telenor

R&I Research Note

R&I N 33/2006

Anders Schürmann, Anne S. Aarbakke, Espen Egeland, Bente Evjemo, Sigmund Akselsen

Let a picture initiate the dialog! A mobile multimedia service for tourists

R&I Research Note	**N 33/2006**
Title	**Let a picture initiate the dialog! A mobile multimedia service for tourists**
Author(s)	Anders Schürmann, Anne S. Aarbakke, Espen Egeland, Bente Evjemo, Sigmund Akselsen
ISBN / ISSN	0809-1021
Security group	OPEN
Date	2006.12.15

Abstract
This document presents an application for mobile phones that supports discovery of services in general, and more specific, provides tourists travelling in a certain region easy access to dynamic updated information about tourist attractions. Image retrieval is fundamental to the solution, and general challenges related to content based image retrieval is discussed.

Keywords
Mobile service, image retrieval, CBIR, MMS, tourism

Telenor R&I N 33/2006

Preface

This note describes an information retrieval service for tourists on the move. The conceptual idea is generated within the MOVE project (www.moveweb.no) and has been inspired by challenges addressed in the CAIM project (caim.uib.no) and discussions with associate professor Randi Karlsen, at Institute of Informatics, University of Tromsø. Student Anne S. Aarbakke made the first draft of this document.

Tromsø/Harstad, December 15th 2006

Sigmund Akselsen (project leader)

Contents

1 Introduction

Images are part of our everyday lives. They are vital to informative tasks, education and entertainment – in professional as well as private spheres. The penetration of digital cameras and camera phones makes it easy to take pictures anywhere and at any time. The costs of image production are minimal, and the total number of images rises tremendously. Images together with audio clips and text are available over the Internet. Still, the image retrieval process is poorly supported. These problems are addressed in the CAIM (Context Aware Image Management)[1] project where it is claimed that much Internet data could be useful for information seekers if it can be located and presented in an integrated way:

> *Currently, Internet search engines provide separate information retrieval support for each media type, requiring that users search one media type at a time. Selection and integration of these results is left to the users. In order to hinder that the multi media collections remain underutilized, efficient information retrieval algorithms must be developed* (Karlsen and Nordbotten 2006).

Within the MOVE project the particular information needs of tourists are examined and mobile services to meet these needs have been developed and tested:

> *The aim of the MOVE project is to enhance the travellers' experiences through new aesthetic expressions mediated by technological solutions that merge people, objects and places with their network digital representations. In parallel, attention is given to the many kinds of tourist providers and their efforts to face the tourists at the right moment with the right information* (Akselsen et al. 2006).

Tourists explore new arenas and they want to understand and learn about the attractions they visit and the surroundings in general. They prepare well before the journey and bring with them travel books and tourist information guides in order to be informed of the possibilities and to find detailed information about already planned activities, accommodations and so on. Attraction theory within tourism holds that various attractions have different meanings to different tourists and groups and therefore can be classified as primary, secondary and tertiary attractions (Leiper 1990). A primary attraction is pronounced and known to the tourists in advance. A secondary attraction is also known in advance, but even if it supports selection of destination it is ranked as less interesting. Tertiary attractions are unknown before departure but can be discovered when touring, based on information underway or on location. Studies show that approximately 50 % of the decisions on which attractions to visit are made during the journey, and further that the active tourists are the most satisfied ones (Viken et al. 2004). This supports the idea on which the MOVE project is founded, namely to ease access to regional and dynamic information and make it accessible on a communication terminal – the mobile phone - that the tourist brings about all the same.

Dynamic information includes, among others, announcements of local concerts or football matches, restaurant menus, unoccupied rooms at lodgings, activities recommended on sunny days only, etc. This kind of information may change,

---

[1] *http://caim.uib.no/*

and therefore cannot be well documented in tourist brochures printed half a year before the tourist season. A mobile information service built upon information sources that are continually updated should grant for access to dynamic information.

In the combination of these two strands – the problems *and* possibilities related to image retrieval management and mobile multimodal information services – the task reported here was defined: *Design and implement a mobile service for tourists that take advantage of the mobile phone's photographing and image transfer facilities.* The following scenario has guided the work:

> *Imagine a tourist travelling in the Lofoten archipelago. She is on the ferry heading for her first destination. To shorten the time she looks at a tourist magazine and finds amusement in the glossy pictures. There are beautiful mountains and picturesque fisherman's villages. And look – there is the funny tail of a whale close to a boat seated by tourists. Perhaps a whale safari could fit into the – so far - loosely composed holiday program. The announcement informs about where to go and when to start, but she is curious about the conditions on board, the movements of the sea in such a small boat and the likelihood of actually meeting the whale on a trip like this.*
>
> *She resolutely takes a picture with her camera phone. The announcement of the whale safari is the object of the picture. She links the picture to an MMS[2] and dials a four digit code which seems pretty familiar to her. Within a minute she receives a MMS containing several pictures from recent safaris, and an inspiring and confident voice invites her to participate. The forecast for tomorrow is promising and the early morning tour is not fully-booked yet. When she finally watches the video clip she cannot think of anything better than a whale safari. One extra keystroke and she is talking to the advertiser. She cannot wait to forward the MMS to her friends with the subscripts: Tomorrows adventure!*

The MMS to Search (M2S) service described in chapter 2 is an example of a service that exploits non-textual queries for more information. It represents services that realise a direct link between physical objects and digitalised information that exists about these objects. No annotation or language interpretation is needed. But, there are great challenges related to image retrieval and image management. These challenges are discussed in chapter 3.

The design of the system was made on the presumptions that both open source system code for image retrieval and efficient implementations of multimodal messaging protocols would be available.

[2] *Multimedia Messaging System*

2 The M2S concept

M2S is a short name given to a concept that enables search in a database dependent on an image request transferred as an MMS. A single picture, taken by a camera phone, initiates an interchange of MMS messages. The burden of formulating the request for information is put on the system. See appendix 1 for an exemplified presentation of M2S[3].

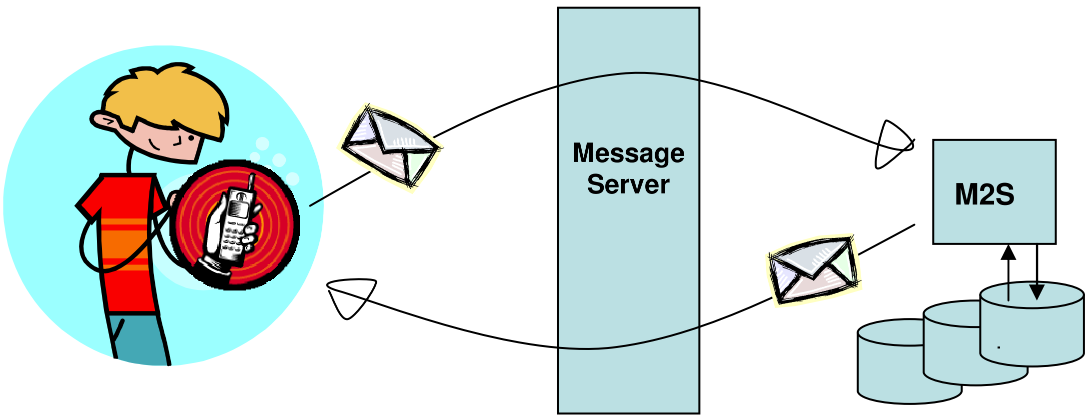


Figure 2.1 M2S enables a multimedia dialog

2.1 Functionality

The M2S service should provide information and experiences beyond those of traditional information sources available when travelling. The service is expected to offer

- More detailed information
- Recent updated information, e. g. extraordinary opening hours, today's menu, special weekend fares etc
- Multimedia information, e. g. links to video clips or WAP sites or other relevant tourist services
- Easy access to booking facilities or direct contact with booking personnel

The intention was to make M2S accessible by anyone carrying a camera phone. However, as the hit rate of the image retrieval system varies with the quality of the image request, a high quality camera is preferred.

