliverables will be made publicly available after the corresponding material will have been accepted for publication in journals/conf.proceedings.

D4.2a	Computational approach to the integration problem	4	UJI	11	R	PU	15
D4.2b	Autonomous generation of object awareness	4	UJI	22	D	PU	27
D4.3a	How to build a global awareness of the peripersonal space	4	UJI	6	R	PU [6]	30
D4.3b	An embodied agent which learns to situate itself in the environment through active exploration	4	UJI	12	D	PU	36
D4.3c	Final robot head-eye/arm set-up featuring the robot eye system developed within WP1	4	UJI	7	D	PU	36
D5.1	Report on neural discharges in the medial parieto-occipital cortex	5	UNIBO	34	R	PU [6]	24,36
D5.2	Report on monkey eye movements and arm movements in the link across fragments	5	UNIBO	33	R	PU [6]	30
D5.3a	Report on the respective influence of motor and visual parameters on fragment location obtained from saccade adaptation data on humans	5	WWU	16	R	PU [6]	15
D5.3b	Report on the respective influence of motor and visual parameters on fragment location obtained from saccade adaptation data on monkeys	5	UNIBO	15	R	PU [6]	30
D5.4	Report on cooperative behaviour in shared workspace	6	WWU	29	R	PU [6]	27,36
D6.1	Launch of the project web-site	6	UG	3	D	PU	3, updates
D7.1	Creation, composition and publication of web pages	7	UG	3	D	PU	3, updates
D8.1	Literature database	8	UG	2	D	PU	3
D8.2	A collection of one page student reports about cooperation work	8	UG	2	R	CO	36
			TOTAL	**398.5**			

B1.3.5 Work package descriptions

Work package number	WP-1		**Start date or starting event:**		Month 1
Work package title	Eye movements for exploration of the 3D space				
Activity type	RTD				
Participant id	**UG**	WWU	UNIBO	UJI	K.U. Leuven
Person-months per beneficiary	56	-	-	6	1.5

Objectives

This Workpackage is devoted to the study of ocular mechanics and oculomotor control, for both single eye and conjugate movements. The target is to investigate how mechanics of the eye plant affects the strategies implemented by the brain to drive typical biological motions ocular motions (including saccades and smooth pursuit). A second goal is the study of the geometric and kinematic effects of ocular motions on image flow, for supporting the estimation of 3D information from ocular motions. Finally, from the engineering point of view the major expected achievement is to develop a bio-inspired stereoscopic robot system capable to emulate the ocular motions to be used during the planned experimental tests.

Description of work

We plan to structure the work in the following tasks:

Task 1.1: *Binocular eye coordination.* [**UG** 8PM, UJI 1PM]
This task is focused on the study of the geometric and kinematic effects of ocular movements on depth vision in the case of single eye and conjugate binocular physiological motions. To this aim also other phenomena, observed in humans an primates, as the dynamic orientation of *Listing's plane* will be considered.
We will study how specific ocular motion strategies, possibly observed in humans, could affect the performance of vision algorithms designed to extract 3D information. Furthermore, since in principle conjugate movements provide also a geometric constraint for the estimation of depth, it is planned to consider both (relative) geometric and kinematic informations adopting specific motions, e.g. based on *switching* between salient features, to provide additional inputs to visual processing algorithm in order to obtain more robust estimate of depth.

Task 1.2: *Perceptual influences of non-visual cues.* [**UG** 12PM, K.U.Leuven 1.5PM]
Direct and indirect perceptual consequences of eye movements will be investigated. First, we will derive directly the relationships between the version and vergence angles and the locations of the fixated point in the 3D space, as well maps of binocular correspondences (horizontal and vertical retinal disparities) around the fixation point. Second, we will introduce specific mechanisms (e.g., gain fields) to modulate the responses of the disparity detectors of the visual system (developed in WP2) on the basis of the vergence and version signals (related to a point on the extended horopter surface). These modulations are expected to influence perception at a more global level, being integrated across the whole image (different eccentricities) and across different depth planes. The results of this task will be taken into consideration in the design flow of WP2, by including the necessary primitives in the visual circuits to exploit the different modulating mechanisms successfully developed and validated in this workpackage. This task also include testing the efficacy/efficiency of neural coding of egocentric distance (e.g., amplitude modulation of the model's cell responses vs. 'compensatory' shift of their disparity tuning, when tested at different fixation distances).

Task 1.3: *Control of voluntary eye movements in 3D.* [**UG** 18PM]
Conjugate saccades are under voluntary control and bring the gaze onto an object of interest. Horizontal vergence, also, is under voluntary control for the same reason. Version and vergence form a unitary search

mechanism for exploring the 3D layout of a scene. This task is focused on the study of the interplay existing between the mechanics of the eye plant and the strategies implemented by the brain to drive typical biological ocular motions. The goal is to understand the control mechanisms adopted by the brain to coordinate the action of the extra-ocular muscles for the single eye and for conjugate ocular movements.The approach proposed to tackle this problem is based on the extension of the eye model proposed in (Cannata and Maggiali, 2006) where it is shown that physiologically correct saccadic motions could be generated by any action of the four recti extra-ocular muscles. This implies that the action of the extra-ocular muscles is only related to the target ocular motion (and not to the way the eye moves), so it can be expected the existence of visuo-motor transformations mapping image related information into motor commands. Furthermore, given the differences among the basic ocular motions, a precise formulation of the tasks (in robotic sense), behind them, could represent also a mean to understand the kind of information required for their implementation.

Task 1.4: *Bioinspired Stereovision Robot System.* [**UG** 18PM, UJI 5PM]
This task is focused on the design of a human sized and bioinspired binocular robot system capable to emulate basic ocular movements. The goal is to validate experimentally the models investigated in Task 1.2. The approach proposed to tackle this problem is to extend the design proposed in (Cannata and Maggiali, 2007a) for the development of a mechatronic robot eye system (i.e. a fully embedded tendon driven robot eye designed to study and emulate saccadic movements). The system development will mostly focus on study of design solutions required to implement a human sized eye hosting a miniature camera. Finally, the investigation for a possible design solution featuring 6 extra-ocular muscles to compensate for torsional ocular components will be considered.
The target mechatronic system will consist of a stereo head-eye system formed by a supporting platform and two tendon driven robot eyes. Each eye will host a miniature camera to provide feedback to the image processing and motion control modules. The size of the whole system is expected to be close to that of an average human, while the kinematics of the eye will be specified to emulate human ocular motions as saccades and smooth pursuit. The dynamic performance of the eye will be limited by the technological solutions adopted[7] and performing a reasonable trade-off between the size specifications defined above and the speed/acceleration limits (i.e. *not too big vs. not too slow*).
The image processing and control hardware will be, in a first stage, based on standard computing hardware in order to make simpler the set-up integration and speed-up the software development. However, during the project the development of an embedded computer system for the motion control modules, and possibly for some of the vision processing modules will be investigated, in order to develop a self-standing device. The integration of the advanced head-eye system with the robotic arm will be pursued during the last part of the project, and exchange of knowledge is planned before the final prototype will be released. The development of an integrated head-eye-arm platform will be conditional on the success of the tests of the anthropomorphic robot vision head. In particular, due to the number of innovative issues expected to arise both at scientific and technical level during the development stage, priority will be given to the tests performed separately on the two systems and before the final integration. In any case, should that integration be technically unfeasible, we will try to minimize the differences between both systems. Though obviously some features of the eye system will not be available in the pan-vergence-tilt head (such as the eye torsional d.o.f.) the control model will try to be based on similar parameters as those of the eye system so as to facilitate the migration from the former to the latter and, in any case, improve its biological plausibility.
Comparisons will be made, by example, from the geometric and kinematic point of view, as these structural features of the anthropomorphic platform affect the system model, the control procedures and performance and possibly the visual processes as well (e.g., negligible influences of the viewing distance on retinal motion information due to eye movements, reduced computation time of stereo matching at reasonable mechanical costs).

[7] As the project will not address the investigation of non conventional actuation and sensing, according to the state-of-the-art the acutation of the system will be based on *commercial* DC motors.

Contingency Planning:

Risk 1:	The objective of Tasks 1.1-1.3 is a scientific advancement. Possible delays or technical obstacles could place limitations to the direct implementation of the results obtained during this task in the final demonstrator.
Contingency plan:	In case of need, existing visual servoing methods could be used to support the experimental testing of the results of the other packages.
Risk 2:	The objective of Task 1.4 is a technological activity. The robot design is based on the experience gained with the development of a large scale system. The major risks could be related with performance limits of the prototype.
Contingency plans:	If performance of the prototype were not satisfactory, scaling down of the requirements for experimental tests could be considered. As extreme solution an existing (commercial) eye-head robot could be used for testing the modules developed e.g. limiting the tests to horizontal ocular motions only, and in order to test the features of the modules developed in the other WPs.

Deliverables

D1.1 Binocular eye coordination and its role in depth vision. Technical report. (Month 12).

D1.2 Non-visual depth cues and their influence on perception. (Month 18, preliminary report, Month 30 final report).

D1.3 Voluntary transfer of fixations to new depth planes. Technical report. (Month 30).

D1.4 Bioinspired Stereovision Robot System. Robot Prototype. (Intermediate version Month 24; Final prototype Month 30).

Work package number	WP-2	**Start date or starting event:**	Month 1		
Work package title	Active stereopsis				
Activity type	RTD				
Participant id	UG	WWU	UNIBO	UJI	K.U.Leuven
Person-months per beneficiary	25	-	-	-	39

Objectives

This Workpackage is devoted to the specialization of disparity detectors at different levels in a hierarchical network architecture to see the effect of learning (higher-order disparity detectors) for the extraction of the binocular stereo and stereomotion features. A vergence motor strategy is learned that, combined with the sparse detectors, optimizes the quality and efficiency of the feature-extraction for the specific tasks (guided by the attention signal) considered here. In a second task, scanning strategies are developed that enable this learned architecture to accurately describe the 3D structure (or headcentric disparity) of a visual fragment. This strategy will be designed to be robust to the limited accuracy of the motor system.

Description of work

We plan to structure the work in the following tasks:

Task 2.1: *Network paradigm for intelligent vergence control.* [UG 13PM, **K.U.Leuven** 20PM]
The extraction of binocular features relies upon a harmonic (i.e., amplitude and phase) representation of the visual signal, operated by a set of “simple cell” units (S-cells). Such representation allows us to reconsider and analyse the flexibility and robustness of the multi-channel perceptual coding, adopted by the early stages of the mammalian visual cortex, for the “atomic” components of early vision. Stereo and stereomotion percepts are made available by layers of “complex cell” units (C-cells) which gathers S-cells outputs according to specific architectural schemes. More precisely:
S-cells provide a flexible joint representation of elementary spatiotemporal binocular features of the visual signal (e.g., 3D (x,y,t) band-pass Gabor-like channels with proper frequency and orientation bandwidths). At this level, the “totipotency” of the representation contains all the necessary basic components to differentiate into several classes of visual descriptors.
Complex cells (C-cells) shall actively eliminate sensitivity to a selected set of parameters, thus becoming specifically tuned to different static and dynamic disparities.
At each level, the punctual operations such as sums, squarings and divisions that yield the resulting percepts implement complex operations on the harmonic content of the visual signal. Though these computations can be supported by neuromorphic architectural resources organized as arrays of interacting nodes, the network does not give completely the way to a pure connectionist approach, rather it allows us to design an architecture in which, at every layer, the “connectionist paradigm” is steeped in “signal processing”. On this basis learning processes could develop the high degree of variability of the cell’s responses towards the specialization of disparity detectors for the control of vergence.

Assumption: Vergence learning is restricted to a single fragment, the size of the foveal region.

Paradigm: LeNet5 (LeCun et al., 1998) with a fixed input layer of initially ‘C-cell’ disparity detectors, later on enriched with the disparity-based feature detectors developed in WP3 (cf. deliverable D3.1a), see also Fig.2. The network outputs a position-invariant vergence (invariant over the visual fragment). Position invariance is important here to arrive at a single vergence value, optimized over the fragment. Although the precise mechanism of learning is not intended to be biological, the system does operate in a self-supervised manner. It learns from the sensory consequences of its own motor actions.

The system consists of two vergence strategies:

1. a *general purpose* strategy that aims to maximize the total activity of the retinocentric map of 'C-cell' disparity detectors in as few vergence steps as possible
   - starting from a fixation point and an initial vergence, we have a 'basic' disparity cells activity pattern
   - the network outputs a new vergence estimate that, when applied increases this activity pattern, which produces a new vergence estimate, etc., until convergence
2. a strategy *specific for object detection*
   - the top-down signal that binds the visual fragments based on attention/selection indicates which descriptors of the fragment are most conspicuous
   - the activity pattern of these descriptors is maximized by iterating on the vergence much as in the previous general purpose strategy

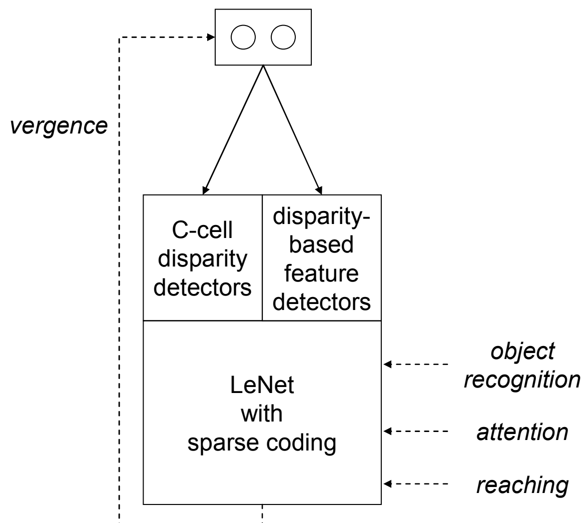


Figure 2: Intelligent vergence control of the anthropomorphic eye system (shown on top).

