2006 – 2009

Final Project Report

HERMES

Human Expressive Representations of Motion and their Evaluation in Sequences

http://www.hermes-project.eu

Information Society Technologies

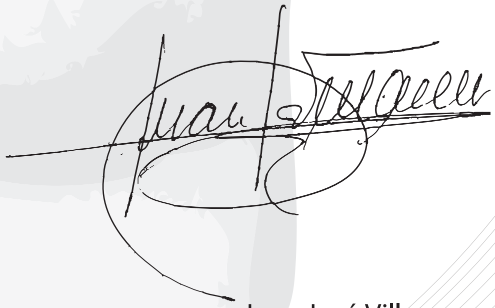
SIXTH FRAMEWORK PROGRAMME

The HERMES Project

"HERMES concentrates on how to extract descriptions of human behaviour from video sequences, in a restricted discourse domain."

- We confront agent, body and face motion evaluation, in scenarios ranging from wide-field-of-view with multiple-agent scenes to specific inferences of emotional state.
- The sensing and reasoning processes are tightly integrated in a perception-action cycle: cooperating pan-tilt-zoom sensors enhance cognition via responses to uncertain/ambiguous interpretations.
- The system has been exposed to video recordings from different parts of Europe. Its explanatory and arguing capabilities assess its strengths and weaknesses.

Project Nº	027110
Duration	36 months (+6 months extension)
Start Date	01/03/06
End Date	28/02/09 (extended to 31/08/09)
European Commission Funding	2 100 000 , 00 €
Hermes homepage	http://www.hermes-project.eu



Juan José Villanueva
Project Manager

Head of Hermes
Istanbul Archaeological Museum, Turkey

Executive Summary

Hermes Project

HERMES is a consortium project that concentrates on extracting descriptions of people behaviour from videos in restricted discourse domains, such as pedestrians crossing inner-city roads, approaching or waiting at stops of buses and even humans in indoor worlds like halls or a lobbies.

These video recordings allow us to explore a coherent evaluation of human movements and facial expressions across a wide variation of scale.

A system has been developed that starts with basic knowledge about pedestrian behaviour in chosen discourse domains, but eventually clusters evaluation results into semantically meaningful subsets of behaviours.

Objectives

The main objective of HERMES is to develop a cognitive artificial system to allow both recognition and description of a particular set of human behaviours arising from real-world events. Specifically, we model knowledge about the environment to suggest interpretations from motion events, and to communicate with people using natural language texts and synthetic films. The events are detected in image data-streams obtained from arrays of multiple active cameras (including zoom, pan and tilt).

The HERMES procedure combines:

- Detection and tracking of agents while they are some distance away from a location.
- If they are even closer and their face can be resolved, facial emotions are checked.

"We base on the three stages of human behaviour: motion of people, their posture and gestures, and their facial characterization."

Hermes, sculpture of Szeroka street
Toruń, Poland

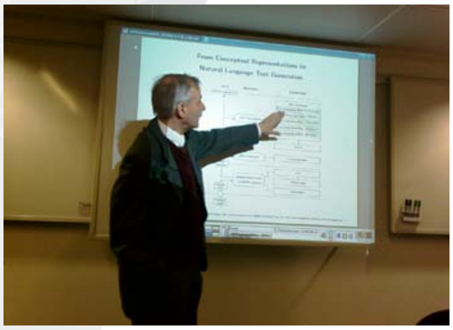
"Natural language texts and synthetic animation are used to communicate with end-users"

Approach

HERMES fulfils two main objectives: on the one hand, the generation of conceptual descriptions based on acquired and analysed motion patterns. On the other hand, the communication using visualization of synthetic motion patterns.

Natural language text generation is accommodated within HERMES based on these considerations:

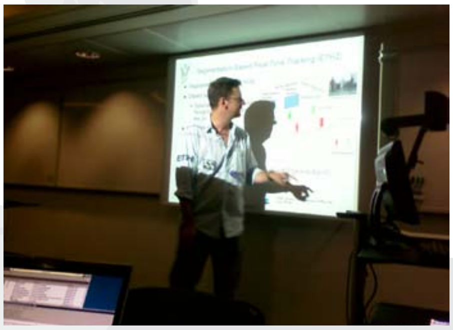
- Semantic descriptions enable researchers to check details of the conceptual knowledge base.
- They also allow communication with end-users in a natural manner.
- They support conceptual abstraction, thereby facilitating the communication of short messages or essential details.



Animation is accommodated within HERMES based on the following considerations:

- Analysis-by-synthesis at the three stages of human behaviour, i.e. motion of people, their posture analysis, and their face characterization.
- Animated computer graphics as a visual language to quickly communicate essential aspects..
- Animated computer graphics, again at three motion categories, for checking the underlying conceptual knowledge base.

Using both approaches, quantitative measures and qualitative descriptions have been developed to analyze the robustness and efficiency of the proposed cognitive system.



