R&I Research Note

R&I N 44/2009

Anders Schürmann, Marit Ånestad, Erik Bræck Leer, Sigmund Akselsen, Bente Evjemo

Tromsø Visual Guide
Finding information by snapping a picture with your camera phone

R&I Research Note	**N 44/2009**
Title	**Tromsø Visual Guide Finding information by snapping a picture with your camera phone**
Author(s)	Anders Schürmann, Marit Ånestad, Erik Bræck Leer, Sigmund Akselsen, Bente Evjemo
ISBN / ISSN	/0809-1021
Security group	OPEN
Date	2009.12.09

Abstract

This research note describes the Tromsø Visual Guide (TVG) – a mobile phone based tourist guide based on image recognition, where the user formulates a query to the system by submitting a picture. We describe initial test made with the image recognition software from Evolution Robotics applied to three dimensional objects and the challenges this poses. We describe how the guide is currently functioning, and look at the possibilities of further development of the guide, including ranking and presentation by context and monetization by marketing income from local service providers.

Keywords

TVG, Image Recognition, Similarity search, Tourist services, CBIR, ViRP, VISEP

Telenor R&I N 44/2009

Preface

This research note describes the Tromsø Visual Guide (TVG) – a visual interactive tourist guide. TVG has been developed by Telenor GBD&R within the CAIM[1] project, partly financed by the Norwegian Research Council. The CAIM project will focus on research and the development of tools for context-aware image management, where image description, query formulation, retrieval from heterogeneous distributed environments, and ranking are designed for using context information. Important application domains are those requiring image capture and multimodal retrieval in mobile environments. Much of the practical work has been performed by students from the University of Tromsø.

[1] *CAIM (Context-Aware Image Management), see http://caim.uib.no/*

Contents

1 Introduction

Tromsø Visual Guide (TVG) is a digital tourist guide which allows the user to search the guide by taking an image of an object of interest. Our vision is that instead of typing a query, the user can simply point her camera phone at what she needs information about. The concept is further described in the following chapters.

1.1 The tourist challenge

As tourists we love to travel to new and unknown places, to explore and experience new countries and cultures. We want to learn about these new places, and an important tool for the tourist in this process is the tourist guide. The book based tourist guide is still very popular, but there is an ever increasing number of digital tourist guides, on the Internet as well as on the mobile phone.

Digital tourist guides offer advantages over the paper based guides in that they can easily be updated, are interactive and searchable, can include many types of media, and utilize context information to increase effectiveness and efficiency as well as enriching the search results. Some of them include user generated content which often holds great value for travellers. On the down side they also require power, network connectivity, and a mobile device to run on.

A common challenge for all these guides, digital and paper alike, is how to make a connection between what you see in the real world and what you read in the guide. If you have a person accompanying you, you can simply point at what you want information about and ask “please tell me about that statue”. TVG aims to make it possible for users to point at what they want information about by taking a picture of it.

1.2 User scenario

Imagine that you are a tourist visiting the town of Tromsø for the first time. Walking along the pedestrian street you come to the town square. In the middle of the square there is a large statue with a man standing in a boat with a harpoon. You find the statue interesting and would like to know more about it. You bring out your mobile phone and take a picture of the statue. You then send the picture as an MMS to the tourist guide service. Soon after you receive a SMS from the tourist service telling you that the statue is named “The Fishing and Hunting Monument”. The message also contains a link to additional information. You click the link and the browser on the phone presents the history of fishing and hunting in the northern waters.

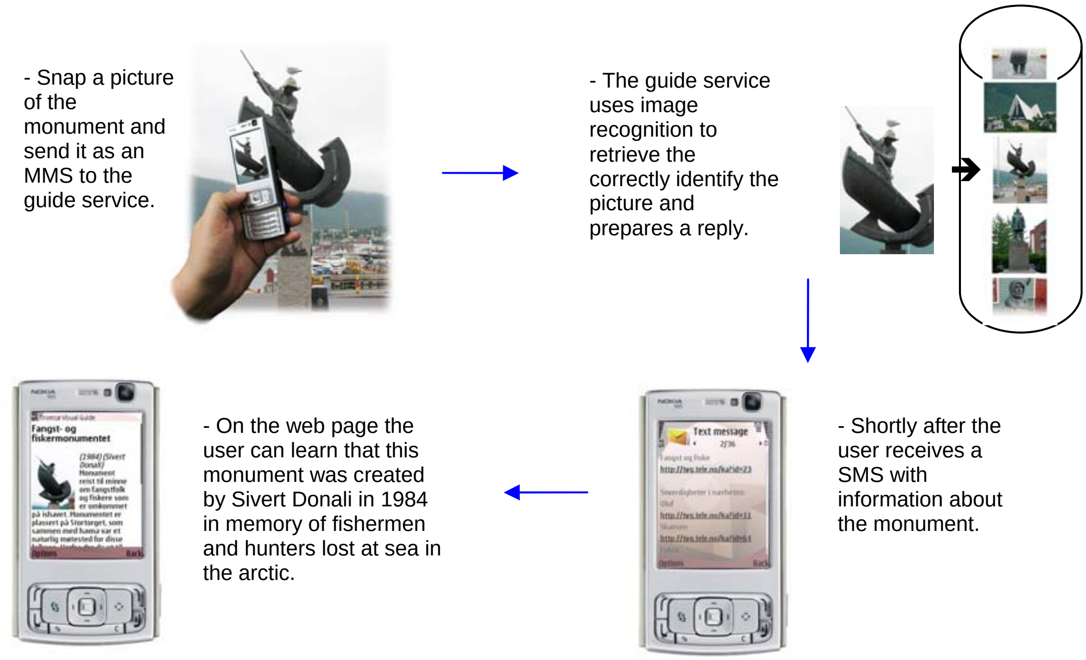

- Snap a picture of the monument and send it as an MMS to the guide service.
- The guide service uses image recognition to retrieve the correctly identify the picture and prepares a reply.
- Shortly after the user receives a SMS with information about the monument.
- On the web page the user can learn that this monument was created by Sivert Donali in 1984 in memory of fishermen and hunters lost at sea in the arctic.

Figure 1: User interaction in TVG. See [Schürmann et al., 2008a] for details.

1.3 History

TVG builds on ideas and concepts in Telenor which dates back to the beginning of the Century, i.e. before the camera phone became a mainstream technology. The initial idea started with the notion of smart binoculars – binoculars which could tell you about what you were looking at. As mobile phones with built-in cameras started to appear the idea again resurfaced, this time using the mobile phone's eye to tell you what you are looking at. In order to achieve this we planned to use image recognition [Datta et. al., 2008]. We started to build a prototype to demonstrate the feasibility of this idea in 2006, when a student on a summer internship surveyed available open source image recognition software, and started work on the initial service. The service addressed the tourist setting, previously investigated through the MOVE project [Akselsen et al., 2006]. It was named MMS 2 Search (M2S) due to its use of multimedia messaging. By means of M2S the user (preferably a tourist) could use the mobile phone to capture interesting images in tourist brochures and through an MMS/SMS exchange retrieve dynamically updated information about the targeted item [Schürmann et al., 2006].

Our rationale for using images of the tourist brochure instead of taking images of the object itself was that it would be easier for the image recognition software to recognize images since they would be quite similar. However, test results showed that the open source image recognition software had a rather low recognition rate, and it was not good enough to perform any trials in a real user setting.