2.2 Design and logical connections between modules

A modular design was chosen to allow easy exchange and plug in/out updates of components included in the image retrieval system and the messaging

[3] *http://move.tele.no/dist/m2s/M2S.pps*

support system. In figure 2.2 the shaded boxes constitute the image retrieval system. As mentioned in the introduction there are great challenges related to image retrieval. These challenges are addressed in chapter 3, together with an outline of how these modules became part of the M2S service as it appears today.

The system is initiated with an HTTP POST request to the RecvMmsServlet. The request contains the contents of an MMS in the form of a SOAP message. The request is forwarded to a SOAPReceiver that parses the message and returns an object that holds the contents of the MMS message. The image-part of the message is retrieved and written to disk. Then the method lookupImage in the M2SController is called, with the originator phone number and path to the image-file as parameters. The M2SContoller then calls the searchForSimilar-Image method in the ImageSearcher to do the actual image recognition.

ImageSearcher returns an id to the most likely image, or NULL if no image with sufficient resemblance is found. If the image id is not NULL, the M2SContoller queries the ContentServer to lookup the database for the answer for this image id. Depending on the nature of the answer, M2SController will call an appropriate method in the ResponseHandler to send the answer back to the originator of the query.

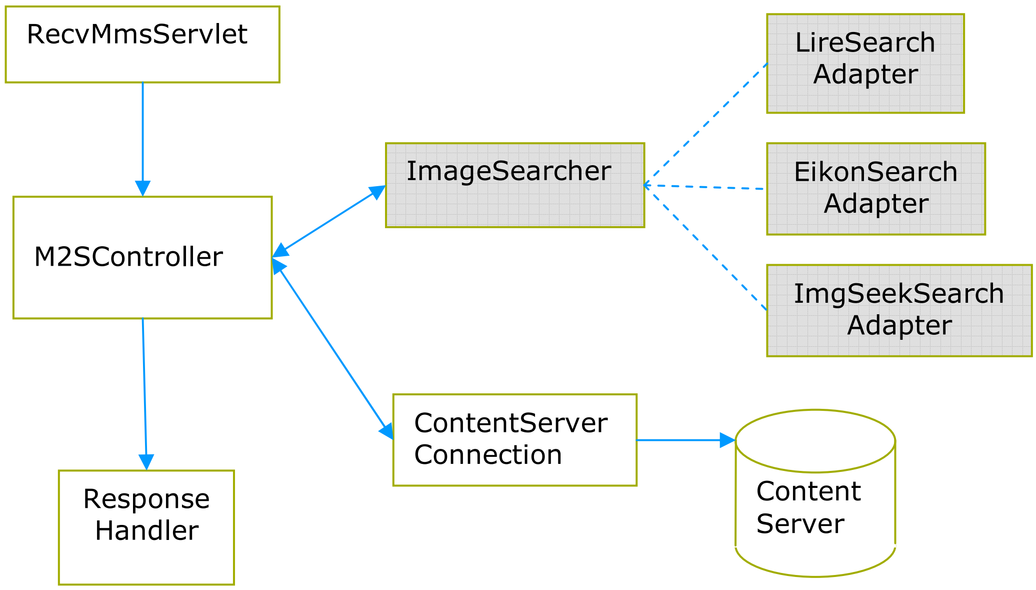


Figure 2.2 Overall design of M2S

The M2S system has been implemented in Java 1.5.0 and with the use of IDEA IntelliJ 5.0 and Eclipse 3.1 IDE tools. The system will be run on a Windows 2000 based server (ACER Aspire T135), with an AMD Sempron 3100 processor and 1024MB of RAM. The server is running Apache Tomcat 5.0 and has a MySQL 4.1 database. The following two devices were used in order to take pictures and send test MMS messages: Nokia 6680 with a 1,3 megapixel camera and Sony Ericsson K700i with a 0,3 megapixel VGA camera.

2.3 Messaging

The M2S service relies on messaging in order to receive input and send output back to the user.

2.3.1 System input

The user takes an image with her mobile phone, includes a service code, and sends it as an MMS to a special four digit phone number, often referred to as a short code[4]. There is no need for further annotations. During testing the service codes "m2s" and "move" were used together with the phone number 2034.

This particular number is controlled by the PATS[5]-lab in Trondheim. The PATS laboratory is partly owned and run by Telenor, and is supposed to facilitate the development of new telecom services. To register and operate such a number is relatively expensive and several services are therefore allocated to the same number. This strategy makes it necessary to identify which service the incoming message belongs to, and here the codes become efficient separators. However, it is not an ideal situation to put this kind of memory burden on the user.

Figure 2.3 An MMS ready for transfer. Here the short number 2177 is used

When the message server at the PATS-lab receives an incoming MMS message it looks for a code in the subject field or in the first text element of the message. The system uses this code to identify a corresponding servlet's URL. If a mapping is found the MMS message will be forwarded to this servlet.

2.3.2 System output

The M2S system will reply with an MMS, a WAP Push or an SMS message, depending on the result of the image retrieval and the responses available in the content database. The ResponseHandler takes care of all aspects concerning outgoing communication.

[4] *http://en.wikipedia.org/wiki/Short_code*
[5] *Program for Advanced Telecom Services*

If there is no MMS available, the system will search for a relevant link to a video stream. A successful match results in a message with the link included. If not, a search for a link to a WAP site is initiated. If a WAP link is found it will be included and sent to the originator, if not, the message will contain a link to a general WAP page with information about Lofoten.

Messages that include links can be sent as either WAP push (SI - Service Indication) or SMS messages. In SMS messages the link is written as ordinary text, but most modern mobile phones will be able to identify the link and highlight it and make in clickable. WAP push messages will have the link embedded in the message in a special field. Dependent on the phone the link might or might not be visible. The keys on the handset are usually connected to the link in such a way that it is easy for the user to activate the link. An example of a WAP Push message can be seen in figure 2.4. For this particular handset WAP Push message and the link enabled SMS message would look very similar. Although the WAP Push message was invented to make it easier to use links in messages, there are indications that users prefer links in SMS messages instead[6] / [7].

To send WAP Push and SMS messages through Telenor's network we have utilized the CPA interfaces supplied by the PATS-lab.

Figure 2.4 WAP Push message on a Nokia 6680

If there is no relevant response available, an SMS message is sent to inform the user. Alternatively the service returns a link to a regional tourist information WAP site. The WAP Push alternative is therefore used in this situation as well.

2.4 Content

During the project period 6 MMS messages and 3 videos in a format suitable for streaming or download to mobile phones were produced. They are all relevant to the tourist scenario presented in the introduction. See appendix 2 for the complete list and short descriptions. The following section outlines some challenges related to production of MMS messages. Finally the content database is described.

[6] *http://mobhappy.com/blog1/2006/07/10/what-a-waste/*
[7] *http://www.tomhume.org/2006/08/wap_push_and_cu.html*

2.4.1 Production of multimedia messages

An MMS can contain text, images, videos, sound, and to a certain degree, also animations. A particular language, called Synchronized Multimedia Integration Language (SMIL[8]), can be used to control the layout and timing of the different elements within an MMS. SMIL is an XML based markup language with a tag structure similar to HTML with a <head> and a <body>-tag. Production of an MMS is a three step process: design, converting and testing.

