INTELLIGENT MAPS FOR VISION ENHANCED MOBILE INTERFACES IN URBAN SCENARIOS

Linde Vande Velde[1*], Patrick Luley[2], Alexander Almer[2], Christin Seifert[2], and Lucas Paletta[2]**

1. Tele Atlas, Gent, Belgium
2. Institute of Digital Image Processing, JOANNEUM RESEARCH, Graz, Austria

* Moutstraat 132, 9000 Gent, Belgium, ph: +32/9 244 89 63, linde.vandevelde@teleatlas.com

** Wastiangasse 6, A-8010 Graz, Austria, ph: +43/316 876-1769, lucas.paletta@joanneum.at

ABSTRACT

This paper focuses on the innovative concept of *intelligent digital maps*. Event and context information are stored as part of the city maps in order to make the maps intelligent and to exploit further context information of objects. In the EU funded project MOBVIS [10], we develop advanced methodologies to aid mobile vision and context awareness tasks using the enhanced Tele Atlas 2D and 3D City Maps. Augmenting the maps with visual features, semantic and context information enables to support within a given context specifically targeted geo-services in order to deliver appropriate geo-referenced map information on demand. We present preliminary results of geo-indexing the map to retrieve map information in a mobile object recognition task.

KEYWORDS

Mobile interfaces, digital city maps, geo-services, object recognition.

INTRODUCTION

This work proposes the investigation of geo-services that would use the exploitation of geo-spatial information for the support of mobile services, in particular, vision based functionalities that enable natural interfaces for situated annotation in a local urban environment. Vision is a rich source of information with the intrinsic potential for the extraction of complex relations in the environment but the disadvantage of usually requiring an infeasible amount of processing in order to discriminate semantics from the given noisy data. We propose to apply geo-spatial information to cut down the dimensionality of brute force search in visual feature spaces for the extraction of useful information. The concept of the Intelligent Map includes positioning information and applies Geo-Services to index into enhanced data of state-of-the-art city maps, for the goal of making mobile vision services feasible.

Location based services are a hot topic today. Everyone wants to find the nearest restaurant etc. For this, the user needs a device which gives access to an application using a digital map containing addresses and points of interest to perform the search function for geo-spatial context. It gives the obtained results by means of a map with an indication of the requested location, and can furthermore map to routing information or navigation functionality.

The first step to get location based services is to provide the position to the system. There are 3 ways to provide position information: (1) manual address input (2) mobile phone cell-id position (3) GPS position. The accuracy of the cell-id position information depends on the position of the user. In rural areas the positional accuracy can be very low (10 km) where in urban areas this can reach up to 50 meter. The GPS position is typically accurate to about 15 meters. DGPS, AGPS & map matching are techniques used in navigation applications to improve the positional accuracy up to a few meters. In some situations there is no position provided due to lost of signal. This especially occurs in urban areas with high buildings, called urban canyons.

There where in 2004 globally 1 billion GSM subscribers [4], of which in 2005 295 million camera phones [6], there where 3,4 million PDA units in Q1 2005 [3] compared to a 10 million personal navigation (PNAV) devices. This means that an alternative positioning technology on GSM is required to improve the accuracy of position information, in order to support LBS applications or to ensure a GSM user and a GPS user is able retrieve accurate positioning information whenever needed.

The main objective of the MOBVIS project [9] is to develop an attentive interface for the implementation of mobile vision technology in urban scenarios. The claim is to make mobile images understanding possible by intelligent interaction between three major functional components, namely, (i) the advanced computer vision for object awareness, (ii) the exploitation of multi modal context, and (iii) the intelligent access to map knowledge.

We will present the potential of this new technology by going beyond LBS with simple signal and co-ordinate based relations, towards visually interpreting the world to object awareness for the future enrichment of mobile LBS.

The object information generated from vision sensing, fed into a compact description of the current context is used to index the geo information of the intelligent map component, in order to retrieve related relevant information on the observed objects, including positioning information of the observed object. [6]

In this paper, we will first introduce in very detail the concept of ‘Intelligent Map Design’ on the basis of state-of-the-art Tele Atlas map technology. Then we settle ‘Geo-Services’ in the context of supporting urban mobile vision services and describe ‘Contextual Indexing’ interacting with 3D city maps. ‘Mobile Object Recognition’ will be part of future enabling mobile technologies. We finally describe the sample Geo-Service functionality in the context of Intelligent Maps to index into visual recognition via position and enhanced map information.

