

# Geo-Services and Computer Vision for Object Awareness in Mobile System Applications

## *Authors*

*Patrick LULEY, Lucas PALETTA, Alexander ALMER*

JOANNEUM RESEARCH Forschungsgesellschaft mbH, Institute of Digital Image Processing, Wastiangasse 6, A-8010 Graz, Austria

*Mathias SCHARDT*

Graz University of Technology, Institute of Remote Sensing and Photogrammetry, Steyrergasse 30, A-8010 Graz, Austria

*Josef RINGERT*

Graz University of Technology, Institute of Geoinformation, Steyrergasse 30, A-8010 Graz, Austria

## *Abstract*

In recent years, location and context aware systems have been presented for the indexing and annotation of both location and user state relevant information to the user. These systems were mostly based on geo-referencing from GPS signals, and driven by changes of the system state in dependence on situations that impact the overall system performance (user pose, energy consumption, responsiveness, etc.).

In this paper we focus attention on a completely innovative aspect of contextual indexing for mobile system applications. We claim that the application of geo-services and vision based sensing enables to determine *object* characteristic information that can be used for *semantic indexing* of information and thereby provides an innovative quality of service. Computer vision can be used to extract object information, such as, the identity of buildings (Figure 1), information signs, persons, etc., and thereby enables object based indexing which can finally be applied to extract a semantic description of the environment. Geo-services are mandatory to support vision based object recognition and semantic indexing in several ways. Firstly, location based sensing provides a prior on the geo-reference of the objects in the field of view, reduces the number of possible object hypotheses and therefore can dramatically simplify the complexity of recognition. Secondly, vision provides an estimate of the user position from the geo-referenced object information. Geo-services can use this to reason more efficiently about a current object and user position.

## *Introduction*

Computer vision is an emerging technology for providing reliable vision competences outdoors, dealing with the immense complexity of the visual input and associated degrees of freedom in image understanding. Mobile imaging as an upcoming technology will become ubiquitous such as in camera phones, vision enhanced handhelds, and wearable cameras. Mobile vision technology would make new approaches possible in scenarios of personal assistance, mobile work, assistive services and also in general location based services.

Services, based on location awareness, require the knowledge about the actual location of the user, the current user context and geo-referenced information about areas and points of interest. Different technologies can be used to fulfill these requirements. Location awareness can be provided based on GPS, using wireless network technologies such as GSM and WLAN or using self-location possibilities e.g. based on street names and house numbers. Location awareness for a mobile service in urban areas can, by the use of GPS only, not be assured everywhere and anytime, because of the known weaknesses of GPS signal availability in

urban areas. Mobile systems operating in urban environments must take advantage of contexts arising from the spatial and situated information at a current location of the pedestrian user. Therefore, location awareness can be realized based on the knowledge of the location of the geo-referenced objects of interest, which allows also determination of the user position.

In frame of the projects MOBVIS (“Vision Technologies and Intelligent Maps for Mobile Attentive Interfaces in Urban Scenarios”, EC funded project) and “Mobile City Explorer”, which is a national funded project (Austrian Space Agency), a scenario of city exploration (see Fig. 1) is proposed [LPA05], where the tourist pedestrian can send visual queries to a server based object recognition system. The goal of the system is to attain situated object awareness, i.e., to get the semantic information about the environment for object and location identification, to get semantically indexed access to knowledge data bases. This methodology has the advantage that the object information is most appropriate to structure perception, and to structure the indexed search into annotation information.

In the following chapters a mobile application system (see Fig. 2) will be described which offers area-wide location awareness based on a common smart-phone and image based object recognition tool in combination with a GPS module.

### *User Scenario*

This Section briefly describes a user scenario, in case of a city-tourist type pedestrian, focusing on the service of image based object recognition.

The common way of doing city sightseeing is using a printed city map with integrated sightseeing-tours leading the tourist along a pre-defined path from one sight to the next. Brief descriptions of the sights can be found at the backside of the map.

By the use of the image based object recognition service the tourist gets the freedom to explore the city without any pre-defined sightseeing tours.