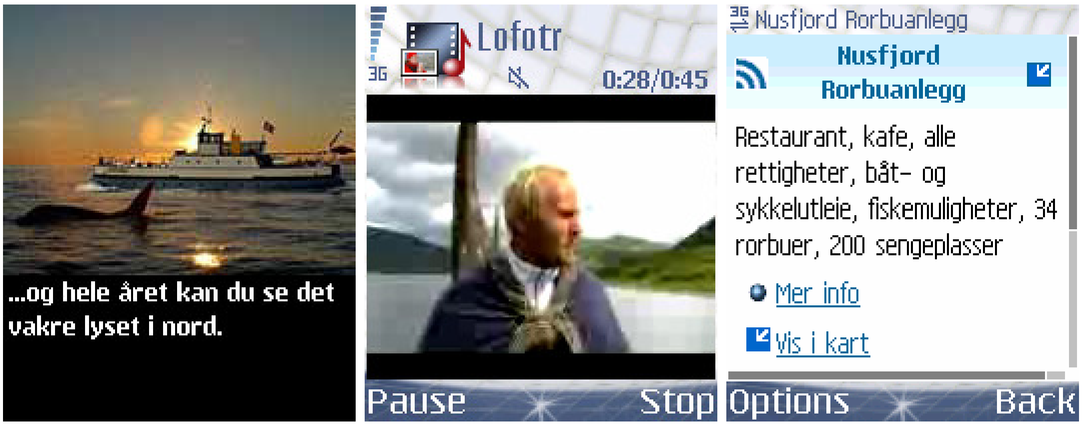


Figure 2.5 A response MMS can be constituted by several elements: a picture, a video clip and textual information with links to WAP sites

Design

The design of an MMS benefits from an intuitive and graphical approach. For web-site production there are several graphical applications available to manipulate multimedia elements. These applications produce HTML code automatically. Similar SMIL applications for MMS development were hard to find. The ones which were available were targeted more towards consumers than developers, had limited functionality and did not provide precise control and preview mechanisms. Among them were Sony Ericsson MMS Home Studio9 and Nokia Mobile Internet Toolkit 4.1[10]. We decided to use Adobe GoLive CS2[11] in our production of SMIL code. In particular the preview function with syntax check and colour highlighting was valuable. However, the final result as it appeared on the mobile phone was not always predicted in the design phase.

Converting

In order for the MMS to be displayed on a mobile phone it has to be converted to a binary MMS-format, and saved as an .mms file. Tools that would assist the developer with this job proved difficult to find. We ended up using Nokia Developer's Suite for MMS[12] version 1.1, a tool first launched in 2003. We later discovered that this functionality was also provided by NOW SMS/MMS as well.

[8] *http://www.w3.org/AudioVideo/*

[9] *http://www.sonyericsson.com/fun/wxhtml/software1/softwaredetail1?campid=2375&type=software*

[10] *http://forum.nokia.com/info/sw.nokia.com/id/d57da811-c7cf-48c8-995f-feb3bea36d11/Nokia_Mobile_Internet_Toolkit_4.1.html*

[11] *http://www.adobe.com/products/golive/*

[12] *http://forum.nokia.com/info/sw.nokia.com/id/dcc1a536-7640-4e0b-8ce6-0e1605974b4f/Nokia_Developer_s_Suite_for_MMS_Version_1_1.html*

Testing
None of the Nokia phone emulators which were tested were able to play the MMSs in a realistic way. Transferring the MMSs to the phone by USB or Bluetooth was not successful as the phone did not recognize the data as an MMS. In order to test the messages on the phone we had to send the messages through the telecommunication network. In order to achieve this we utilized a 60 days free trial of MMS gateway called NOW SMS/MMS[13]. This gateway, running locally on the developer's laptop, was connected to Telenor's MMSC through a mobile phone, which was connected to the laptop via Bluetooth. In this setting the phone was both the sender and the receiver. This procedure was somewhat cumbersome (and expensive), but still satisfying as the MMSs were received in the same manner as an MMS normally is received by a phone. Mobile phone manufacturers support the SMIL standard differently, and there are also different versions of the SMIL specification (1.0, 2.0 and 2.1). This means that the final result on the phone might be surprising even though these challenges have been addressed in both design and converting phases. To avoid such discrepancies the best strategy is to use simple and basic SMIL elements only.

The display area that is available for display of images also varies a lot on different phones, and it is not necessarily equal to the screen size of the device. Usually too large images will be scaled down in order to fit the screen of the device. This might result in images of low quality and make them unreadable if they contain text. To avoid this, images have to be adapted to each device, which was not an option for this project.

For a service to send SMIL controlled MMSs there is also a need for support for this in the telecom networks and in the mechanisms that give third parties access to these networks. We have experienced that this is not always the case.

2.4.2 Content database

In cases where the image recognition part of the system finds a resembling image, an appropriate answer has to be created and sent back to the originator of the MMS. The mapping of images and answers is kept in a simple database.

We chose to use the freely available MySQL DBMS[14] for our database. The reason for this choice was that we had some previous experience with it and that it was already installed on the server the system was going to run on. Since the functionality required by our database is limited, we could probably have used a much smaller and simpler DBMS.

The database contains only one table (see table 2.1). The field *Image ID* is the name of the image file and uniquely identifies the image. *Name* is a name associated with the image, usually the name of a company, attraction or location. The *MMS Zip-file name* contains the unique name of the zip file containing an MMS presentation connected to the image. *WAP URL* contains a link to a WAP site containing more information in relation to the image. *Video stream URL* contains a link to a streamable video presentation related to the image.

Table 2.1 Data fields in the M2S content database

Image ID	Name	MMS Zip-file name	WAP URL	Video stream URL

[13] *http://www.nowsms.com/*
[14] *Database Management System*

Although the last three fields might all contain relevant information, this will typically not be the case. Since the system is only supposed to give one answer, the table could actually contain just one field and an id to identify what kind of content it contained. We have chosen to use three fields since it varies which type of content certain mobile phones can accept. For instance streaming is not recommended for phones without 3G capability. However the system in its current version does not test the capabilities of the phone before sending the answer.

The database should probably be broken up into two tables according to normalization theory. This would result in one table for the image id and associated name, and one table for the content. The database would then become extensible and it would be easy to add other categories of content later on.

At the time of writing the database holds 54 rows, meaning that there are 54 images that the system could match the queries to. For these 54 images 12 have MMS answers, 37 have a WAP link and 9 have a link to a streamable video. For 9 entries in the database there are no answer content. In these cases the system will return a WAP link with general information of the Lofoten region. The WAP sites referred to are a subset of approximately 200 sites that were produced for the J2ME-based[15] service *KartGUIDE* (Evjemo et al. 2005). Some of them are dynamically generated and contain a short description (exemplified in figure 2.5), while others hold extended information, including images. The 54 images were extracted from a PDF version of Lofoten Info-Guide 2006[16].

[15] *Java 2 Mobile Edition*
[16] *http://www.lofoten.info/InfoGuide_06/index.htm*

3 Image retrieval – challenges and tests

M2S was developed to illustrate the potential of such services, and not to meet requirements stated for real use services. This principle influences the selection of system modules and also the test procedures. We still want to pay some attention to the major problems related to image retrieval.

3.1 Overall challenges

The result of a query depends on the precision of the query, the organisation of the information and the search mechanisms. A successful information search is characterised by *high recall* (proportion of relevant information in the search compared to the sum of relevant information in the source) and *high precision* (proportion of relevant information in the search). These characteristics are therefore useful in an evaluation of the strength of a search mechanism or a search engine.

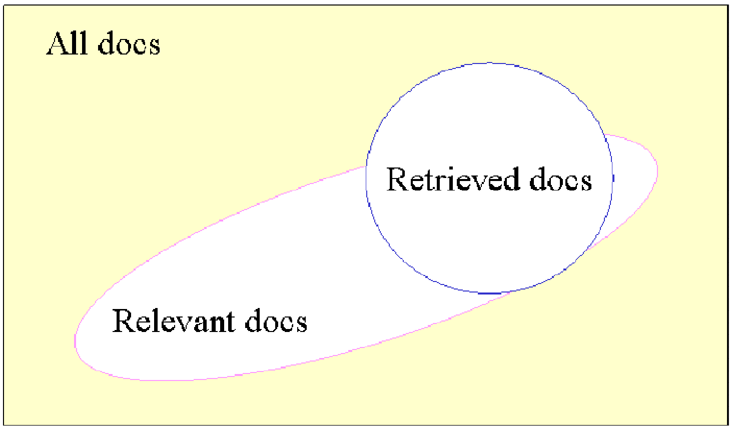


Figure 3.1 This search has high precision and medium recall, modified after (Nyland 2006)

Image retrieval might be initiated by a text based query, a content based query or a context based query, or a mixture of these strategies. A text based query presupposes that the images are annotated and that the seeker is able to construct a query that matches the vocabulary of the annotations, both syntactically and semantically. Manual annotation results in subjective interpretations of the content or the message of the image. When different persons take part in this work they interpret and annotate differently – which of course will affect the retrieval results. In addition, annotating is resource demanding and therefore a strenuous and not always feasible strategy.