Other departments within Telenor had also been working with similar ideas for mobile use of image recognition, e.g. [Canright et al., 2007] and in 2008 Telenor bought a trial licence for the Evolution Robotics ViPR[2] system. This system performed well above the other image recognition systems tested so

[2] *ViPR - Visual Pattern Recognition http://www.evolution.com/core/ViPR*

far, and was good enough to be used in public trials. Under the “Film fra sør”[3] film festival in Oslo 2008, a trial was conducted building on the same idea as in M2S - with users taking images of the festival catalogue in order to receive additional information or stream a trailer of the movie [Policroniades et al., 2009]. This trial also built on experiences gained from a test of the TIFF (Tromsø international film festival) event assistant [Schürmann et al. 2008b].

Since the ViPR software proved to perform so well on images of catalogues we decided to try it out on “real world” 3D objects. We believe this to be a more challenging task since there can be more significant differences in the query images in relation to orientation, rotation, scale, angle and perspective. Other objects might be occluding the object of interest, and especially for outdoor images the light and the surroundings might change very much with time of day and season. The TVG prototype was developed in order to explore these challenges.

1.4 Visual pattern recognition

Evolution Robotics originally developed their ViPR technology in order to create a vision solution for robots so they could navigate and interact with the real world.

ViPR is based on Scale Invariant Feature Transform (SIFT), an algorithm first developed by David Lowe [Lowe, 2004]. ViPR works by finding descriptors it uses to encode unique visual patterns such as the corner of an object. As the most distinct regions (called features) are localized in the image, unique descriptors are computed for each of them. Features in the image are invariant and can be matched between images. These features are like “fingerprints” within the image. Each SIFT feature is described by its location, orientation, scale, and a keypoint descriptor (accounts for shape distortion and changes in illumination). Having many features for each image it is possible to recognize the image even when 50-90% of the object is occluded. According to [Munich et al., 2006] ViPR has a success rate of 80-95% in uncontrolled settings and 95-100% in controlled settings. This is similar to what we have experienced in our own experiments.

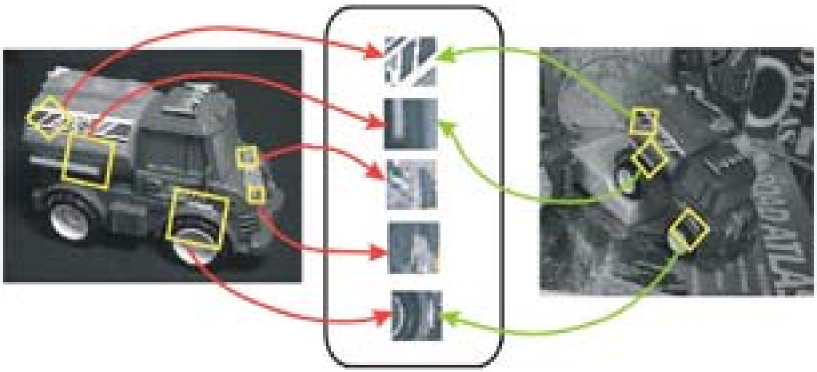


Figure 2: Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale and other imaging parameters[4].

[3] http://www.filmfrasor.no/en/

[4] http://itp.nyu.edu/RepresentingEarth/?p=503

2 Tromsø Visual Guide

The Tromsø Visual Guide consists of two parts: An image recognition service which matches photos to a point of interest (POI), and the actual guide service which brings the information and content to the end user (Figure 3). These parts will be described in the following subchapters.

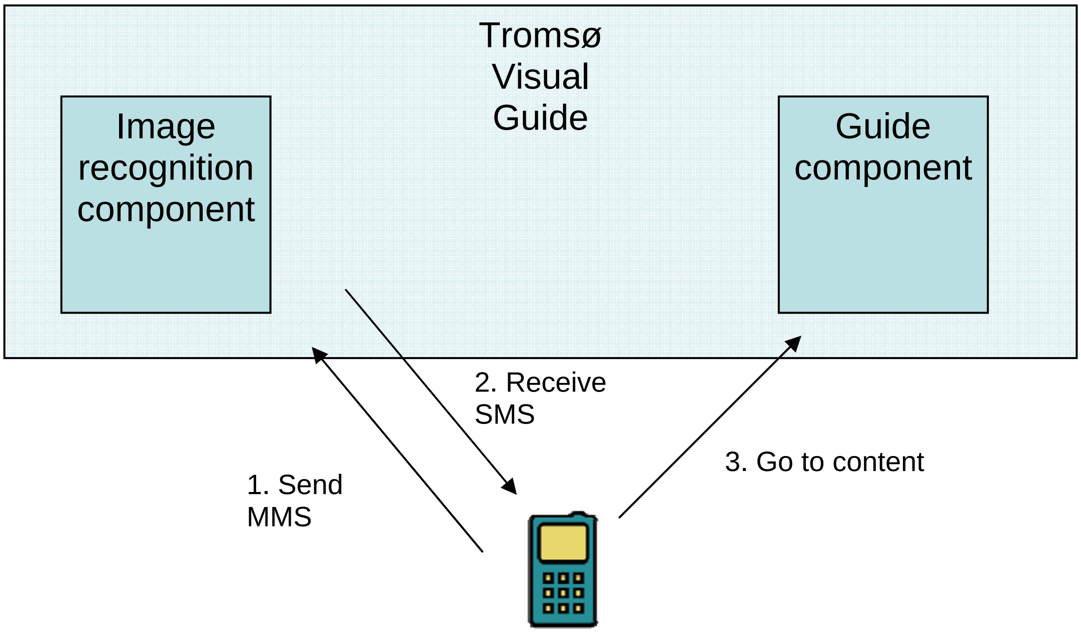


Figure 3: Overview of TVG

2.1 Image recognition component

The image recognition component is responsible for matching query images from the end user against images of actual POIs and returning an answer back to the user. An overview of the interaction can be found in Figure 4. In the current version of TVG we use MMS and SMS to communicate with the end user. This choice is based on the fact that MMS is a wide spread service available to most camera phones on the market today. This means that in a trial it would be easy for users to try out the service on their own phones.

We also considered making a dedicated client for this service, but this would require installation on the end user's handset, which could be difficult for many users. However, a TVG client could provide a better user interface for the service than MMS, and would probably also be a bit faster as the client could communicate directly with the TVG server.

When the user has taken a photo of the object of interest she will send this image with the code word "*Ka*"[5] to the five digit service number. This number is not dedicated to TVG, but is controlled by Telenor Playground[6] labs, which offer

[5] *"Ka" means "What" in northern Norwegian dialects.*

[6] http://playground.telenor.com

easy ways to set up mobile services. The TVG server polls the Playground server every minute to download any new MMSs which might have been sent to the service.

