mobvis

Vision Technologies and Intelligent Maps
for Mobile Attentive Interfaces in Urban Scenarios

www.mobvis.org

Summary

The MOBVIS project identifies the key issue for the realisation of smart mobile vision services to be the application of context to solve otherwise intractable vision tasks. In order to achieve this challenging goal, MOBVIS claims that three components,

(1) multimodal context awareness,
(2) vision based object recognition, and
(3) intelligent map technology,

should be combined for the first time into a completely innovative system – the attentive interface.

In multimodal interface design, "Attentive User Interfaces" are an emerging paradigm with the emphasis to generate interface control directing attention in human and machine. MOBVIS conceptually and functionally extends this framework to "Attentive Interfaces" that involve machine attention mechanisms, i.e., to selectively index into information spaces dependent on a given user relevant context.

The attentive interface would be the only possible way to cut down the numerous system's hypotheses on the real world, by first aggregating context information, then applying context to make mobile vision based object awareness feasible, and incremental updating of map based geo-information to provide a knowledge base for future context exploitation.

The project goal is research to identify and investigate on key challenges for the development of mobile vision interfaces, and providing a demonstrator interface that is attentive to objects of interest in urban scenarios, such as, buildings, infrastructure, informative icons and text.

The detected objects will together define contextual situations, they cue to expected places, objects, information and events, and they feed into an augmented digital map representation, as a basis for enriched smart services in personal assistance for the mobile and automotive industry.

Mobile vision will become a fundamental technology for enhanced perceptive presence, context aware and attentive interfaces, and urban environments provide the scenarios for emerging applications. The MOBVIS project overtakes the impetus of upcoming mobile imaging, and encounters the challenges to integrate computer vision and intelligent map technology, in order to enable mobile vision services become reality in the near future.

Objectives

The main objective of the MOBVIS project is to explore and exploit the concept of attentive interfaces for the design and implementation of mobile vision technology in urban scenarios. The claim is to make mobile image understanding possible by intelligent interaction between three major functional components, i.e., advanced computer vision for object awareness, exploitation of multi-modal context, and intelligent access to map knowledge, where each can boost performance of the individual components and the system as a whole. That involves making attentive interfaces a first-class concept for mobile situated intelligence, involving fundamental research on Artificial Intelligence enabled computer vision methodology, which will be a decisive step forwards in mobile perceptual presence.

By interfacing these components under the decision making of the attention control interface, we will provide a new way for mobile multimodal context awareness to connect with advanced computer vision. We will show the potential of this new technology by going beyond location based services with simple signal and co-ordinate based relations, towards visually interpreting the world to emerge object awareness for the benefit of future enrichment of mobile services on multimodal interfaces.

Information Society Technologies

Project information: MOBVIS is a Specifically Targeted Research Project (STReP) funded by the European Commission | IST-2002-2.3.4.1 FET open | Project Nr: FP6-511051