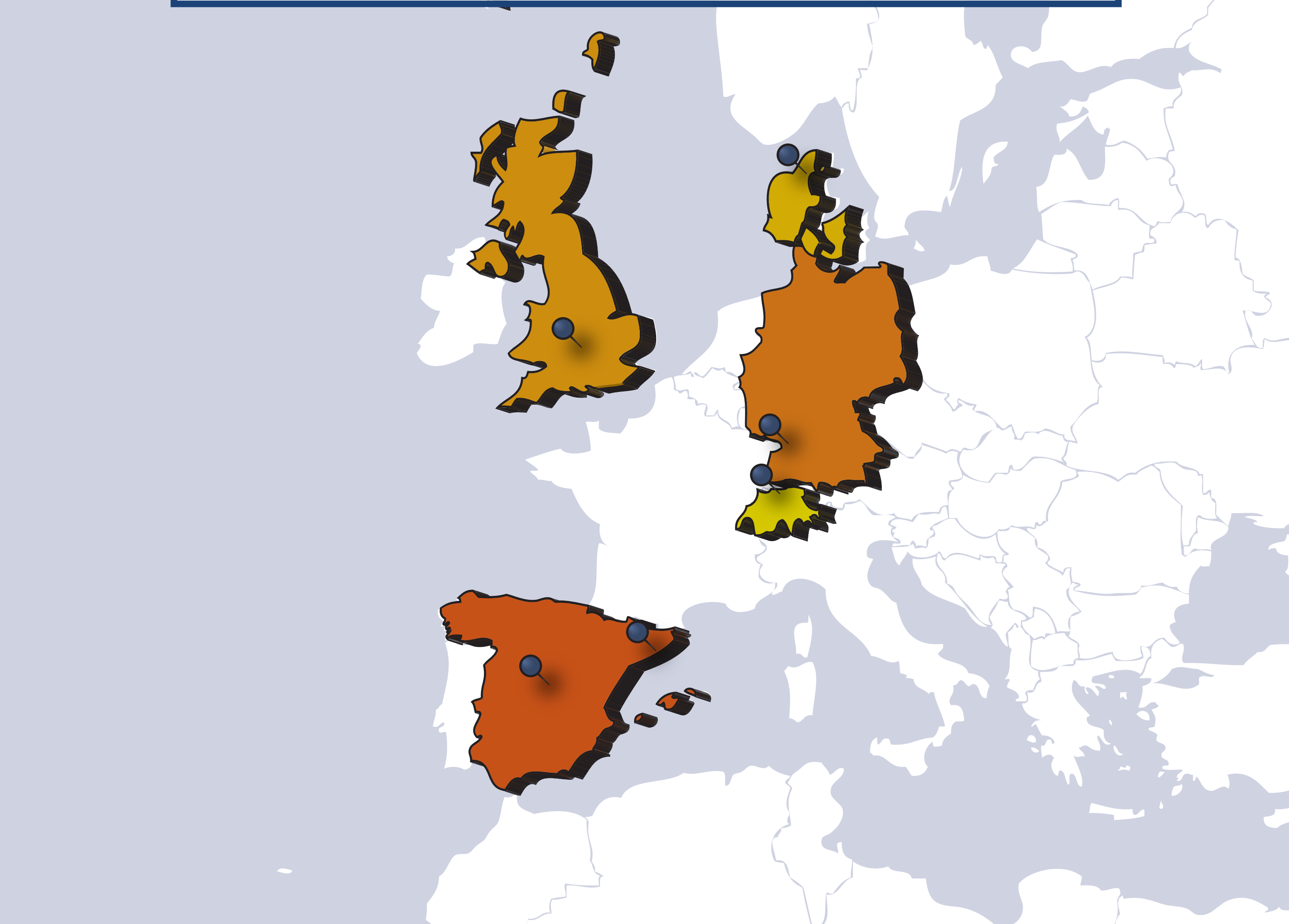
Sandalbinder Hermes untying his sandal
Louvre museum, France

Contents

Caduceus (Wand of Hermes)

HERMES Consortium

	Computer Vision Center **CVC** Universitat Autònoma de Barcelona	Spain
	Institut für Algorithmen und Kognitive Systeme **IAKS** Universität Karlsruhe	Germany
	Computer Vision and Media Technology Laboratory **CVMT** Aalborg University	Denmark
	Computer Vision Laboratory **BIWI** ETH Zürich	Switzerland
	Active Vision Laboratory **AVL** University of Oxford	United Kingdom
	Answare Technologies	Spain



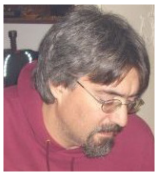
WP0 Project Management

Project Manager

Juan José Villanueva

Steering Committee

Management and Exploitation Office



Technical Coordinator

Andrew D. Bagdanov

Management Coordinator

Tonny Velin

Scientific Coordinator

Jordi Gonzàlez

Work Package leaders

WP0 WP1 WP2 WP3 WP4 WP5 WP6 WP7 WP8 WP9

Objectives

The objective of WP0 is the project organization and the management of the project activities. The main objectives are to initiate the project phases, to complete all the necessary administrative tasks required by the Commission, and to provide short progress reports and the agreed deliverable results to the EC on due time.

WP0–1 leader

Juan José Villanueva
juanjo@cvc.uab.cat

Definition of requirements

WP1

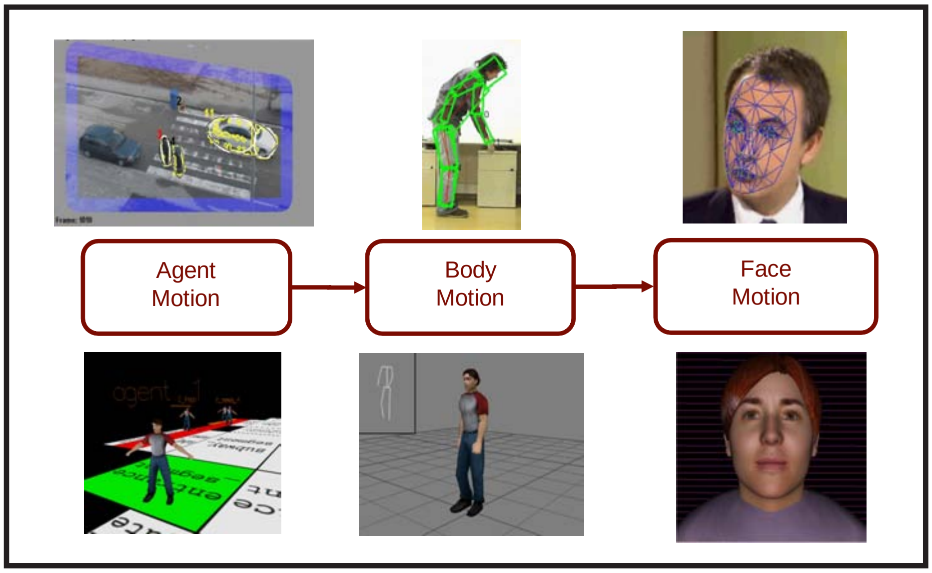
Objectives

WP1 synthesizes the requirements that define the technology to be developed in the context of HERMES. Due to the risk inherent to the complexity of the developed system, the work in WP1 establishes a realistic set of requirements, assumptions and restrictions to fulfil the project objectives and also to allow potential applications (of interest to partners) while taking into account the limited resources allocated.

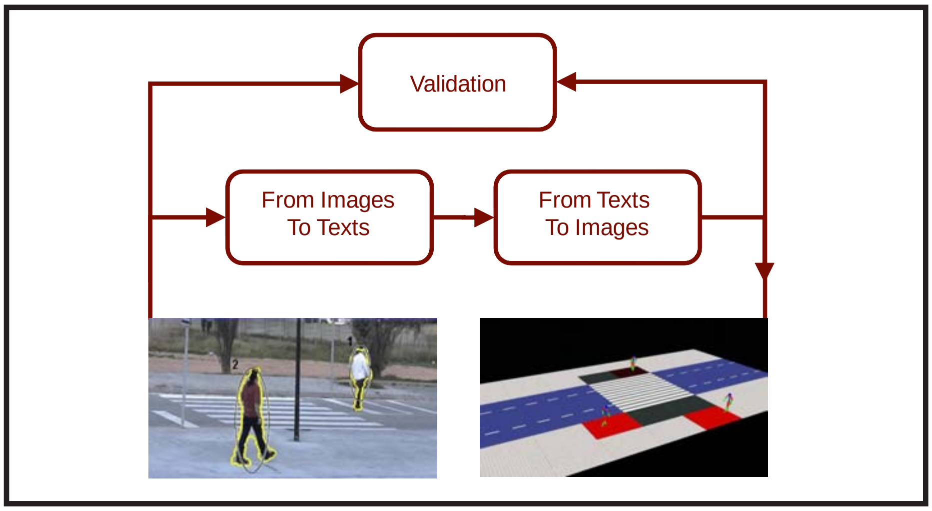
Activities & achievements

The defined requirements were elaborated, completed, synthesized, and harmonized to assure a complete and consistent output for use in the development, design, and implementation phases. An ontology was also defined to determine what the system is eventually going to say; which videos were to be recorded; what had to be selected, tracked, and converted into natural language output; and which questions were to be answered by the system.