Reflex-like vergence. As a first step, the network will be trained off-line, using a set of binocular images obtained from an anthropomorphic mechatronic robot eye system available by UG. In this phase, eye movements are simulated. We will train the network given the current vergence as an additional input, and with the additional learning constraint that only a limited set of detectors can be selected, thus forcing the learning procedure to come up with its own optimal, yet sparse set of near- and far detectors. The network outputs are two-fold: 1) the network outputs a retinocentric disparity map that closely resembles one obtained using a large, uniform filter-bank of disparity detectors, and 2) the network should generate a vergence signal (direction + magnitude) with which the next vergence will be determined using a leaky integrator. As a second step, we will test the obtained vergence network in the robot eye system and assess its performance. We will also assess the specialization of the disparity detectors at different levels in the network architecture to see the effect of learning (higher-order disparity detectors).

Voluntary exploration. Here we will consider an "attention" signal coming from the saliency map developed in (WP3), for locating and recognizing objects in the scene, to direct the eyes. This imagery will then be used to train the vergence network, thus, conditioned on the presence of an attention signal. We will compare both vergence networks, i.e., for reflex-like vergence and vergence in the presence of voluntary exploration.

Relevance of dynamic stereoscopic information. The apparent object's motion-in-depth (MID) due to eye

movements yields to different image trajectories in the right and left monocular images of the corresponding point in the 3-D scene. From a control perspective, eye movements can be guided by the outputs of C-cells' populations specialized for phase-based 2nd-order motion features (temporal variations of disparity). It is worthy to note that, thanks to the metamerism of the first layer of the cortical architecture (S-cells), both the disparity and its temporal variations can be potentially available for each pixel at the same computational cost. By hierarchical combinations of the same signals provided by spatio-temporal frequency channels, cortical units can indeed actively eliminate sensitivity to a selected set of parameters, thus becoming specifically tuned to different features, such as disparity but not MID, or MID but not disparity.
From this perspective, different cortical circuits will be specifically designed to come up with specialized (and possibly adaptive) dynamic disparity detectors for their direct use in the control loop.

Task 2.2: *Interactive depth perception.* [UG 12PM, **K.U.Leuven** 19PM]
We will explore/develop scanning strategies to accurately describe the 3D structure (or headcentric disparity) of a visual fragment (with the ultimate goal that the entire fragment is processed by populations of disparity detectors). The fragment, by definition, spans a limited 3D region. Gain modulation (by eye position information) of the disparity information obtained during these saccades will allow for the construction of a higher-level representation that codes for absolute position. In this representation, 3D positional uncertainty will be implicitly encoded and fed-back to instruct and guide the scanning mechanism towards the next 3D saccade. Since the aim is to obtain a very accurate description and the precision of the disparity-detectors exceeds the precision of the motor system, a sub-goal will be to develop a mechanism that renders this transformation (from disparity to depth) robust to the limited accuracy of the motor system. Specifically, the mapping from disparity and eye position to 3D depth will be made insensitive to the effects (on disparity) of small vergence/version errors (cf. Van Ee and Erkelens, 1996). Also, vertical disparities (as derived e.g. from the reduced correlation of the horizontal-disparity sensors) will be exploited to obtain more accurate eye position information (Read and Cumming, 2006).

Contingency planning:

Task 2.1: If the sparse coding turns out to be a goal that is difficult to achieve, we can resort to a uniform (dense) coding. If the attentional signal for controlling voluntary exploration is lacking or difficult to obtain, we will resort to eye movements or pre-labeled objects and loci in the imagery.
Task 2.2: If the mapping turned out no to be robust, then we will resort to more classic algorithms from computer vision. Initially, the system will be built using a conventional binocular vision platform (D2.2a at month 24). We will use representations that are general enough to allow for the increased degrees of freedom of the anthropomorphic system. The results obtained in this first stage will provide a baseline for measuring the performance of the mechatronic system (D2.2b at month 36).

There are several arguments why a more complex paradigm is needed for vergence control:

1. Vergence control could be done by a simple servoing mechanism (by using disparity detectors to generate an error signal that needs to be minimized), but this would not go beyond reflex-like behavior. Indeed, we wish to support voluntary explorations of the peripersonal space which, eventually, should enable the system to learn to see in 3D through coordinated eye movements.
2. Vergence has to be decided for the whole visual fragment (foveal region), which could contain a small object (fixation point on the centroid) or a part of a larger object (fixation on the edges) or other high contrast regions. This calls for an optimized choice of the vergence so that the relative depths in the fragment fall as much as possible in the range of the available disparity detectors. (Note that it also implies position invariance of the vergence model.)
3. Besides 'C-cell' disparity detectors, also disparity-based feature detectors (developed for detecting and recognizing objects) serve as input for our vergence control. These detectors will be developed based on image statistics (WP3) and can take complicated forms. Due to the complexity of their

response, the coupling with vergence is nontrivial, and hence a sufficiently complex network architecture is required.

4. Vergence will be made dependent on attentional signals and other feedback signals, visual and/or motor, from which we also are not able to directly infer the vergence angle. But these signals can be used for influencing the choice of the vergence.

The selection of the network paradigm:

1. The LeNet5 model has a biologically-inspired architecture (it uses simple and complex cell layers), which it shares with the Riesenhuber and Poggio model, and which was originally proposed by Fukushima. It is now widely accepted as a model of object recognition for the ventral pathway (V1,V2,V4, IT).
2. The LeNet5 model uses supervised learning at every stage; the Riesenhuber and Poggio model, and also Hamker's model (this proposal) use unsupervised learning except for the last stage which is supervised. We have selected the LeNet paradigm for this application since it is expected to have a much better performance being such a network optimized at every level of the hierarchy. LeNet5 has been shown to be a powerful paradigm not only for object recognition, also using binocular images, but also for robot navigation and non-linear dimensionality reduction of images.
3. Our model is inspired by LeNet, but differs from it in that (among other differences) the input layer is not learnt (it consists of disparity detectors, later on including the binocular feature detectors developed in WP3).

Deliverables

D2.1 Convolutional network for vergence control. Technical report. (Month 18, preliminary report, Month 30 final report)
D2.2a Algorithm for 3D scene description through interactive visual stereopsis adaptation using a conventional binocular vision platform. Software module. (Month 24)
D2.2b Algorithm for 3D scene description through interactive visual stereopsis adaptation using the mechatronic system. Technical report and Demonstrator. (Month 36)

We will use benchmark data (known ground truth) and imagery obtained from scenes used in the project (such as a cluttered desk). In the first case, performance will be quantified by using the ground truth, in the second case, in terms of recognition performance. In both cases the proposed active system will be compared to a passive system in terms of accuracy using the same computational resources, or resources required to achieve the same accuracy. It is expected that the size of the filter bank required to achieve the same detail with a passive system will be larger.

Work package number	WP-3	**Start date or starting event:**		Month 1	
Work package title	Selecting and binding visual fragments				
Activity type	RTD				
Participant id	UG	WWU	UNIBO	UJI	K.U.Leuven
Person-months per beneficiary	6	58	-	-	2

Objectives

This Workpackage is devoted to develop novel concepts of selecting and binding within a fragmented 3D scene representation. One of those fragments is object identity. Object identity will be obtained from a bi-directional, hierarchical representation of learned feature detectors. The development of the appropriate learning rules will be an essential part of this project, since the learned connections will be used for the selection of a fragment. Beyond object identity, a distributed representation requires to actively bind and represent the relevant visual fragments for the task at hand. We will develop a novel framework for the task relevant binding of fragments in a global workspace using reward-based learning.

Description of work

We plan to structure the work in the following tasks:

Task 3.1: *Defining visual fragment: object identity.* [UG 4PM, **WWU** 19PM, K.U.Leuven 1PM]
It is well known that object identity is determined along the pathway from V1, V2, V4 and IT. Classical models of biological object recognition use a cascade of filters in which the complexity of representation along with the receptive field size increases with increasing filter stage (Fukushima, 1980; Riesenhuber & Poggio, 1999). Algorithms of learning the computation along this pathway are still under debate (Simoncelli, 2003). We have recently developed a promising learning algorithm. It uses presynaptic inhibition and Hebbian learning to simultaneously learn feedforward and feedback weights. The weights converge to localized, oriented and bandpass filters similar as the ones found in V1. In this task we will use the information of both cameras, or initially from pre-computed stereo images, such that the model neurons can tune themselves to disparity according to the statistics of the visual scene. Since the learning algorithm is not limited to early visual areas we apply learning at different levels of the hierarchy to develop tuned receptive fields with increasing selectivity to complex features and disparity cues. We expect that this learned representation provides a solid basis to determine object identity since it reflects the properties of the visual input within a sparsely coded framework simultaneously at different levels of complexity. As a first step, the network(s) could be trained using off-line binocular image sequences obtained by a currently available binocular head in UG.

Task 3.2: *Selecting visual fragment.* [UG 2PM, **WWU** 19PM, K.U.Leuven 1PM]
This workpart is devoted to develop a strategy to bind and select the distributed visual fragments on the basis of the momentarily existing task. The selection/attention process will be dynamic and task-contingent. We will extend our concept of a population-based inference approach (Hamker, 2005). It provides a framework to integrate feedback from different sources, which has lead to the prediction that phenomena of attention could arise by interactions between different brain areas (Hamker, 2005). As a result of these interactions, the computation in one representation immediately influences the computation in another and modifies in parallel the conspicuity of the processed and coherent visual fragments, such that the conspicuity of each descriptor represents the accumulated evidence. The binding process will operate continuously, but it can roughly be illustrated by two processes. One operates in parallel over all fragments and increases the conspicuity of those that are relevant for the task at hand, independent of their location in the visual scene. This provides salient locations which can be used for eye movement selection in WP1 and WP2. The other is linked to action plans, as developed in WP 1 for eye movements, and binds those fragments together, which

are consistent with the action plan, typically by their location in the visual scene.
The final model should be able to localize a particular target object in a visual scene. This will be evaluated by a benchmark involving a number of different scenes and different target objects. The reference model to which our model will be compared with is a standard, bottom-up saliency model.

Task 3.3: *Selecting between behavioral alternatives.* [**WWU** 20PM]
Vision requires high-level cognitive control in form of visual-visual and visual-reward associations, specifically when vision is embedded into a task that requires to interact with the environment. The representation of fragments are coordinated by the thalamus and basal ganglia in order to resolve conflict between competing systems and to schedule tasks in time related to the idea of a central selection device (Redgrave et al., 1999). We suggest the cortex being more involved in learning specific correlations and exact computations, whereas the basal ganglia with the thalamus coordinates these activations by activating specific loops in time. Such a model is inspired by the discovery of largely independent basal ganglia-thalamocortical circuits (Alexander et al., 1990), which build specific topographically organized loops (Middleton & Strick, 2002). Within each circuit, specific cortical areas send excitatory projections to selected portions of the striatum, the input of the basal ganglia. Since the basal ganglia exert an inhibitory influence over the thalamus the activation of the ganglia-thalamocortical circuits requires a disinhibition in time. This task coordination in the basal ganglia will be learned based on stimulus-reward associations.

Almost every task requires to hold previously visible information or the recall of memory as context information, but especially when the sharing of workspace is required. Such properties of cognitive vision systems are often referred to establishing a global workspace for specific task such as planning and action control. Models for establishing a global workspace are still rare (Van der Velde & de Kamps). It has been generally suggested that this is persistent activity requires the release of dopamine. However, the dopamine system is not specific enough to schedule events in time. We propose that the dopamine system is an arousal system that induces competition and supports short-term memory of salient stimuli by increasing the efficiency of recurrent interactions, but over time the basal ganglia will learn to schedule a task by activating specific thalamocortical loops.
We use an actor-critic model of the basal ganglia (Suri et al., 2001) to allow the dynamic representation of context information for the task at hand. For example, if a visual stimulus is necessary for a forthcoming operation leading to reward, the basal ganglia will learn to support its active memorization by disinhibiting a representative perirhinal-thalamic loop. The actor learns to take over control of the reverberatory loops through the thalamus. The critic learns to associate events with rewards and thus learns to predict reward. It is part of the dopamine system which computes a reinforcement signal for learning and working memory.

Contingency planning:
Task 3.1: The risk attached to this task is limited. However, additional resources will be devoted to this task in case it is unsuccessful.
Task 3.2: If the learning of appropriate bi-directionally connected feature detectors (D3.2) is delayed, we will use solely simple color cues for determining the next target location of the eye movement.

Deliverables
D3.1a: Demonstration of learning disparity-tuned feature selective cells. Software module. (Month 12).
D3.1b: Demonstration of object selective cells at intermediate complexity showing properties of disparity. Technical report. (Month 24)
D3.2: Object-based top-down selection using learned bi-directional connections between feature detectors to localize the object of interest in a cluttered 3D scene. Software module. (Month 36)
D3.3a: A model of working memory that allows to activate context information for the task at hand based on the association of previous events leading to reward. We will use the learned feature responses (WP2) on real world scenes if the feature-detectors are available, otherwise artificial representations will be used. Software module. (Month 24).
D3.3b: Final, fully tested version of the Working Memory Model. Technical report. (Month 36).