3.2 Content based image retrieval

The scenario presented in chapter 1 describes an image based query using a picture taken of an announcement in a brochure. Content based image retrieval (CBIR) is no trivial task. Pictures that are regarded as equal by human beings might be regarded as unequal by computer programs. The retrieved information depends on the level of precision, which is set according to the purpose. In medical environments, for instance an examination of x-ray images, the precision demands are extremely high. In contrast a travel agency's inquiry for a picture of northern lights may have a relatively low demand for precision. Another example; pictures of a house taken from slightly different angles are still regarded pictures of this particular house – judged by a human being. A computer program would probably conclude differently. Most matching algorithms assess equality and similarity due to proportions of colours or match of contours and shapes (pattern recognition).

There are many content based image retrieval systems, and several are made available as open source. Some systems require that the query is posed as an example image provided by the user, others require a sketch or contour lines given by the user and other systems again require specifications of salient colours in the retrieved image sample. The Java based system Octagon[17] presupposes for instance a database structure and some preliminary analysis before processing, and proceeds incrementally based on previous relevance of search results. The GIFT[18] system presupposes an indexed database and performs therefore an indexing task as the first step.

The precision and the recall of the search vary of course according to the queries and additional searching mechanisms. Further, some CBIR systems presuppose user feedback in the image retrieval and matching process - an iterative process which is not recommended in a mobile setting where efficiency and simplicity are important factors.

The chosen image recognition system for M2S is described in section 3.3 and a brief survey of some open source image based query systems is found in appendix 3.

3.3 Requirements and selection of CBIR system

To build a content based image retrieval system from scratch is challenging and resource demanding, and far beyond the scope of this project. Therefore a search for open source systems was initiated. However, open source systems – and commercial available systems as well – are often poorly documented. Some systems are published as a web demo only, with no possibilities to test the system with relevant image collections. These circumstances made the search for an appropriate image retrieval system more difficult than anticipated.

The following requirements to the CBIR-system were stated:

- It must be possible to create a database or collection of relevant images for the system to query against.
- A query should result in a list of images ordered by relevance.
- The system must be able to produce a valid result in a single operation without relying on user feedback.
- The system should not presuppose any other information but a single picture (neither annotations nor context information) to make the query.
- It should be possible to programmatically interact with the system. Preferably through a Java API.
- The system should be able to give a valid result within 30 seconds.

[17] *http://users.utu.fi/jukvii/octagon/*
[18] *http://viper.csse.monash.edu.au/~davids/cgi-bin/PerlMRMLClient/PerlMRMLClient.pl*

After reviewing several CBIR systems we selected a few for further examination (see appendix 3 for a brief overview of several open source systems). The most promising one at the time being was LIRE (Lucene Image Retrieval)[19]. LIRE utilizes Caliph[20] to create an index of all images in its database. The images are indexed from scalable colour, colour layout, dominant colour and edge histogram. To search through this index and retrieve the most resembling images LIRE relies upon the information retrieval API Lucene[21].

In addition to LIRE a CBIR system called Eikon[22] was tested, both independently and together with LIRE. Eikon is a multiresolution image querying algorithm based upon wavelets, described by Derose et al. (1996). Eikon converts the image to a 128x128 pixel thumbnail before decomposing each colour by means of a Haar wavelet[23].

As LIRE did not provide optimal results (see next section) another system, named imgSeek[24], was tested. imgSeek is a photo collection manager and viewer with a CBIR capability. It is primarily a so called stand alone program, but it can also interact with other systems. Like Eikon, imgSeek relies on multi-resolution Haar wavelet decomposition.

Recall and precision
In general there is an inverse relationship between recall and precision. Comprehensive retrieval implicates use of synonyms, related terms etc. As a consequence, precision will suffer. Unfortunately, if the searcher does not use these techniques he will not achieve high recall. Knowing the goal of the search to find everything on a topic, just a few relevant papers, or something in-between - determines what strategies the searcher will use. The goal of M2S is to retrieve relevant information related to the attraction that the tourist is aiming at, and it is very important that the tourist is given relevant information (i.e. not info about other attractions than the one aimed). If a match is not found with high probability, a link to general tourist information about the area should be replied. M2S thus puts a high demand on precision and less on recall.

3.4 Test results

The messaging part of the service was not completed within the project period due to problems with the access mechanisms to the telecommunication network that were uncovered during the development phase. More precisely the challenges were tied to lack of support for receiving MMS, for sending SMIL enabled MMS, and application library conflicts between different modules of the system. Subsequently these challenges have been solved, but there have not been enough time for extensive testing of the M2S service as a whole. The handsets and the image retrieval systems, however, have been tested more thoroughly.

3.4.1 The handset and the photographer's conditions

The precision of the image retrieval system seems to be highly dependent upon the quality of the images that are used as query images. The easiest way to

---

[19] *http://www.semanticmetadata.net/lire/*
[20] *http://www.semanticmetadata.net/features/*
[21] *http://lucene.apache.org/*
[22] *http://eikon.recursion.org/*
[23] *Alfred Haar, a Hungarian mathematician (1885-1933) who defined a measurement and a methodology – a wavelet – used in mathematical analyses and number theory.*
[24] *http://www.imgseek.net/*

specify quality is to examine the resolution of the images. Initial testing indicated that the precision is higher for images taken with the Nokia 6680 (1,3 megapixel) than the Sony Ericsson K700i (0,3 megapixel). However, images sent with Nokia 6680 are scaled to VGA resolution before sending. Consequently there are properties other than the initial resolution of the image, which should be taken into consideration.

The light sensitivity of the camera seems to influence precision of the system. The quality of the lens and the ability to focus on close items (macro focus function) are other properties of importance. The actual light conditions influence the resulting picture. Little light could result in a dark and unclear image. Too much light might give reflections in the surface that is to be photographed. White balance will also vary with light conditions resulting in distortion of the colours in the image.

Another problem is that the format of the advertisement may not match the format of the image taken by the camera, so the image will include other information besides the advertisement making it look very different from the image which it is supposed to match as shown in figure 3.2. Also, the angle used between the camera and the object might distort proportions and make recognition even harder. Finally, the orientation of the image might cause problems. See figure 3.3.

Figure 3.2 Example of an MMS image where the area of interest does not match the format of the image itself

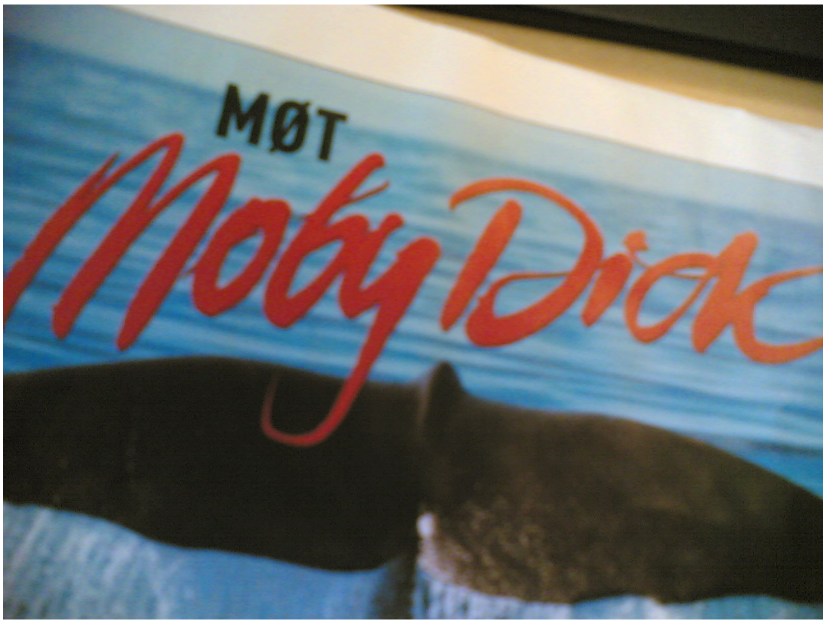


Figure 3.3 Image taken of an advertisement where the image is not correctly orientated

3.4.2 Response time and message transfer

The transfer of MMS messages in the network varies significantly with network quality and the size of the message. A very rough estimate indicates that the time needed for sending an image as an MMS with size 30 kB would be around 1.2 seconds (240 000 bit/200 000 bit/s) over a 3G network and around 6 seconds over a GPRS network (240 000 bit/40 000 bit/s). A few simple tests of the same, using a stop watch and two mobile phones, give a somewhat higher estimate. Sending an image of 30kB as an MMS from a Nokia 6680 over the UMTS network in an urban environment took on average approximately 15 seconds. Receiving the same MMS on a Sony Ericsson K700i over the GPRS network took on average 17 seconds. In addition to this processing in the network took on average 9 seconds, resulting in a total message transferral time of 41 seconds.