Since the output of this WP was input for all other WPs, it detailed and contextualized the global requirements for each workpackage scope. The discourse domains were specified and also the locations for recording human agents. The considered scenarios provide a common test input for all partners, but deliberately allowing them for leeway to record additional videos for testing.



Human expressive representations of motion...

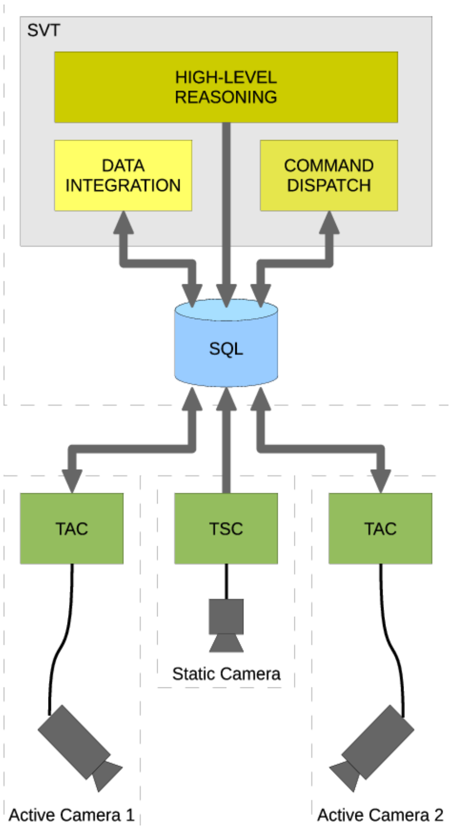


...and their evaluation in sequences

WP2 Distributed Camera System

Objectives

WP2 provides a platform to acquire video data for use in WP3, WP4, and WP5, and investigates the link between high-level understanding and sensing action. Its goals are: to provide an architecture from which to serve video data for WP3, WP4, and WP5; and to couple high-level scene interpretation with sensing actions.



HERMES multi-camera architecture

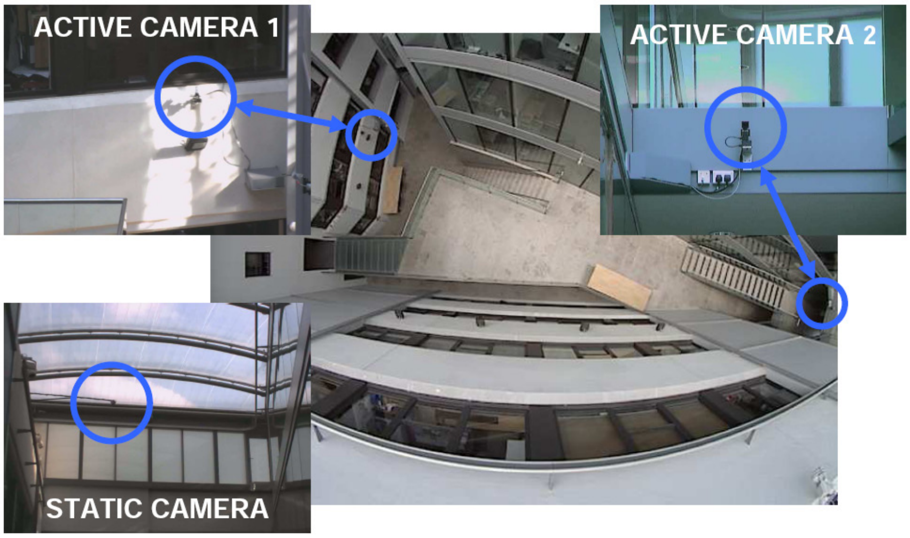
Activities & achievements

In addressing these two goals, the consortium has contributed on three ways:

- We developed an architecture for multi-camera systems comprising asynchronously communicating, heterogeneous cameras. An SQL database is used for data archiving and inter-camera communication.
- We interfaced the multi-camera system to WP6's inference engine and collaboratively created Situation Graph Trees to describe interesting activities in scenes.
- We made progress on information theoretic means to control a set of cameras. The key idea is that limited sensing resources should be directed to where they will yield the highest "information" gain.

Future directions

We have just scratched the surface of possibilities in the feedback from high-level inference to sensing process. An immediate step will be to link our information theoretic control into a live system as an internediate layer between inference and action. We also expect our ideas to extend naturally to other cognitive systems.



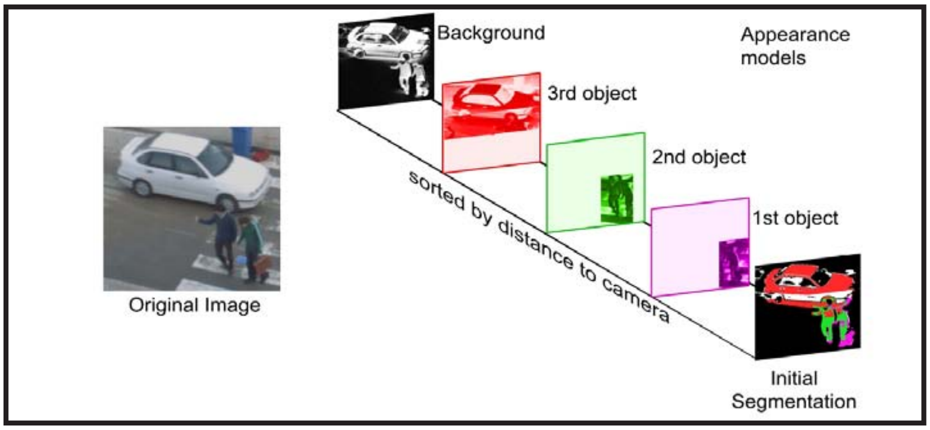
Collaborative distributed cameras (static and active)

WP2–3 leader

Ian D. Reid
ian@robots.ox.ac.uk

Evaluation of Agent Motion in Video Sequences

WP3



Tracking by segmentation

Objectives

WP3 aims to track agents and note their trajectories and other coarse scale features, in order to generate conceptual descriptions about the agents and their relationships among them and with other scene objects.