INTELLIGENT MAP DESIGN

The object information and the context retrieved from the vision sensing is compared with geo-information present in the area of the GPS position or the cell-id where the object is selected. An object can be recognized by the object recognition tools as eg. a church located in a park etc. The MOBVIS server as such has the task to find all hotels located in a park near the GPS position or in the cell-id, in the geo database located. The more details of the real world are present in the digital maps the better the change on recognizing the object and the surrounding of the object provided with the vision sensing and by retrieving the correct position of the object and meta information on the object to the user.

Standard Digital City Maps

The standard Tele Atlas MultiNet digital map product contains following information: (1) Geometry capabilities: centrelines for all roads and streets, railways, ferry lines, and rivers; polygons for lakes, land cover, administrative areas, and postal districts; point features for city centers, junctions, points of interest, etc. (2) Geocoding attributes: administrative structure (country, municipality, province), street names, alternate street names, house-number ranges, postal codes, address areas, etc. (3) Routing Attributes: functional road class, form of way, route numbers, direction of traffic flow (one-way, two-way, divided motorway), road condition, network classifications, blocked passages, special restrictions, restricted manoeuvres, toll roads, etc. (4) Navigation Attributes: information on complex intersections, z-levels for bridges and tunnels, signpost information, tourist roads, TMC locations and paths, etc.

The description of the current context is compared to all these map features and attributes and as such can give a first rough indication of the location of the selected object. However, in order to give an exact location of a building, also the building information and other indicative reality information should be present in the geo-database.

Enhanced Maps

During the MOBVIS project an analysis was made on which reality features could give added value to improve the accuracy of the location of the selected object. Two feature groups are defined: road furniture features and map display features.

The road furniture (traffic lights, sign posts, traffic signs) are standard necessary for large scale maps [7]. Other road furniture such as kiosks, public transport stoppoints, trees and poles are defined as required features in the MOBVIS geo-database. The exact location of these road furniture features in the geo-database are very valuable for the retrieval of the accurate position of the in the field selected object. Pedestrian features such as zebra crossings and stairs are also in this context valuable enhanced features.

The second group contains features supporting the enhanced map display. There are the 2D city maps, containing building footprints, related meta-information and sidewalks. There are the 3D city maps containing the 3D building geometry inclusive the roof geometry and building façade information. Depending on the importance and the complexity of a building a different level of details (LOD) is modeled. Buildings modelled in (LOD1) have generic facades, buildings modelled in LOD2 have a ground floor which is photorealistic, while the other floors are generic modelled and buildings modelled in LOD3 have a complete photorealistic texture [8]. Complex important tourist buildings are always modelled in LOD3 and are called 3D landmarks.

One of the aims of the MOBVIS project is to define the minimum required LOD to ensure the object and its related context information is recognized in the geo-database. The research on this topic is ongoing, and no final results can be published yet.

Tele Atlas has developed together with GTA Geoinformatik GmbH [11] its own production methodology to produce enhanced 3D maps (Figure 1). The imagery gathered with the Tele Atlas mobile mapping vehicle equipped with geocoded cameras, are in a semi automatically way processed and result in the cost efficient production of 2D & 3D city maps. [8]

In the MOBVIS project all previous mentioned enhanced features are produced for a selected test area in Graz. It is obvious that the more features are recognized around the selected object, the higher the change the correct position of the object can be obtained. However one of the goals is to understand the added value of each individual feature or a group of features to define with which features the digital maps need to be enhanced in the future to support in an optimal way the position detection making use of vision technology.

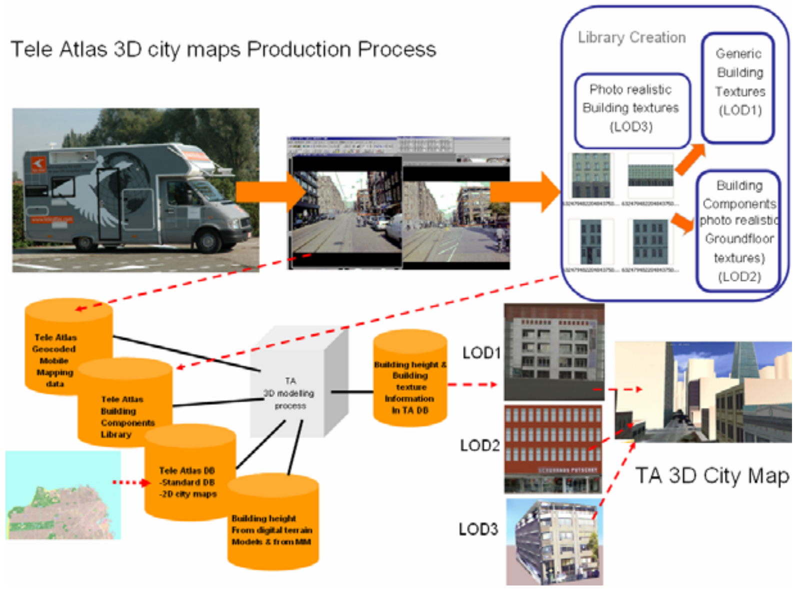

Figure 1: Tele Atlas 3D city maps production process.