The reason for the variation between the test and the estimate is probably due to processing time on the phone, time needed to establish connections, and overhead involved with sending the actual message. Even though the image is 30kB the MMS would probably be somewhat larger.

It must be emphasized that this was not a scientific test of network capacity, but just to give the reader an idea about the time involved when using the system. Time taken to send a message through the network may vary considerably.

The time needed to process the query in the M2S system, from the image is received in the servlet[25], until the answer is sent from the M2S system, is about 3 seconds. Well within the 30 second requirement listed in section 3.4.1. This processing time depends on the selected query module (Lire, Eikon, imgSeek) and the content of the image database.

[25] *Servlet is Java based technology used to generate dynamic content for web sites*

The time taken to deliver the answer to the originator will vary considerably depending on several conditions, among other whether the requested information is presented as an MMS, video or WAP-page. Video and WAP will be quite fast as only a link to the content has to be sent to the originator as an SMS or WAP push message. A few simple tests indicated that it would take about 6 seconds to send a WAP push message from the M2S service.

With MMSs the content itself will have to be sent, and this will take more time. The respons MMSs used varied from 24kB to 81kB in size. Transferring the smallest MMS from the M2S server to the phone over a 3G network took on average 16 seconds. Transferring the largest message over the GPRS network took 44 seconds.

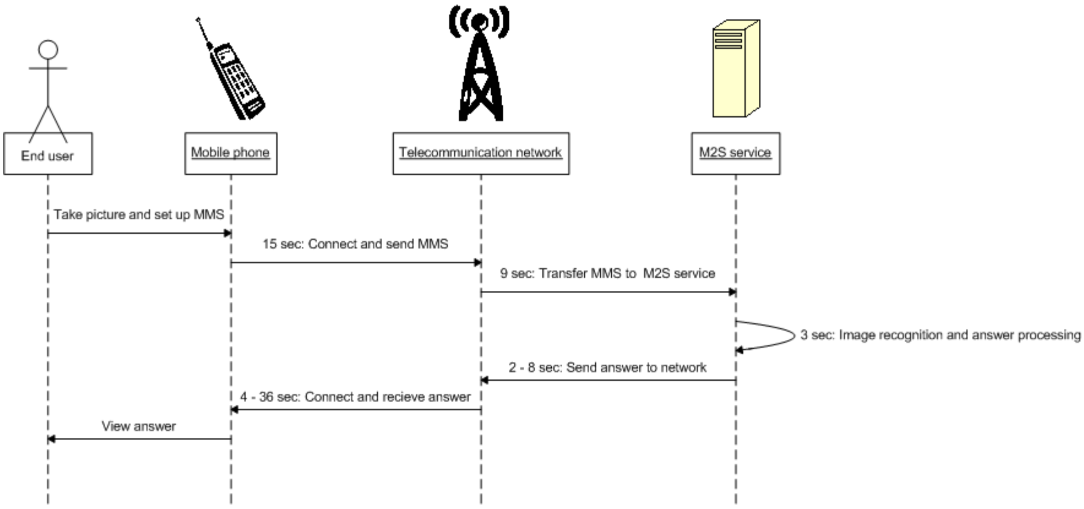


Figure 3.4 Sequence diagram of the M2S service with time estimates.

3.4.3 Image recognition system

The image recognition system was tested independently from the rest of the M2S system. The query images used in these tests had been transferred directly from the handset to the test computer. Thus these images might be of higher quality than those that would be received through an MMS service, since the images are often automatically scaled down for use in MMS messages.

Initial testing with LIRE showed a hit rate of approximately 20 %, meaning that LIRE will detect a wrong image in 4 out of 5 cases. This performance is clearly too poor to produce any service of value.

In response to a request, LIRE produces a list of result images, where the similarity to the query image – the precision - is denoted with a number between 0 and 1. An image given the number 1 means that a perfect match has been made. However, as the performance was rather poor, the top ranked image was not necessarily an image related to the actual attraction.

To increase the precision of the image recognition (cf. requirement related to precision, section 3.3) a threshold - a minimum score - was added to the algorithm to rule out non-correct images. The effect turned out to be limited as the minimum score had to be set very low in order to not rule out relevant matches.

On some query images Eikon preformed better than LIRE, but on average it was less precise. A combination of Eikon and LIRE was considered as well, but this did not improve precision of the image recognition.

Initial testing with imgSeek on the same image samples gave a hit rate of approximately 80 %. This is very promising, as it has a precision high enough to provide a service of value for the end user. For the time being LIRE has not been replaced by imgSeek, but this will probably be attempted later on.

4 A next version of the M2S service

Multimedia queries based on pictures and sounds may establish strong links between physical objects and digital information related to these objects when combined with efficient CBIR methods and tool. The M2S service is a modest example only.

4.1 Alternative searching keys

In the M2S service an image is used in the query for more information - the image is the searching key. We know – and the problems described in chapter 3 underline this issue – that image retrieval is challenging and that fundamental problems should be solved before retrieval systems with sufficient high hit rates are available. Until then, alternative searching keys should be examined.

To label actual artefacts with visual markers is one such alternative strategy. Bar codes, for instance, are well known from retail stores and retail distribution, and further matrix codes as shown in figure 4.1 have been used. Visual markers, like the CybStickers (Rahlff 2005), are more artistically designed. They are used as visual links between persons that casually visit the same location. Visual markers are also used as the initial step to receive audio based information at outdoor locations as part of a tourist's guide[26].

These symbols might take the image's place in the search query, and the searching phase of information retrieval is based upon pattern recognition or in case of direct number labelling on a straightforward one-to-one link. In the latter case an SMS might be used for transition of the request, instead of an MMS.

Figure 4.1 Examples of matrix codes[27]

Search based on visual markers is less time consuming than content based retrieval. However, the use of codes and visual markers presuppose a prior manifestation of the link between the physical and the digital world. This is not the case when the searching key is an image shot by the user. Services based upon CBIR systems have no prior connection between the physical source (here the tourist brochure) and the digital information source (here the M2S service and the MMSs that are produced to fit quires based on the brochure). In a fast changing world - both the physical and the digital – this feature is of great value. Thus the potential strength of services based upon CBIR systems is far beyond those of visual markers.

[26] *http://www.bbc.co.uk/coast/mobile/*
[27] *http://www.sintef.no/content/page1____3896.aspx*

A third and even less sophisticated strategy is to formulate textual key words for the search or pick up key words that are annotated in the tourist brochure. The corresponding searching procedure should be a simple text based image retrieval system. This implies a fast search, but at the other hand the query formulation will be more cumbersome for the user and no correction support in case of poor writing abilities or spelling problems in general.

Combining several information sources and searching strategies might be profitable. In a description of a patented system[28] the author says:

> *…Moreover, advances in image recognition allow a greater degree of automated recognition of objects, strings of letters, or symbols in digital images. This makes it possible to convert the graphical information into a symbolic format, for example, plain text, in order to then access information about the object shown.*

This suggestion implies that future image retrieval systems might benefit from combining several strategies, both content based and textual based search, and collect information from several sources in an incremental retrieval process or also to strengthen and refine the query before the retrieval process is initiated.