Activities & achievements

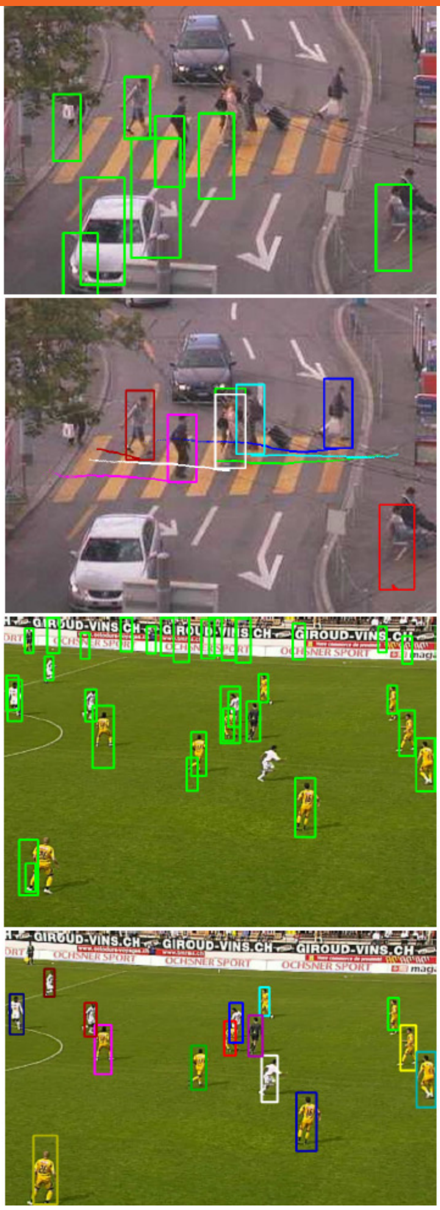
At the heart of most surveillance systems is the requirement to track targets. We developed new algorithms for figure-ground segmentation-based tracking, and integrated a real-time implementation into the final demonstrator system. We have also been at the forefront of tracking-by-detection methods, publishing at IEEE CVPR, ICCV, and BMVC, and winning a best-paper prize at CVPR 2007.

WP3 is about more than tracking: close collaboration allowed us to integrate an inference engine with the agent-based tracking data, to conduct simple real-time reasoning about the scene. This inference has, in turn, been used for natural language generation (WP6), generation of virtual agents (WP7), and camera selection/control (WP2, WP8).

Future directions

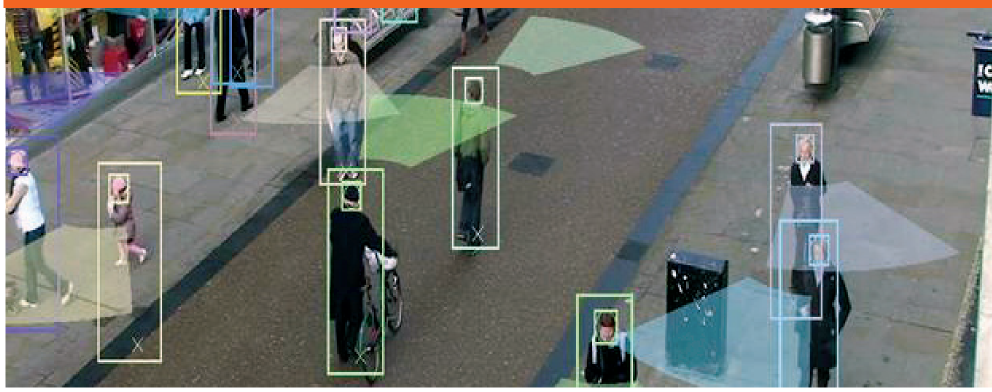
Incorporating high-level knowledge with low-level tracking data is challenging. Current tracking systems typically operate in states of perpetual surprise; algorithms try hard not to lose the target, and little intelligence is used in reacquistion. By contrast, if an activity pattern emerges, a system –like a human– can afford periods of deliberate inattention, having a sufficiently rich model to be confident of reacquisition.

Enumerating all possible situations in a SGT rapidly becomes infeasible for even small numbers of agents or vaguely complex scenes. A key challenge will be to capture scene knowledge by learning, instead of hand-crafting the SGTs.



Multiple target tracking using person detectors as input

WP4 Evaluation of Body Motion in Video Sequences



Head-pose estimation

Objectives

This WP is the core of action evaluation: already knowing the rough position of the body, derived from WP2 and WP3, it supports both the recognition and the animation of body postures. The pose and motion of the body is modelled, estimated, and associated to verbs such as *walking*, *standing*, or *bending*.

Recognized action: 'Running'

Activities & achievements

WP4 activities concentrate on:

- collecting relevant test data sets,
- pose estimation, and
- action recognition.

Pose estimation is relevant on its own (*e.g.*, attentional focus via head pose), and also a basis for action recognition. A multitude of actions such as *waving*, *boxing*, or *running* are recognized by algorithms that have different foci regarding their goal: view-invariance, real-time, recognition rate, etc., leading to distinct approaches based on features, silhouettes, or depth. The algorithms closely integrate with WP3's segmentation and tracking, yielding comprehensive systems.

Future directions

Future methods on segmentation and tracking should be more robust against partial occlusion, background clutter, and problematic lightning conditions. It is also a challenge to enhance robustness by combining multiple cameras and using depth data. Lastly, context in action recognition has shown to be a promising avenue to follow.

WP4 leader

Thomas B. Moeslund
tbm@imi.aau.dk

Evaluation of Face Motion in Video Sequences

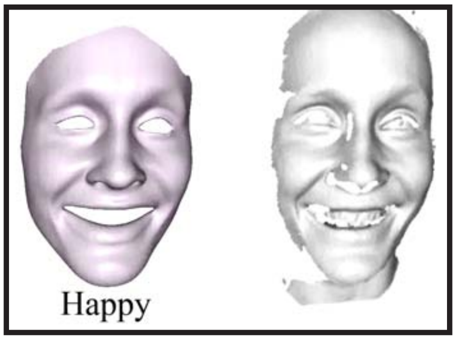
WP5

Objectives

The focus of WP5 is on the development of algorithms for facial expression processing and the analysis of facial motion in continuous image streams. The main issues that we addressed are automatic face processing, real-time performance, and several levels of detail to meet the various requirements of end-users. Other aspects are the modeling and generation of emotion descriptions which serve as input for the animation of virtual faces.