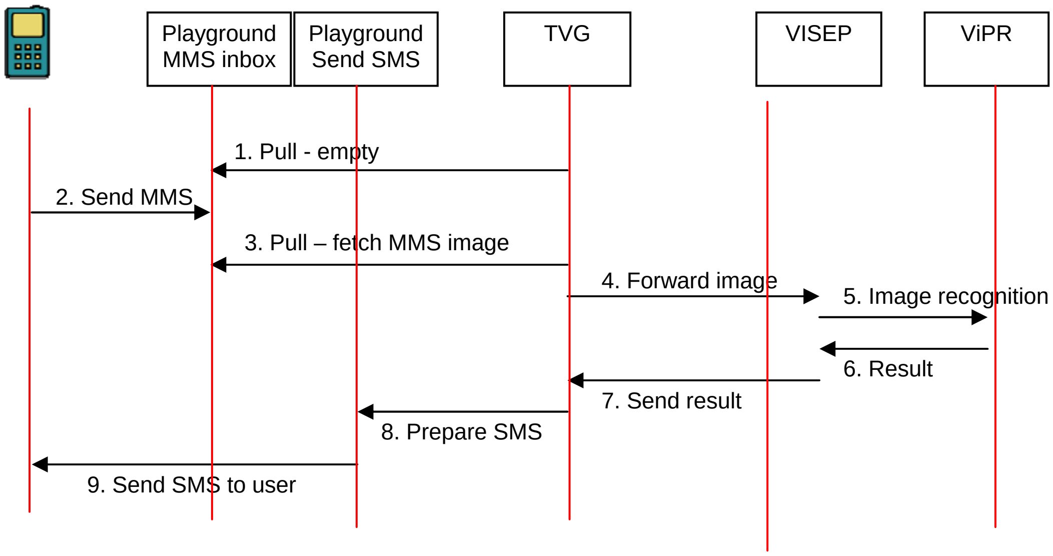
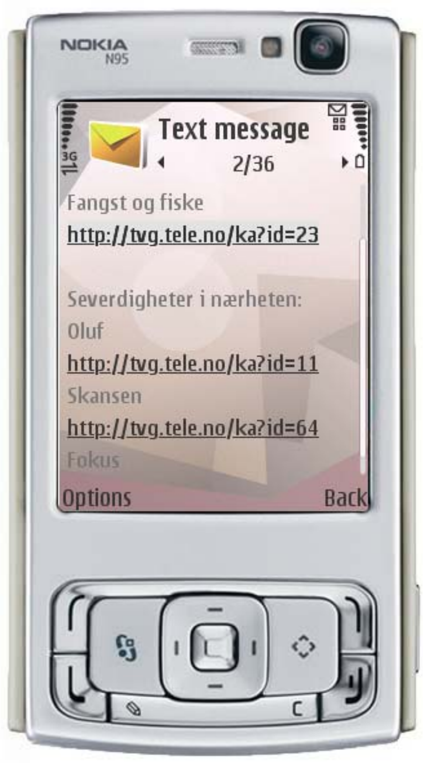


Figure 4: Sequence diagram of interaction in TVG.

When TVG receives a new message it will extract the image part of the message and forward it to a service named VISEP [Policroniades et al., 2009], using HTTP POST upload. VISEP is a wrapper service for the ViPR technology from Evolution Robotics, and is located on a different server from the TVG software. VISEP makes it possible for several services to utilize the license we have acquired for ViPR.

ViPR contains a database of images upon which it can perform image recognition methods. When it receives an image it will search its database and output a ranked list of the images which are the most similar, and which are above a minimum threshold for similarity. For each image registered in the ViPR DB, VISEP has registered some information (e.g. key/name, an associated web page, etc). VISEP will append this information to the query result before sending the reply back to the caller in the form of a XML document.

Upon receiving the reply from VISEP, TVG will check if the result contains any matching objects. If no matches could be found, TVG will send a SMS to the user informing her of this. The message will also include an option to browse the guide manually. If the result from VISEP does contain hits, TVG will pick the most similar object, resolve the key to a POI in the database and prepare a reply message to the user. This message will include the name of the object and a link to the object’s page in the guide. An example of a SMS message can be found in Figure 5.



Fangst og fiske
http://tvg.tele.no/ka?id=23

Severdigheter i nærheten:
Oluf
http://tvg.tele.no/ka?id=11
Skansen
http://tvg.tele.no/ka?id=64
Fokus

Figure 5: Example of a SMS message received from TVG. The image deviates from the current version in that it also shows links to related POIs in the proximity.

The image recognition component of TVG has been developed using Java EE version 5 and Enterprise JavaBeans 3.0. The application is running on a Sun Glassfish[7] Enterprise Server v2.1.

2.2 Guide component

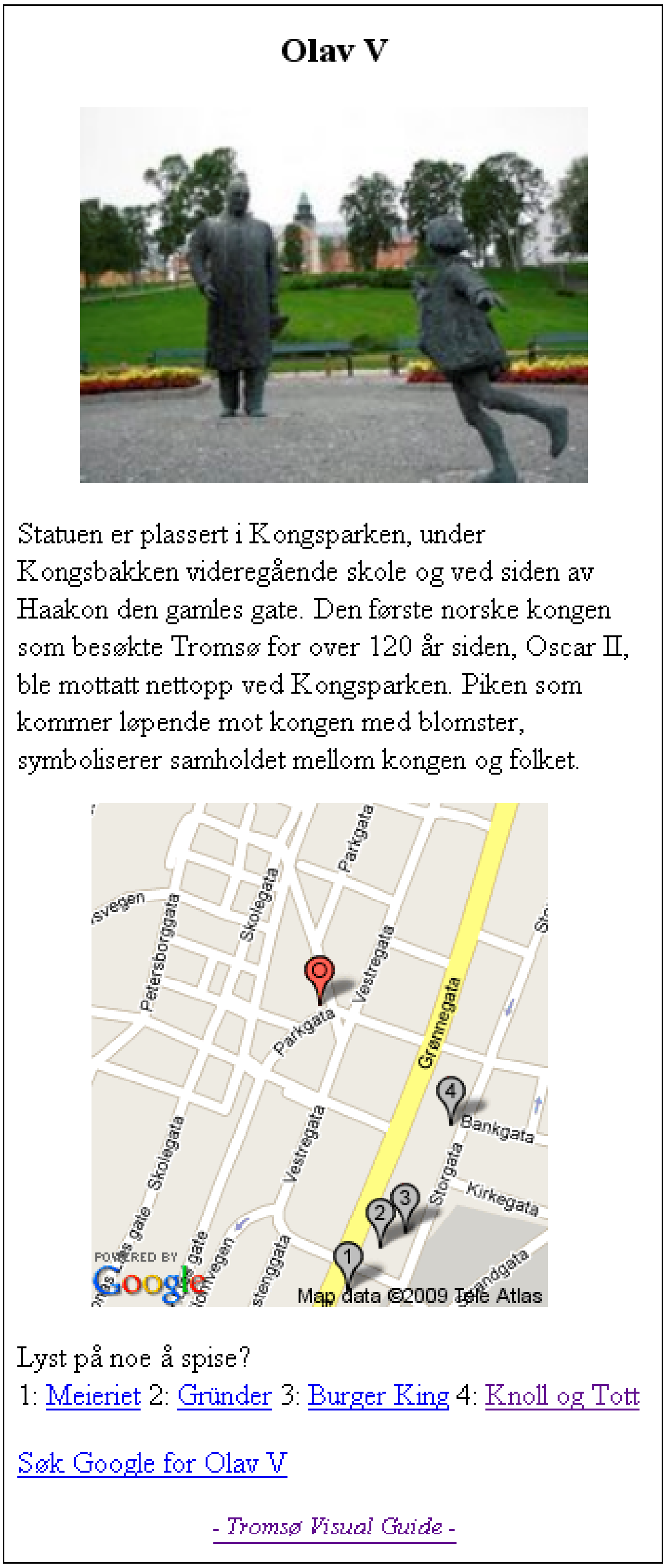
The guide component is the part of the system which provides the end user with information about the POIs in Tromsø. A page can be generated for each of the POIs containing an image of the POI, a description, and a map of where the POI is positioned. See Figure 6 for an example. The page also presents the user with some options. She can search Google for more information about the POI. She can post the image she used in the query on a blog with a personal message. Or she can view places to eat in the vicinity of the POI. These options are described in more detail below.