4.2 A context sensitive service

The concept may be further developed to include context sensitive mechanisms. If for instance the *date and time* of request is added, the quality or the fitness of the response MMS might be improved. One might anticipate that dynamic information, for example information of number of vacant rooms at hotels, special offers at nearby restaurants, ferry queues and so on, would be valuable information for people on the move. As indicated in the scenario in the introduction the user received a weather report together with detailed information of the next day's whale safari. Updated and enriched information services are most valuable in the setting described.

Figure 4*.2 Will future systems inform this photographer of the myths represented by these wall decorations?*

Next to the time, *location* is an obvious context source. A mobile phone service might benefit from the position information available from the network cell that the phone is linked to, or also from satellite based positioning systems.

[28] The system "Image base inquiry system for search engines for mobile telephones with integrated camera" was claimed by Hartmut Neven to the US patents authorities in February 2004 and approved in August 2005. His company Neven Vision was bought by Google in August 2006.

Considering the M2S service and the challenging image retrieval process, it is easy to see that added positioning information could be exploited as an information filter and thereby contribute to more efficient search processes and to a better hit rate. Beeharee and Steed (2005) have suggested that additional information of what is likely to be visible from the user's position – based on geometry of the real world, can be used to reduce the number of recommended sources. Such techniques could be used to gain valuable context information that increases hit rates.

The optimal situation occurs when a tourist takes a picture of any artefact or building or landscape and gets immediate response on his handheld terminal containing multimedia information that is relevant to the person's interests, the point of time and other significant conditions. These services presuppose considerable improvements of coming context aware image based retrieval systems, and the retrieval systems should be able to *merge several information sources* into efficient searching keys (filtering non-relevant information) or use stepwise retrieval processes that gradually refine the search for actual responses.

The CAIM project, mentioned in the introduction, addresses these aspects (Karlsen & Nordbotten 2006; Tøndersen 2006). The project puts forward a wide range of multiple contexts that could possibly help to identify the semantics of images and thus establish supportive mechanisms for CBIR systems. So far the following context variables are suggested: *spatial information* (e.g. location, orientation and speed), *temporal information* (e.g. time of the day, date and season of the year), *device information* (e.g. camera model, lens specifications), *environmental information* (e.g. temperature, air quality and light level) and *information of the photographer*. The M2S service, or similar services, will probably be further developed to test some of these parameters.

5 Similar and supporting initiatives

5.1 Similar services

During the project period similar systems have shown up. Researchers in England, for instance, have prototyped a mobile phone service that creates an interactive map from snapshots of a paper based map[29]. Recently search engine actors like Google have intensified their effort in image retrieval oriented projects. The following scenarios (and services) are launched:

> *… people can use a regular camera cell phone to take a picture of a movie billboard, and then send the image to a special database that returns a film trailer, locates a theatre showing the advertised movie, or let the person buy tickets to the film."* (SmartMobs 2006).
>
> *… the next time (within 10 years) you backpack through a foreign country, take a picture of that statue to find out how old it really is. Or snap a picture of that German road sign and translate it into English in order to finally understand what the Germans really mean by "Fahrt."* (MOPOCKET 2006).

These scenarios illustrate situations where pictures initiate a dialog or a request for more information, and a situation where the system and not the user is given the assignment to formulate an efficient request and collect information necessary to execute a successful information retrieval process. The burden is placed upon the system, which is the qualified party to do fast and extensive collection and integration of different information sources.

The commercial actor Mobot[30] has chosen the slogan *Point. Click. Connect* in their campaigns. They offer a wide range of services to connect consumers using camera phones to mobile content and commerce. For instance will pictures of posters or labels, CD covers, advertisements activate responses as ring tones, good jokes, where to buy and so on. Another example is the service MyClick[31] that enables a mobile phone user to shoot an image on a magazine and receive information from the advertiser responsible for the published picture.

> *… MyClick is a patent-pending system that allows a mobile user to take a specific picture of an image on a billboard, magazine or even on-screen image. An on-handset client then converts that image to data, before delivering it over the air to a network server, which recognises it as a specific request for data. It then pushes back to the handset whatever information or message the advertiser wishes to deliver.*
>
> *… The underlying airtime and data charges are paid for by the advertiser, with MyClick taking its revenue from the mobile operator in a revenue share agreement. The benefit for advertisers is that they can capture a user's identity automatically, and use the process to initiate any process, whether for sales, promotion or other message. They also get a clear*

[29] *http://www.newscientisttech.com/article/dn10416-phone-creates-interactive-maps-from-snapshots.html*
[30] *http://mobot.com/*
[31] *http://www.click.hk*

> *indication of which campaigns are working, and which aren't* (Mobile Europe 2006).

This quotation illustrates how these services enable new business models to occur, and further, how rather casually requests from users make the advertiser able to direct future advertisement more accurately. These ideas are followed up by Evjemo et al. (2006) where the prototype MMS-Dialog is launched. It is based upon the following scenario:

> *Mrs. Smith spots an advertisement for the service email on the mobile She does what the advertisement tells her to do: Shoot a picture and send an MMS to a short number. She receives an MMS with pictures and detailed information. The MMS states:"Your phone is capable of handling this service. Choose the main button to download and buy".*
>
> *Mr Carpenter spots the same advertisement, and sends an MMS as well. However, he has an older mobile phone. This is recognized by the service and he receives an MMS that informs about the device based limitations of the email service. He is offered renewed communication with the provider.*

Here the M2S service put into another setting and merged with the coming Mobile Device Management systems to identify the sender's device and check whether the device and installations cope with the announced product or service. This functionality will probably help to overcome the challenges related to missing standards of mobile device and ease the diffusion of mobile services.

It is also timely to mention the recent Telenor R&I initiatives (Canright & Engø-Monsen 2006):

- *Shoot'n'Buy.* Here the focus is on snapping a picture with a mobile device, and sending the *picture* in as a query for more information (possibly leading to 'buy'). This concept is inspired by ideas launched by Neven Vision (now part of Google).
- *DELIS:* EU IST project, running till Q4.2007; focus on peer-to-peer search engines.
- *SAPIR:* EU IST project, running from Q1.2007 – Q2.2009; focus on multimedia search (pictures, music, etc) from mobile devices.
- *Commercialization:* a spin-off company *T-Rank AS* has been in operation since May 2006. The aim is to take novel relevance technologies, developed in R&I and covered by several patents, to the level of finished products, which may then be sold, licensed, etc.

When everyday situations are considered there are lots of possibilities. Yeh et al. (2005) have made an example application where images of particular artefacts were input to the image retrieval process that finally pointed to relevant web sites. This service could be helpful to the person in a shopping situation:

> *Among several fancy articles a small purse catches the customer's interest. But will the colour fit the dress? She should like to see the alternatives. The mobile phone makes it easy to compose an MMS query for similar purses in the nearby located shops. The picture enclosed grants for a search beyond this particular brand. Immediately she receives a complete list of purses available within walking distance. A purse, not the same brand but – according to the enclosed picture –*

> *with a delicate dull ink, appears a better alternative. Even the price is beneficial. Smiling she turns to the shop across the street.*

So far this concept has been closely linked to images as the very link between the physical world and the digital world, and image retrieval as the corresponding searching mechanism. Doubtlessly audio based queries could be beneficial to certain situations:

> *Eva has found a seat on the buss and turns on her Mp3 player. A lovely piece of music is fading out. Oh – there it is! She has heard part of this tune several times, but still the singer and title is unknown to her. She makes a decision; this song should belong to her private music library. As the Mp3 player is part of her mobile phone the switch to MMS and construction of a query for the tune is easy. Before she reaches her destination she has received a full scale version of the song. Yes – it fulfilled her expectation. Perhaps a perfect present for her sister? She has all the information needed: The MMS reply contained beautiful pictures of the vocalist, the entire list of his productions and one touch access to further purchase. She enjoys his latest music video when leaving the bus.*

This scenario was inspired by Wang (2006) and a somewhat similar service named MobiQuid[32] that was piloted in connection to Telenor's MobilHandel project in 2001.