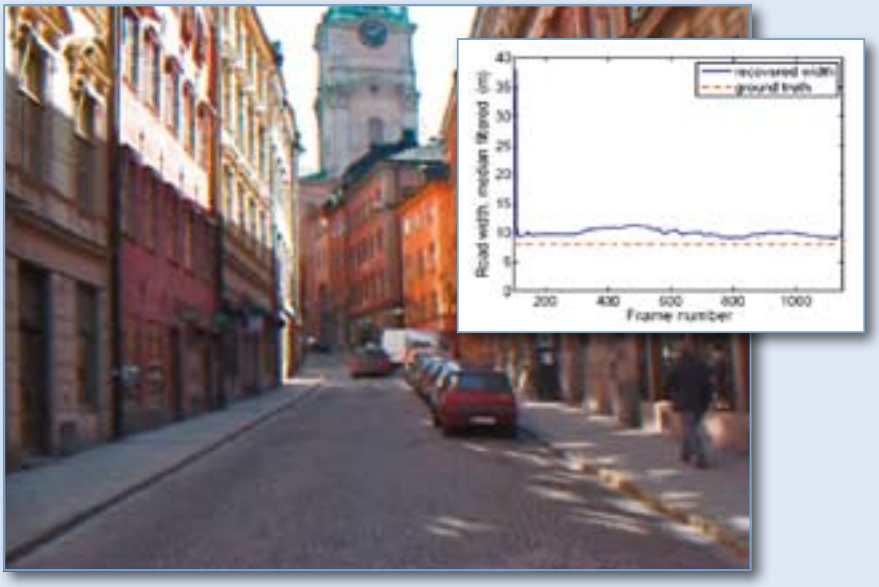
Visual Localization

The user image is automatically localized by relating the image to the MOBVIS image database. Via triangulation the user's position and orientation is determined, yielding accuracies comparable to GPS. In addition, image-based localization enables novel services, like hyperlinking reality or geo-referenced object detection.

The illustration shows a query image (blue frame) and some reference images (green frames) used to position and orient the query image and consequently the user.

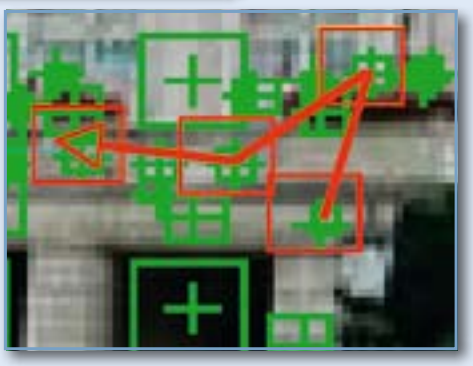
Some geometric relations relating the query image with one of the reference images are indicated by the dark green lines.

Visual Attention



Strategies of attention naturally refer to a cascaded processing of – potentially – different visual features, each indexing to a certain coverage of an associated search space. A first step in the cascaded processing is to localise categorical visual features, those that would relate to a specific set of objects, or, inversely, to relate to background information, such as vegetation and cobblestones.

MOBVIS developed a multi-cue attention system that combines bottom-up and top-down influences. Sequential attention was developed to exploit geometrical constraints for object recognition by a concept that is inspired from human attention and eye movements.



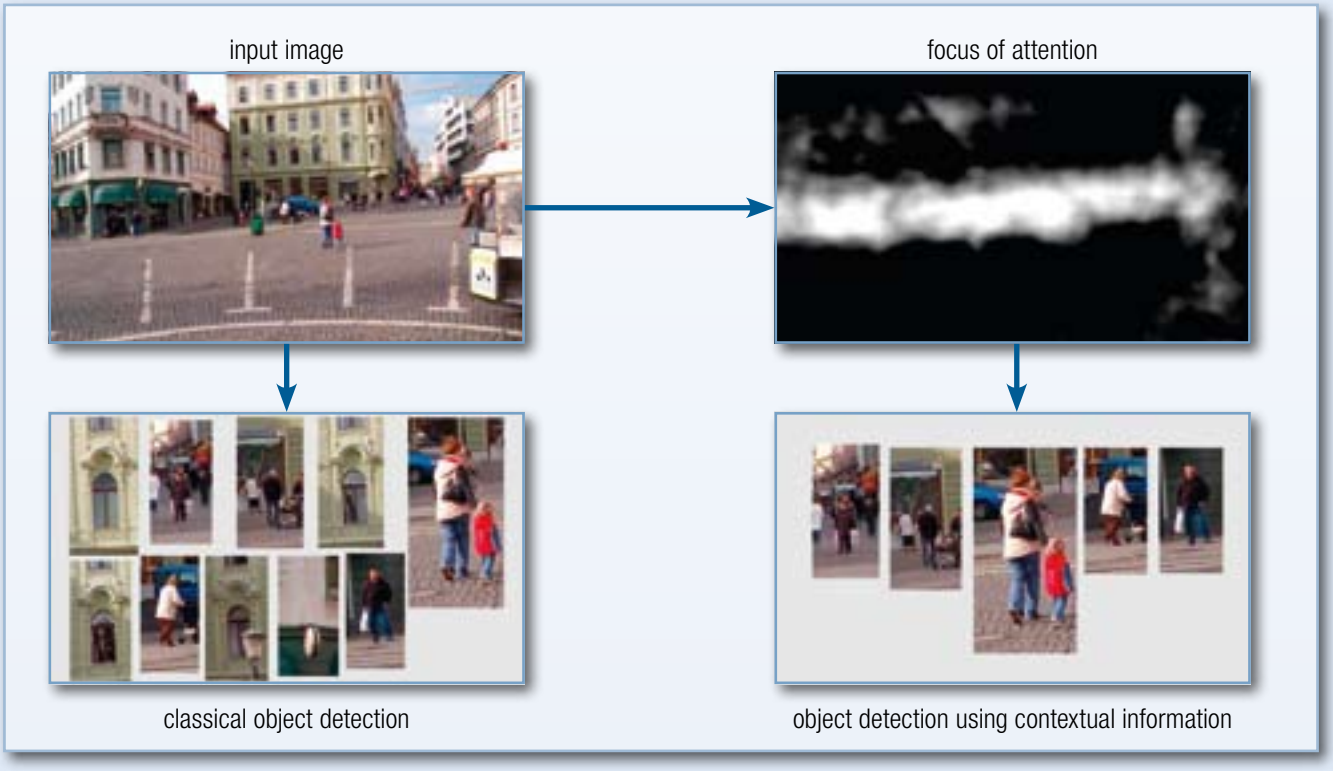
In addition, the extraction of street profiles from 3D information recovery supports indexing into city maps for location awareness.