The tourist moves completely free through an unknown area and if he is interested in any object (e.g. a historical building or a statue) he just has to take a picture of it with his camera-phone, with or without GPS device connected, and pressing the “Identify” button. As result he gets a detailed description of the object containing multi-media tourist information. As a second achievement he also gets the position of the identified object which can be used for navigation. As the position is well known from the GPS device or the object recognition the Geo Services are able to render a map showing the surrounding area. This map is also sent to the tourist as part of the response of the object recognition and gives him an overview of the near ambience.

Our user scenario and the role of image recognition in mobile tour guides was well-investigated in an usability study by Nigel Davies, Keith Cheverst, Alan Dix and Andre Hesse [DCDH05]. The result was that there is a very high acceptance for image based object recognition by tourists.



**Fig. 1.** Scenario: capturing an image about a building object of interest that is to be identified by the system.

### ***System architecture***

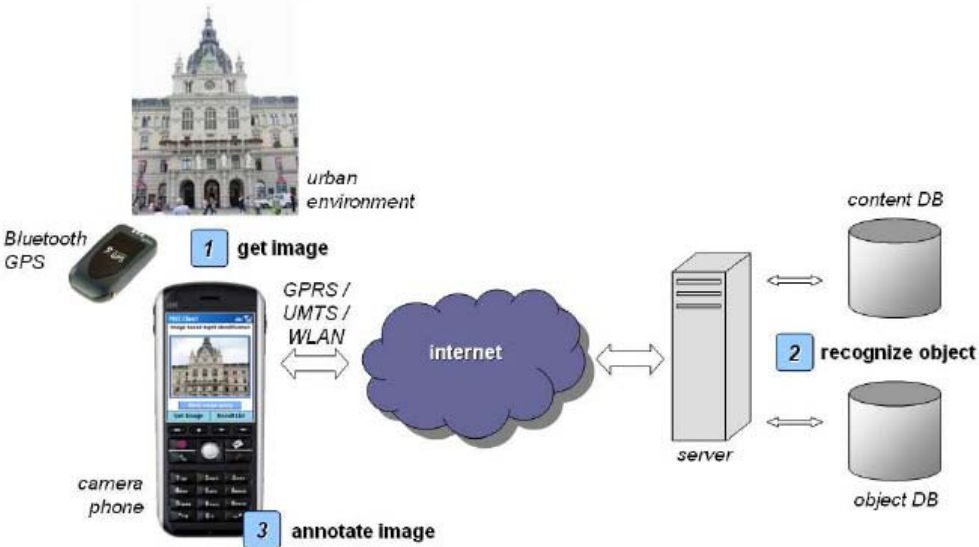
We will briefly describe how a common smart-phone with built-in digital camera can be used for image based object recognition. A GPS device, built-in or connected to the phone via cable or Bluetooth can help to accelerate the recognition process considerably. The overall concept consists of three main phases.

In the first phase, a software client is activated by a user on his personal smart phone. The software can be directly downloaded from a website to the internet enabled smart-phone and will then be installed on the device. The software-client offers functionality to capture an image of an object the user intends to identify. Next, if available, the smart-phone reads the actual position from the GPS device. If the GPS cannot obtain a position for any reason, the cell information of the phone-network provider can be used to approximate the user location instead. The image of the object and the position of the snapshot-location are combined together into a SOAP Message and send to an image-recognition web-service, running on a dedicated server, over a common wireless internet connection like GPRS or UMTS.

In the second phase, the web-service reads the request from the client (smart-phone) and extracts the image and the GPS position. The image is then analysed by a dedicated recognition algorithm to obtain representative image features like local descriptors that characterize a specific object. Next, these features are compared with reference features stored in an object-database. The database contains images of different objects, with the pre-processed features and the position of the snapshot-point. The object is then identified by matching the features of the user-picture with the features of in the database. This matching process can become, in the case of a large database with many objects, very time consuming and that is where the GPS position can help. To accelerate this process the set of objects under investigation in the database can be selected with the users GPS position. Only those database