GEO-SERVICES

Ongoing work in the frame of MOBVIS investigates the interaction of context with Geo-Services that deliver appropriate augmented map information on demand. Therefore, the above mentioned Intelligent Map consists of functionalities for executing the Geo-Services and providing access to retrieve appropriate information from the augmented map. Within the final proto-demonstrator software the results will be demonstrated and visualized.

Concept of Geo-Services

The Geo-Services are responsible for the interaction with the map based geo-information knowledge. A complex functional interface to the digital map information will be defined for the capability to appropriately response 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 perform object hypotheses (see also mobile object recognition). Based on intelligent maps the content and interactive functions could be logically queried and re-structured, map symbols and user interface could be dynamically adjusted, and navigation functionalities could change its orientation strategy based on the actual user behavior.

Context information is represented by geo-referenced data of spatial-temporal character, such as, events (object information with time stamp) or abstractions of spatial layout of objects. The basis for the augmented map is the Tele Atlas standard and enhanced map. Access to the augmented map (see Figure 2) will be performed by the Geo-Services functionality which can be thought of an interface that will be using a simple programming language to access map data on demand, connecting the map information with purposively distributed databases consisting of meta-information. Therefore, Geo-Services access the object, event and context information within the augmented map. This map can be viewed as consisting of several context layers (public, user, private) that extend the standard and enhanced TA digital map. Meta-information is linked to the individual data via object specific links to purposively distributed databases. Figure 2 gives an overview about the information flow between context information and augmented map (Intelligent Map concept).

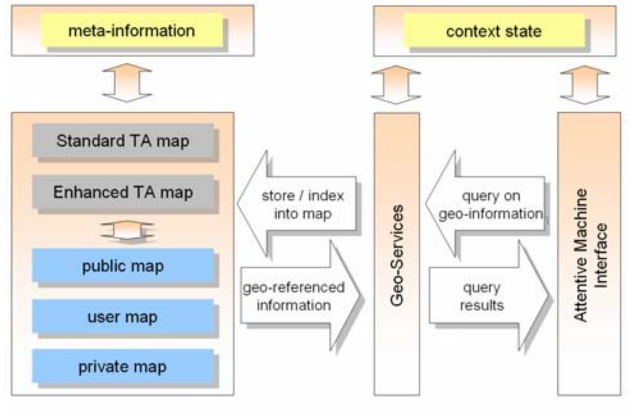

Figure 2: Overview of the Geo-Services as central module within the Intelligent Map concept

In general, the Geo-Services will be involved in a variety of elementary perceptual tasks, which can be sketched as follows. (1) Geo-referenced (static) objects relate object position with associated information and thereby mediate location awareness. This kind of infrastructure objects includes buildings, places, landmarks, road signs, etc. Their use is primarily in positioning tasks, annotation for augmented reality, in mobile historical culture learning and civil engineering directed mobile inspection tasks. (2) Dynamic objects relate object identity with associated information. Their global position might dynamically change or even remain unknown. Sample objects are people, faces, cars, company brands, and text on information boards. Their use lies in associating annotation information or text translation. (3) Context situations refer to collections of object identities and activities, binding scenes (e.g., 'the airport') with goal-driven behaviors (e.g., 'searching counter') into a related spatio-temporal context. Their use lies in tracking and storing personal diary information, and in associating information related to a specific context.

The following objects are relevant in the MOBVIS scenario and have to be visualized in the intelligent map, which will be handled by the Geo-Services: (1) Landmarks that discriminate a location in positioning tasks (2) Infrastructure objects, such as, places, buildings, zebra crossing, and road signs (3) Dynamic objects, such as, people, cars (4) Text and logos on information boards.

Contextual Indexing and 3D City Maps

Contextual Indexing substantiates the attentive interface in a concrete interaction system for the communication between object awareness, context, and geo-services. 'Map indexing' represents an object or context based request on the intelligent map to return geo-information specifically related to object or context features, such as, an object list indexed by a currently extracted visual feature (edge, plane slant, color, etc.), such as, 'find position of local yellow colored objects'.

The following context situations are relevant in the MOBVIS scenario and are applied for contextual indexing into geo-information: (1) Sidewalks (a context of road and cars, facade, and infrastructure) (2) Buildings (a context of windows, entrances) (3) Meeting people (people, possibly known faces, communication activity).