Work package number	WP-4	**Start date or starting event:**		Month 1	
Work package title	Sensorimotor integration				
Activity type	RTD				
Participant id	UG	WWU	UNIBO	UJI	K.U.Leuven
Person-months per beneficiary	10	-	16	41	1

Objectives

This workpackage is devoted to define (task 1) and generate (task 2) an action-perception integrated representation of objects in the peripersonal space in a dynamical way. Such integrated representation is achieved through the practical interaction of the robotic system with the environment, using both visual input and proprioceptual data concerning eye and arm movements. Task 3 is aimed at achieving a global awareness of the peripersonal space which allows to simultaneously learn to reach towards different visual targets, demonstrate binding capabilities through active exploration and build an egocentric, 3D "visuomotor map" of the environment.

Description of work

We plan to structure the work in the following tasks:

Task 4.1: *Merging perception-related and action-related visual information.* [UG 1PM, **UJI** 6PM, UBO 3PM]

Although the two streams of the human visual cortex evolved for different purposes, being the ventral stream dedicated to perceptual vision, and the dorsal stream to action-oriented vision, they need to collaborate in order to allow proper interaction of human beings with the world. We conceived a modelling framework of such action-perception integration to be applied on a robotic setup (Chinellato et al., 2006). The design of the different brain areas has been performed taking into account not only biological plausibility, but also practical issues related to engineering constraints. The final purpose is to mimic the coordination between sensory, associative and motor cortex of the human brain in vision-based grasping actions, through the coordination between the two visual streams of the human cortex. We plan to extend our model including movements of the proximal joints, such as in reaching actions. This implies a less focused attention on a single item of our world, but a more distributed and complete awareness of the whole environment (Cisek and Kalaska, 2005). The main goal of this task is thus to obtain an integrated representation for describing objects belonging to the peripersonal space. Such representation has to include both cognitive knowledge (identity, material, meaning of the object, ...) and visuomotor data (mainly affordances and all data related to hand-object interaction).

The goal will require our expertise in both the computational neuroscience of vision (Chinellato & del Pobil 2007) and biologically inspired vision-based robotic grasping (Chinellato et al., 2005; Chinellato et al., 2006b).

Task 4.2: *Generating visuo-motor descriptors of reachable objects.* [UG 4PM, UNIBO 8PM, **UJI** 20PM, K.U.Leuven 1PM]

Once the integrated representation of task 4.1 is available, the ambitious step of integrating separate but related visuomotor activities can take place. The goal of this task is to generate the above representation in a dynamical way, through the practical interaction of the robotic system with the environment, and using both visual input and proprioceptual data concerning eye and arm movements. The main principle to follow is that arm reaching direction and gaze direction modulate each other through gain fields (Salinas & Thier, 2000). On the one hand, these mutual modulations facilitate a potential action towards a focus of optic attention. On the other hand, the activity of approaching a certain goal object increases the drawing of attention toward it,

and thus the creation of perceptual awareness regarding the object.
For what concerns the practical implementation of such principles, radial basis function networks have shown to be biologically plausible and computationally effective in coordinate transformations and simultaneous multisensory representations (Pouget & Snyder, 2000; Pouget et al., 2002).

Following this approach, the robotic arm should naturally achieve very good open-loop reaching capabilities towards known targets. As it is often the case with humans, visual feedback, and thus a closed-loop control, would be necessary only towards moving targets or in the case of finer movements (as e.g. manipulation, or reaching in presence of obstacles). Such control can be achieved through 3D visual servoing, a technique in which our group has a long date experience (Cervera et al. 2003).

Task 4.3: *Constructing a global awareness of the peripersonal space.* [UG 5PM, UNIBO 5PM, **UJI** 15PM]
Considering a peripersonal space including different kinds of objects, a global awareness of the environment can be incrementally built through repetition of goal-directed actions toward perceived objects. Perception-action binding would results from the reinforcement of co-occurrent stimuli, even in a relatively crowded environment. In fact, the above mentioned gain modulation principle would bias the identification of competing stimuli towards the one that best explain the visuomotor pattern and proprioceptual eyes and arm data. In this way, the agent will simultaneously learn to reach towards different visual targets, achieve binding capabilities through active exploration and build an egocentric, 3D “visuomotor map” of the environment.
Test on functionality and performance under controlled situations. The system will be able to reach a limited set of artificial objects in a controlled real environment as the result of its perceptual awareness about an object and the modulation of the reaching and gaze directions.

The experimental activity will be supported by the development of an integrated head-eye/arm set-up. As a preliminary stage the system will be based on a standard robot arm coupled with a standard head-eye system. Starting from month 24 (corresponding to deliverable D1.4), the integration of the prototype developed within WP1 within the set-up will be pursued. A preliminary coordination activity has been planned [1 PhD student from UJI will spend a period in UG lab] at an early stage during WP1 activities (month 3 to month 9), to properly assess the system specifications useful for the final integration: in particular sw and hw interface requirements.

Contingency planning:

Risks:
Task 4.2: risks relative to the application of computational principles to a real robotic setup.
Contingency plan: Simplify the representation and suit it to the robot.
Task 4.3: delay in Task 4.1; problems in dealing with a real environment; final set-up not operational.
Contingency plan: Simplify the tasks; Adapt the environment; Use the preliminary set-up.

Deliverables

D4.1 Description of integrated representation. Technical Report. (Month 9).

D4.2a Computational approach to the integration problem. Technical Report. (Month 15).
D4.2b Autonomous generation of object awareness. Demonstrator. (Month 27).
D4.3a How to build a global awareness of the peripersonal space. Technical report. (Month 30).
D4.3b An embodied agent that learns to situate itself in the environment through active exploration. Demonstrator. (Month 36).

D4.3c Final robot head-eye/arm set-up featuring the robot eye system developed within WP1 (Month 36).

Work package number	WP-5	**Start date or starting event:**	Month 1	
Work package title	Human behavior and neural correlates of multisensory 3D representation			
Activity type	RTD			
Participant id	UG	WWU	UNIBO	UJI	K.U.Leuven
Person-months per beneficiary	2	43	78	3	1

Objectives

This Workpackage is devoted to the definition and the execution of specifically-designed psychophysical and neurophysiological experiments to study the human behavior of active perception and to find neural correlates of multisensory 3-D representation. The experiments are intended to provide architectural guidelines for the organization of perceptual interactions and will guide the production of artificial intelligent systems able to explore and interact with the 3D world.

Description of work

We plan to structure the work in the following tasks:

Task 5.1: *Role of visual and oculomotor cues in the perception of 3D space.* [UG 2PM, WWU 2PM, **UNIBO** 29PM, K.U.Leuven 1PM]
2D eye-position fields and visual responses gain fields have been found in the medial parietal occipital cortex, by testing how changes in the direction of gaze in the frontal plane influence the spontaneous activity of the cell in darkness, and the visual response of the cell (Galletti et al., 1995). Recently, it has been suggested that this region of the brain could be a crucial node for processing information related to stereopsis (Parker, Nature Rev, may 2007).
In this task UNIBO will extend in 3D the results already obtained in 2D, previously varying the fixation point on a tangent plane (Galletti et al., 1999a ; 1999b). The proposed tasks will be achieved by using a set of fixation points placed at different distances with respect to the monkey. We will study the influence of vergence eye movements in the perception of the 3D space. We will also study the binocular tuning of the visual cells: by measuring the cell's response to visual stimuli presented at various distances respect to the fixation point (i.e., at different disparities). 3D mapping of visual receptive fields: size and depth of visual receptive fields will be achieved using visual stimuli in depth. Visual stimulations will be performed using 3D objects or surfaces with different orientations in 3D.
We will furnish to WP1 and WP2 results on the "3D eye-position fields" and "gain fields" of neurons of medial parieto-occipital cortex (i.e., the 3D parts of the visual space -defined in terms of distance and direction- that influence (1) the spontaneous activity of the cell in darkness, and (2) the visual response of the cell).This characterization could provide indications on the role of non-visual cues (such as he eyes' version and vergence angles) on the perception of the 3D space as well as the role of visual cues in mastering the 3D peripersonal space.
The experiments will test different hypothesis on the structural basis for establishing a coherent internal reference frame for visuospatial representation and maintaining the integrity of this frame during eye movements. In general, the role of oculomotor information *vs* binocular visual cues (e.g., absolute and relative disparity) will be assessed. At a *functional* level, an important question is whether comparisons between absolute retinal disparities (i.e., relative disparity measurements) or variations of eye vergence signals are used to update the visuospatial representation during eye movements in depth. At an *architectural* level, the spatial distribution of the visual response modulation in retinotopic coordinates will allow us to qualify the nature of the mechanism by which visual cells encode a reference to the current fixation point. Two different hypotheses will be tested: (1) a multiplicative interaction (i.e., gain fields) depending on the absolute positions of the eyes; (2) an explicit coding/representation of global properties of the disparity fields

(elicited by the specific geometrical configuration of the eye system and of depth surfaces in real-world situations). The results will influence the computational resources on which the performance of the artificial system will be based (e.g., reducing the number of units in the population necessary to span the multidimensional representation space).
Moreover, all the results obtained from the experiments of Task 5.1 are important for the detailed implementation of the computational model that will be developed in the robotic application (see WP4). Taking as a starting point a previous model developed by UJI [Chinellato and del Pobil 2007], which addresses the problem of estimating distance, shape and orientation of simple objects through the combination of perspective and stereoscopic cues, more elaborated neural mechanisms can be reproduced exploiting the outcome of such experiments. More specifically, proprioceptual vergence data affects the modulation between concurrent cues in distance and pose estimation, and the new experiments can help clarify in which way. Similarly, the stereoscopic modules of the existing model can be improved by taking into account the new insights offered by the planned experiments regarding binocular tuning and the characteristics of stereoscopic receptive fields. The influence of monocular and binocular cues in different conditions will also be tested and included in the model.

Task 5.2: *Link across fragments.* [**UNIBO** 30PM, WWU 2PM, UJI 1PM]
Neurophysiological experiments will guide the production of artificial intelligent systems able to explore and interact with the 3D world. This will be achieved in this task through a multisensory representation of 3D space obtained through active ocular and arm movements. The basis of this approach is the conviction that a complete and practical cognition of visual space can be achieved only through active exploration of it. The natural effectors of natural cognition are the eyes and the arms. UNIBO will perform electrophysiological recordings in the awake monkey performing different tasks, all with the head restrained and freedom to perform arm movements as well as eye movements. Thus the effectors studied in this experimental approach will be saccadic eye movements and ballistic reaching movements used to explore and represent the 3D space.
Link across single visual fragments will be obtained by measuring covert attention toward different parts of the visual world while the monkey performs reaching movements in the three dimensional space. Reaching movements to visual objects will be evoked in different positions of the 3D peripersonal space. The monkey will move its hand from the body space to several positions of the peripersonal. Being the monkey freely looking, arm reaching movements will be typically preceded by saccadic eye movements towards the same position. Ocular and arm-reaching movements in the 3D space will be studied both during the planning phase and during the execution phase.
We will evoke reaching movements toward a foveated target and toward a peripheral target, by controlling the monkey focus of attention and thus addressing it toward different positions of the workspace.

In both Tasks 5.1 and 5.2 the behavioral performance of the monkey will be monitored by acquiring kinematic information on the monkey arm-movements, and also measuring reaction times and performances of eye movements and arm movements to the different behavioural events in the 3D space. Both tasks will be performed in conditions of stereo-vision and in conditions of monocular vision. This will allow us to furnish to the artificial system information on the strategies used by biological systems in conditions of stereopsis and in monocular conditions.

The results will be passed to WP4 (Task 4.1) in order to allow an extension of the UJI model to the realization of proximal arm movements, as those performed in reaching movements and (Task 4.2) contribute in generating visuo-motor descriptors of reachable objects. More precisely, focusing on the relation between vision and reaching movements, the experiments of Task 5.2 will provide valuable information on how fixation can affect the execution of ballistic movements, for example through gain fields, which can be measured and thus emulated in the computational model. The facilitation of the motor command due to covert attention in the artificial and the natural systems can thus be compared and analyzed.

Task 5.3: *Motor description of fragment location.* [**WWU** 19PM, UNIBO 11PM, UJI 1PM]
To study the role of motor parameters in fragment location WWU will investigate perceived peripersonal space in the saccade adaptation paradigm. In saccade adaptation, the target of a saccade is displaced during the saccade. Because vision is suppressed during a saccade, subjects do not see the displacement. When this procedure is applied consecutively subjects gradually adapt their saccade such that it directly reaches the displaced position. It is believed that adaptation takes place on the motor side of the saccade generation pathway but we have shown that saccade adaptation shifts the perceived position of objects around the saccade target location in the direction of the adaptation, hence that adapting motor space also adapts sensory space. We will use saccade adaptation to study the role of motor commands in fragment location. Fragments will be displayed while gaze of the participant is tracked. Saccades to a particular fragment location will be adapted by displacing this fragment whenever a saccade is initiated. After adaptation is achieved, perceived fragment location will be tested. This will be done by displaying a fragment in question only briefly before the adapted saccade and afterwards asking the subjects to report the location of the fragment with a pointing device. After setting the parameters of the task on saccade adaptation in the human, the same task will be used to train a monkey in order to test whether fragment location descriptors are motor or sensory. For this purpose, a collaboration between UNIBO and WWU is foreseen: 1 PhD student from WWU will spend a period in UNIBO lab in order to set up the stimulations and run the task and will participate in the analsis of experimental data.
As done in the human being, saccade adaptation will be used to study the role of motor commands in fragment location. Fragments will be displayed while tracking the gaze of the monkey. Saccades to a particular fragment location will be adapted by displacing this fragment whenever a saccade is initiated. After adaptation is achieved, perceived fragment location will be tested.