Activities & achievements

A system for emotion modeling has been specified for distributed camera systems (WP2) with pan-tilt-zoom sensors controlled by the tracked agent motions (WP3). For evaluation, a database containing image sequences of human emotions has been built and made available to the scientific community.



Recognition and transfer of an expression to a facial avatar

Several tools and methods for 2d and 3d face tracking and face alignment have been developed that allow for automatic processing of facial motion in real-time. The systems are designed to process a wide range of input streams from low-resolution 2d image data to 2.5d data acquired by active or passive stereo systems. A head pose tracker has been developed in collaboration with WP4.

Several levels for the generation of facial motion descriptions have been considered, ranging from low-level speaking and non-speaking classification over speech recognition to high quality facial emotion estimation.

In cooperation with WP7, a virtual face model has been designed based on the MPEG-4 standard to visualize the generated emotion descriptions. Furthermore, a system has been developed that accurately aligns a 3d template face to 2.5 data. By pre-recording the basic facial expressions, the system tracks and transfers the expressions of a human to a facial avatar in real-time.

Face alignment for the analysis of 2d face images

WP5 leader

Luc Van Gool
vangool@vision.ee.ethz.ch

WP6 Software Integration for Generation of Natural Language Texts


```
3@630   A pedestrian labelled as 'actor 3' appears
        in the field of view.
3@632   He moves on the southeastern sidewalk.
3@663   He walks within it.
3@766   He walks on the zebra.
3@900   Actor 3 stands near by another pedestrian.
3@928   Actor 3 moves on the zebra.
3@1085  Actor 3 stands near by another pedestrian.
3@1147  Actor 3 moves on the zebra.
3@1250  He walks on the street.
3@1254  Actor 3 chases another pedestrian.
3@1289  Actor 3 runs on the street.
```

A pedestrian chases another pedestrian.

Objectives

Imagine a user communicating with a set of distributed PTZ-cameras, as if they were humans reporting what they see. This requires converting a video into a textual description of temporal events. The user should be able to request summaries of recent developents in chosen languages; to obtain responses to his questions for details; and to send commands, *e.g.*, to zoom in onto a particular body.

Activities & achievements

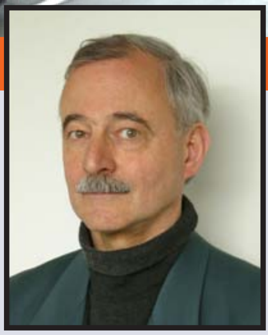
The HERMES-consortium has designed, implemented, and tested various prototypical systems to cover the different requirements set out. Currently, a slimmed-down, fixed- and PTZ-camera demonstrator system generates natural language text on activities of a particular agent (human or road vehicle) from a schematic conceptual representation expressed as a set of Fuzzy Metric-Temporal Horn Logic formulas, based on the agent's trajectory data.

Future directions

The capability to detect, represent, and describe more involved agent behavior is desirable, including stylistically adequate formulations for a larger range of behaviors. Also, a combination of the algorithmic rigor associated with a formal logic approach and the flexibility expected from a ML-oriented approach poses a challenge for future research.

Pedestrian crossing a road via a traffic island.

```
0@210  A pedestrian labelled as 'id 0' comes in
       from the north sidewalk.
0@220  He stops.
0@222  He walks on the north sidewalk.
0@231  He stops.
0@234  He stops to due a car labelled as 'id 1'.
0@239  He walks on the north sidewalk.
0@266  He walks on the crosswalk.
0@371  Now he has reached the center island.
0@401  He walks on the crosswalk.
0@503  He walks on the south sidewalk.
```

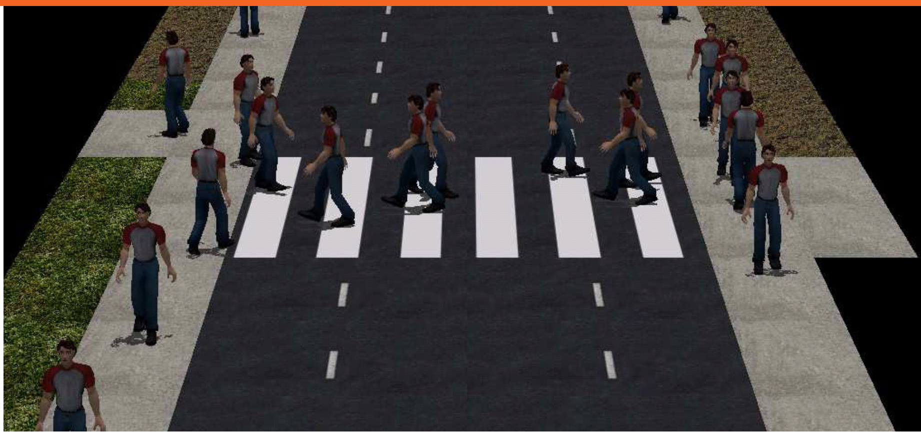


WP6 leader

Hans–Hellmut Nagel
nagel@iaks.uni-karlsruhe.de

Software Integration for Generation of Virtual Environments

WP7



Objectives

The goal of WP7 is to visualize the conceptual descriptions generated for behaviour (WP3), action (WP4) and emotion (WP5), joining them into a common animation framework. WP7 generates synthetic animations reproducing sequences of predefined semantic primitives, provided beforehand or derived from end-user's textual descriptions.

Activities & achievements

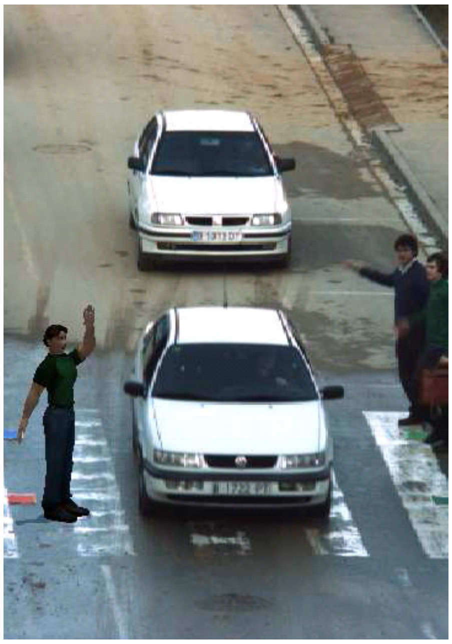
Producing virtual components includes:

- Conception, creation and visualization of a virtual scene;
- Its texturing and illumination;
- 3D-modeling of virtual objects and human characters; and
- Their inclusion into the virtual scene.