The pages of the guide are generated dynamically by using Java Server Pages[8] (JSP) together with JavaBeans which hold the data fetched from the POI database. The guide component runs on the same Glassfish Enterprise Server v2.1 as the image recognition component. The POI database runs on a MySQL 5.1 Server.

A request to the service could look like this:
http://move.tele.no:8080/MyTVGClient/index.jsp?ID=3011&NR=4790296474
The request contains two parameters. *ID* is an identifier for the POI. *NR* is an

[7] *https://glassfish.dev.java.net*
[8] *http://java.sun.com/products/jsp*

optional parameter which identifies the user of the service by utilising her phone number. The NR parameter is used for a.o. the blog feature.



Olav V

Statuen er plassert i Kongsparken, under Kongsbakken videregående skole og ved siden av Haakon den gamles gate. Den første norske kongen som besøkte Tromsø for over 120 år siden, Oscar II, ble mottatt nettopp ved Kongsparken. Piken som kommer løpende mot kongen med blomster, symboliserer samholdet mellom kongen og folket.

Lyst på noe å spise?
1: Meieriet 2: Gründer 3: Burger King 4: Knoll og Tott

Søk Google for Olav V

- Tromsø Visual Guide -

Figure 6: Screen-dump from how a guide page looks in TVG.

2.2.1 Personalization and marketing

Behind most mobile services there is a business plan or an idea of how one could capitalize on the service. Many of the business models we see on the Internet today are built on advertising, where a marketing message from an advertiser is displayed on the same page as the content of interest to the user. We believe that advertising could be part of a business plan for a commercial version of TVG.

For advertisers it is important that the targeted audience is receptive to their marketing message. The ad must be of relevance to the user for it to have any value. If an ad for a concert in Oslo is displayed to a person in Tromsø with a music taste not matching the music played at the concert, chances are high that this will be a waste of money for the advertiser. The more targeted the audience, the more the advertiser is willing to pay for the ad. A more targeted audience can be achieved by utilising information about the users' context, i.e. position, time, gender, age, interests etc. Further details on personalisation of mobile services are found in [Natvig et al., 2008].

A mobile phone based tourist guide like TVG fits well with tailored marketing. Much of the content in the guide is location specific, and although we don't use GPS to track the user, we can identify the user's position from the image queries to TVG. We assume that users are more interested in ads related to their current location. From monitoring which links the user is clicking on it is possible to say something about the users' interest. Info from the operator could also be used, like the user's age, gender and home address. From the mobile browser it is also possible to detect language preferences.

The TVG prototype demonstrates how tailored marketing can be used in a tourist setting. We chose to use restaurants and fast food services for this demonstration. Everyone needs to eat, and tourists especially tend to eat out. Eleven restaurants in the centre of Tromsø were included. Each entry consists of name, opening hours, position, price level and link to the web page of the place. The marketing is tailored in relation to three parameters: proximity, opening hours and price level.

Opening hours: There is no need to advertise for a nearby café if it is already closed for the day. TVG will only show places that are open at the time of the request. The opening hours of the different places to eat vary during the week. It is, however, possible to do a rough classification of the opening hours into morning, daytime, afternoon and evening, to get decent results for demonstration purposes.

Proximity: Only ads for the closest restaurants will be displayed. The distances are calculated in a naïve, trigonometric fashion which gives some odd results due to the high latitude of Tromsø. Using the Haversine[9] formula would improve this inaccuracy.

Price level: In the TVG prototype we let the user manually select her preferred price level when registering a profile. But this information could also be learned by studying user behaviour over time.

If a user profile is available, the application selects the open restaurants that match the user's price level sorted by the distance from the attraction. If user information is not available, the application selects the open restaurants sorted by the distance from the attraction. Currently up to 4 restaurant ads are shown on the map and on the web site of the attraction.

[9] http://www.movable-type.co.uk/scripts/latlong.html

In the future, the marketing can be extended to other types of business than restaurants, and the targeting might use additional criteria. Some places are very central to many attractions, while others just a little bit away gets less exposition. The selection algorithm might consider this to give a more "fair" exposure of the different participating places.

2.2.2 Map

The POI information page also contains a map showing the position of the POI itself as well as the four restaurants in the vicinity of the POI (See Figure 6). The map is generated by using the Google Static Map API[10]. This API allows you to embed a map image without requiring JavaScript or any dynamic page loading as the normal Google Maps API does. All parameters used in generating the map are defined in an URL, and the map is returned as an image of an optional format. An example of such an URL could be:

> http://maps.google.com/staticmap?center=69.65004,18.95207&markers=69.65004,18.95207,midredo|69.64777,18.95271,midgray1|69.64811,18.95344,midgray2|69.64823,18.954,midgray3|69.64909,18.95502,midgray4&ZOOM=14&size=240x320&MAP_TYPE=roadmap&FORMAT=png32&MAP_KEY=ABQIAAAAFCUZQha48M_p1njxKoi40hRBnt1AnJa--Je23nvuqS8BrX1H7xRZXPAfUoTNpLbhLZuOnVFpvAIgkQ

During the implementation of the service there have been some challenges regarding the implementation of the maps on pages for mobile phones. These challenges have not been fully resolved, and the map image may not appear correctly in all mobile browsers.

2.2.3 Search

In order to find more information about a POI we have added a search option. This is simply a link to Google Mobile Search[11] with the search parameters already specified. A search can give more and up to date information than the gathered information. It is possible to reach the home page of institutions with updated opening hours and prices. It could also be interesting to see if the attraction has been in any recent news reports.

A challenge with search is to find good search terms that yield relevant hits. For a monument it might be more relevant to find information about the people related to the monument than the actual monument itself. For an institution, it might be more relevant to get information about that building in Tromsø than general information about similar institutions other places.

We experimented a bit with this to find good search terms. For most attractions, using just the name of attraction as the search term yielded good search results. For some attractions, it was necessary to use the full name of the attraction and to be careful with the spelling. "Tromsø Domkirke" gave good results, whereas "Domkirka" did not give the desired results. Some names are still a challenge like "Skansen" and "Vår Frue kirke". This could probably be improved by adding the name of the location to the search (i.e. Tromsø) or by creating a separate entry in the database for the search terms.

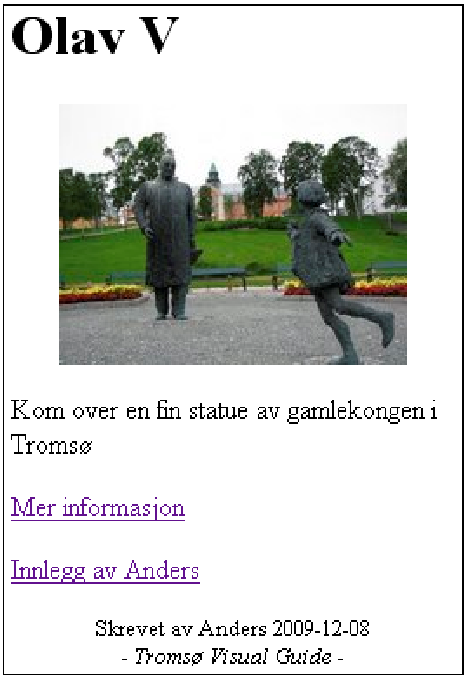
[10] *http://code.google.com/intl/nb-NO/apis/maps/documentation/staticmaps*
[11] *http://www.google.com/m/*

By using the mobile search feature of Google we assure that the pages in the result are viewable on the small screens of mobile phones. This is done by Google reformatting the pages before sending them to the phone.