The results from Task 5.3 will be used in WP 4 for the generation of an action-minded representation of 3D workspace. More precisely, the saccadic mechanism studied in Task 5.3 can be reproduced in the artificial system by emulating saccade adaptation through a learning process (e.g. by reinforcement learning). Hypotheses will be formulated on the influence of motor and visual components on fragment location, and they will be tested on the robotic setup. As described in WP4, the link between motor commands and visual information would thus emerge by observing the co-occurrence of stimuli in different behavioural conditions. The guidelines for how to actively pursue such emergence through a purposeful saccade will be provided by the experiments of this task.

Task 5.4: *Predicting behaviour and cooperation in shared workspace.* [**WWU** 20PM, UNIBO 8PM, UJI 1PM]
The main purpose of this Task is to equip the model with an understanding of the sequence of allocation of attention, direction of gaze, and movement of the arm of a human cooperation partner. This involves the prediction of the partner's behavior by the model. WWU will study the above-mentioned aspects of human behavior in a constrained task setting with established psychophysical methods of attention measurement, and eye- and arm-tracking. The goal of this study is to achieve a model description of the human's behavior that can be used for prediction by the model. The results will be made available to WPs 2.2 and 3.2 for modeling interactive eye movement strategies, to WP 3.3 for modeling task decisions, and to WP4 for modeling arm movements.
In order to work towards the possibility to unravel the neural mechanisms of behavior prediction we will perform similar studies with monkey subjects in a collaboration between UNIBO and WWU: 1 PhD student from WWU will be responsible of setting the parameters and the measures of the task in order to interpret monkey intentions. This will need a period to be spent in UNIBO in order to collect the data necessary to use the model to predict behaviour.
In this work package WWU will furthermore study the cooperative behavior of two human subjects in the same task (both with eye tracking) in order to (a) learn about the expectations that a human actor in this task has about the cooperation partner and (b) learn about the modifications of the human's behavior in a cooperative setting compared to a single actor setting. Both of these informations will be very important for

the behavior of the model in WP-3 and WP-4. When the model is developed far enough, the cooperation behaviour between the human and the model in WP 4 will be studied. To do this, the eye tracker will be set up in the UJI lab to allow an interfacing with the robot. In this setup, the human subject will work in the same task as in the human-human interaction experiment but the human partner will be replaced by the robot. This approach will allow us to measure the performance of the model in the closed loop situation. If the human-robot interaction turns out to be problematic (for instance for safety reasons or because the subjects are more cautious and restrained than in interaction with other humans) we may envisage instead a test in which the behavior of the model will be simulated and presented to the subject in a virtual reality scenario.

Contingency planning:

Since this Work package is mostly experimental, problems arising may either be related to the experimental approach not working or the data being inconclusive. In either case we will modify the experimental design, possibly with the requirement to modify the target results. However, since the applicants have ample experience in psychophysical and physiological research methods, no problems are being foreseen.

Deliverables

D5.1. Report on neural discharges in the medial parieto-occipital cortex (tasks 5.1 and 5.2). Month 24, preliminary report; Month 36 final report.

D5.2. Report on monkey eye movements and arm movements in the link across fragments (task 5.2). Month 30. It directly relates to Milestone M8 (preliminary analysis of data available to modelers, at month 24)

D5.3a. Report on the respective influence of motor and visual parameters on fragment location obtained from saccade adaptation data on humans (task 5.3). Month 15.

D5.3b. Report on the respective influence of motor and visual parameters on fragment location obtained from saccade adaptation data on monkeys (task 5.3). Month 30. This deliverable directly relates to Milestone M9.ante (analysis of experimental data on the allocation of attention and behaviour prediction in monkeys, at month 18).

D5.4 Report on cooperative behavior in shared workspace (task 5.4). Month 27, preliminary report; Month 36 final report.

Work package number	WP-6	**Start date or starting event:**		Month 1	
Work package title	Project coordination and management				
Activity type	MGT				
Participant id	UG	WWU	UNIBO	UJI	K.U.Leuven
Person-months per beneficiary	3+12	-	-	-	-

Objectives

The main objective of this Workpart is to implement and maintain an effective administrative and management infrastructure of the overall EYESHOTS throughout the phases and management of project resources for their best usage towards the overall project objectives.

Details of project management have been given in Section B2; this WP will implement these structures. Lead group for this WP will be UG. UG will be supported in Technical Managing by K.U. Leuven (see Section B2.1.1).

Description of work

The following tasks will be carried out:

- Overall management and coordination of the project
- Preparation of all project management documentation for the EC, including financial reports,
- Ensure a clear, common project vision and overall requirements definition,
- System and subsystem level quality assurance,
- Ensuring on-time schedule of the project and request of approvals.
- Development and maintenance of the EYESHOTS project web-site (with a public section and a private section with restricted access to the consortium's members). Set-up and maintenance of *e*-Services for sharing and broadcasting documents and data**.**

Deliverables

D6.1 Launch of the project web-site (Month 3, regular updates).

Periodic project reports and other contractual deliverables (e.g., cost statements) are not included in the list of deliverables.

Work package number	WP-7	**Start date or starting event:**		Month 1	
Work package title	Knowledge Management, Dissemination and Use, Synergies with other projects				
Activity type	RTD				
Participant id	UG	WWU	UNIBO	UJI	K.U.Leuven
Person-months per beneficiary	1.5	1	1	1	1.5

Objectives

Dissemination: The EYESHOTS project explicitly recognises the importance of the Knowledge Management of our project results. The goal of dissemination activities is to make the project results known to the community of interested researchers and automation industry as one of the potential developers of next-generation robotic systems. The research community targeted includes areas of robotics and automation, machine learning, neuronal computation, neuroscience, perception and cognitive science, computer vision, image processing.

IPR handling: The objective of this set of tasks is to create attention and awareness for IPR in EYESHOTS and to file the necessary patents. IPR arrangements will be covered by internal agreements between the partners.

Synergies with other projects

Description of work

The work will be divided into three parts:

Task 7.1: Regular publication of research news, events, research results, and demos on the project web-site. The site will also host relevant test results and reports including links to research publications, presentations, and software demonstrations. The project web-pages serve as a means for continuous dissemination of information about the project for the public awareness as well as internally for the project participants.

Task 7.2: In coincidence with the annual project's meetings, to ensure a smooth set-up of an interdisciplinary consortium, brainstorm events - as open workshop sessions - will be organised and used to establish a common ground. In these sessions especially young researchers will be invited to present their current work related to the project.

Task 7.3: For external dissemination three major channels will be used:

1. All research will be published in the best scientific journals, such as IEEE Trans. on Robotics and Automation, Int. J. of Robotics Research, Int. J. of Computer Vision, Vision Research, Neural Networks, Cerebral cortex.
2. Participation to workshops, conferences and other forums and events, as appropriate, taking place at a national, European or international level which are relevant for the projects information dissemination activities. Representative conferences include: ICRA (Int. Conference on Robotics and Automation), ECCV (European Conference on Computer Vision, NIPS (Neural Information Processing Systems), ECVP (European Conference on Visual Perception), Society for Neuroscience).
3. EYESHOTS will set up a mailing-list which will be used instead of a Newsletter to disseminate results to interested parties.

Task 7.4: Synergies with other FP6 or FP7 projects: Representatives of the consortia of relevant projects will be invited to participate to the open workshop session organized in coincidence with the Kick-off meeting, to explore synergies and sharing of experience. Follow-up meetings will be foreseen, during the course of the project, in coincidence with major milestones, possibly as specific events in the framework of periodic thematic workshops organized by CogSys. Links with the following projects have been identified:

POP (**Perception on Purpose**) in the domains of:

- Human-machine interaction
- Multimodal integration of sensorial data
- Active stereopsis

Decisions-in-motion (Neural Decision-making in Motion) in the domains of:

- Control of voluntary eye movements
- Saccadic decision tasks and attention
- Ability of distinguishing the motion components that are passively induced by objects of the outer world from others that are self-induced by the motor system of the eyes.
- Robotic platform for binocular eye movements.

Robotcub in the domains of:

- Humanoid robotic platforms (i-Cub binocular head)
- Multisensory interactive exploration of the environment
- Human-robot interaction.

Deliverables

D7.1 Creation, composition and publication of web pages (Month 3 + major regular updates at Months 12,24,36 + minor continuous update)

Some non-technical deliverables will be issued in the web-site or they will be part of the periodic progress reports.

Work package number	WP-8	**Start date or starting event:**		Month 1	
Work package title	Training, Education and Mobility				
Activity type	RTD				
Participant id	UG	WWU	UNIBO	UJI	K.U.Leuven
Person-months per beneficiary	1.5	1	1	1	1.5

Objectives

Training and Education: We realise that in this multi-disciplinary proposal, training and teaching will be of major importance. The objective of this work package is to make local education, training activities and knowledge of the partners accessible for the entire consortium and extend and combine these efforts wherever necessary. A successful exchange of knowledge will be a primary requirement for a successful collaboration. Intra consortium initiatives are in fact considered as a prerequisite to any external training initiative.

Mobility: The objective of this task is to foster the exchange of personnel and to promote collaboration at every level of the consortium, including student, junior researcher and professor level. The exchange will always be tied to a project activity at the hosted partner. A special budget will be reserved for this activity.

Description of work

Task 8.1 Due to the highly multi-disciplinary nature of EYESHOTS, one of the first tasks will be to compile an extensive bibliography list and source/access information of the basic and relevant literature from computer science, biocybernetics, sensing and motor control as well as learning that will provide a common basis for teaching and education of students. This will help to acquaint students with the terminology used in the different fields and thus ensure good communication across partners.

Task 8.2 All partners will ask their own students to prepare half-yearly a seminar not about their own work but instead about the work of one or two most closely related cooperation partners of EYESHOTS. Their own work should only be touched upon in highlighting the links.

Task 8.3 Extensive exchange of professors and researchers among the partner institutions is planned in this phase. They will include medium- and long-term visiting periods by young researchers (with a special focus on PhD students) as well as short-term visits of principal investigators.

Task 8.4 In order to deepen the mutual knowledge between all partners and also spread the scientific and technological insights achieved along the realization of the project, we will organize a Summer School which will be especially dedicated, although not exclusively reserved to, students and senior researchers of EYESHOTS. The Summer School will be organized within the frame of the well-known IURS Summer School of the UJI partner, this year at its 7th edition.

Deliverables

D8.1 Literature Database (Month 3)

D8.2 A collection of one-page student reports about the work of their cooperation partners and an assessment statement of their own supervisors assessing the depth of understanding of the cooperations (Month 36).

B1.3.6 Efforts for the full duration of the project

Project Effort Form 1 - Indicative efforts per beneficiary per WP

Project acronym: EYESHOTS

Workpackage	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	TOTAL per Beneficiary
UG	56	25	6	10	2	15	1.5	1.5	117
WWU	-	-	58	-	43	-	1	1	103
UNIBO	-	-	-	16	78	-	1	1	96
UJI	6	-	-	41	3	-	1	1	52
K.U.Leuven	1.5	39	2	1	1	-	1.5	1.5	47.5
TOTAL	63.5	64	66	68	127	15	6	6	415.5

Project Effort Form 2 - indicative efforts per activity type per beneficiary

Project acronym: EYESHOTS

Activity Type	UG	WWU	UNIBO	UJI	K.U. Leuven	TOTAL ACTIVITIES
RTD/Innovation activities						
WP1 - Eye movements for exploration of the 3D space	56	0	0	6	1.5	63.5
WP2 - Active stereopsis	25	0	0	0	39	64
WP3 - Selecting and binding visual fragments	6	58	0	0	2	66
WP4 - Human behavior and neural correlates of multisensory 3D representation	10	0	16	41	1	68
WP5 - Sensorimotor integration	2	43	78	3	1	127
WP7 -Knowledge Management, Dissemination, Synergies with other projects	1.5	1	1	1	1.5	6
WP8 - Training, Education and Mobility	1.5	1	1	1	1.5	6
Total 'research'	**102**	**103**	**96**	**52**	**47.5**	**400.5**
Demonstration activities						
-	-	-	-	-	-	-
Total 'demonstration'	-	-	-	-	-	-
Consortium management activities						
WP6 - Project coordination and management	3+12	0	0	0	0	3+12
Total 'management'	**15**	**0**	**0**	**0**	**0**	**15**
Other activities						
-	-	-	-	-	-	-
Total 'other'	-	-	-	-	-	-
TOTAL BENEFICIARIES	**117**	**103**	**96**	**52**	**47.5**	**415.5**

Note: In the Consortium Management Activities the Person Months displayed by UG in the table is divided into: A+B, where A is "Person Months of part-time personnel for secretarial and technical support, to be hired specifically for EYESHOTS" and B is "Person Months of permanent staff", including the project Coordinator.

