

For ICS/LIS seminar, Tues., 19.1.2010

# Multi-modal Information Retrieval experiences from Context-Aware Image Management

Joan C. Nordbotten

Jan. 19, 2010

## Abstract

Today's tourists commonly use mobile phones, most with cameras and some with GPS functionality. As a tourist, they are presumably new to the area and unfamiliar with local place and historic names that could be used to search for background information about the landmark sites they encounter. The [CAIM: Context-Aware Image Management](#) project seeks to develop an Information Retrieval system that is capable of retrieving multi-modal (text, audio and image) information about the object(s) in an image sent from a mobile-phone camera enhanced with gps coordinate information.

This talk presents experiences from the CAIM project and the various tools required to support multi-modal information retrieval (MMIR). Since MMIR problems are not solved, topics for future research and development will be highlighted.

## Presentation

### Outline:

#### *Multi-modal Information Retrieval - MMIR*

1. *Modalities of information & data*
2. *Mobile Multi-modal information retrieval*
3. *Image context*
4. *The Information Retrieval (IR) process*

#### *The CAIM project framework*

1. *The technology*
2. *MMIR from a Nokia 5800 phone*
3. *MMIR from a PC using [VISI3](#) demonstration*

#### *Status of Multi-modal IR in CAIM*

1. *Evaluations*
2. *Future research & Development*

**Multi-modal information – MMI****Definitions**

Multi-modal: having or involving several modes, modalities

Mode: 1) a particular form or variety of something  
2) a manner of doing something (Webster, 1974,2003)

Information: "The meaning that a *human* extracts from data by means of known conventions of the *representation* used." (IFIP. Gould, 1971)

Multi-modal Data: The *representation* of information using multiple modes (formats)

**Notes:**

1. Representation is both a noun and a verb, i.e. Satisfying the 2 definitions given of "mode".
2. MM information is more correctly MM data

**Data Modalities****examples:**

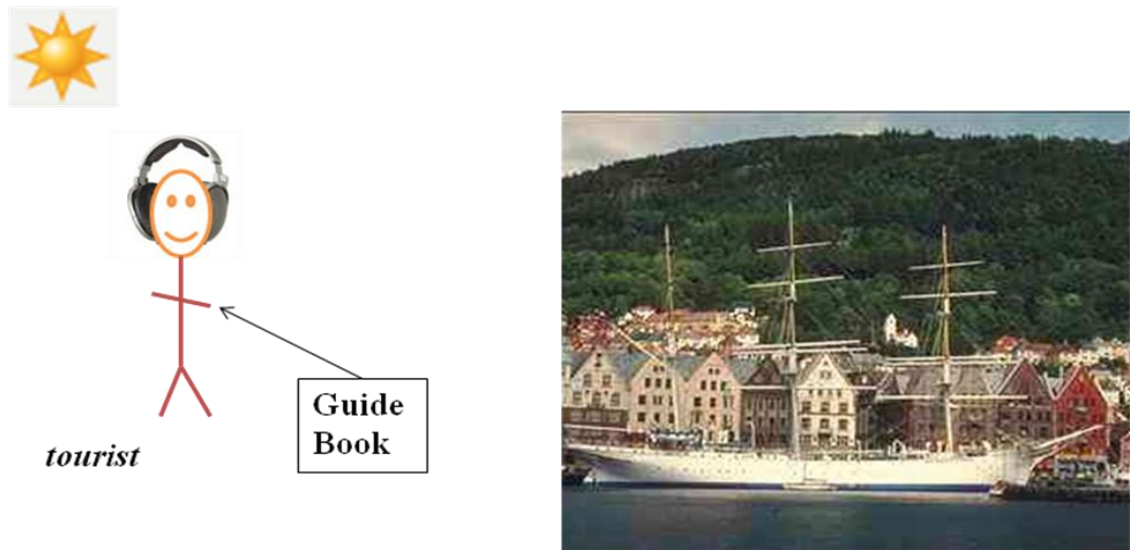
- Numeric count, size, age, ...
- Text document, title, keyword, ...
- Image painting, drawing, photo, sketch, ...
- Audio speech, music, signal, ...
- Spatialpoint (gps), line, area, direction, ...
- Temporal date, time, duration (season, period, ...), ...
- Sensory? Touch, taste, smell