Object Awareness

Object awareness is investigated to detect and recognise objects of high interest in urban scenarios, such as, buildings, infrastructure, people, and signs. MOBVIS demonstrates how geo-indexing significantly improves performance in mobile object recognition by exploiting the information of augmented digital city maps. Query image and GPS based position estimate are sent to the server which responds with results from the geo-indexed object recognition. Furthermore, visitors might be informed with annotation, including history, event and shop relevant information, about the point of interest.

Visual Context Awareness

MOBVIS provided a concept of vision based context on how to extract, learn and use contextual features to guide object detection. Three complementary types of contextual features are proposed: viewpoint prior, geometrical context and textural context. The concept aids the detection process, yielding speedup and increasing detection accuracy. Examples are shown for pedestrian detection.



Multimodal Context

Activity is an important source of context information. MOBVIS explored methods for unsupervised activity modelling based on signals from multiple body-worn sensors, including accelerometers. For a given set of long-time captured information it was possible to build models that correspond to different everyday activities, including eating and shopping, and without requiring a prior training, user annotation or information about the number of tasks involved.

Multimodal Positioning

MOBVIS introduced new outdoor positioning possibilities that are offered by combination of GPS and WLAN positioning, as well as motion estimation by dead reckoning and state-of-the-art vision positioning. The combination of vision-based technology with incremental positioning has found to enable continuous position estimates, making it directly comparable to standard techniques such as GPS and WiFi. Interestingly, computer vision has shown to enable localization accuracies comparable to GPS.

Red: Groundtruth

Blue: GPS, WiFi, DRC

Yellow: GPS

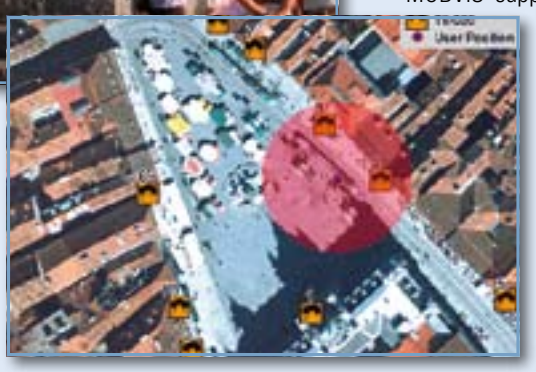
Green: Vision & DRC

Augmented Digital City Maps

Vehicles are collecting data about urban infrastructure for the definition of map features and points of interest, including geo-referenced images, traffic infrastructure and tourist sight information. Map features are stored in and provided to mobile vision services by the Mobile Mapping Data Warehouse of Tele Atlas. Standard digital city maps are augmented with these data as a support of mobile vision services. User track and image reference data are visualised and can be interactively accessed in the MOBVIS user interface.