Producing virtual human motion implied animating the characters with the motion descriptions used thorough WP3-5. The integrated motion was examined for consistency using parameters of emotion and style, to find stable links between both kinds of animations. Particularly, we investigated whether a given emotion may influence the animation of particular actions or behaviours.

Additionally, generating virtual animations based on conceptual descriptions can be seen as inverting WP6's natural language generation: here, the semantic primitives related to behaviours, actions, and emotions are visualized using computer graphics. Since a virtual environment contains all motion and position information on all agents and objects, interactions can be revisited.

This system was evaluated for accuracy of visualization and degree of interaction. As a result, behaviour descriptions are visualized by means of virtual actors.

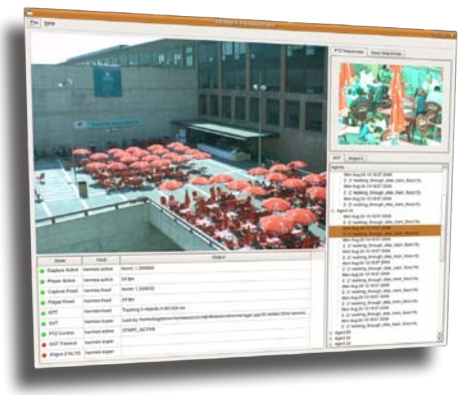


Virtual actors generated from users' descriptions and reacting to real scenes.

WP7 leader

Jordi Gonzàlez
poal@cvc.uab.cat

WP8 Validation and Demonstration



Objectives

WP8 focuses the integration, demonstration and validation of the other WPs' outputs. A major goal of HERMES is to establish how the analysis of agent, face and body motion, plus synthesis of these outputs in the form of inferred, high-level interpretation should be integrated into a cohesive platform for experimentation, validation and demonstration. In WP8, the outputs of all WPs were pooled and integrated to build integrated hardware and software systems to meet such objectives.

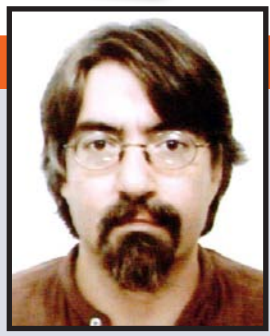
Activities & Achievements

In the final year of the project, the consortium turned its attention in earnest towards realizing both real-time and offline demonstration and validation systems.

Building on the prototype indoor demonstrator developed in WP2 and WP3, an outdoor scenario was designed to probe the effectiveness of the HERMES approach to image visual cognition in multi-sensor networks. An integrated hardware platform was designed, built and installed on the rooftop of the CVC building in Barcelona. This hardware platform consists of two high-speed cameras (one fixed and one pan-tilt-zoom), and three dedicated servers to host HERMES systems for analysis of agent motion, for active camera control, and for inferring high-level descriptions of agent behaviour in the scene. To further support demonstration activity, an integrated and modular software platform was designed, allowing modules to be developed off-site and later integrated into the platform with minimal effort.

Real-time tracking and analysis of agent motion results from WP3, active camera control from WP2 and inference of agent behaviors from WP6 and WP7 have been incorporated into the demonstrators.

Real-time demonstrator interface controlling the fixed camera (top), active camera (bottom left), and split view of both cameras (bottom right).



WP8 leader

Andrew D. Bagdanov
bagdanov@cvc.uab.cat

Exploitation and Dissemination

WP9

Objectives

WP9 promotes the dissemination of the project results to the wide public, making them well known throughout in the scientific community and the European industrial companies. It also promotes the industrial and market exploitation of the project results.

Activities & Achievements

Many interchanges of research people among the HERMES partners took place during the whole project duration. The research partners have disseminated the results of the project with publications and presentations in events and conferences. Technical papers and articles in scientific and professional journals and specialized magazines have been published demonstrating the work that has been carried out during the project. Presentations at national and international conferences and forums have been to be given (e.g. ICPR, CVPR, ECCV, ICCV, and other international and national conferences and scientific journals, like IEEE Transactions on Pattern Analysis and Machine Intelligence, Computer Vision and Image Understanding, Pattern Recognition Letters, etc).

The construction of the HERMES website, **http://www.hermes-project.eu**, has greatly contributed to disseminate the project results. The website contains demonstrators of the project highlights and other results supplied by the partners. It has also been used to archive reports among the partners.

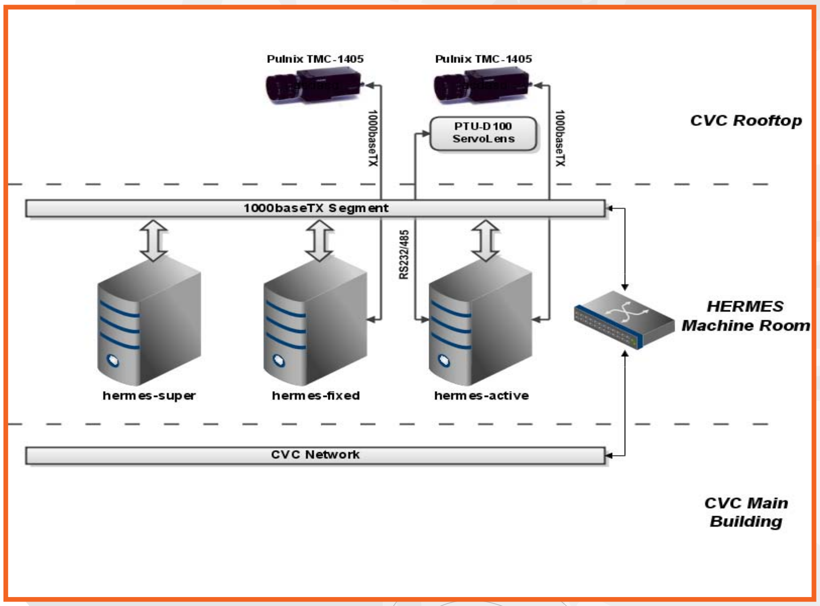
The project partners will exploit the project results by using the basic principles of HERMES in commercial projects (e.g. smart surveillance for security projects) and other research activities.