Furthermore, 3D information based content in the enhanced city map will provide several benefits for vision based mobile services via contextual indexing. (1) Matching features from mobile imagery and stored 3D information (object recognition from mobile imagery) (2) Providing 3D feature for the search of appropriate information in the urban environment (positioning from landmarks) (3) Displaying 3D objects in the user interface for user based interaction (object based navigation).

MOBILE OBJECT RECOGNITION AND GEO-CONTEXTUAL INDEXING

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, thus providing 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 [11]. 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. Here we describe geo-context to focus object recognition on a specific set of object hypotheses, and demonstrated that the challenging problem of identifying building objects in urban environments can become feasible.

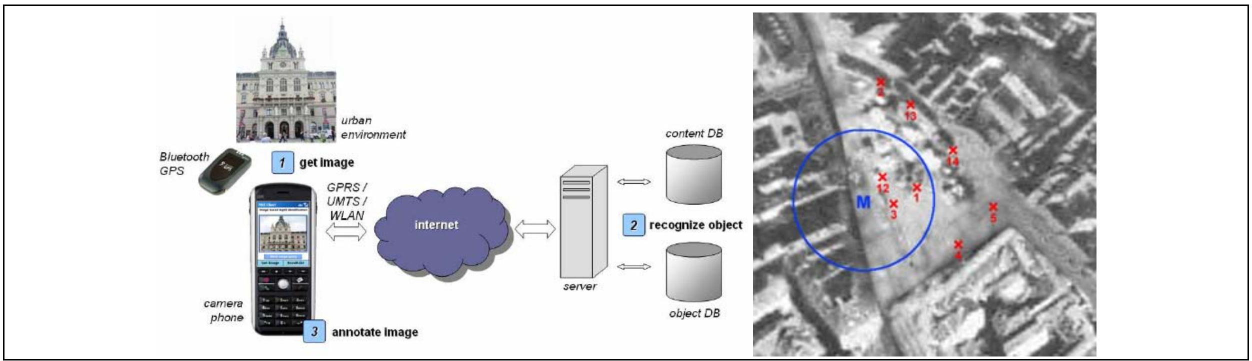

Figure 3: (a) System architecture of the mobile object recognition service. (b) Geo-Context, e.g., from GPS or cell-id based position estimates (‘M’ with blue uncertainty radius), can set priors by geographically indexing into a number of object hypotheses (‘X’s are coordinates of user positions while capturing images about objects of interest).

MOBVIS Mobile Interface

Figure 3 depicts the technical concept and the three major stages in situated mobile object recognition and annotation. The system consists of an off-the-shelf camera-equipped smart phone, a GPS device (built-in, e.g., A-GPS, or Bluetooth externally connected), and a server accessible through mobile services that runs the object recognition and annotation software. This specific client-server architecture enables large-scale application of urban object awareness, using GPS to index into the geo-referenced object database, and leaving object recognition restricted to a local urban area on the server. In the future, mobile clients might run the application even faster. The user captures an image about an object of interest in its field of view, and a software client initiates submission of the image to the server. The transfer of the visual information to the server is performed either via GPRS, UMTS, WLAN (PDAs), or MMS (multimedia messaging service). If a GPS device (bluetooth or built-in A-GPS) is available, the smart phone reads the actual position estimate together with a corresponding uncertainty measure, and sends this together with the image to the server. In the second stage, the web-service reads the message and analyzes the geo-referenced image. In future implementation stages, it is planned to compare the incoming images with the 3D maps which are present on the server. The geo-coded 3D maps will in this way support the vision tools to provide accurate position information. In order to get position the user should take at least 2 images of any object of interest, such as, buildings, to extract the accurate position. Based on a current quality of service and the given decision for object detection and identification, the server prepares the associated annotation information from the content database and sends it back to the client for visualization.

Position Indexed Visual Recognition

Global object search in urban environments – comprising thousands of buildings – is a challenging research issue. However, within most application scenarios, positions would be available from GPS or cell-id based geo-referencing which can be used to index into an otherwise huge set of object hypotheses. Geo-reference indexing for selected object hypotheses first requires a database containing on-site captured geo-referenced imagery about objects. In the future implementation, 3D models of Tele Atlas will be used as well to support the generation of object hypotheses. We will compare the mobile imagery that were locally captured in the urban environment and match them to the appearance retrieved from the stored 3D object models. ’Ground truth’ geo-referencing can be performed manually, e.g., on corresponding air-borne imagery (3b). From the differences between ’true’ and on-site measured positions we can determine an average positioning error. Based on this quantity, we partition the complete set of object hypotheses into subsets of hypotheses (’neighborhood cell’) of local context within a neighborhood for further processing. For each of these neighborhoods, we would learn the informative features and an attentive mapping to saliency.