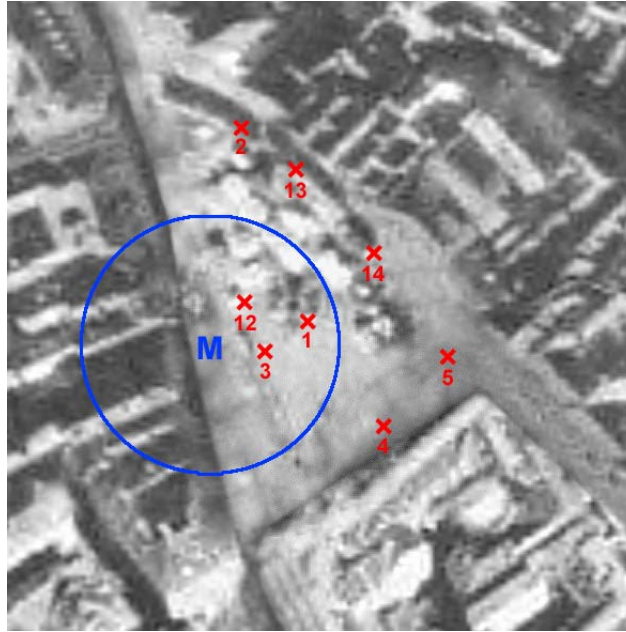
objects come into account for the matching process, which have corresponding images with a snapshot-point near the users-position. (Fig. 3) Once the object is identified the object-database is queried for some information containing text and pictures. This information is integrated into a SOAP-message and is then send to the smart-phone client as response.

In the third phase the web-service response is presented to the user on the smart-phone. Additional to the object quick information, containing text, pictures and a location based city-map, there is an URL which can be used by the user to obtain detailed information about the object. The URL can be viewed in a common smart-phone internet browser like “Opera” or “Pocket Internet Explorer“. The URL contains, as a parameter, the unique identification number of the desired object and links to a dynamic website, which is generated at runtime on the server. The layout of the website is optimized for the requesting hardware platform – because different smart-phones have different display resolutions. Fig. 2 illustrates the complete technical concept and its three phases.



**Fig. 2.** Overview of the mobile application systems concept

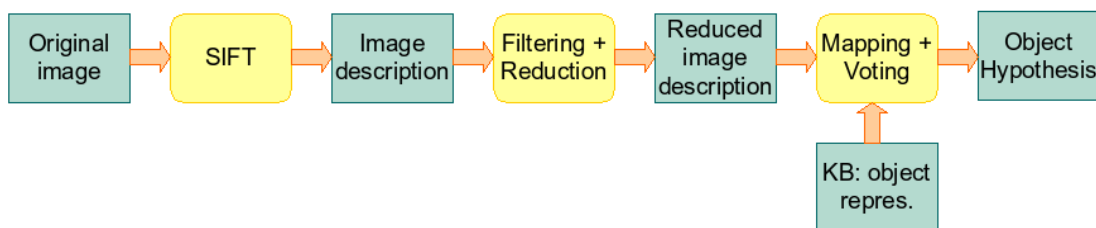
The aim of this client-server architecture is to bring the image based object recognition service to any person using a common camera-phone and to gain scalability in reference of the number of objects in the database and complexity of the image-recognition algorithms.



**Fig. 3.** Geo-Context, e.g., from GPS based position estimates ('M' with blue uncertainty radius), can set priors by geographically indexing into a number of object hypotheses ('X's are coordinates of user positions while capturing images about objects of interest).