2.2.4 Blog

When the user finds something of interest it is not uncommon that she would like to share this with others or maybe just store it for herself. For this reason the Tromsø Visual Guide also comprises a blog function. This is envisioned to be used as a part of a journey diary.

Having made a query to TVG and viewing the subsequently returned guide page, the user can select the "Blog this" option on the lower part of the page. This will start a dialogue which allows the user to enter the text of the blog post in a web form. Upon submitting the form the text will be combined with the users query image to form a blog post in our basic travel blog[12] (See Figure 7). The post will also contain a link to the attraction in the guide, and a link to all posts made by the user.



Olav V

Kom over en fin statue av gamlekongen i Tromsø

Mer informasjon

Innlegg av Anders

Skrevet av Anders 2009-12-08
- Tromsø Visual Guide -

Figure 7: Example of a blog post in our basic travel blog.

Many people today keep their own blog or have their own page where their post their photos. For these people it would probably be preferable if it was possible to post to their preferred blog. Based on this assumption we have integrated the blog feature in TVG with the blogging service *Blogger*[13] and the photo site *Flickr*[14]. The user will have to enter the services' login credentials and log into

[12] *In cases where the user simply navigates the guide without making a query, the image of the attraction from the guide will be used in the blog post.*

[13] *http://www.blogger.com*

[14] *http://www.flickr.com*

the 3rd party blogging service, in order to post to these services. Social services like Facebook and Twitter could also be added as possible post options.

The phone number is used as the identifier of the user. This phone number is included in the link that is returned as the answer to an image query by TVG. All subsequent links will also include this id in order to track the user. This is however a very insecure scheme for linking content to the user. A commercial solution would probably need to use proper authentication methods in order to verify the user id before being able to create content in the system.

2.3 Content

The information made available in TVG is stored in a database referred to as the POI database. The same database has been used for storing information in several other prototypes in the CAIM and MOVE[15] projects. It currently holds information about 31 attractions (POIs) in Tromsø (see appendix 1). Each POI has a name, a description and a geographic position.

In addition to the attractions the database also contains basic information about places to eat in Tromsø. This information includes name, short description, price level, opening hours and geo-position. It currently holds 19 restaurants and fast food places. This table is used to demonstrate the marketing part of TVG.

The content of the database has been added from various online and offline sources during the project period. The content is not necessarily up to date.

The image database used by ViPR has been generated from a set of 290 images of 29 POIs in Tromsø. The images have been taken by summer interns during the summer 2008, using a 8 megapixel Canon PowerShot S5 IS camera. Most of the images are automatically geo-tagged from GPS by using the program RoboGEO[16] to extract the correct position from the tracklog of a Garmin eTrex Legend Cx, and write it to the image's EXIF header. The images were resized to 640 x 480 pixels, using Google Picasa, before they were submitted to ViPR. The resizing was done in order to increase the performance of ViPR.

2.4 Testing ViPR on tourist attractions

Most of the initial testing of ViPR was done on two-dimensional objects like CD-covers or poster ads [Canright et al., 2007]. But in the TVG scenario the image recognition would also have to perform well on three dimensional objects in an outdoor environment.

During the summer 2008 we did some test in order to check how well ViPR would perform on these "real world" objects. At the time of this test the TVG service was not operational, so the images, taken with a Nokia N95 phone, were transferred to a computer and then used in queries to the ViPR system using the command line tool provided by Evolution Robotics. The image database consisted at the time of 72 images of 6 different points of interest (POI) in the town of Tromsø (see appendix 2 for details). The images in the database and those taken with the mobile phone were all taken under similar weather conditions (clouded). A total of 27 mobile phone images were used to query the database. These were mainly taken in front of or from the side of the POI - a point of view we think tourists would prefer for their pictures. 22

[15] *http://move.tele.no*

[16] *http://www.robogeo.com*

pictures were taken of the 6 objects present in the database, while 5 pictures were taken of objects not present in the database. These 5 pictures were taken in order to test ViPR on false positives (i.e. match registered on an image when no correct match is available in the database).

2.4.1 Results

When a query is posted to the ViPR system it will return a ranked list of hit images. If the image on top of the list is of the same POI as the query image we consider it to be a hit – correct identification of the POI.

Out of 22 images of taken from different angles of the 6 POIs in the ViPR database, 21 images were correctly identified. Only one query did not return any hits. This picture was taken at a very low angle in front of a statue (Figure 8), a position most tourist would not use.

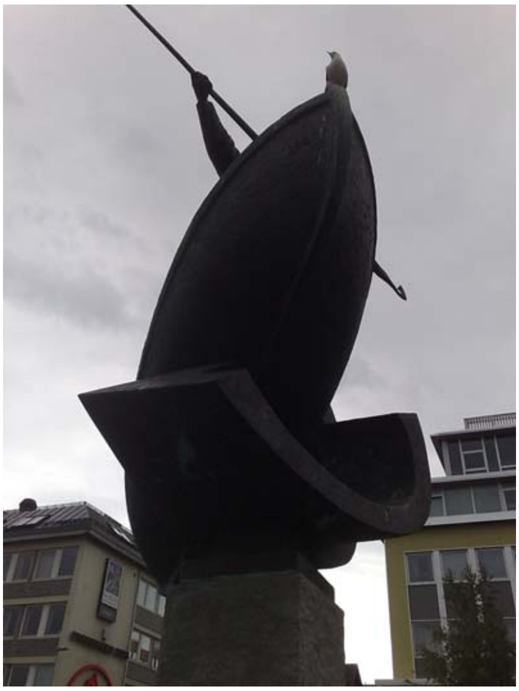


Figure 8: The only image not to find a correct match. The image is taken very close and below the monument.

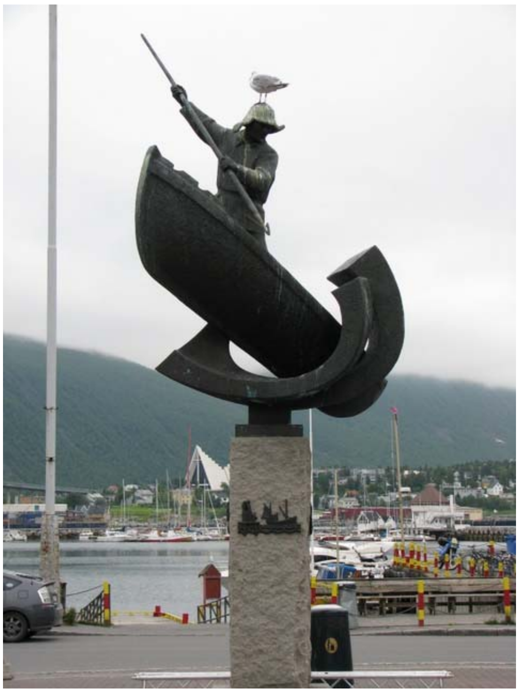


Figure 9: An image of the same monument taken from a more likely angle for a tourist.

