MOBVIS FP6-511051
Deliverable Report

D 2.1
"Report on Specification of Scenarios, Demonstrators and User Requirements"

Vision Technologies and Intelligent Maps for Mobile Attentive Interfaces in Urban Scenarios
Project co-funded by the European Commission
Sixth Framework Programme (2002-2006)
Information Society Technologies
FP6-2002-IST-C / FET Open
STREP

Due date of deliverable:	May 31, 2006 (month 12)
Actual submission date:	Dec 30, 2006 (revised)
Start date of project:	May 1, 2005
Duration:	36 months

Work package	**2**
Task	**1**
Lead contractor for this deliverable	**TUD**
Editor	**Bernt Schiele**
Authors	**Tâm Huynh, Bernt Schiele, Kristof Van Laerhoven, Ulrich Steinhoff, Ales Leonardis, Linde Vande Velde, Christin Seifert, Lucas Paletta, Eric Hayman, Mårten Björkman**
Quality reviewer	

Project co-funded by the European Commission within the Sixth Framework Programme (2002–2006)		
Dissemination Level		
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)	X

CONTENT

1. INTRODUCTION AND SCENARIOS

The aim of this document has been formulated in the work plan as follows: ”Early in the project, the final demonstrator scenarios will be specified to allow for a timely set up of its integrated version to run experiments with the mobile devices. The scenarios specify use cases in urban environments, assuming the major application domains regarding

- a visitor (tourist) requiring assistance in unknown environments (in the following called the ’visitor’ scenario), and to
- a personal diary task (in the following called the ’personal diary’ scenario), requiring time stamped event descriptions to build up individual memory, also including incremental map updating. ”

Furthermore the work plan specifies that the scenarios should be determined by user requirements derived from potential users. In order to follow this specification we conducted a series of informal interviews with various potential users (ranging from students, over sales-people to elderly people). However during those interviews it became evident that the users had a hard time to think and become stimulated beyond the rather obvious scenarios and applications as for example the navigation-aid in the visitor scenario. During various discussions during project meetings and among various project partners and based on the informal interviews with the potential users we identified a number of different scenarios:

first-time visitor scenario : in this scenario the most obvious functionality is pedestrian navigation. A person visits e.g. a town for the first time in his live and needs information to navigate around, to find restaurants, to obtain general or specific information about the town or particular buildings, etc.

n’th-time visitor scenario : in this scenario the user has been in town X before a few times - often for a different purpose. As the user has been in town X only a few times he still might want to get or even need navigational aid. But at the same time he might want to know where the excellent restaurant is he was in before or wants to get propositions what to visit based on what he has seen during previous visits. He might also like to get support to remember events or people linked to his last visits or linked to town X.

regular visitor scenario : a user might visit a town or places (such as a friend’s house, the airport or the library) on a regular basis so that in most cases, no navigational support is required. However, he might want to link information or reminders with any of those locations and may want to obtain just-in-time and just-at-location information.

permanent visitor scenario : most users have several places (such as their workplace, home, etc.) they ’visit’ nearly every day. In such a scenario the navigational support will be minimal and other functionalities such as just-in-time or just-at-place information will become more important. At the same time the system may help to organize and help with daily work and leisure as a memory aid and context aware assistant. The personal diary (assistant) can be seen as part of this scenario.

The above list of different scenarios represents a spectrum from a purely navigational support in unknown environments to the context-aware and personal diary system that helps with the daily work and leisure of the user. One might discretize this spectrum in a different way but the above list of scenarios proved helpful for discussions within the consortium as well as with potential users.

2. USE CASES

In order to identify potential use cases as well as useful functionalities we used the scenarios described in the previous section. Based on these scenarios and through another series of informal interviews with a range of users we developed the following use cases as a guideline to define and develop useful functionalities. Those use cases are presented in the following as a compiled list integrated into an overall use case scenario. Within the descriptions of the use cases, particular sections are underlined to clarify what is aimed for within the MOBVIS project specifically.

The following scenario describes a person's visit to the city of Graz. The person is on a business trip and has an appointment later during the day. Before the meeting he wants to take the opportunity to visit some tourist locations in the city.