### *Visual Object Recognition for Object Awareness*

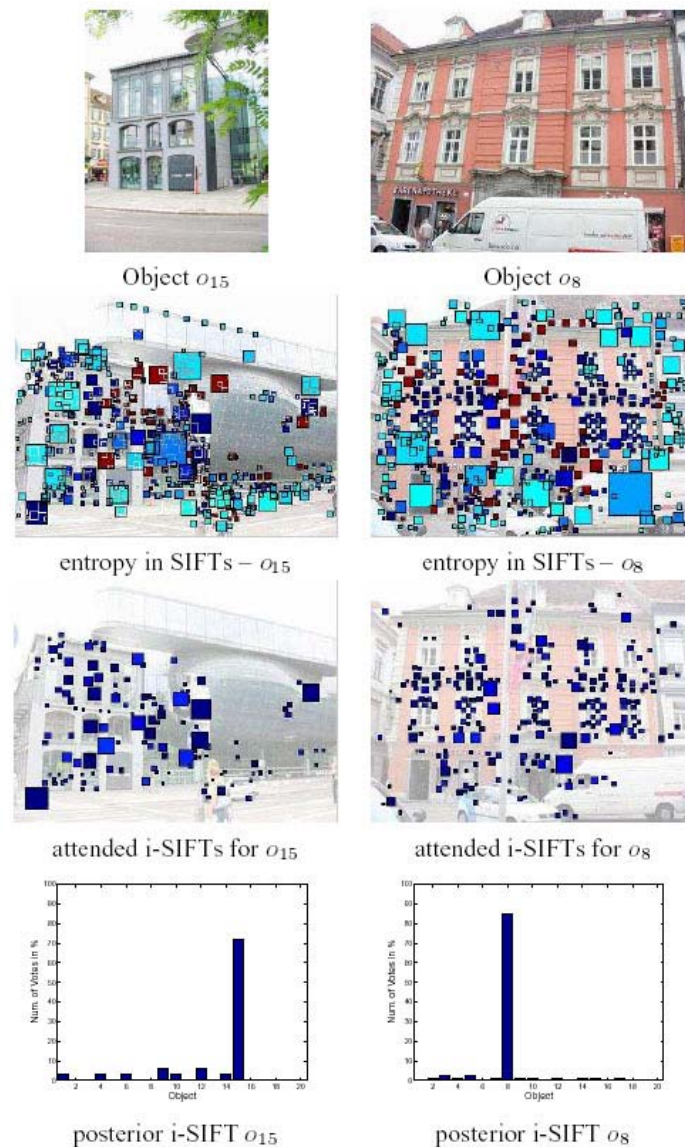
In contrast to signal based approaches as proposed in most location based services, object awareness points towards semantic indexing which will enable much more flexible interpretation of the information from a local environment. Object awareness relies on a structural matching process of comparing the situated information extracted from the sensor streams with prototypical patterns that were developed from experience. In the case of computer vision, object recognition methods enable to identify characteristic patterns of visual information in the field of view, such as, infrastructure objects (tourist sights, traffic signs, information boards, etc.), people, or objects of every days use (chair, lamp, mobile phone, etc.).



**Fig. 4.** Sketch of the methodology used for object recognition in a robust framework. Local patterns (SIFT) are detected for an image description, filtering using only informative descriptors leads to a reduced description. This sparse description is matched towards a reference description to determine object hypotheses.

Our proposed image based service for object awareness requires both, robust and fast visual object recognition of typically low-quality outdoor images. We therefore applied a methodology that is highly suited for mobile vision applications, i.e., the Informative Features approach [FSPB04]. This method extracts state-of-the-art local image features, i.e., SIFT

(Scale Invariant Feature Transform) descriptors [LOW04], from the greyscale image, in a first step. These texture patterns are then analysed to identify among them the most informative descriptors which are retained to establish a kind of attention filter. Responses in terms of most informative features are then matched to reference descriptors to provide a probability distribution on a pre-determined set of object hypotheses. In addition, the Informative Features approach enables to detect any object presence in the image, and background vice versa [FSPB04]. We also provide a confidence value to quantify the degree of uncertainty in the final object decision, and thereby enable to define a ‘quality of service’ measure for usability analyses.



**Fig. 5.** Sample recognition of objects  $o_{15}$  and  $o_{18}$  (“Kunsthau”, “Bärenapotheke”, out of the TSG-20 image reference database<sup>1</sup>). Top down: (i) reference images, (ii) location of SIFT descriptors with colour coded quality of service, (iii) selection of high quality descriptors, (iv) distribution on object hypotheses with correct decision based on maximum confidence.

<sup>1</sup> Download at <http://dib.joanneum.at/cape/TSG-20>.

### ***Geo Services***

As already mentioned, the knowledge of the tourist position is a very important fact within the system, not only for routing and guidance purposes but also to support the matching algorithm for the object recognition. Primarily GPS is used for positioning. If GPS is not available, either the cell information of the phone-network provider, a successful accomplished object recognition, or a manual input of the street name and number can be used for position determination. In the later case the street name and number is matched against a database where all buildings are stored with their according coordinates. This is especially in urban areas a method where the user can determine the own position with minimal input an effort.