Two of the images which were correctly identified additionally returned false positives (i.e. wrong images were identified as hits, but not ranked as the best hit). A closer look at the hit images revealed that these images were not so “wrong” after all. The POI we were trying to identify was the town hall (Figure 10), while the false positives were of a statue of King Håkon VII (Figure 11). However, the statue is placed not far from the town hall, and on some images in the database the town hall is visible in the background. So while these hits were registered as false positives, the correct POI was also in these images.

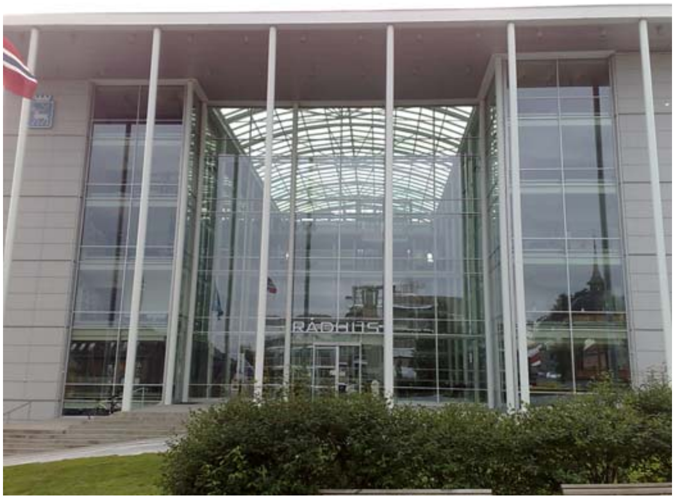


Figure 10: The query image of the town hall.

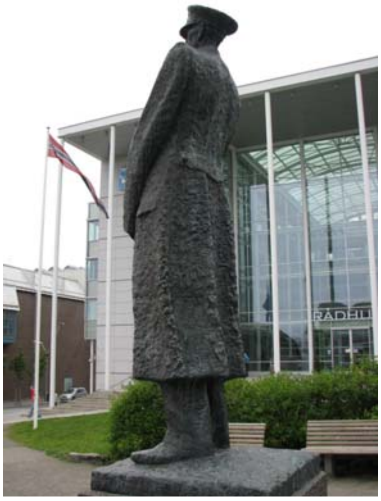


Figure 11: The result image of the statue with the town hall visible in the background.

Of the five “trick” images used to query the database, one returned a hit – a false positive. The query image was taken of some scaffolding under a tarpaulin cover which had the name of the entrepreneur printed on the side (Figure 12). The hit image was of the Tromsø Cathedral (Figure *13*). However, a closer look on the hit image revealed that the entrance of the cathedral was covered in scaffolding from the same entrepreneur. Again ViPR proved to be better than what our tests had anticipated. Although registered as a false positive, the match between the images was actually correct.

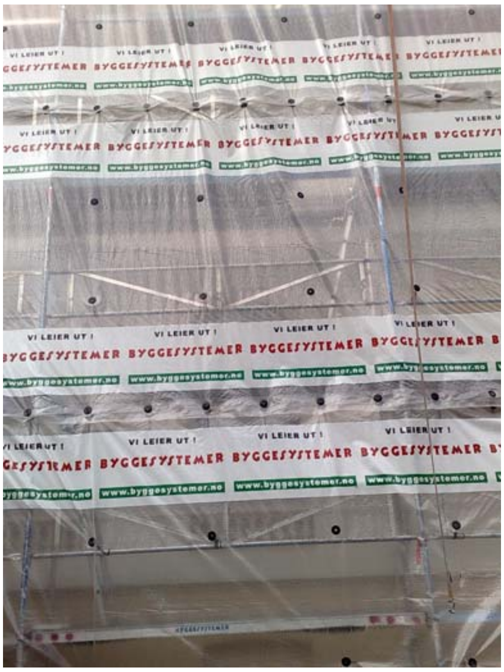


Figure 12: The query image of the scaffolding.

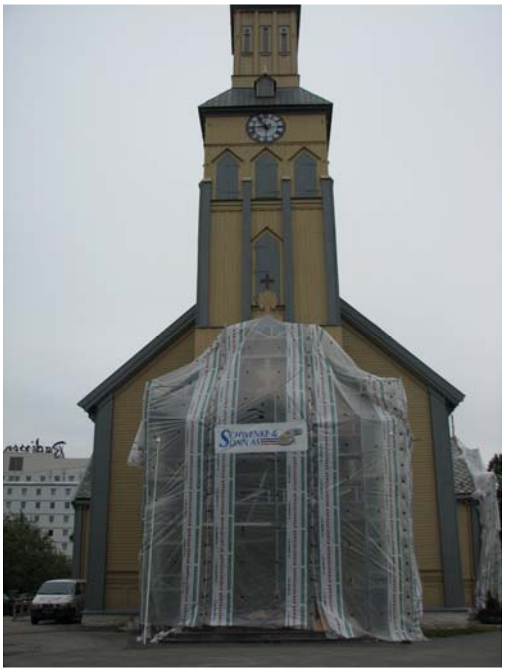


Figure 13: The image which came up as a match for the scaffolding image.

This test was intended as an early indication on the ViPR's feasibility for use in recognition of three dimensional objects, and cannot be used to say anything conclusive about ViPR. There are many limitations to our test. The number of images in the database is low and of POIs even less, meaning that there are relatively many images of each POI. These factors make the task much easier for ViPR. However, the result from this test is very promising and exceeded our expectations.

An overview of the test results can be found in appendix 2.

3 Discussion

3.1 Messaging or dedicated client

For this demo we chose to rely on the messaging and web capabilities of the mobile phone in order to communicate with the end user. This choice makes the service available to most modern camera phones. The user interface would be familiar to the user, although it could be considered cumbersome. The interface would also vary according to phone models, making it more difficult to create a unified description of the user interaction.

An alternative to utilize the built in services would be to create a dedicated client for TVG. This would ensure a similar user interface across handsets, and also a more logical user interaction in that the user would not have to switch between the messaging clients and the web browser. The user would also receive feedback when the system was searching for similar images, and the search process would be faster due to fewer systems involved and no need to wait for a poll timer to time out. However, the user would have to install the client herself, a task which would be difficult and unfamiliar for most mobile phone users. A dedicated client would also probably take longer time to develop than a messaging based client, and would reach fewer mobile phones since the same client could not be used on all phones due to different operating systems.

The best option would probably be to have the service available through both a message service and a dedicated client.

3.2 Service discovery

The challenge of how to make the user aware of the service remains independent of whether the service is based on a dedicated client or a solution which utilises the messaging services in the phone. For the dedicated client the user would need a link to where one could install the service. For the message client the user would require the phone number of the service. Today the service also relies on a code word, but this could be made redundant if there were only one service using the specific number.

In a tourist setting one could use the printed tourist information catalogues to disseminate information about the guide service. This would, however, probably just reach a part of the potential users of such a service. Informing tourists about a service through SMS when they reach a new country or town could be a tempting way to increase the awareness of the service. But in Norway this kind of SMS marketing is not allowed.