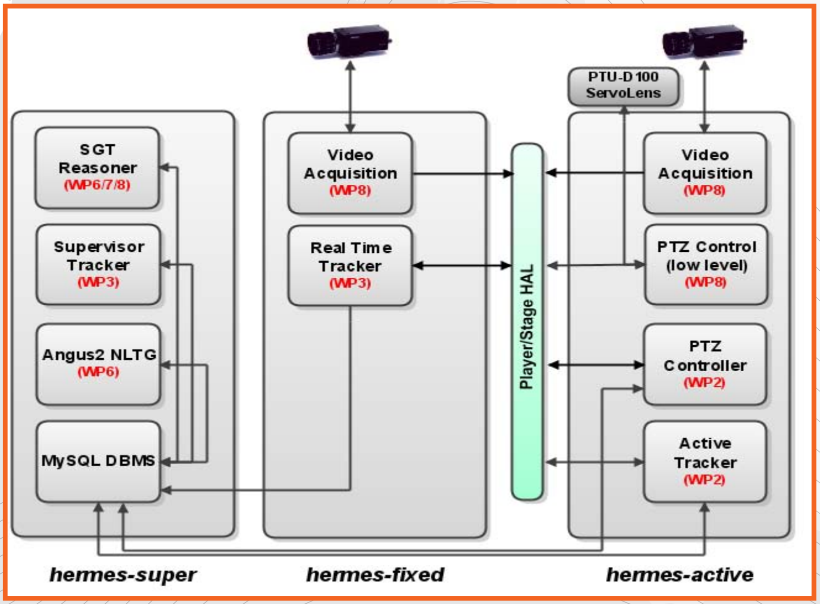
WP9 leader

Tonny Velin
tvelin@answare-tech.com

Showcasing activity

"A hardware and software platform integrates the components of HERMES"





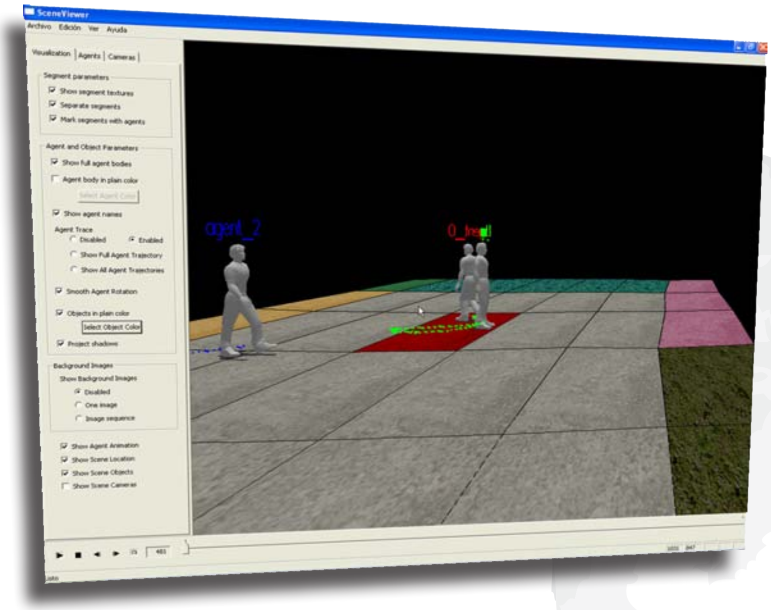
HERMES showcase activities have been concentrated in the form of an integrated hardware and software platform that consolidates the various components of the HERMES approach.

In concert with the demonstration and integration activities of WP8, two integrated graphical user interfaces have been implemented to demonstrate and showcase the performance of HERMES modules:

- An interface for interacting with the **real-time demonstrator** was built that allows the user to configure, administer and monitor the components of the HERMES demonstrator platform. This interface also allows the user to visualize streaming video from both cameras in real-time, as well as tracker output in the form of 3D annotations on live video. Inference results from analysis and reasoning about agent motion is also presented to the user in this interface, along with captured detail images from the PTZ camera.
- To complement the initial real-time demonstration system, a second **offline demonstrator** for animation of virtual environments, based on the outputs of WP7, was constructed that bases its virtual reconstruction of agent behaviour on the outputs of the real-time demonstrator.

The end result of these efforts is a flexible and modular system for showcasing some of the major developments from the HERMES project.

Head of Hermes from a hermaic stele
Ancient Agora Museum of Athens, Greece



"Two demonstration systems, one real-time and one offline, showcase the major developments of HERMES"

Hermes bearing the good person by Praxiteles
Archaeological Museum of Ancient Olympia, Greece

Expertise of the Partners

CVC

The CVC is a non-profit institution founded in 1995 by CIRIT (*Consell Interdepartamental de Recerca i Innovació Tecnològica*), CIDEM (*Centre d'Innovació i Desenvolupament Empresarial*) and UAB (*Universitat Autònoma de Barcelona*). The strategic objective of CVC is to do both Research and Development of quality on Computer Vision. From a scientific point of view, the CVC wants to contribute to increase the knowledge in this field. From a technological point of view, the CVC aims to contribute to innovation and industrial competitiveness collaborating with companies to develop technological projects.

CVMT

The Computer Vision and Media Technology Laboratory (CVMT) is part of the technical faculty at Aalborg University, Denmark. CVMT conducts research in areas such as 1) computer and robot vision, 2) computer graphics, 3) motion analysis, 4) color vision, 5) autonomous systems and agents, 6) multimedia and virtual reality (VR) systems and interfaces, and 7) operational vision systems for both industrial and medical applications. CVMT has been involved in about 15 European projects and a number of national projects. Currently CVMT is a part of and coordinating one EU project and three national project.

IAKS

The *Fakultät für Informatik der Universität Karlsruhe* (TH) is one of the leading faculties for Computer Science in Germany. The research group *Kognitive Systeme (KOGS)* lead by H.–H. Nagel is affiliated with the *'Institut für Algorithmen und Kognitive Systeme' der Universität Karlsruhe*. This group addresses all problems encountered in the algorithmic transformation of image sequences into representations of the semantics of the depicted scenes and their temporal development, covering data-driven treatment of video signals via model-based approaches towards detection and tracking up to logic-based representations and their conversion to natural language text.