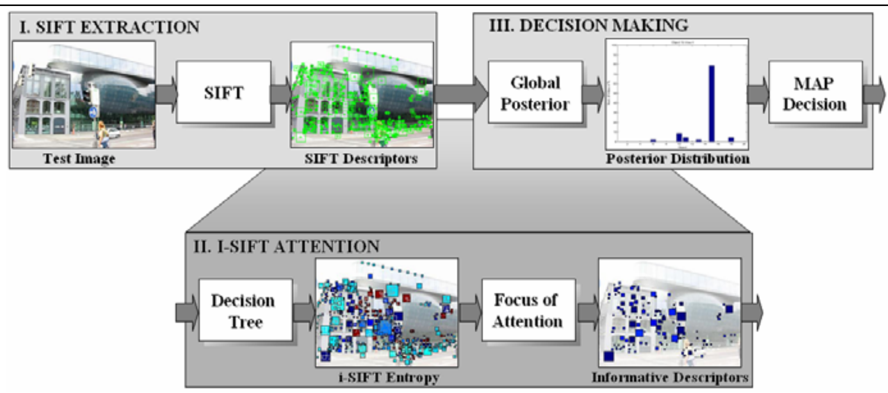

Figure 4: Concept of robust mobile visual object recognition used in MOBVIS [10,11]. (I) Local descriptors are extracted from the image. (II) Only informative (i-SIFT) are retained for classification. (III)A posterior distribution over object hypotheses from local classifications provides then a Maximum A Posterior (MAP) decision for recognition.

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 (Figure 4). 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 (buildings, tourist sights, traffic signs, information boards, etc.), people, or objects of every days use (chair, lamp, mobile phone, etc.).

CONCLUSIONS

In this paper, we presented and discussed the concept of the Intelligent Map. The intelligent interaction between the mobile vision technology for object awareness, the exploitation of multi modal context and the access to digital enhanced 2D and 3D map knowledge shall result in the offering of second generation location based services, allowing the retrieval of relevant information on observed objects, including accurate position information, independent the positioning technology used in the mobile device.

Enhanced 2D and 3D maps play already an important role in map display in navigation applications. In addition Tele Atlas learned that these enhanced maps will also be necessary to support the second generation of LBS, as such Tele Atlas is offering already 2D and 3D city maps and is extending its coverage over the next years: (1) 2D City Maps: Today: Europe (99), US (60), Q4 2007: Europe (120), US (80) (2) Full textured 3D City Maps: Q4 2006: Europe (4) mid 2008 (48). (3) 3D Landmarks: Today: Europe (518), 2007: Europe (1000), 2008: US (1000).

Future work will be specifically focused on the interaction between 3D city map information and Geo-Services for the mobile service of object recognition.

ACKNOWLEDGEMENTS

This work is funded by the European Commission's project MOBVIS (FP6-511051), and by the FWF Austrian joint research project Cognitive Vision under sub-projects S9103-N04 and S9104-N04.

REFERENCES

[1] Evensen, K.,. CALM versus DSRC - complementary technologies. *Proc. of 10th World Congress on ITS*, ITS 2003, Madrid, Spain, 2003.
[2] Füßler, H., J. Widmer, M. Käsemann, M. Mauve, and H. Hartenstein, Contention-based forwarding for mobile ad hoc networks, *Ad Hoc Networks*, vol.1, pp.351-369, 2003.
[3] http://www.cs.ucsb.edu/~malvarez/FR.ppt
[4] http://weco.net/course/wi/3-13Presetation.ppt
[5] http://people.csail.mit.edu/trevor/recentcolloq.ppt
[6] MOBVIS Technical Annex.
[7] http://www.ertico.com/en/links/links/gdf_-_geographic_data_files.htm
[8] Vande Velde, L., Navigation in 3D, Enhanced map display for car navigation and LBS, Proc. *of 11th World Congress on ITS*, ITS 2004, Nagoya, Japan, 2004.
[9] http://www.mobvis.org
[10] Fritz, G., Seifert, C., and Paletta, L., A Mobile Vision System for Urban Object Detection with Informative Local Descriptors, *Proc. IEEE 4th International Conference on Computer Vision Systems*, ICVS 2006, IEEE CD-ROM, January 3-5, 2006, New York, NY, USA.
[11] GTA Geoinformatik GmbH - http://www.gta-geo.com/eng/index_e.html