Geo-Services & Incremental Map Updating

Geo-services are responsible for the interaction with the map based geo-information knowledge. A complex functional interface to the digital map information has been defined in MOBVIS for the capabilities to realise appropriate responses to requests from the MOBVIS system components e.g., under variation of the spatial scope and the quality of the request on geo-information, and for the provision of specific information to the vision module to generate object hypotheses. Geo-services enable intelligent user position and orientation based filtering of surrounding objects for geo-indexed object recognition and analysing of map features for real-time context detection.



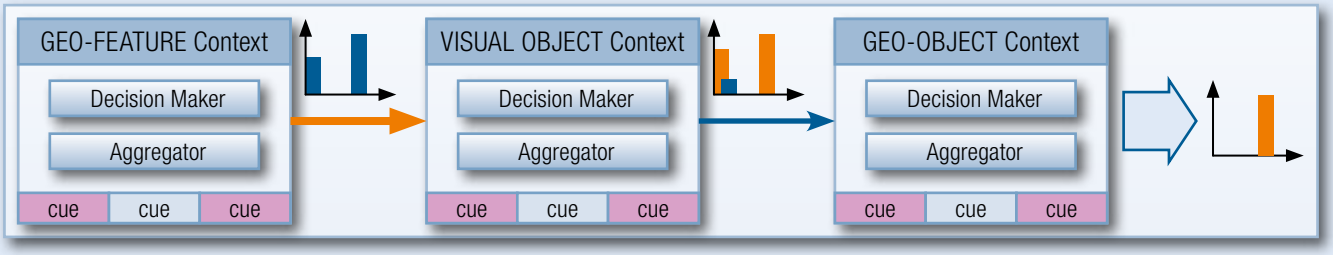
MOBVIS supports incremental updating of maps and therefore automated authoring of urban infrastructure, including road furniture, public transport, and public objects, such as coffee shops.