#	Use Case	Functionality
	A reviewer travels for the 2nd time to Graz. He arrives at the train station. He has his mobile with him, which is equipped with a camera and various other sensors. GPS is integrated as well, but the reception in some of the narrow streets of Graz is very poor, so that it's doesn't work most of the time. He has still two hours before the meeting starts. He forgot the name of the street, where his meeting will take place - but he knows the name of the company. He opens his navigation application on his mobile and types the name of the company. The system doesn't recognize the name -- possibly because it is a small company.	
1	He remembers that he took a picture of the company building's facade the previous time he was there. He opens his personal diary, which has recognized where he is and knows that he has already visited Graz once, six months earlier. The first pictures that appear as he browses his pictures folder are from his previous time. He immediately finds the name of the company and the related picture.	*Personal Diary*
2	The reviewer sends the picture to the "recognize & position me tool". A few seconds later he gets the address and a snapshot of the map on which the location of the company is indicated. Good, but where is he now? He takes a picture of the front of the railway station and sends this picture to the "recognize & position me tool" as well. A few seconds later he gets again a map with his current position.	*Single Snapshot Localization*
3	In order to make sure that he knows how to find his way back to the station, he stores the picture of the facade of the building of the railway station in his personal diary.	*Personal Diary*
4	He wants to start his trip. So he sends a request to the server with request provide "route to destination", and he adds the option "tourist route with a duration of max 1 hour". He gets a map back with the starting point and destination of his trip and four extra waypoints, indicated as landmarks. He also receives a route to follow.	

#	Use Case	Functionality
5	At any time during his trip, he can request a list of interesting objects nearby.	
6	During the trip, he tags certain objects he sees: an ice cream stand with *"really good ice cream"* and an antique shop *"I have to check that later"*, as well as parts of buildings *"that's neo-baroque at its best"*. As a result, objects that are similar or related then pop up automatically.	*Personal Diary, Situation Dependent Display*
7	Here only objects are displayed that are interesting for the current user, defined by the user profile (the user is interested in architectural interesting buildings). Alternatively the user can give certain constraints that have to be met by the objects returned, e.g. all Chinese restaurants that offer lunch menus. He can then select an object from the list to get further information.	
8	The user's position is updated on the map in a continuous and reliable way, by fusing information from different sensors. However, at one point during the trip, the system has difficulties as he wanders into a part of town that makes it hard to pinpoint his position. The mobile device consequently notifies the user, and he has the option to take a snapshot and recover his position.	*Multi-Sensor Positioning, Single Snapshot Localization*
9	The reviewer starts walking and suddenly he sees a bit further a landmark which was not indicated on his route. He leaves the predefined route and walks towards it - (it seems to be some kind of tower, located just a few streets away). He is interested in its architecture and wants to get information on it. He takes a picture of the tower, and a few seconds later he gets the tourist information on this landmark.	*Recognition & Retrieval of Related information*
10	To assist the object recognition, all available multimodal sensor data (for instance WiFi position) and context information are used to make the recognition as reliable as possible.	*Multi-Sensor Positioning*
11	The system thus delivers a list of likely recognized objects (this time the list fortunately contains only one item) and the user selects the correct object from the list and gets further information about it (the name, the year of construction, building style, and also current usage).	*Recognition & Retrieval of Related information*
12	The reviewer wants to get back to his original route, but suddenly he realizes he lost his orientation. He walks to the corner of the street and takes a picture of the street name plate and a second picture of the nearest house number sign. He sends both pictures to the "recognize & position me tool". A few seconds later he gets a map with the indication of his new position and with indications on how he has to go back to his originally planned route.	*Street name Recognition for localization*
13	Back on his route, he arrives at a large square with a lot of cafes. He is searching for a local Graz specialty. The two coffee shops on the square that happen to offer it are highlighted on his map. As it happens, in one of these cafes there is a person he had planned to meet. She has tagged the coffee shop with her name, and on	*Recognition & Retrieval of Related information*

#	Use Case	Functionality
	his map, the shop's entrance is highlighted as "Martha is here".	
14	He doesn't find his exact position on this big square, however. He sends again two pictures of two impressive landmarks on the square with the request "send new position" & "display in 3D". He gets immediately his position and a line of sight to both cafes, on a 3D image of the square.	*Single Snapshot Localization*
15	The square has fortunately plenty of WLAN stations in the surrounding buildings that could assist in determining an accurate position for the vision-based positioning.	*Multi-Sensor Positioning*
16	The reviewer's personal diary has noticed that the user has chosen to visit a coffee shop, and that this is getting quite the habit around 4pm, already the 5th time this month. There also seems to be a particular brand of coffee shop that is preferred.	*Personal Diary*
17	His personal diary notices that an open wireless internet connection is in place for the coffee shop's customers, and decides to fetch his latest emails, as it predicts that it will be a while before it can access the internet again, and he usually stays at least half an hour.	*Personal Diary*
18	After a coffee, he decides to call a taxi and spells the name of the company where he has his appointment. The taxi driver brings him to the company. While he's in the taxi, he has some time to take a look at the path of his walk through the city. His mobile device has reliably recorded the path he has taken.	*Multi-Sensor Positioning*
19	On arrival at the company, he suddenly remembers that he forgot the name of the person he has a meeting with. Fortunately, he has still a group picture from the previous time he was there in his personal diary. He shows the picture to the receptionist, and	*Personal Diary*
20	Finally he has an appointment with Mr Paletta.	