B1.3.7 List of milestones and planning of reviews

List and schedule of milestones

Mile stone no.	Milestone name	WPs no's.	Lead beneficiary	Delivery date	*Comments: Success criteria/expected functionality and means of verification*
M1	Low-level automatic servos based on primary disparity information	WP1	UG	6	**Expected functionality:** Structural mechanisms for binocular coordination of version and vergence movements. Strategies for visuomotor binocular control. **Means of verification:** Test on functionality and performance of conjugate eye movements between two assigned targets on different depth planes. Cooperation of vergence and version movements.
M2	Merging of action and perception	WP4	UJI	9	**Expected functionality:** Intermediate result. Integrated representation model for describing objects belonging to the peripersonal space. Towards a distributed and complete awareness of the 3D environment at arm-reaching distances. **Means of verification:** The model will describe bidirectional associations between visual and motor information to provide a (vision+behavior)-based description of the peripersonal space.
M3	Learning algorithm for bi-directionally connected disparity tuned feature-selective cells	WP3	WWU	12	**Expected functionality:** Learning should lead to multiple feature selective cells that show a tuning to different disparity values, similar as observed in real neurons.. **Means of verification:** The tuning must cover the spectrum of disparity values as determined by the input scenes to allow the usage of these cell responses at later stages.
M4	Experimental data on fragment location in humans obtained	WP5	WWU	12	**Expected functionality:** Intermediate result. Psychophysical bases for respective influence of motor and visual parameters on fragment location obtained from saccade adaptation data (on humans). **Means of verification:** Hypothesis that saccade adaptation modifies perceived location of saccade goals confirmed by experiment.
M5	Convolutional network algorithm for sparse near/far coding	WP2	K.U. Leuven	15	**Expected functionality:** A sparse near/far binocular disparity representation will be generated in a biologically-plausible multi-layered architecture for vergence control. **Means of verification:** Test on network convergence to a single vergence value, optimised over the visual fragment. We will use benchmark data (known ground truth) and imagery obtained from scenes used in the project (such as a cluttered desk).
M6. ante	Experimental set-up	WP5	UNIBO	8	The visual stimulation system, the new eye-movement recording system and the device for reach-in-depth will be ready.
M6	Monkey training for neural recording completed	WP5	UNIBO	15	End of monkey training.

M7	Target location (saliency map) for the next eye-movement in goal-directed search	WP3	WWU	22	**Expected functionality:** The saliency map should indicate the location to be searched for a target object by an increased activity as compared to other values at locations of distractor objects. **Means of verification:** The validity of the approach will be demonstrated with different target objects. This model should be superior to a model without goal directed, top-down signals.
M8. ante	Experimental data on fragment location in monkeys obtained	WP5	UNIBO	22	**Expected functionality:** Intermediate result. Neurophysiological bases for respective influence of motor and visual parameters on fragment location in monkeys. **Means of verification:** Hypothesis that the basis for a coherent frame of reference for visuo-spatial representation is organized in retinotopic or spatial coordinates confirmed by experiment.
M8	End of single cell recording session.	WP5	UNIBO	24	**Expected functionality:** Intermediate result. Preliminary analysis of neuronal data available for EYESHOTS partners. **Means of verification:** Hypothesis that the purposesul arm movements in 3D space are organized in retinotopic or spatial coordinates confirmed by experiment.
M9. ante	Experimental data in single actor setting obtained	WP5	WWU	18	**Expected functionality:** Intermediate result. A model description of allocation of attention and behaviour prediction in humans (and monkeys). Action-minded representation of the 3D workspace. Preliminary data available for EYESHOTS partners. **Means of verification:** Hypothesis that gaze tracking can be used to predict allocation of attention and behaviour prediction confirmed by experiment.
M9	Experimental data on human-human interaction obtained	WP5	WWU	27	**Success criteria:** Success of Objective 3. Psychophysical bases for cooperative human-robot interactions in a shared workspace. **Means of verification:** Hypothesis that gaze tracking can be used to predict cooperative behavior confirmed by experiment.
M10	Eye position gain fields	WP1	UG	30	**Expected functionality:** Link between disparity measurements and eye vergence/version angles. Towards head-centric depth perception. **Means of verification:** Testing the efficacy/efficiency of neural coding of egocentric distance (e.g., amplitude modulation of the model's cell responses vs. 'compensatory' shift of their disparity tuning, when tested at different fixation distances).
M11	Algorithm for robust head-centric 3D description of visual fragments	WP2	K.U. Leuven	30	**Expected functionality:** Scanning strategies to accurately describe the 3D structure of a visual fragment. **Means of verification:** Robustness of the transformations between different spatial frames of reference to the limited accuracy of the motor system.
M12	Arm reaching gain fields	WP4	UJI	30	**Expected functionality:** Sensorimotor transformations. Computational models of joint visuo-motor representation of the 3D space based on neurophysiological/psychophysical findings. **Means of verification:** Test on functionality and performance under controlled situations. The system will be able to reach a limited set of artificial objects in a controlled real environment as the result of its perceptual awareness about an object and the modulation of the reaching and gaze directions.

M13	Anthropomorphic eye system	WP1	UG	30	**Success criteria:** Final prototype of mechatronic robot eye system released. **Expected functionality:** (i) Ocular motion range closer possible to 90 deg. (ii) Maximum speed and acceleration according to system size constraints and actuators available on the market. A least maximum target ocular speed of 90 deg/sec is expected. (iii) Due to the experimental level of the mechanical design no accuracy performance figures can be fixed. (iv) Mechanical characteristics of motion should approximate Listing's Law as observed in humans. **Means of verification:** Test of functionality of the performance of the prototype in terms of: i) Ocular motion range. ii) Maximum speed and acceleration. iii) Accuracy. iv) Characteristics of the motion for saccadic and smooth pursuit behaviours. **Remark:** points (iii) and (iv) depend on the performance of the control and vision strategies as specified in M1.
M14	Interactive stereoscopic system	WP2	K.U. Leuven	36	**Success criteria:** Success of Objective 1. Development of a robotic system for interactive visual stereopsis. **Means of verification:** Test on functionality and performance in real-world situations. An example functionality in a real-world situation is the active exploration of a 'real' cluttered desk or a scene used in the human/computer interaction tasks (e.g. block sorting). This involves switching between reflex-like vergence movements and voluntary 3D exploration and results in a description of the peripersonal space in terms of the fragments present and their 'location' (or eye movements required to reach them).
M15	Construction of a global awareness of the peripersonal space	WP4	UJI	36	**Success criteria:** Success of Objective 2. Development of a model of a multisensory egocentric representation of the 3D space. **Means of verification:** Test on functionality and performance in the same real-world situation illustrated in M14. The test involves pointing and reaching momentarily assigned targets (objects located in the peripersonal space) while keeping a continuous visual exploration of the scene. The system will be able to build a multisensor egocentric 3D representation of an unknown natural real-world environment involving a set of everyday objects, as the result of its multimodal interaction in its peripersonal space. On request it will be able to change the gaze or reach towards a particular object.

Reviews will be synchronised with ends of project reporting periods – which in general coincide with the major milestones of the project.

B2. Implementation

B2.1 Management structure and procedures

This section describes the management structure of EYESHOTS, which is organized in three levels:

1. The Strategic Management level, which consists of the EYESHOTS Council, and the Scientific Advisory.
2. The Executive Management level, which consists of the Executive Management Committee (Coordinator/Scientific Manager, Technical research Manager, and Administrative Manager).
3. The Technical Management level, which consists of the Work Package Leaders.

- Strategic Management Level -

This level consists of the **EYESHOTS Council**, which has at least the Coordinator and one member per partner. At the beginning, it will be composed of the following persons:

Coordinator:	Silvio P. Sabatini
UG:	Giorgio Cannata
WWU:	Markus Lappe
UNIBO:	Patrizia Fattori
UJI:	Angel P. del Pobil
K.U. Leuven:	Marc M. Van Hulle

The EYESHOTS Council will meet at least annually. These meetings will contain an open section where all EYESHOTS partners will be able to contribute, and a management section where only the Council Members will meet. The EYESHOTS Council has as its main task the strategic management.

Contingency handling: One of the major tasks of the EYESHOTS Council is to handle project contingencies. Basic contingency plans had been described in the WPs of Section B.1.3.5. These plans will be discussed in the Council, augmented according to the state of the project and set into operation, if need arises.

Scientific Advisor (SA): One scientific advisor from the outside whom we will invite to our annual meetings. The purpose of the SA will be to provide outside views and advise on the implementation of the project. The SA shall help the consortium to prepare for the review meeting by assessing the quality of the achieved goals from the outside.

- Executive Management Level -

The **Executive Management Committee** consists of the Coordinator and Scientific Manager (Dr. Silvio P. Sabatini), Technical manager (Prof. M. Van Hulle), and the Administrative Manager (Ms. Laura Garbaglia, UG). The Executive Committee will hold video-conferencing meetings (3 times per year) and a physical meeting per year.

Tasks of the Executive Management Level: The Executive management level will be concerned with the administrative aspects of the day-to-day scientific, technical and financial management. Such as:
<u>Scientific</u>: Publication coordination in shared papers, workshop planning, advising of PhD student cooperations, coordination of mobility and training, etc.
<u>Technical</u>: Technology transfer from WP results to (industrially accessibly) real-world robotic applications; IPR handling; Valorisation of the neurophysiological and psychophysical experimental scenarios with respect to real-world situations of operational anthropomorphic machines.
<u>Administrative</u>: Reporting and financial management.

For financial management, contractual management, knowledge management and management of technology transfer, Intellectual Property Rights & Patents, the Administrative Manager has assistance from UG's staff.

- Technical Management Level -

Most of the day-to-day scientific and technical management is delegated to the Scientific and the Technical research Manager who will be assisted by the five Work Package Leaders and their deputies.

The Work Package Leaders are:

WP1: Giorgio Cannata
WP2: Marc M. Van Hulle
WP3: Fred Hamker
WP4: Angel P. del Pobil
WP5: Patrizia Fattori

Within their sphere of expertise, each WP leader will coordinate all scientific, technical and local collaborations necessary for the advancement of the project. Each WP leader will develop a management structure that is tailored for the tasks to be undertaken. If it is necessary to re-allocate resources in the Work Packages, or to make adjustments in them which result in other and/or delayed milestones and deliverables, then such changes may be requested by the WP leader, but can only be permitted by the EYESHOTS Council. The Technical Management Committee should hold regular meetings at approximately 6-month interval.

- Meetings -

The EYESHOTS Council is scheduled to meet at the beginning of the project to define the detailed work plan of the first year and take the necessary strategic decisions on the overall project development. The kick-off meeting will contain an interdisciplinary tutorial session to bring together background knowledge and establish bilateral collaborations. Further meetings of the EYESHOTS Council are scheduled at approximately 12 months interval. At Council Meetings the Work Package Leaders and their deputies should present workpackage reports, which address the following issues:

- Current progress of the workpackage in general and for each task in particular, and its adherence to pre-defined workplan and timetable.
- Unresolved issues and suggested actions to solve them.
- Provisions on the next activity steps, open questions.

Each Council Meeting will define the detailed workplan for the next period, and will handle contingencies and other strategic issues.

These meetings will contain a session (in a form of "open workshops") where all EYESHOTS partners will be able to contribute (e.g., by ways of presentations of PostDocs and PhD students) in order to maintain an appropriate level of interaction among the partners.

Technical meetings of the Work Package Leaders are scheduled at approximately 6-month interval.

Additionally, ad-hoc meetings (e.g., sub-group technical meetings) will be organised as required at the working level.

B2.1.2 Project monitoring and reporting

The project monitoring is carried out by the Project Coordinator and addresses concrete tasks:

- Checking the interaction between the consortium members during the work execution.
- Checking the progress of the work, on a regular period.
- Detailing how and when the partners must exchange the documentation.
- Setting out editorial standards for report contents.

The progress of the work will be monitored in relation to the scientific and technical deliverables.

B2.1.3 Planning and control

EYESHOTS management and decision-making will be centralised and execution of the Work Packages will be contracted out to project partners. The management will be responsible for setting project milestones and the partners will be responsible for meeting those milestones. Specific partners will be alerted (as required) to perform specific tasks required by the project. The control of each Work Package is based on the tracking of the project through communications with the Work Package Leaders, short quarterly progress reports submitted to the Coordinator and the screening of material submitted to the project website.

B2.2 Beneficiaries

Partner 1: Università degli Studi di Genova (UG)

There will be two research units involved in the EYESHOTS project:

- The **PSPC-Group** at the **Department of Biophysical and Electronic Engineering** (DIBE)
- The **MAC-lab** at the **Department of Computer, Communication, and Systems Science** (DIST).

Contribution to EYESHOTS: Scientific and technological expertise for the design and development of the anthropomorphic binocular vision system and for the control of eye movements. Development of dynamic binocular vision strategies based on populations or biologically-plausible neural units.

Key Personnel involved in EYESHOTS:

Dr. Silvio P. Sabatini: 12+3 PMs (URL: http://pspc.dibe.unige.it)

Prof. Giorgio Cannata: 12 PMs (URL: http://www.dist.unige.it/Users/cannata/)

Role in EYESHOTS: Principal investigator. Project Manager. Coordinator of WP1 and WP6-8. Contribution to WP2 and WP4 Leader in D1.1, D1.2, D1.3, D1.4. Links between Information Technology and Neuroscience. Main characteristics: shared expertise in robotics, stereo vision, theoretical neuroscience, cortical modelling.