The actual position is also used to present maps of the surrounding area of the tourist on the mobile device. The maps show a city map for orientation purposes, points of interest with continuative links, the GPS Track of the device and routing information. For the creation a map server application running on the server is used, cf. Fig. 2. In the current projects the map server originally developed by the University of Minnesota (UMN MapServer) is used. The map server primary function is to read data from various sources, which not necessarily need to be stored on the same computer, and pull these layers together into a graphic file. One layer may be the city map, another the points of interest, or the routing information. Each layer is overlaid or drawn on top of the others and then printed into a graphics for the tourist to see on the mobile device. Due to the fact that the map content depends on the user profile and interests, a dynamic creation of the content is necessary. This is done by using PHP MapScript, a dynamically loadable PHP module that makes the MapScript functions and classes of the map server available in a PHP environment.

The map server application is able to easily integrate data stored in databases. For applications which bear strong relation to positions on earth the use of databases with support of geographic features is advisable. In a lot commercial and open source databases this support can be found. In the context of the Mobile City Explorer project PostGIS is used. PostGIS adds support for geographic objects to the PostgreSQL object-relational database and follows the OpenGIS “Simple Feature Specification for SQL”. PostGIS enables the execution of spatial queries by offering spatial functions, such as Distance(), Intersect(), and Within(), among others, to narrow down the result of the search.

Cartography on mobile devices is not comparable with traditional or digital cartography. The small display size, limited processor power of the device, and varying lightning conditions have to be taken into account. The map must show only the relevant degree of details for the current task to avoid a too noisy overall impression on the small display. The dedicated use of colour and font size shall support the readability under different lightning conditions.

### ***Conclusions***

Object awareness provides a concept for semantic indexing into huge information spaces where standard approaches suffer from the high complexity in the search processing otherwise, and provide a means to relate the mobile agent to a semantic aspect of the environment. Visual object recognition using innovative and robust pattern recognition methodologies is an emerging technology to be applied in mobile computing services. Due to the many degrees of freedom in visual object recognition, it is highly mandatory to constrain object search by means of context based attention. In this paper we used geo-context to focus object recognition on a specific set of object hypotheses, and demonstrated that the challenging problem of identifying tourist sights in urban environments can become feasible.

In the projects MOBVIS and Mobile City Explorer our work will target at further exploiting contextual relations in order to cut down the complexity in otherwise unconstrained visual

object recognition and make mobile services feasible. This challenging research goal will be achieved by identifying and tracking spatio-temporal aspects of the user's task context, improving the representation of geo-context by multi-sensor information fusion, and by applying attention strategies on the visual input. In addition, future research results will be achieved in more developed demonstrators, using camera equipped phones and wearable interfaces as mobile platforms for presentation.

### ***Acknowledgments***

This work is funded by the EU-IST project MOBVIS (FP6-511051) and the ASA via project "Mobile City Explorer" co-ordinated by Arsenal Research.

### ***References***

- [LPA05] Luley, P., Paletta, , Visual Object Detection from Mobile Phone Imagery for Context Awareness, Proc. 7<sup>th</sup> International Conference on Human Computer Interaction with Mobile Devices and Services, mobileHCI2005, Salzburg, Austria, September 19-22, 2005, pp. 385-386.
- [DCDH05] Davies, N., Cheverst, K., Dix, A., Hesse, A., Understanding the Role of Image Recognition in Mobile Tour Guides, Proc. 7<sup>th</sup> International Conference on Human Computer Interaction with Mobile Devices and Services, mobileHCI2005, Salzburg, Austria, September 19-22, 2005, pp. 191-198.
- [LOW04] Lowe, D., Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60(2), pp. 91-110, 2004.
- [FSPB04] Fritz, G., Seifert, C., Paletta, L., and Bischof, H., Rapid Object Recognition from Discriminative Regions of Interest, Proc. 19<sup>th</sup> National Conference on Artificial Intelligence, AAAI 2004, San Jose, CA, July 25-29, 2004, pp. 444-449.