3. FUNCTIONALITIES AND TECHNOLOGIES

3.1 FUNCTIONALITIES

Based on the above list of use cases we have identified the following useful functionalities for the range of scenarios:

1. **Recognition & Retrieval of Related Information – UL (& JRS)**
   The system recognizes landmarks and buildings in the city from pictures taken by the user (and possibly by using other context information). E.g., the user takes a picture of a building and sends it to the server, which matches the façade of the building to façades stored in its geospatial database.

2. **Localization**
   a. **Single Snap-Shot Localization – KTH**
      The system localizes the user from a single image taken with the mobile device.

   b. **Street Name Recognition for Localization – JRS**
      The system recognizes the names of street signs which are present in pictures taken by the user.

   c. **Multi-Sensor Positioning – TUD**
      Data from the available vision- and non-vision sensors is fused in order to localize the user.

3. **Situation Dependent Display, Change Detection and Reaction – TUD**
   Based on the current situation and context of the user, the system behaves differently and displays different kinds of information. E.g., at night time when the user is taking a leisurely stroll through the city, the system might display nearby bars and music events nearby.

4. **Personal Diary – JRS**
   The system includes a personal diary to store and retrieve personal information. The user can manage his personal diary at any time, which includes creation, deletion of a diary, as well as adding, editing and removing entries. Entries can be pictures, comments, etc.

3.2 TECHNOLOGIES BEHIND THE FUNCTIONALITIES

The previous section has introduced a range of functionalities that have been identified to be useful within the overall scenario. This section then discusses the technologies that the Mobvis-consortium wants to research and develop in order to support and enable those functionalities.

1. **Recognition and Retrieval of Related Information**
   In order to facilitate recognition and retrieval of related information e.g. with respect to buildings, the Tele Atlas data will be enhanced with information about textures of facades and 3D landmarks. Using such image information, different approaches will be explored such as local and global feature based recognition (UOL), informative

feature based recognition (JRS) and building recognition (KTH). Also, specific urban infrastructure will be detected (UOL).

2. **Localization**

   a. **Single Snap-Shot Localization**
   In order to enable localization, the Tele Atlas map data will be further enhanced with terrain models, mobile mapping data, pedestrian features and traffic signs. Those will enable to store and use such features within a geoservice engine (JRS). For 3D localization we plan to explore weak geometric constraints (UOL) as well as strong geometric constraints (KTH), both based on epipolar geometry and homographies. Whenever possible and available, crude localization will be also used to e.g. restrict the search space (TUD).

   b. **Street Name Recognition for Localization**
   Based on the street-name data contained in Tele Atlas map data street-name recognition will enable localization within such data (JRS) aided by geometric constraints (KTH). Again, crude localization may help to restrict the search space (TUD).

   c. **Multi-Sensor Positioning**
   To enable alternative position technology various multi-sensor based techniques will be explored (TUD). Besides GPS and vision based localization, technologies such as WLAN and inertial sensors will be explored and fused together depending on availability. For this, the Tele Atlas map data will be enhanced by pedestrian features (such as entry-points for buildings, paths in parks, etc).

3. **Situation Dependent Display**
   Depending on the situation of the user the information content and the way to display information on the interface can be changed (JRS). For this the modelling and recognition of the context and situation of the user is important and will be based on a multi-sensor context-aware system (TUD).

4. **Personal Diary**
   The personal diary scenario aims to store relevant information such as objects, context and other data to help the user to access, sort and retrieve important information later. For this various components will be developed ranging from the acquisition of object representations (UOL, KTH) e.g. of personal objects or 'bookmark' objects, visual attention modelling (JRS), personal context modelling and recognition such as routine and personal activity (TUD) as well as map visualization of the personal diary (JRS).

4. DEMONSTRATORS

The various functionalities listed in the previous section will be researched within MOBVIS and successively tested in a series of demonstrators. D 7.2.1 describes in more detail which of the above functionalities will be demonstrated at the year-1 review and those functionalities that have been tested.