3.3 Response time

In the current version of TVG we rely on a timer which polls for messages every minute from the MMS repository. This is very ineffective and in the worst case scenario adds a minute to the response time of the service. This could be improved by implementing a web service which could receive a notification each time a new message arrives in the repository. Our MMS service provider, Telenor Playground, would like to support such a solution, but it is currently not operational.

3.4 Image recognition

Doing image recognition on three dimensional objects require many images from different angles and in different surroundings in order to yield good results. Collecting and adding these images manually could prove to be a challenge. A way to overcome this could be to add images used in queries, which have been recognized as similar (above a certain threshold), to the image recognitions pool of images. In doing so the image recognition would probably become more accurate the more people use it.

A similar scheme could be used to add tags to images. For instance if people tagged the images before adding them to their blog, these tags could be presented as possible tags for the next person using the service and wanting to blog about it. This idea is already being used by the experimental auto annotation system *SpititTagger*[17].

For a large scale guide system, containing many thousand of images and POIs, a pure image recognition based service will face performance challenges. A way to improve the response time of the system could be to only perform image recognition on a subset of the images in the system. By utilizing context information about the images in the system, one could narrow down the collection of relevant images. For instance position could be used to only search through images in the vicinity of the query image. GSM positioning through the mobile phone network would provide adequate precision for this application. Another type of context would be time of year, so only images from the same season would be searched. And a third example could be weather conditions. These types of context don’t necessarily have to be used as a filter for the search, but could just as well be a factor contributing to the ranking of the search result, making the overall rank more relevant and accurate.

3.5 Related work and commercial efforts

In the later years there has been a rapid development in content based image retrieval (CBIR) and similarity search. From only being of academic interest, we are now seeing usable services being launched for commercial use. Google revealed their interest in image search in August 2006 when they acquired the image recognition company *Neven Vision*[18]. In April 2009 Google Labs launched a beta of their service *Google Similar Images*[19] which in part is based on CBIR. The Microsoft search engine *Bing* has also included a similarity search feature for image search. Both these search engines only allows you to perform similarity searches on images already in the search index. I.e. it is not possible to upload images in order to perform an image search. But this option is available from niche search engines like *TinEye*[20] and the visual fashion search service *Like.com*[21]. Both these search engines have also recognized the potential the mobile phone holds for visual search, and created search clients for the Apple iPhone.

Some mobile handset manufacturers have also recognized the potential of the mobile phone as the primary device for making visual queries. Nokia have launched a beta of the service *Point and Find*[22]. They offer a search client for

[17] http://water.ece.ucsb.edu/jkleban

[18] http://google.about.com/od/n/g/nevenvisiondef.htm

[19] http://similar-images.googlelabs.com

[20] http://www.tineye.com

[21] http://www.like.com

[22] http://pointandfind.nokia.com

their Symbian S60 mobile phones, but also a portal for marketers wanting to run a marketing campaign based on visual search. The scenario where the user snaps an image of a movie poster, CD-cover or book cover has also been acknowledged by several other visual search services like *Kooaba*[23], *SnapTell*[24] and *SnapNow*[25]. Recently Google launched a beta of their new visual search client for the Android operating system, named Goggles[26].

3.5.1 Alternatives to image recognition

TVG uses images to create a connection between the information in the digital world and the attraction in the physical world. This could also be achieved in other ways. For instance could 2D barcodes [Schürmann et al., 2008b], or in the future NFC/RFID [Hardy et al., 2008], be used to make the connection between the physical and digital world. This would however require each object to be physically tagged with a barcode or a tag, therefore such a solution is best suited in scenarios with a limited collection of objects, e.g. in museums.

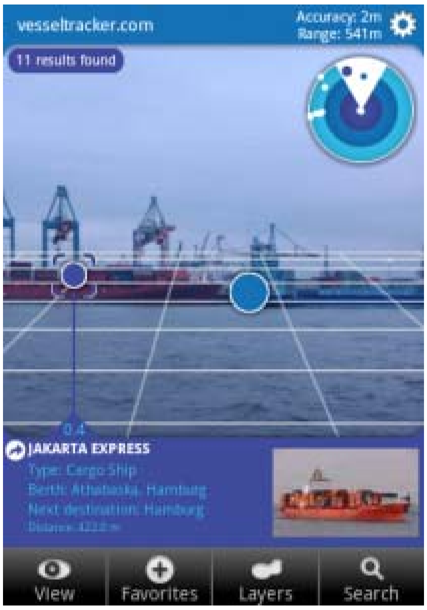
For tourist guides covering a large area of outdoor attractions, geo-positioning of the user is in many cases a better solution than tagging. In cases where the attraction covers a larger area cell positioning utilizing the mobile phone network could be used. But with smaller, well defined attractions a more precise position is needed. This could be achieved by using GPS, a sensor which is included in many modern smart phones. Much research has been done on using positioning in mobile guides [Kray and Baus, 2003]. In combination with a digital compass (magnetometer) the guide does not only know where you are, but also in which direction you are focusing your attention. In Telenor we have previously explored this idea with the Smart Binoculars concept [Bakken, 2007]. A guide using GPS and compass would be able to make the physical digital connection in a similar way to what we are doing with image recognition today. Point your phone at the object of interest and the phone tells you what you see. Since no computational intensive image recognition is needed, a query based on position and direction could be faster, leading to better user interaction. However, there are still few phones that come with digital compasses, and these are often inaccurate and vulnerable to interference. We do expect this to improve in the coming years, making GPS and compass a good option for tourist services.

3.5.2 Augmented reality

Combining the GPS, compass, inclinometer[27] and camera in a mobile phone makes in possible to present information to the user in a new and intuitive way – through augmented reality (AR). With AR you see the real physical world through the lens of the camera. But superimposed on what you see is information from the digital world, making it easy to see which objects the information relates to (see Figure 14). The sensors (GPS, compass and inclinometer) will always keep track of what the camera is pointing at and so display the correct information. Several applications are already on the market for the Apple iPhone 3Gs and Android phones. These include the mobile

[23] http://www.kooaba.com
[24] http://www.snaptell.com
[25] http://www.snapnow.co.uk
[26] http://www.google.com/mobile/goggles
[27] *An instrument for measuring angles of slope/tilt, elevation or inclination of an object with respect to gravity.*

augmented reality browser *Layar*[28] and the world browser *WikiTude*[29]. Google Goggles also includes AR features to display information.



JAKARTA EXPRESS
Type: Cargo Ship
Berth: Athabaska, Hamburg
Next destination: Hamburg
Distance: 422.2 m

Figure 14: Screenshot from Layar showing information about ships superimposed on a view through the camera lens.

A problem for these browsers is the inaccuracy of the compass resulting in a poor user experience when the information in the browser doesn't match up to what the user is actually looking at. The user experience could probably be improved by utilizing image recognition, making the phone aware of the actual object you are looking at, and not just the position, direction and angle. Preferably the image recognition should be done on the phone to secure a swift user response, but this can be a challenge for phones with little computational power. For handsets with a high capacity network connection server side processing is probably also a good option.

3.5.3 Creating content

Both Layar and WikiTude relay on 3rd parties to supply the content for their browsers. Sources include geo-tagged information from Wikipedia, Flickr, Panoramio and many others. Relaying on user generated content or crowdsourcing is probably the best way to create and maintain information in a large scale tourist guide. There are already several tourist sites, like *TripAdvisor*[30] and *VirtualTourist*[31], which relies on users to create much of the content.