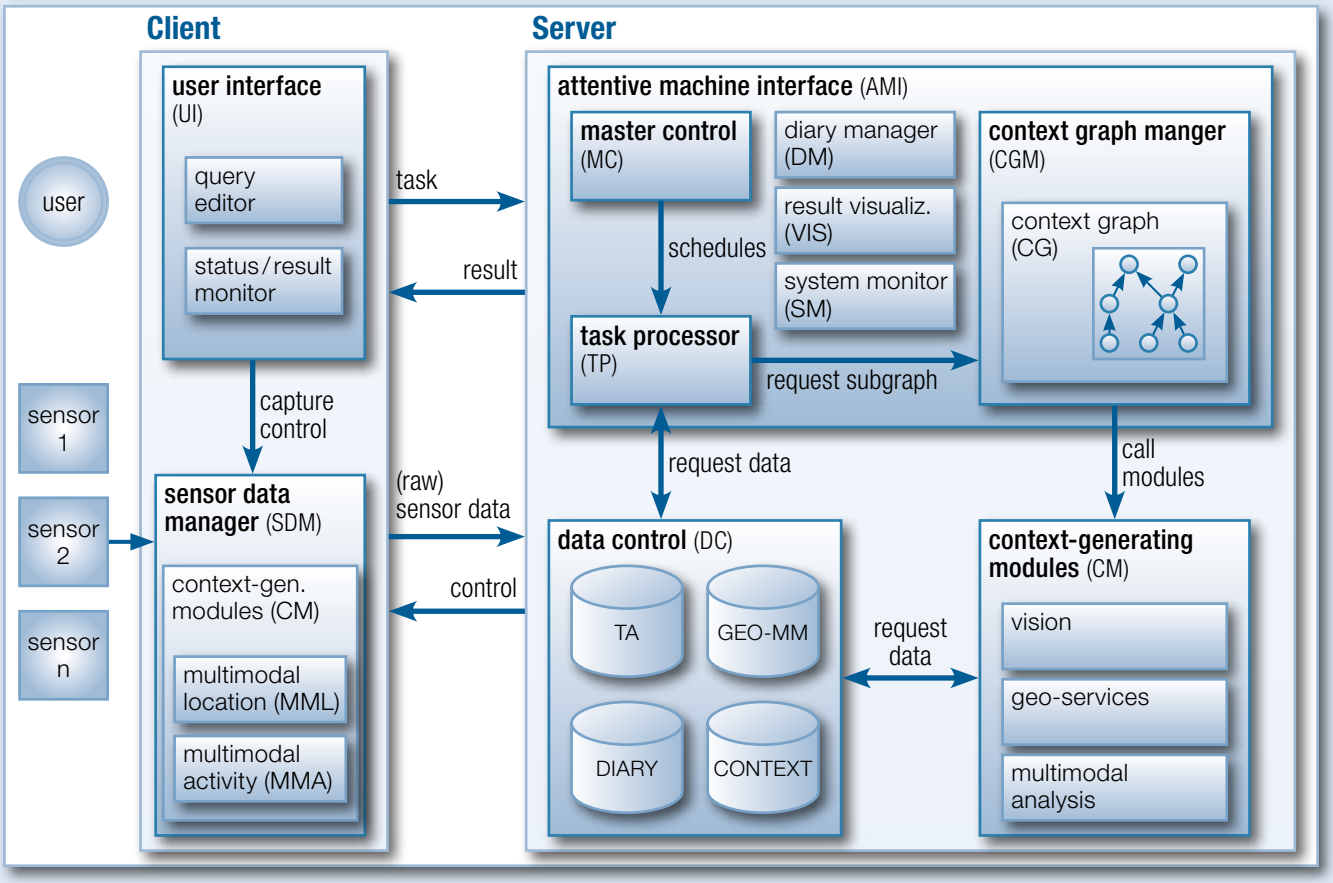
Attentive Interface

The context framework used in the Attentive Machine Interface (AMI) defines a cue as an abstraction of logical and physical sensors which may represent a context itself, generating a recursive definition of context.

Sensor data, cues and context descriptions are defined in a framework of uncertainty. The architecture of the AMI reflects the enabling of both bottom-up and top-down (attention driven) information processing.

Attention enabled by the AMI means focusing operations on a specific detail of a situation that is described by the context.





Partners & Expertises

JOANNEUM RESEARCH Forschungsgesellschaft mbH, Institute of Digital Image Processing, Austria
Lucas Paletta (coordinator), Alexander Almer
Computational Perception / Geo-Visualization and Mobile Computing | Mobile computing, geo-services, object recognition, attentive systems
Wastiangasse 6, 8010 Graz, Austria | Phone +43 (316) 876-1769 | Cell +43 (699) 1876-1769 | Fax +43 (316) 8769-1769
www.joanneum.at | dib@joanneum.at

University of Ljubljana,
Cognitive Systems Lab, Slovenia
Aleš Leonardis
Computer vision, object recognition, visual positioning, visual context
http://vicos.fri.uni-lj.si

Royal University of Technology, Computer Vision and Active Perception (CVAP), Sweden
Jan-Olof Eklundh
Computer vision, 3D information recovery, attention
www.nada.kth.se/cvap

Technical University of Darmstadt,
Multimodal Interactive Systems, Germany
Bernt Schiele
Multimodal systems, fusion and recognition, multimodal context
www.mis.informatik.tu-darmstadt.de

Tele Atlas N.V.,
Strategic Research, The Netherlands
Linde Vande Velde
Digital maps, geo-information
www.teleatlas.com

www.mobvis.org