Expertise in EYESHOTS: Biomechanical models, Robotics, Depth vision, Robot vision, Theoretical neuroscience.

Partner 2: Westfaelische Wilhelms-Universitaet Muenster (WWU)

At the **Department of General and Applied Psychology** two research units will be involved:

1. Contributing on computational modelling and psychophysical experiments on saccade adaptation and shared attention.
2. Contributing on models of attention, object recognition, learning, and the control of visual perception.

Key Personnel involved in EYESHOTS:

Prof. Markus Lappe: 15 PMs
(URL: http://wwwpsy.uni-muenster.de/Psychologie.inst2/AELappe/en/index.html)

Dr. Fred Hamker: 15 PMs
(URL: http://wwwpsy.uni-muenster.de/Psychologie.inst2/AELappe/personen/hamker.html)

Role in EYESHOTS: Dr. Hamker is the coordinator of WP3 and leader of D3.1a, D3.1b, D.3.2, D3.3a and D.3.3b. Prof. Lappe is Leader in D5.3, D5.4.

Expertise in EYESHOTS: Links between Information Technology, Computational Neuroscience and Cognitive Psychology. Main characteristics: shared expertise in theoretical neuroscience, cortical modelling, oculomotor function, human perception.

Partner 3: Alma Mater Studiorum – Università di Bologna (UNIBO)

The **Laboratory of Vision and Action** at the **Department of Human and General Physiology** will be the main Neuroscience partner contributing with neurophysiological experiments in awake behaving monkeys while performing arm-movements and visual fixation tasks. Contribute on the development of models of the cortical mechanisms involved in visuomotor integration processes.

Key Personnel involved in EYESHOTS[8]:

Prof. Patrizia Fattori: 5 PMs
(URL: www.gallettilab.unibo.it/Laboratories-Staff.html/staff%20fattori.html)

Prof. Claudio Galletti: 3 PMs
(URL: www.gallettilab.unibo.it/Laboratories-Staff.html/staff%20galletti.html)

Role in EYESHOTS: Principal investigator. Coordinator of WP5. Leader in D5.1, and D5.2. Involved also in D5.3 and D5.4. Main characteristics: Neuroscience.

Expertise in EYESHOTS: experimental approach to the link between perception and action; neurophysiological and psychophysical investigation of the neural substrates for eye-hand coordination and of brain processes lying between vision and eye movements, attention, arm movements.

Partner 4: Universitat Jaume I de Castellon (UJI)

The **Robotic Intelligence Laboratory** at the University Jaume I will contribute theoretical and technological expertise for visually-based sensorimotor coordination of the robotic arm. Development of models for visuo-motor representations of the 3D space.

Key Personnel involved in EYESHOTS:

Prof. Angel P. del Pobil: 9PMs (URL: http://www.robot.uji.es/people/pobil/)

Prof. Enric Cervera: 4PMs (URL: http://www.robot.uji.es/people/ecervera/)

Role in EYESHOTS: Coordinator of WP4 (Sensorimotor integration). Contributions to WP1. Main characteristics: shared expertise in implementing visuomotor coordination in robotics systems for eye-hand coordination, ocolumotor systems and robot heads (stereo, pan, tilt, vergence).

Expertise in EYESHOTS: Spatial cognition and motion planning, perceptual learning for manipulation, visuomotor coordination, visually-guided grasping, control architectures for multimodal sensorimotor coordination, perceptual grounding, cognitive neuroscience models for prehension.

Partner 5: Katholieke Universiteit Leuven (K.U.Leuven)

The **Laboratorium voor Neuro- en Psychofysiologie** of the **K.U. Leuven Medical School** will be involved in the development of neural models for vergence control and for learning receptive fields for interactive stereo processing.

Key Personnel involved in EYESHOTS:

Prof. Marc Van Hulle: 6 PMs[9] (URL: http://simone.neuro.kuleuven.be)

[8] In addition, Prof. Fattori and Prof. Galletti will also participate to the supervision of day-to-day experimental research activities conducted by young researchers. Own resources will support this effort estimated about 6PMs.

[9] In addition to that, Prof. Van Hulle will also participate to the supervision of postdoctoral students through a general tutoring program that cuts across the research activities conducted in his lab, with an effort that will not be directly charged to the EYESHOTS project, but based on own resources.

Role in EYESHOTS: Technical Manager. Coordinator of WP2. Leader in D2.1, D2.2a, and D2.2b. Responsible for Active stereopsis through reflex-like vergence and voluntary exploration.

Expertise in EYESHOTS: Computer- and biological vision (multiscale motion and stereo algorithms), machine- and neural network learning (including information-theoretic learning), computational and experimental neuroscience.

B2.3 Consortium as a whole

The expertise of the consortium is summarised in the following table:

Expertise	UG	WWU	UNIBO	UJI	K.U.Leuven
Robotics	X			X	
Biomechanical models	X			X	
Motor control	X			X	
Machine Learning		X		X	X
Computer vision	X				X
Experimental neuroscience		X	X		
Theoretical Neuroscience	X	X			X
Cognitive Psychology		X		X	
Psychophysics		X	X		
Neurophysiology			X		

B2.4 Resources to be committed

- *RTD budget overview* -

A total of 400.5 **person months** are allocated to this project for RTD activities.

The effort corresponds to approximately 2 Research Assistants per group. In addition, of the participating universities will provide subsidiary funding to support additional PhD students and undergrads.

We have allocated about 120K€ for **equipment**. This sum is mainly required in order to engineering the robotic experimental set-ups and for the psychophysical and neurophysiological activities:

Robotics set-ups:
- Simulation computers
- Commercial small size robot arm
- Pan-tilt-vergence robotic head

Psychophysical and neurophysiological set-ups:
- Eyetracker Eyelink 1000: needed for the human-human experiments in WP 5.4. We currently have only one eye tracker available so that a second eye tracker is requested.
- System to record eye movements from 2 eyes simultaneously.

The following existing resources of a value of about 0.8 M€ are own resources provided by the consortium partners (in addition to the operating theatre fully equipped for chronic surgery in the monkey, and animal room equipped to house macaque monkeys) all of which are required in the different project parts:

Robotics/computing:
- Robot eye and hand prototype
- Development tools for microcontroller based control systems and MATLAB/Simulink
- Two 7-degrees-of freedom arms with 3-finger Hand and stereo camera set (eye-in-hand configuration); Approximately value 170K€
- Parallel jaw gripper for one of the arms (9K€)
- Stereo pair (15K€)
- PCs (total: 30,8K€) and access to Cluster Computer of K.U.Leuven (1K€/node/year)
- Standard equipment for electronic and mechanical hardware development.

Psychophysical apparatus:
- Virtual Reality Lab with head-mounted display and optical position tracking
- Large-scale projecting screen for high resolution near full-field optic-flow stimulation
- Two different eye trackers (expertise in EyeLink® and iView®)
- Motion Star Body tracking (24 sensors, cableless).

Neurophsyiological equipment
- Minimatrix Thomas Recording for multiple electrode recording
- Alpha Omega Multi Spike Detector (software + computer)
- Cyber amp 380 Axon Instruments Signal conditioner 8 channels
- Primate chair, Crist instruments
- System for recording in the awake behaving monkey bioelectrical activity from 1 electrode
- Room for computerized off line analysis of electrophysiological data

Then there are about 119K€ for **Travel,** which represents about 2,000 € per person and year. The groups, however, are committed to provide additional support for their students.

Finally there are about 85K€ for **Consumables**. The major costs for this item mainly concern to the expenditures for neurophysiological experiments and on robot parts for prototyping and experimental set-ups:
- Consumable electronics and electro-mechanical components
- Electrodes Thomas recording (47 euro each) for 400 electrodes
- Alpha Omega software for off-line sorting
- Software to perform 3D stimulations
- Laboratory supplies
- Mechanical and electronic components to home build the device to perform 3D reaching movements and vergence eye movements

The RTD budget will also comprise the costs for dissemination and training activities.

B3. Impact

B3.1 Strategic impact

The engineering principles and the advanced robotic head platform developed in EYESHOTS, which are expected to provide a more reliable and adaptive representation of the 3D space than existing systems - on the basis of vision/motor associations that reinforce and disambiguate the visual cues - , will contribute to progress in computer vision and robotics in general. More precisely, we expect such results to be applied/used in all situations in which a full 3D awareness of unstructured environments is necessary, especially when cooperation between robots and humans are foreseen (e.g., in disassembling tasks).

The research results will contribute (1) to the definition of a strategy to achieve a global perception of the 3D spatial relations for controlling spatially directed actions (e.g., reaching), and, in general, visually-guided goal-directed movements in the whole peripersonal workspace; (2) to an assessment of the computational benefits of a human-like mechanical structure of eye-plant on vision processing; (3) to visual neuroscience research by helping improve current understanding of cerebral mechanisms of human attention, active stereopsis and eye movement control through its application to real-world examples provided by an operational anthropomorphic machine.

B3.2 Plan for the use and dissemination of foreground

The dissemination activities are described in the WP7.

- Knowledge management and handling intellectual property -

The Coordinator and the EYESHOTS Council will undertake a number of specific measures to assure maximum awareness of the available knowledge and foster communication in the consortium. These measures shall be implemented by the Project Coordinator and include:

- Setting up and maintaining a project web page that will have a public and a private access,
- Setting up and maintaining a project repository for publications, software, etc.,
- Organisation of regular internal workshops with the project partners.

Assessment of the possibility for patenting individual results and giving advice to individual partners or sub-groups. The actual IPR and patent handling, however, is governed by the consortium agreement.

- Exploitation plan of the results -

This project has an explorative, pre-industrial character. Therefore, short or medium term industrial exploitation is not its primary objective. However, in case our findings would lead to important industrial innovations, we will consider protecting our inventions (e.g. through patents) and will develop its commercial exploitation through selected industrial partnerships.

In particular, we envisage important innovation on

1. Robotic platforms and systems as well as software platforms and frameworks in prototype form for visual and visuomotor processing.
2. Software components and applications which instantiate mechanisms for adaptive perception action learning.

- Plans for disseminating the developed platform outside the consortium -

The potential use of the EYESHOTS anthropomorphic mechatronic binocular platform outside the consortium as an alternative to more conventional robotic binocular heads will be promoted though:

1) Advertising the EYESHOTS platform worldwide and presenting it in demo sessions of conferences on “Robot vision” and “Visual perception”.
2) The creation of a dedicated section of the EYESHOTS website containing:
    a. a repository for descriptive material about the platform (public);

b. the design sheets of the mechanical and electronic components, and the basic control software libraries (under free formal registration).

Furthermore, for a large part of the project the integration of sensorimotor behaviours (fixation + reaching) will be based on conventional platforms (binocular pan/tilt/vergence heads and commercial robotic arms), this will allow us to systematically compare the performance of the EYESHOTS binocular system (available at month 30) with those obtained with the more conventional platforms. Such comparisons will be made publicly available on the website and possibly discussed during joint events with consortia of related projects.

The possibility of disseminating the developed vision systems as an “open platform” will be assessed at M24 and M36.

B4. Ethical issues

Ethics on Animal Research

UNIBO: Neurophysiological studies on non-human primates

The research on non-human primates (NHP) has been planned and thought following all the guidelines proposed by European Commision, Research DG, Unit L3 Governance and ethics:
ftp://ftp.cordis.europa.eu/pub/fp7/docs/research.animals.doc
ftp://ftp.cordis.europa.eu/pub/fp7/docs/ethics-animal1.pdf
ftp://ftp.cordis.europa.eu/pub/fp7/docs/ethics-animal2.pdf
ftp://ftp.cordis.europa.eu/pub/fp7/docs/ethics-animal3.pdf
Specifically, we have adopted the EC directive on the 3Rs (replacement, reduction, and refinement), in compliance with Directive 86/609/EEC.

JUSTIFICATION of THE USE OF NON-HUMAN PRIMATES

In EYESHOTS we aim at developing an anthropomorphic artificial intelligent system able to achieve a full awareness of 3-D environment for interaction and motor control. This highly cognitive skill is typical of the human brain: the ability to internally reconstruct the 3D world by manually interacting with the environment, by reaching and manipulating objects and by directing toward them eye movements or attention. Such complex abilities are common only to human brain and to that of NHP. For this reason some of the neuroscience research will be conducted on human beings and on NHP.

WHY NOT USING OTHER ANIMALS

The choice of NHP has been done following the Refinement rule (3 Rs). Actually, given the recalled above integrated aspects of the issues addressed in the present project, the relevant information for EYESHOTS could not be obtained using in vitro systems, nor cell coltures or isolated slices and preparations. Other animal species (such as rodents or other mammals) cannot be used for this study because high level cognitive tasks, such as that required for linking together the different visual fragments, the orchestration of eye-movements, attention, precise arm reaching movements, hand-object interaction to master the 3D world, is an exquisite ability of the primate brain. For all these reasons Replacement alternatives cannot be used for this study.