**Multi-modalities:**

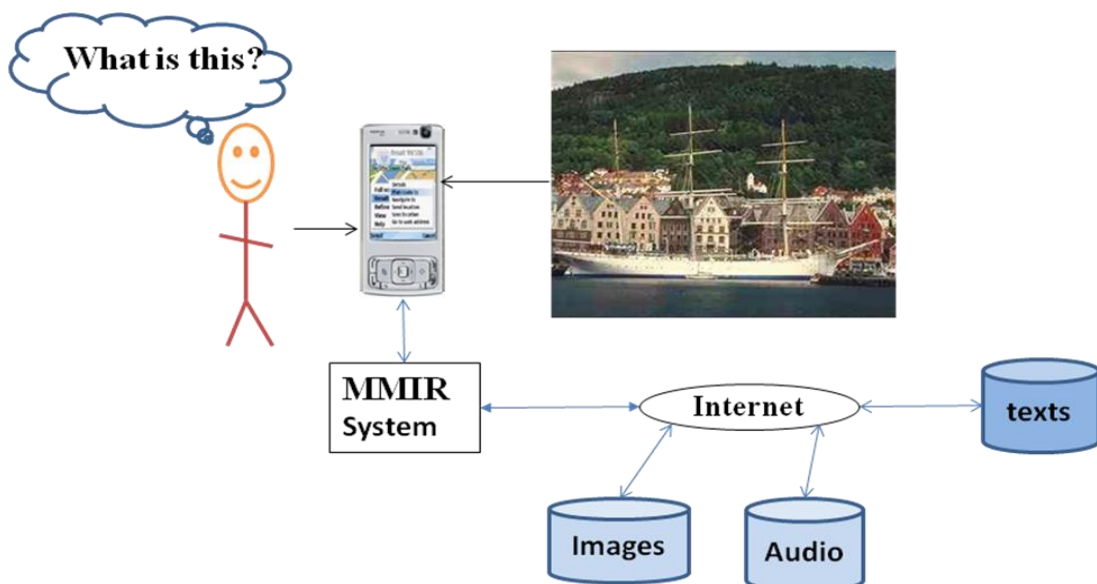
- Static illustrated text: newspaper, web-page, map, chart, table, ...
- Dynamic video, film, TV

**Notes:**

1. Numeric and text/character data are, with binary data, the oldest digital data types,
2. Information (text) retrieval dates from 1957.
3. Geographic Information Systems date from 1962
4. Temporal data is central to Statistical Information Systems
5. Video, film, TV are serial modalities, while web-sites can have a star structure and use.

**Multi-modal information examples**

Multi-modal Information: image, text, audio, senses,  
 The **Statsraad Lehmkuhl** is a three-masted barque rigged sail training vessel  
 The fastest sail training vessel in the world. (by nature of winning many races)

**Mobile MMI Retrieval****Multi-modal Information (text, audio) retrieval from an image + gps query**

1. User: tourist without knowledge of local names
2. Mobile phone with camera + gps functionality
3. Image databases linked to text & audio

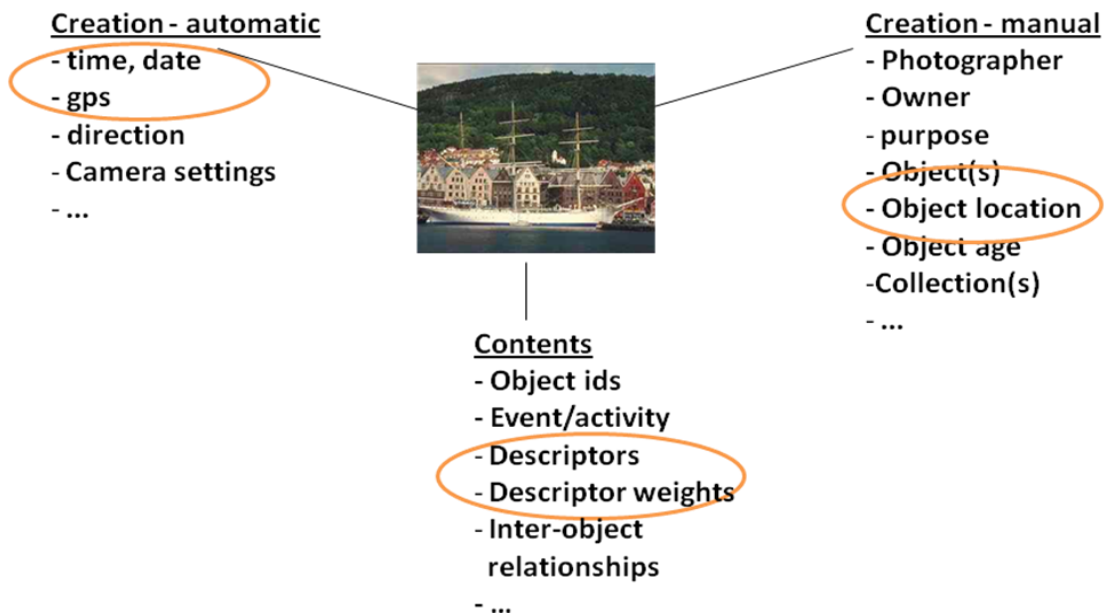
## Image context

1. **"Context** any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, ...". (Dey & Abowd, 2000).
2. **"Context** the information that must be common knowledge between the user and system for communication between them to succeed." (Elgesem & Nordbotten, 2007).