The ongoing terminal convergence between mobile phones, cameras, radio, and audio players and recorders provides easy and immediate access to non-textual queries. Combined with the communication capabilities of the mobile phone the user holds an efficient device to initiate non-textual requests. At the moment the image retrieval systems represent the most salient hurdles for such services to be realised.

5.2 Bottom up content production

In chapter 4.2 we discuss how context information can improve the image retrieval process. To avoid enormous work loads this kind of context information should be added automatically, both when the image is shot and when it is used. There are of course images "out there" already that has no context information added and there is context information that not easily can be added automatically. How to deal with this problem? It is actualised with the exponential growth of digital images and the many producers of images (everybody has a digital camera), and the use of image publishing and image sharing applications.

In other settings resource consuming tasks are left for machines to do. Annotation of images implies subjective assessment of the content of the image. To conduct such tasks the human being is doubtlessly superior to the machine. Thus, due to the resources needed annotation of images is done for specific purposes only.

Now the situation is about to change as Google has released an application that makes it fun to label (tag) images. Google Image Labeler[33] is a real-time collaborative application, where users work with an online partner, assigned by Google, to look at the same image and decide on some labels together. This is

[32] *http://www.dinside.no/php/art.php?id=33701, December 3rd 2002*
[33] *http://images.google.com/imagelabeler/*

voluntary work that will increase the amount of images with tags and geographical position. Also sites like Flickr[34] and YouTube[35] thrive on tagging. Contributors of uploaded images tag their material as well as they can, and simultaneously other members of the community are invited to add comments. It is a rough and ready form of classification, and it has attracted much interest. This kind of bottom up production of context information seems promising.

There are also geotagging initiatives where anyone might map their photos. See for instance the websites Panramio[36] and Tagzania[37].

5.3 RFID and connected objects

The falling cost of short range communications will ensure that we will be surrounded by small processors in all kinds of devices. It is not unreasonable to assume 100 million online devices in Norway in the future according to the Telenor Strategic Research project termed Connected Objects (Thorstensen et al. 2006). This also gives an opportunity for a wide range of new services providing object and information access. New cell phones with increased net access, processing power and local communications links will provide excellent platforms for services based on these connected objects in both business and consumer segments. One such application would be search for digital information based on RFID tags in nearby physical objects. In this case the RFID tags may play an essential role in the formulation of queries.

[34] *http://www.flickr.com/*
[35] *http://www.youtube.com/*
[36] *http://www.panoramio.com/*
[37] *http://www.tagzania.com/*

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 F. 2006. *MOVE – a summary of project activities and results (2004-6).* Fornebu, Telenor Research and Innovation (R&I R 17/2006.)

Beeharee, A, Steed, A. 2005. Filtering location-based information using visibility. *Location- and context-awareness: first international workshop*, vol 3479, 306-315.

Canright, G, Engø-Monsen, K. 2006. *Search hitting the mobile*. Fornebu, Telenor Research and Innovation, R&I Newsletter No 3-2006.

Derose, T D, Stollnitz, E J, Salesin, D H. 1996. *Wavelets for Computer Graphics*. Morgan Kaufmann.

Evjemo, B, Akselsen, S, Egeland, E. 2006. *MMS initierer kundedialogen.* Fornebu, Telenor Research & Innovation. (CSS project Short Note 4/2006.)

Evjemo, B, Akselsen, S, Bergvik, S, Schürmann, A, Ytterstad, P. 2005. Turister gjør det med mobilen! Utprøving av mobile tjenester for turister i Lofoten sommeren 2005. Fornebu, Telenor Research and Development. (R&I R 26/2005.)

Karlsen, R, Nordbotten, J. 2006. *CAIM. Context Aware Image Management. A project proposal.* Project site at http://caim.uib.no/

Leiper, N. 1990 Tourist attraction systems. *Annals of Tourism Research* 17, 367–84.

Mobile Europe. 2006. *http://www.mobileeurope.co.uk/news/news_story.ehtml?o=2375, July 27th 2006.*

MOPOCKET. 2006. URL: http://www.mopocket.com/2006/01/biometric_cellphones_hyperlink.php, 12. januar 2006.

Nyland, K J. 2006. *Bildesamlinger og kontekst. Masteroppgave i informatikk.* Institutt for Informatikk, Det matematisk-naturvitenskapelige fakultet, Universitetet i Tromsø.

Rahlff, O-W. 2005. CybStickers – Simple Shared Ubiquitous Annotations for All. UbiPhysics workshop. *UbiComp 2005*. Tokio, Sept.11-14. 2005.http://ap.hpl.hp.com/UbiPhysics/papers/CybStickerUbiComp2005paper_Short_Rahlff.pdf

SmartMobs. 2006. URL: http://www.smartmobs.com/archive/2006/08/16/google_takes_a_.html, 16. august 2006.

Thorstensen, B, Bjørnvold, T A, Ytterstad, P, Lopez J C C. 2006. *Connected Objects in the business market.* Fornebu, Telenor Research and Innovation (R&I R 06/2006.)

Tøndersen, A-H. 2006*. Image retrieval based on location and time.* Masteroppgave i informatikk. Institutt for Informatikk, Det matematisk-naturvitenskapelige fakultet, Universitetet i Tromsø.

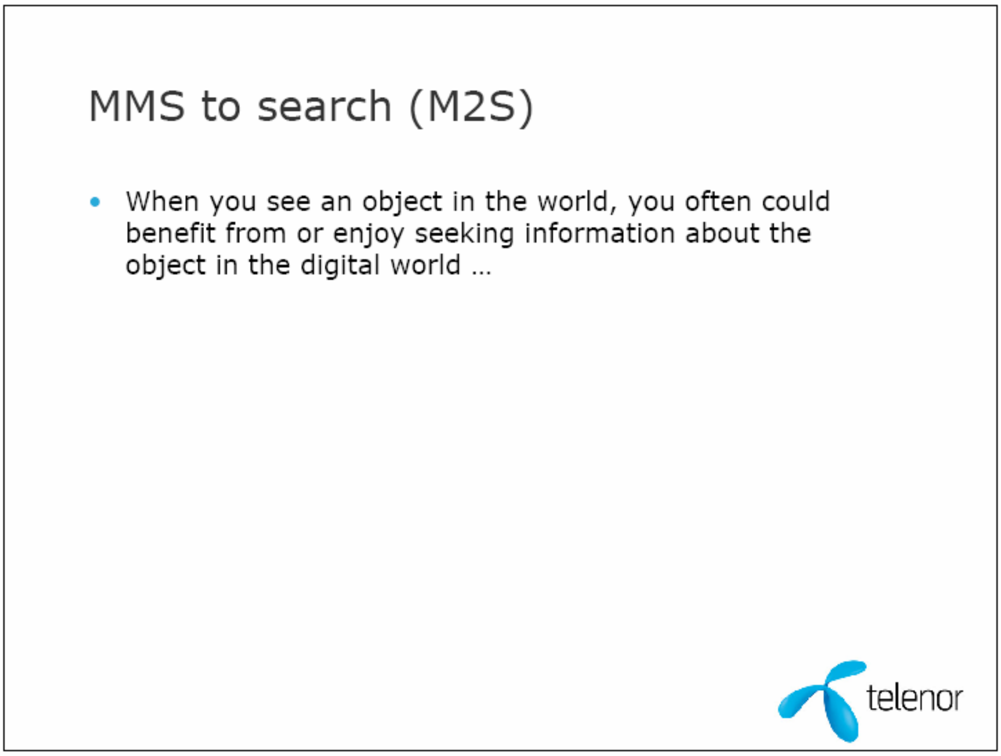
Viken, A, Akselsen, S, Evjemo, B, Hansen A A. *Lofotundersøkelsen 2004.* Fornebu, Telenor Research & Development. (R&D R 27/2004.)

Wang A. 2006. The Shazam music recognition service. *Communications of the ACM*, 49 (8).