3.5.4 Other applications for imager recognition

There are also other interesting applications of image recognition for services in the tourist domain. It can be used to organise the layout of images in relation

[28] *http://layar.com*
[29] *http://www.wikitude.org*
[30] *http://www.tripadvisor.com*
[31] *http://www.virtualtourist.com*

to each other, so that many images can be made into one large image which is intractable in three dimensions. This idea has been demonstrated by Microsoft in their service *Photosynth*[32]. One could also imagine adding images of different time periods in order to create a forth dimension (time) like done in the *4D Cities*[33] project. By combining these features in a mobile service with context sensor support the mobile phone could be a looking glass into an unknown city or even a known city, but at an unfamiliar point in time.

3.6 Conclusion

With the TVG prototype we have explored the opportunity of using the ViPR image recognition system for a visual search based tourist guide. Our initial research show that the system work satisfactory on three dimensional objects, and thus we have the critical component we need to create a prototype of the service. The prototype is a tool to do further research into the field of visual search. It would for instance be interesting to do user trials with this prototype to investigate if this type of interaction appeals to users in a tourist setting.

For a commercial version of TVG a few challenges will have to be met. The system will have to scale well if it is to reach a large market. This means that the image recognition system will have to be able to search through large amounts of images. Utilising context could be a key to overcome this challenge. The system will also have to scale in terms of content. Relying on the users themselves to create the content in the service could be a solution, but it is then important to get a large and dedicated user base. There is also the question on how to monetize on the service. We have here pointed to the option of monetizing on marketing income from local service providers.

We are starting to see many similar services appearing in the market. An option for a realization of the TVG service could be to build on these services. Both Layar and WikiTude provide open APIs for developers, while TinEye has a commercial API. We expect more open APIs to appear in the time to come.

[32] *http://photosynth.net*

[33] *http://www.cc.gatech.edu/4d-cities/dhtml/index.html*

References

Akselsen, S, Bjørnvold, T A, Egeland, E, Evjemo, B, Grøttum, K J, Hansen, A A, Høseggen, S, Munkvold, B E, Pedersen, P E, Schürmann, A, Steen, T, Stikbakke, H, Viken, A, Ytterstad, P. 2006. *MOVE – a summary of project activities and results (2004-6).* Fornebu, Telenor Research and Innovation. (R&I R 17/2006.)

Akselsen, S, Evjemo, B, Schürmann, A, Egeland, E,Elgesem, D, Horsch, A, Hove, L-J, Karlsen, R, Nordbotten, J. 2007. *Scenarios for CAIM (Context-Aware Image Management)*. Fornebu, Telenor Research and Innovation. (R&I N 26/2007.)

Bakken, T. 2007. *Smart Binoculars – a context aware mobile application.* Fornebu, Telenor Research and Innovation. (R&I N 29/2007.) Company CONFIDENTIAL

Canright, G, Engø-Monsen, K, Policroniades, C. 2007. *Shoot and Buy: Exploring Commercial Opportunities of Mobile Visual Search.* Fornebu, Telenor Research and Innovation. (R&I N 42/2007.) Company INTERNAL

Datta, R, Joshi, D, Li, J, Wang, J Z. 2008. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv*. 40, 2, Article 5 (April 2008), 60 pages. DOI = 10.1145/1348246.1348248 http://doi.acm.org/10.1145/1348246.1348248

Hardy, R, Rukzio, E, Wagner, M, Paolucci, M. 2008. Touch & Interact: Applied to a Tourist Guide Prototype. In: *Demonstration at NFC Forum Global Competition at 2nd European NFC Developers Summit (WIMA 2008)*, 28 - 30 April 2008, Monaco.

Kray, C, Baus, J. 2003. A survey of mobile guides. *Workshop on HCI in mobile guides*, 5th Int. Symposium. 2003

Lowe, D. 2004. Distinctive image features from scale-invariant keypoints, *Int. J. Comput*. Vis., vol. 60, no. 2, pp. 91–110, 2004

Munich, M E, Pirjanian, P, Di Bernardo, E, Goncalves, L, Karlsson, N, Lowe, D. 2006. SIFT-ing through features with ViPR. *Robotics & Automation Magazine, IEEE*. Sept. 2006. Volume: 13, Issue: 3. On page(s): 72-77. ISSN: 1070-9932. http://people.cs.ubc.ca/~lowe/papers/06robotics.pdf

Natvig, E, Evjemo, B, Prøitz, L. *Personalisation of mobile services – challenges and opportunities*. Fornebu, Telenor Research and Innovation. (R&I N 30/2008.) Company INTERNAL

Policroniades, C, Nøkleby, C, Andersen, H, Rugelbak, J, Myksvoll, K, Geers, R. 2009. *Mobile Visual Search Pilot Report*. Fornebu, Telenor Research and Innovation. (R&I R 1/2009.) Company CONFIDENTIAL

Schürmann, A, Aarbakke, A S, Egeland, E, Evjemo, B, Akselsen, S. 2006. *Let a picture initiate the dialog! A mobile multimedia service for tourists*. Fornebu, Telenor Research and Innovation. (R&I N 33/2006.)

Schürmann,A., Evjemo,B., Akselsen,S. 2008a. *Tromsø Visual Guide*. Forskningsdagene, Tromsø, September 26-27, 2008. http://caim.uib.no/publications/TVG_poster_9-2008.pdf

Schürmann, A, Egeland, E, Akselsen, S, Evjemo, B. 2008b. *Exploring non-textual interaction. User responses on the TIFF event assistant*. Fornebu, Telenor Research and Innovation. (R&I N 21/2008.)

Appendix 1: Name of POIs in Tromsø

Roald Amundsen

Latham-monumentet

Jødisk minnesmerke

Tromsø Domkirke

Adolf Thomsen

Mor og barn

Tromsø Bibliotek

Tromsø Rådhus

Fokus Kino

Kulturhuset

Vår Frue katolske kirke

Håkon VII

Olav V

Arthur Arntzen

Fiske- og fangstmonumentet

Tromsø Sparebank

Tromsø Kunstforening

Passasjer

Alberte

Verdensteatret

Roald Amundsen

Polarmuseet

Skansen Tromsø

Ishavskatedralen

Nord-Norsk Kunstmuseum

Polaria

Polstjerna

Isbjørn

Fjellheisen

Fridtjof Nansen

S.A.T.D.

Appendix 2: Results from ViPR test

Image	**Hits**	**False positive**	**Reference images available**	**Percentage %** **(hit rate)**
Amundsen1	1		12	8.33%
Amundsen2	2		12	16.67%
Amundsen3	5		12	41.67%
Amundsen4	4		12	33.33%
Amundsen5	6		12	50.00%
Domkirka1	6		27	22.22%
Domkirka2	5		27	18.52%
Domkrika3	4		27	14.81%
Fangst1	2		11	18.18%
Fangst2	2		11	18.18%
Fangst3	0		11	0.00%
Fangst4	2		11	18.18%
Haakon1	4		10	30%
Haakon2	4		10	40%
Haakon3	3		10	30%
Haakon4	3		10	30%
Haakon5	2		10	20%
Letemps1	7		7	100%
Letemps2	7		7	100%
Letemps3	7		7	100%
Rådhuset1	2	1	5	20%
Rådhuset2	3	1	5	40%
Fake1	0		0	-
Fake2	0		0	-
Fake3	0		0	-
St_fake1	1	1	0	-
St_fake2	0		0	-