BIWI

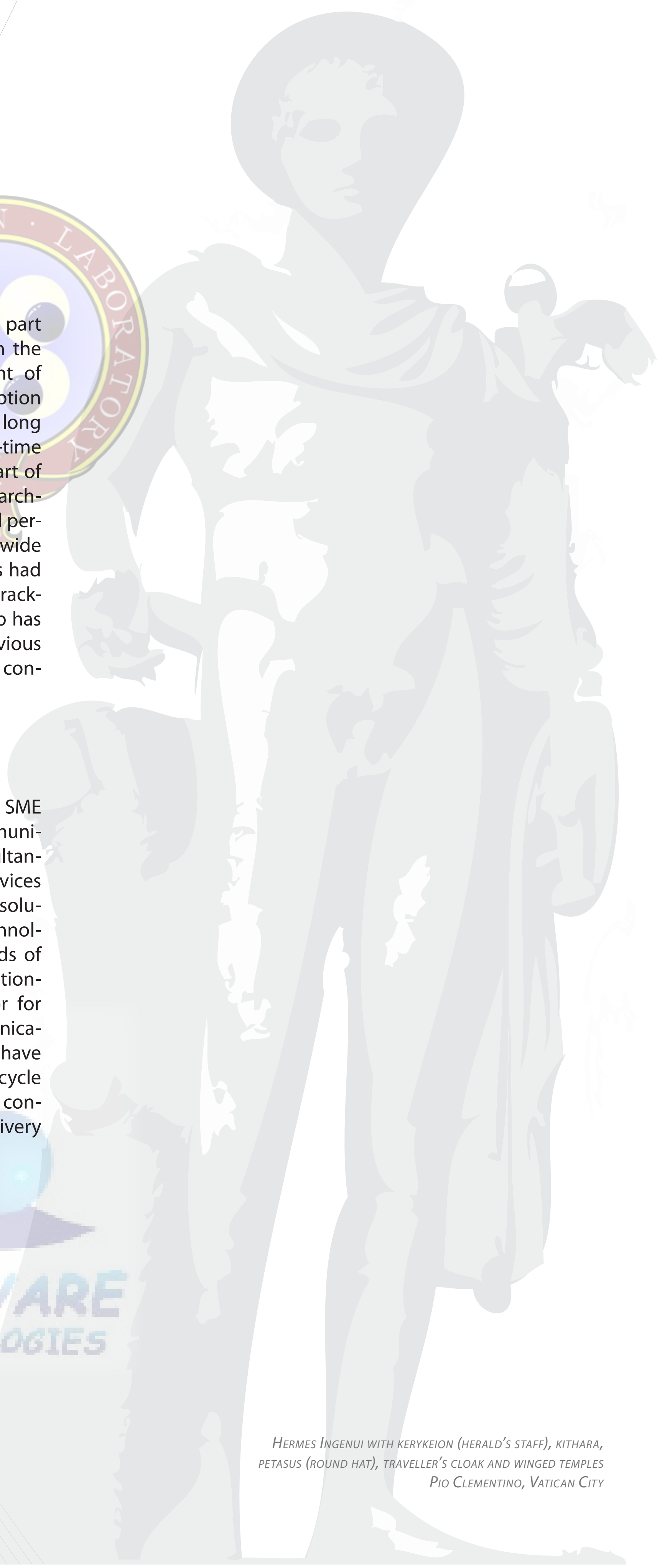
At the Eidgenoesssiche Technische Hochschule (ETH) Zurich the participating team is the Computer Vision Laboratory (BIWI), which is part of the Dept. of Electrical Engineering D-ITET. The work of the group focuses on scene understanding, medical imaging, and remote sensing. Here mainly the expertise in the area of scene understanding is relevant. This includes work on tracking, gesture analysis, and motion capture. The group has developed multiple trackers, ranging from generic, fast blob trackers up to trackers for articulated structures like hands or full bodies. The group has been involved in many European and Swiss projects.

AVL

The Active Vision Laboratory is part of the Robotics Research Group in the University of Oxford's Department of Engineering Science. Since its inception in 1991 the lab has established a long record of research in active, real-time and robotic vision. A substantial part of the group's effort is invested in researching the theory of visual sensing and perception and its application to a wide variety of problems. The group has had substantial success in both visual tracking and in 3D geometry. The group has been involved in a numerous previous European projects. Its faculty have consulted widely for major companies.

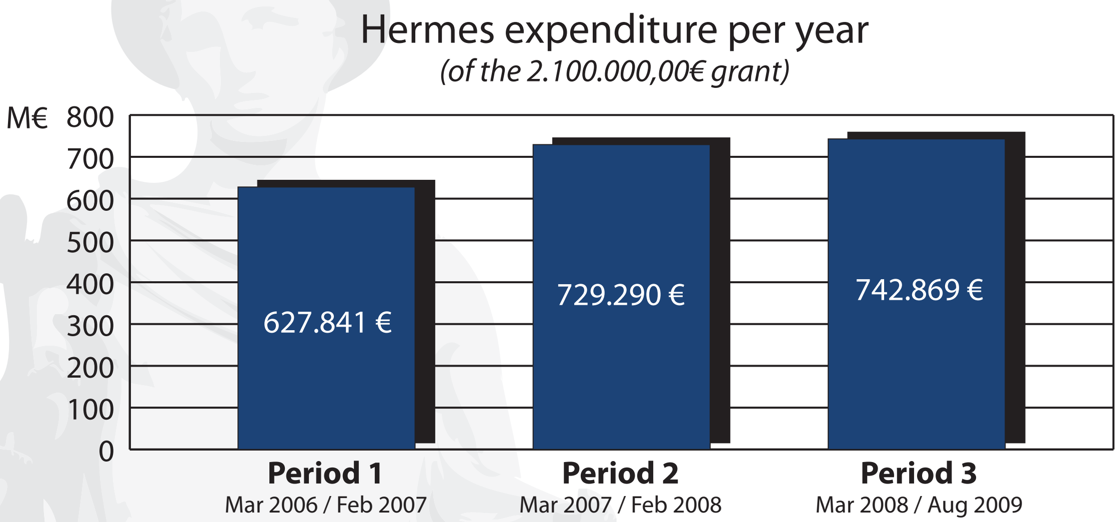
ANSWARE

Answare Technologies is a Spanish SME specialist in Information and Communication Technology, providing consultancy and projects development services and offering a variety of high-tech solutions based on state-of-the-art technologies, tailored to the specific needs of customers. Answare operates in national and international market sector for Aerospace, Defence, Telecommunications and Internet. Answare's staff have participated in the complete life cycle for software systems projects from conception through development, delivery and into maintenance.



Hermes Ingenui with kerykeion (herald's staff), kithara, petasus (round hat), traveller's cloak and winged temples
Pio Clementino, Vatican City

Finances



The HERMES project has been allocated a budget of 2.100.000 € by the European Commission to fulfil the objectives of its work programme. This grant was distributed over the 42 months period of the project (36 months + 6 months extension), from February 2006 to August 2009. The grant was divided among all the partners, according to their activities and commitment in the project.

"The HERMES project has been allocated a budget of 2.100.000 € by the European Commission"

Hermes wearing the petasus, a voyager's cloak, the caduceus and a purse
Museo Chiaramonti, Vatican City

Authors:
Written with the contribution of all HERMES Partners

Designed by:
Carles Fernández Tena

http://www.hermes-project.eu

September 2009

Information Society
Technologies

SIXTH FRAMEWORK
PROGRAMME