Yeh, T, Grauman, K, Tollmar, K, Darrell, T. 2005. A picture is worth a thousand Keywords: image –based object search on a mobile platform. *Proceedings of CHI 2005*,

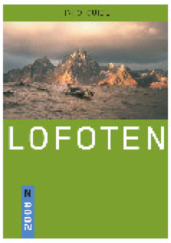
Zhao, R, Grosky, W. 2002, Bridging the semantic gap in image retrieval, Distributed Mulitmedia Databases: Techniques and Applications, T.K. Shih (ED.), Idea Group Publishing, Hershey, Pennsylvania.

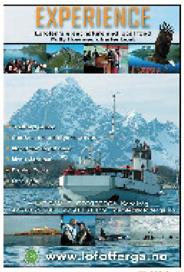
Appendix 1 Presentation of M2S



Scenario for M2S

- A tourist in Lofoten would like to find attractions and activities in the region.
- She comes across the printed Lofoten Info-Guide from Destination Lofoten.
- Inside the Info-Guide she discovers an ad for *Lofotferga*. A Trollfjord cruise seems interesting ...





telenor

Send image-MMS to get more information

- In order to get more information about Lofotferga, she takes a picture of the ad with her cameraphone ...
- ... and sends it as an MMS to a shortnumber.

telenor

Searching for similar images

- At the receiver a computer at Telenor searches a database for an image that matches the received image.
- If a match is found, a reply is sent.
- The reply can take different formats ...

telenor

The reply can be an MMS presentation ...

- The presentation consists of several sheets with images and text.
- It could also include music or a narrator voice.
- Inside the MMS one can also include links to WAP-sites with additional information

telenor


M/F GAMLE LOFOTFERGA
Gamle Lofotferga ble bygd i 1965 og er 25m lang og 8m bred.
I dag går Gamle Lofotferga som charterbåt fra mars til november.
Ombord har vi en variert meny, som tilberedes og serveres fra vår buffet på dekk.
Forfriskninger for en hver smak kan nytes til turen da Gamle Lofotferga har alle rettigheter.
telenor


Vi arrangerer turer for alle typer grupper.
Mange av våre oppdrag går til sagnomsuste Trollfjorden.
Fra oktober til desember kan vi tilby spekkhoggersafari...
...og hele året kan du se det vakre lyset i nord.
For se en video fra Lofotferga kan du åpne koblingen på denne siden.
Besøk også www.lofotferga.no for informasjon
telenor

Appendix 2 Content produced for M2S

Name	Description	Format
Galleri Harr	Presentation of the Gallery Lofotens Hus with paintings of Karl Erik Harr. The presentation also includes some information about Henningsvær, where the gallery is situated. The 10 slide presentation includes 8 images, 7 texts and WAP link to a page with more information.	MMS
Gamle Lofotferga	Presentation of the Old Lofoten Ferry in 10 slides, including 10 images, 8 texts and a WAP link. Shows images and tells about possible excursions with the boat.	MMS
Havaktiviteter	Presentation about different activities in relation to the sea but not connected to one specific attraction. The presentation consists of 11 slides including 12 images and 8 texts.	MMS
Lofotakvariet	Presentation of the Lofoten Aquarium, telling about activities at the aquarium and opening hours. The presentation constitute 7 slides, including 6 images, 4 texts, a WAP link to a site with more comprehensive information, and an audio track with background music.	MMS
Lofotkatedralen	Presentation of the Lofoten Cathedral on 5 slides with 4 images, 4 texts, a WAP link and a narrating audio track telling about the Sea Church – *Sjøkjærka*.	MMS
Skulpturlandskap Nordland	Presentation of the 5 art installations that constitute the Lofoten part of *Artscape Nordland*. The 8 slide presentation includes 6 images and 6 texts.	MMS
Lofotr	This 45 second long streamable video gives an example of a tour with Lofotr, a Viking ship replica, located at the Viking Museum at Borg.	Video (3gp)
Lofotakvariet	This 29 second long video shows the different sites and activities at the Lofoten Aquarium. The video is a shortened version of a video available at website of Lofotakvariet.	Video (3gp)
Lofotferga	This video (1 minute and 33 seconds) gives the viewer an impression of a trip with the old ferry, to the Trollfjord in Lofoten, can be like. The video is produced by Aurora Borealis Multimedia AS, but has been adapted by us to a format suitable for video streaming to mobile phones.	Video (3gp)

Appendix 3 Reviewed CBIR systems

NAME	Technical features
APORT	Downloaded demo, but was not able to test due to system instability/failure http://staff.science.uva.nl/~mliempt/downloads.php
BLOBWORLD	Research project form 1999. No published API. Article available. http://portal.acm.org/citation.cfm?id=714153&dl=GUIDE,ACM&coll=GUIDE&CFID=11111111&CFTOKEN=2222222
BolinOS	Open source project, but not focusing on CBIR. Not found suitable for our needs. http://www.bolinos.com/com/downloads/
CIRES	Research project, but no published API. Utilizes relevance feedback. Not impressive results with demo. http://amazon.ece.utexas.edu/~qasim/research.htm
COMPASS	Promising, but no published APIs. No commercial, might get access if we contact them. It can utilize relevance feedback. http://compass.itc.it/
generation5	No published API http://generation5.org/content/2004/aiSomPic.asp?Print=1
GIFT	The GNU Image-Finding Tool. Open source. Promising, but need more investigation. http://www.gnu.org/software/gift/
Idée Espion	Commercial product used in, among other, Adobe Photoshop Elements 4.0. No testing facilities available. Provides an SDK for C++. Can run as a service or can be embedded in other systems. http://www.ideeinc.com/espion-sdk.php
Image-Seeker	Promising demo, but no published APIs. Demo with search in Corbis royalty free images. http://corbis.ltutech.com/
IMatch 3	Commercial product. API unknown, but unlikely. 30 day trial available, but not tested. http://www.photools.com/
imgSeek	Promising. Open source based on Python, but comprehensive. Interaction possible through command line interaction. http://imgseek.net/
In-Two – (in]2	Commercial product. No demo available http://www.in-two.com/
ISTORAMA	CBIR for Internet. Relies on user feedback during retrieval process. No API. http://uranus.ee.auth.gr/Istorama/
LIRE	Open source with published API. Successful search when images are similar sized and when target image has distinct contours. Based on Caliph and Emir. http://www.semanticmetadata.net/lire/
MascoT	Research project, focusing on object recognition in images. Promising, but no published API. Java based. http://seco.asa.cs.uni-frankfurt.de/Seco/mascot/mascot_eng.html
MUVIS	Video retrieval system. Not feasible for our needs. http://www.iva.cs.tut.fi/COST211/MUVIS/muvis.htm

OCTAGON	Promising, but not open source. No API available, but the developer of the system is considering this.
	http://octagon.viitala.eu/
IIS-0097329	"Object and Concept Recognition for Content-Based Image Retrieval", Research project by University of Washington. Several demos, but no published API. Contact could be considered.
	http://www.cs.washington.edu/research/imagedatabase/
PictureFinder	Research project by Technologie-Zentrum Informatik (TZi) in Germany. Most information in German. Demo, but no published API. CBIR functionality might be provided by In-Two
	http://www.tzi.de/bv/projects/picturefinder
QBIC	IBM's Query By Image Content. Not open source. No testing facilities available.
	http://wwwqbic.almaden.ibm.com/
riya	Promising, but limited functionality in demo so far. No published API. Web based.
	http://www.riya.com/
SCHEMA	Joint venture for CBIR. Developed a reference system, but relies on user feedback. Not promising tests with web demo.
	http://www.schema-ist.org/SCHEMA/
SIMBA	Web-demo only
Viper	Relevance feedback from user necessary. Not applicable for our need. Based on GIFT and MRML.
	http://viper.csse.monash.edu.au/~davids/cgi-bin/PerlMRMLClient/PerlMRMLClient.pl
VIRAGE	Commercial product, aims at video, audio and rich media retrieval. Probably not well suited for our needs.
	http://www.virage.com/content/products/
VisualSEEk	Research project with collection of demos. No published APIs found so far. Might consider further investigation.
	http://www.ee.columbia.edu/ln/dvmm/researchProjects/MultimediaIndexing/VisualSEEk/VisualSEEk.htm
WALRUS	Research project. No published API or demo. Article available.
	http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1262183