WHY THE EXPERIMENTS PROPOSED IN EYESHOTS CANNOT BE ALL PERFORMED IN HUMANS

The electrophysiological technique proposed in WP5 of this project (single neuron recording) will let us know how single neurons code these complex behaviors and thus will provide essential information for achieving the Objectives of EYESHOTS. Single cell recording provides information of high temporal resolution about the signals carried by neurons as well as about the synaptic interconnections between neurons housed in selected areas, and is the most suitable technique to bring information on how these high level functions are accomplished by our brain. This technique cannot be used in the human being because it is invasive. The use of this technique is unjustified in the healthy as well as unhealthy human being, except for curing specific neural

diseases or as a clinical investigation before invasive brain surgery performed for curing the human being her/himself.
Currently available less invasive techniques that give information on the neuronal coding of integrated cognitive processes are brain imaging techniques (MEG, EEG, PET, fMRI and its direct applications as DTI, DCM….). Some of these imaging techniques have a temporal resolution comparable to single cell recording (msec) but not the spatial one (some cm3 against 50 microns); the others do not have both the temporal (some secs) and the spatial (some cm3) resolution of single cell recording. The predominant neural behaviour of one part of the brain emerges from these imaging techniques, but they smooth out the neural behaviour that, though less prevalent, is indicative of how the brain operates (see at this regard: Galletti C., Battaglini P.P. and Fattori P. (1993). Parietal neurons encoding spatial locations in craniotopic coordinates. *Exp. Brain Res.*, **96**: 221-229). In order to model the phenomena under study in EYESHOTS, it is necessary to reach both spatial and temporal resolutions as those offered by the technique propose. Moreover, the interpretation of data coming from brain imaging technique is facilitated, sometimes only possible, when previous single cell recording studies have described the behavior of the activated region (see at this regard: Culham JC, Cavina-Pratesi C, Singhal A. The role of parietal cortex in visuomotor control: what have we learned from neuroimaging? Neuropsychologia. 2006;44(13):2668-84).

Considering all these arguments, the choice of NHP is the most appropriate one in order to support the neuroscientific research of EYESHOTS.

NUMBER OF NHP USED, TRANSPORT, HUSBANDRY, CARE

The experiments will be carried out on 2 monkeys (*Macaca Fascicularis)*. This number has been chosen based on previous experience of our lab in similar experiments, in accordance with the Reduction rule (3Rs). The number of NHP used is the minimum standard for neurophysiological international scientific papers on behaving monkeys, and will allow to achieve statistical significance of collected data, compensating for inter-individual variability.
Neuronal discharges from populations of neurons will be compared using commonly-used statistical parametric and non-parametric methods, in order to correlate neural activation with visual stimulations, eye movements, attention, arm reaching movements. More details on the statistical methods used can be found in some papers from our group, listed in the Bibliography Appendix (see for example Fattori et al., 2005 European Journal of Neuroscience).
Every effort has been made to minimize the use of non-human primates, choosing to plan the use of the minimum number of animals in order to meet the scientific objectives of the project. For this reason, the NHP will not be used only under ICT domain, but also for studies in the domain of HEALTH (Re-use). For instance, after training and single cell recordings for EYESHOTS, the same animal will be used for neuroanatomical studies in order to acquire knowledge on the anatomical organization of posterior parietal cortex and of the pattern of cortico-cortical connections.
The electrophysiological results obtained in WP5 of EYESHOTS will concur to increase our present knowledge of the neuronal coding accomplished by the brain and will be directly re-used in the neuroscience field. As introduced in the state of the art, there is no knowledge at present on how the mastering of 3D world by biological sensori-motor system is accomplished, and thus this research will increase our knowledge in that field. This knowledge could also help the finding of new recovery strategies to be used to rehabilitate patients after stroke or brain lesions that alter the same functions studied in this research project. In addition, topographical aspects of the neuroscience results obtained with EYESHOTS and not included in neural models, will be used to map the distribution of cellular properties of different sectors of the posterior parietal cortex. These topographical data will contribute to the advancement of our knowledge in the neuroscientific field.

The monkey transport, environment, health, housing, enrichment and care, training of personnel and transport, follow the standards recently adopted by the European Commission (207/526/EC).
The Recommendation (adopted on 18 June 2007) can be found at: http://ec.europa.eu/environment/chemicals/lab_animals/legislation_en.htm . It aligns EC legislation

with the revised Council of Europe guidelines (Appendix A of Convention ETS 123): http://conventions.coe.int/Treaty/EN/Treaties/PDF/123-Arev.pdf)

The animals (captive bred) are supplied from an authorized European animal centre (R.C. Hartelust B.V., Netherlands). UNIBO complies with the Italian and European Union laws for the use of animals (Italian law D. Lgs. n. 116 of 1992, in compliance with the European Union Directive 86/609/EU). All the experimental procedures have been approved by the Bioethics Committee of the University of Bologna and authorized by Ministero della Salute (Decreti Ministeriali N° 94/2005-C signed by the Direttore of the Dipartimento Sanità Pubblica Veterinaria, 22/7/2005). In addition, all procedures used have been approved and are controlled by the Central Veterinary Service of the University of Bologna. The primates used by UNIBO group are located in the monkey stabularium at the Dept. of Human and General Physiology, authorized by Ministero della Salute with Decreto Ministeriale N° 52/2004-A, dated 27/5/2004. The monkey stabularium houses the monkeys in social environment that enables the animal to carry on a daily programme of activity. Cages are of adequate size in order to let the monkey display a normal motor and behavioural repertoire and have room for suitable environmental enrichment.

TRAINING and MONITORING OF HEALTH

Monkey training will adopt positive reinforcement techniques, for which the personnel involved in the research has a decennial experience. No deprivation, suffering, punishment will be inflicted to the NHP, not only to safeguard animal's safety (physical, psychological and neurological) but also to perform good physiological research (punishments could be detrimental to the experimental design).

As NHP are spontaneously curious and interested in exploring the environment around them, there is no need to stress the animals in order to induce them to do in the lab what they spontaneously would do in real world. In the laboratory they receive the individual daily amount of food, water and other liquids as reward. Daily weight control is performed in order to monitor the wellbeing of the NHP under training. Particular attention will be paid by all researchers involved in the research to signs of stress (like overreaction to stimuli or persons, fear reactions, abnormal behavioral repertoires) that the monkey should present in any moment. Veterinary competence of Servizio Veterinario dell'Università di Bologna will detect, if present, the clinical signs of pain or distress.

Because of the above-mentioned reasons, the monkey training is not a stressful experience for NHP and the percentage of succesful training reached in 20 years time in our laboratory is 100% .

Following the Refinement rule (3Rs), we have been using for years the research methods that alleviate or minimize potential pain, suffering or distress and enhance NHP welfare.

Veterinary advice has been taken when setting the surgery procedure hereafter summarized. It is worthy to note that surgery is not a repetitive experimental approach, because it is performed only once in the monkey. Surgery is performed uniquely to implant, once forever, the recording apparatus. UNIBO veterinaries are present at the surgery and ensure an high health status of animals. A collaboration with Neurosurgeons of our University has allowed us to find the best surgical procedure and post-operative care (Refinement rule). Surgery will be performed in asepsis and under general anaesthesia (sodium thiopenthal, 8 mg/kg/h, *i.v.*). Adequate measures will be taken to minimize pain or discomfort. Specifically, analgesics will be used postoperatively as usually done in clinical practice (ketorolac trometazyn, 1mg/kg i.m. immediately after surgery, and 1,6 mg/kg i.m. in the following days; as an alternative: buprenorphine 0.01 mg/kg i.m. every 12 hours).

All persons involved in the research proposed in EYESHOTS will be appropriately educated and trained to the standard recommended by the European Council and recently adopted by EU, as well as to the above-mentioned EU legislation at this regard.

Ethics on Humans

WWU- Psychophysical studies on healthy human subjects.

All psychophysical experiments on human subjects will be carried out according to the declaration of Helsinki (*Brit Med J* 1991; **302**: p. 1194).

The experiments involve only healthy adult volunteers and do not pose risks to the participants. Informed consent will be obtained from each participant prior to entering the experiment in full accordance with the Declaration of Helsinki and the provisions for research on humans given by the Council of Europe:
http://www.coe.int/t/e/legal%5Faffairs/legal%5Fco%2Doperation/Bioethics/
Full information is given to participants as to the nature of the research. They are informed that they are under no obligation to take part and they are free to leave at any point in the experiments. No discomfort, stress, or invasive procedures are involved in the research. Data from the participants is subject to privacy protection and access to it is restricted to the experimenter. After the end of the study the data will be encrypted and the original data will be deleted. Scientific publications of the data will not give the name of the participant.
Up to 12 volunteers will participate in each experiment. The total number of volunteers recruited for all experiments will be about 40. Volunteers are recruited mainly from members of the department and students. Volunteers do not receive financial benefits, but students may receive obligatory course credit for participating.

Criteria for inclusion in the experiment are:
- healthy male or female adults
- age between 18 and 60

Criteria for exclusion from the experiment are:
- physical, mental, or neurological disorder
- impairment of vision unless corrected by spectacles
- pregnancy.

Processes and actions taken, and planned to be taken during the project on the ethical issue will be reported in the Periodic Progress Reports.

Literature References

Alexander, G.E., Crutcher, M.D. and DeLong, M.R. (1990) Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. Prog. Brain Res. Vol.85:119-46.

Anzai, A., Ohzawa, I. and Freeman, R.D. (2001) Joint-encoding of motion and depth by visual cortical neurons: neural basis of the Pulfrich effect. Nature Neuroscience. Vol. 4: 513-518.

Awater, H., and Lappe, M. (2006) Mislocalization of perceived saccade target position induced by perisaccadic visual stimulation. Journal of Neuroscience. Vol. 26:12–20.

Awater, H., Burr, D., Lappe, M., Morrone, M. C. and Goldberg, M. E. (2005). The effect of saccadic adaptation on the localization of visual targets. Journal of Neurophysiology. Vol. 93:3605-3614.

Bahcall, D. O. and Kowler, E. (1999). Illusory shifts in visual direction accompany adaptation of saccadic eye movements. Nature. Vol. 400, 864-66.

Bell, A.J. and Sejnowski, T.J. (1997) The "independent components" of natural scenes are edge filters. Vision Research. Vol. 37:3327-38.

Biamino, D., Cannata, G., Maggiali, M. and Piazza, A. MAC-EYE: a tendon driven fully embedded robot eye. Proc. of IEEE Humanoids 2005 Conference, Dec. 5-7 2005, Tsukuba, Japan.

Breveglieri, R., Kutz, D.F., Fattori, P., Gamberini, M. and Galletti, C. (2002) Somatosensory cells in the parieto-occipital area V6A of the macaque. Neuroreport. Vol. 13:2113-2116.

Bridgeman, B., van der Heijden, A.H.C., and Velichkovsky, B.M. (1994) A theory of visual stability across saccadic eye movements. Behavioral and Brain Sciences. Vol. 17(2):247-292.

Büker, U., Drüe, S., Götze, N., Hartmann, G., Kalkreuter, B., Stemmer, R. and Trapp, R. (2001) Vision-based control of an autonomous disassembly station. Robotics and Autonomous Systems. Vol. 35(3):179-189.

Cannata, G. and Maggiali, M. Design of a humanoid robot eye in humanoid robots (Mathias Hackel ed.). ARS Vienna (to appear April 2007)a.

Cannata, G. and Maggiali, M. Implementation of listing's law for a tendon driven robot eye. Proc. 2006 IEEE IEEE/RJS Int. Conf. on Intelligent Robots and Systems IROS 2006, Beijing (China), Oct. 9-15, 2006.

Cannata, G. and Maggiali, M. Models for the design of a tendon driven robot eye. Proc. IEEE-RAS Int. Conf. on Robotics ICRA 2007, Rome (Italy), Apr. 10-14, 2007b.

Cannata, G., D'Andrea, M. and Maggiali, M. Design of humanoid robot eye: models and experiments. Proc. 6th IEEE-RAS Int. Conf. on Humanoid Robots HUMANOIDS 2006, Genova (Italy), Dec 4-6, 2006b.

Cannata, G., D'Andrea, M., Maggiali, M. and Monti, F. Implementation of listing's law for a robot eye. 8th International IFAC Symposium on Robot Control SYROCO 2006, Bologna (Italy), September 6 - 8, 2006a.

Cervera, E., del Pobil, A.P., Berry, F. and Martinet, P. (2003) Improving image-based visual servoing with three-dimensional features. The International Journal of Robotics Research. Vol. 22:821-840

Chinellato, E. and del Pobil, A.P. (2007) Integration of stereoscopic and perspective cues for slant estimation in natural and artificial systems. In Mira J. & Alvarez J.R. (Eds.), LNCS 4528: 399–408

Chinellato, E., Demiris, Y. and del Pobil, A.P. (2006) Studying the human visual cortex for achieving action-perception coordination with robots. IASTED Intl. Conf. on Artificial Intelligence and Soft Computing (ASC 2006): 184-189

Chinellato, E., Morales, A., Fisher, R.B. and del Pobil, A.P. (2005) Visual features for characterizing robot grasp quality. IEEE Transactions on Systems, Man and Cybernetics (Part C). Vol. 35(1): 30-41.

Chinellato, E., Recatalá, G., del Pobil A.P., Mezouar. Y. and Martinet. P. (2006b) 3D grasp synthesis based on object exploration. IEEE Intl. Conf. on Robotics and Biomimetics (ROBIO 2006).

Cisek, P. and Kalaska, J.F. (2005) Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. Neuron. Vol. 45(5): 801-814.

Clark, R. A., Miller, J.M., Demer, J.L. (2000) Three-dimensional Location of Human Rectus Pulleys by Path Inflection in Secondary Gaze Positions. Investigative Ophthalmology and Visual Science. Vol. 41:3787-3797.

Culham, J.C., Gallivan, J., Cavina-Pratesi, C. and Quinlan, D.J. (2007). fMRI investigations of reaching and ego space in human superior parieto-occipital cortex. In R. L. Klatzky, M. Behrmann, & B. MacWhinney (Eds.), Embodiment, Ego-space and Action. Madwah NJ: Erlbaum.

DeAngelis, G.C. (2000) Seeing in three dimensions: the neurophysiology of stereopsis. Trends in Cognitive Sciences. Vol. 4:80-90.

DeAngelis, G.C. and Newsome, W.T. (1999) Organization of disparity-selective neurons in macaque area MT. Journal of Neuroscience Vol. 19:1398-1415.

Demer, J.L, Ho, S.Y. and Pokens, V. (2000) Evidence for active control of rectus extraocular muscle pulleys. Investigative Ophtalmology & Visual Science. Vol. 41: 1280-1290.

Demer, J.L., Miller, J.M., Poukens, V., Vinters, H. V. and Glasgow, B.J. (1995) Evidence for fibromuscular pulleys of the recti extraocular muscles. Investigative Ophthalmology and Visual Science. Vol. 36:1125-1136.

Deubel, H., and Schneider, W.X. (1996) Saccade target selection and object recognition: evidence for a common attentional mechanism. Vision Research. Vol. 36:379-391.

Deubel, H., Bridgeman, B., and Schneider, W.X. (1998) Immediate post-saccadic information mediates space constancy. Vision Research. Vol. 38:3147-3159.

Driver, J., Davis, G., Maxwell, E., and Baron-Cohen, S. (1999) Gaze perception triggers reflexive visuospatial orienting. Visual Cognition. Vol. 6(5):509–540.

Fattori, P., Gamberini, M., Kutz, D.F. and Galletti, C. (2001) 'Arm-reaching' neurons in the parietal area V6A of the macaque monkey. Europena Journal of Neuroscience. Vol. 13:2309-2313.

Fattori, P., Kutz, D.F., Breveglieri, R., Marzocchi, N. and Galletti, C. (2005) Spatial tuning of reaching activity in the medial parieto-occipital cortex (area V6A) of macaque monkey. European Journal of Neuroscience. Vol. 22:956-972.

Flanagan, J.R., and Johansson, R.S. (2003) Action plans used in action observation. Nature. Vol. 424(6950):769–71.

Fujii, N., Hihara, S., and Iriki, A. (2007) Dynamic social adaptation of motion-related neurons in primate parietal cortex. Vol. 2(4) e397:1-8.

Fukushima, K. (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biological Cybernetics. Vol. 36:193-202.

Furman, J.M. and Schor, R.H. (2003) Orientation of Listing's plane during static tilt in young and older human subjects. Vision Research. Vol. 43: 67-76.

Galletti, C., Battaglini, P.P. and Fattori, P. (1995) Eye position influence on the parieto-occipital area PO (V6) of the macaque monkey. European Journal of Neuroscience. Vol. 7:2486-2501.

Galletti, C., Fattori, P., Battaglini, P.P., Shipp, S. and Zeki, S. (1996) Functional demarcation of a border between areas V6 and V6A in the superior parietal gyrus of the macaque monkey. European Journal of Neuroscience. Vol. 8:30-52.

Galletti, C., Fattori, P., Kutz, D.F. and Battaglini, P.P. (1997) Arm movement-related neurons in the visual area V6A of the macaque superior parietal lobule. European Journal of Neuroscience. Vol. 9:410-413.

Galletti, C., Fattori, P., Kutz, D.F. and Gamberini, M. (1999) Brain location and visual topography of cortical area V6A in the macaque monkey. European Journal of Neuroscience. Vol. 11:575-582.

Goodale, M.A. and Milner, A.D. (1992) Separate visual pathways for perception and action. Trends in Neuroscience. Vol. 15:20-25.

Hadsel, R., Chopra, S., and LeCun, Y. (2006). Dimensionality reduction by learning invariant mapping. In Proc. Computer Vision and Pattern Recognition Conference (CVPR'06). IEEE Press.

Hamker, F. H. (2005)b The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. Cerebral Cortex. Vol. 15:431-447.

Hamker, F. H., (2005)a The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision. Journal for Computer Vision and Image Understanding. Special Issue on Attention and Performance in Computer Vision. Vol. 100:64-106.

Hamker, F.H. and Wiltschut, J. (submitted) Hebbian learning in a model with dynamic rate-coded neurons: an alternative to the generative model approach for learning receptive fields from natural scenes. Submitted to Network

Hamker, F.H., Zirnsak, M. (2006) V4 receptive field dynamics as predicted by a systems-level model of visual attention using feedback from the frontal eye field. Neural Networks. Vol. 19:1371-1382.

Hancock, P.J.B, Baddeley, R.J. and Smith, L.S. (1992) The principle components of natural images, Network. Vol. 3:61-70.

Harpur, G. and Prager, R. (1996) Development of low entropy coding in a recurrent network. Network: Computation in Neural Systems. Vol. 7:277-84.

Henderson, J. (2003) Human gaze control during real-world scene perception. Trends in cognitive sciences. Vol. 7(11):498–504.

Himmelbach, M. and Karnath, H. (2005) Dorsal and ventral stream interaction: contributions from optic ataxia. Journal of Cognitive Neuroscience. Vol. 17:632-640.

Hinkle, D.A. and Connor, C.E. (2005) Quantitative characterization of disparity tuning in ventral pathway area V4. Journal of Neurophysiology. Vol. 94:2726-37.

Itti, L., Koch, C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research. Vol. 40:1489-1506.

Johansson, R.S., Westling, G., Bäckström, A., and Flanagan, J.R. (2001) Eye-hand coordination in object manipulation. Journal of Neuroscience. Vol. 21:6917–6932.

Jonides, J., Irwin, D.E. and Yantis, S. (1982) Integrating visual information from successive fixations. Science. Vol. 215:192-4.

Koornneef, L. (1974) The first results of a new anatomical method of approach to the human orbit following a clinical enquiry. Acta Morphol Neerl Scand. Vol. 12(4):259-282.

Kutz, D.F., Fattori, P., Gamberini, M., Breveglieri, R. and Galletti, C. (2003) Early- and late-responding cells to saccadic eye movements in the cortical area V6A of macaque monkey. Experimental Brain Research. Vol. 149:83-95.

Lappe, M., Awater, H., and Krekelberg, B. (2000) Postsaccadic visual references generate presaccadic compression of space. Nature. Vol. 403:892–895.

LeCun, Y, Boser, B., Denker, J.S., Henderson, D., Howard, R.E. Hubbard, W. and Jackel, L.D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, Vol. 1(4):541-551.

LeCun, Y. and Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition. proceedings of the IEEE. Vol. 86(11):2278-232.

LeCun, Y., Muller, U., Ben, J., Cosatto, E. and Flepp, B. (2005). Off-Road obstacle avoidance through end-to-end learning. In Advances in Neural Information Processing Systems (NIPS 2005).

Linsker, R. (1986) From basic network principles to neural architecture: emergence of orientation columns PNAS 1986. Vol. 83: 8779-8783.

Lu, Z.L. and Sperling, G. (2001). Three-systems theory of human visual motion perception: Review and update. Journal of the Optical Society of America A. Vol. 18:2331-2370.

Marzocchi, N., Kutz, D.F., Breveglieri, R., Fattori, P. and Galletti, C. (2005) Gaze modulations on arm movement-related neural responses in area V6A of the macaque. Program No 3631 2005 Abstract Viewer/Itinerary Planner Washington, DC: Society for Neuroscience, 2005 Online.

McLaughlin, S. (1967) Parametric adjustment in saccadic eye movements. Perception & Psychophysics. Vol. 2: 359–362

Middleton, F.A. and Strick, P.L. (2002) Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. Brain and Cognition. Vol. 42:183-200.

Milanese, R. (1993) Detecting salient regions in an image: from biological evidence to computer implementation, Ph.D. thesis, Univ. of Geneva.

Miller, J. M. (1989) Functional anatomy of normal human rectus muscles. Vision Research. Vol. 29: 223-240.

Mozer, M.C. (1991). The perception of multiple objects: A connectionist approach. Cambridge, MA:. MIT Press/Bradford Books.

Olshausen, B.A. and Field, D.J. (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature. Vol. 381:607-9.

O'Regan, J.K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences. Vol. 24(5):939-1011.

Pack, C.C., Born, R.T., Livingstone, M.S. (2003) Two-dimensional substructure of stereo and motion interactions in macaque visual cortex. Neuron. Vol. 37:525-535.

Pouget, A. and Snyder, L.H. (2000) Computational approaches to sensorimotor transformations. *Nat* Neuroscience. Vol. 3:1192-1198.

Pouget, A., Deneve, S and Duhamel, J.R. (2002) A computational perspective on the neural basis of multisensory spatial representations. Nature Review Neuroscience. Vol. 3:741-747.

Qiu, F.T. and von der Heydt, R. (2005) Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. Neuron. Vol. 47:155-166.

Read, J.C.A. (2004) Early computational processing in binocular vision and depth perception. Progress in Molecular Biology and Biophysics. Progress in Molecular Biology and Biophysics. Vol. 87:77-108

Read, J.C.A. and Cumming, B.G. (2006) Does depth perception require vertical-disparity detectors?. Journal of Vision. Vol. 6:1323-1355.

Redgrave, P., Prescott, T.J. and Gurney, K. (1999) The basal ganglia: a vertebrate solution to the selection problem? Neuroscience Vol. 89:1009-23.

Riesenhuber, M. and Poggio, T. (1999) Hierarchical models of object recognition in cortex. Nature Neuroscience. Vol. 2(11):1019-1025.

Ristic, J., and Alan Kingstone A. (2005) Taking control of reflexive social attention. Cognition. Vol. 94:B55-B65.

Rizzolatti G. and Craighero L., (2004) The mirror-neuron system. Annual Review Neuroscience, vol. 27, pp. 169-192.

Sabatini, S.P. and Solari, F. (2004) Emergence of motion-in-depth selectivity in the visual cortex through linear combination of binocular energy complex cells with different ocular dominance. Neurocomputing. Vol. 58–60:865-872.

Sabatini, S.P., Solari, F., Andreani, G., Bartolozzi, C. and Bisio, G.M. (2002) A hierarchical model of complex cells in visual cortex for the binocular perception of motion-in-depth . In Proceedings of Advances in Neural Information Processing Systems 14 2002 Conference, pages 1271-1278, Vancouver, Canada.

Salinas, E. and Thier, P. (2000) Gain modulation: a major computational principle of the central nervous system. Neuron. Vol. 27:15-21.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. and Poggio, T. (2007). Object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 29(3):411-426.

Simoncelli, E.P. (2003) Vision and the statistics of the visual environment. Current Opinion in Neurobiology. Vol. 13:144-149.

Singhal, A., Culham, J.C., Chinellato, E. and Goodale, M.A. (2007) Dual-task interference is greater in delayed grasping than in visually-guided grasping. Journal of Vision. Vol. 7(5)

Straumann, D., Zee., D.S., Solomon, D. and Kramer, P.D. (1996) Validity of Listing's law during fixations, saccades, smooth pursuit eye movements, and blinks. Experimental Brain Research. Vol. 112:135-146.

Sumi, Y., Kawai, Y., Yoshimi, T. and Tomita, F. (2002) 3D Object recognition in cluttered environments by segment-based stereo vision. International Journal of Computer Vision. Vol. 46:5-23.

Suri RE, Bargas J, Arbib MA. (2001) Modeling functions of striatal dopamine modulation in learning and planning. Neuroscience, 103:65-85.

Tweed, D. and Vilis, T. (1987) Implications of rotational kinematics for the oculomotor system in three dimensions. The Journal of Neurophysiology. Vol. 58(4): 832-849.

Tweed, D. and Vilis, T. (1988) Rotation axes of saccades. Annals of the New York Academy of Sciences, Vol. 545: 128-139.

Tweed, D. and Vilis, T. (1990) Geometric relations of eye position and velocity vectors during saccades. Vision Research. Vol. 30(1): 111-127.

Van der Velde, F. & de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. Behavioral and Brain Sciences, 29, 37-70.

Van Ee, R. and Erkelens, C.J. (1996) Stability of binocular depth perception with moving head and eyes. Vision Research. Vol. 36(23):3827-3842.

von der Heydt, R., Zhou, H. and Friedman, H.S. (2000) Representation of stereoscopic edges in monkey visual cortex. Vision Research. Vol. 40:1955-1967.

Walther, D. and Koch, C. (2006) Modeling attention to salient proto-objects. Neural Networks. Vol. 19:1395-1407.

Wersing, H., and Körner, E. (2003). Learning optimized features for hierarchical models of invariant recognition. Neural Computation. Vol. 15(7):1559-1588.

Yoshiyama, K., Uka, T., Tanaka, H. and Fujita, I. (2004) Architecture of binocular disparity processing in monkey inferior temporal cortex. Neuroscience Research. Vol. 48:155-67.