**Note:** Information retrieval is dependent on 'correct' communication

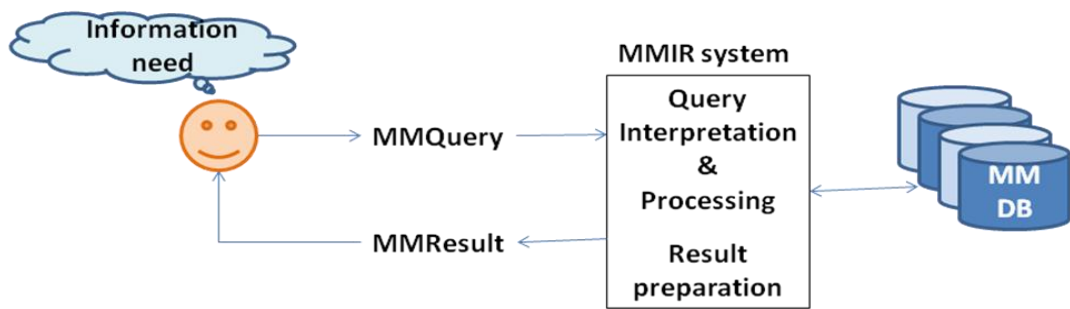
---

## Image context examples



Circled context types are used in the CAIM project as an aid to image interpretation.

**CAIM MMIR – query processing**

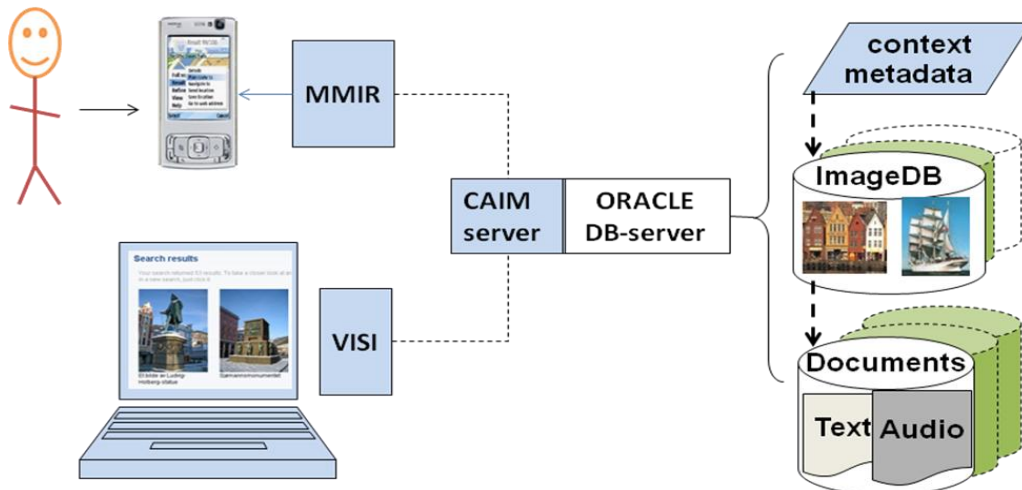


MM-Query: <image, gps-camera, time, date>  
 MM-DB: <image, gps-obj(s), text, audio, <context data>>  
 MM-result: <image-set, text, audio, ...>

**NB: there is no text in the query**

- Since CAIM-MMIR is (should be considered as) an internet-based system,
- the DB should be considered distributed
- multi-modality occurs in each phase: information need, query, DB and result

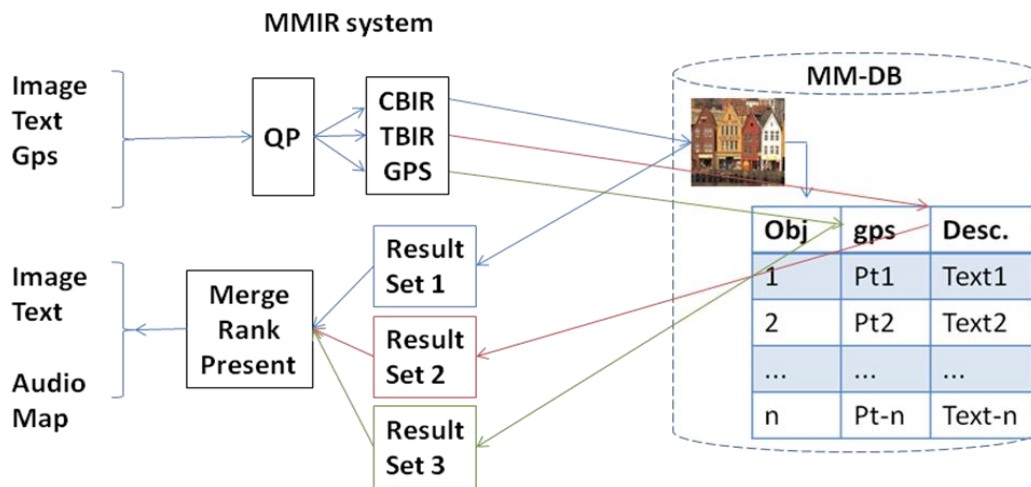
**CAIM prototypes**



**Components:**

1. Relevant information on the Internet, but in separate systems
2. Need to link images to related text/audio/video information
3. Need for a mobile information retrieval system
4. Need for a back-end DB server and query processing system
5. Need for a development environment

## MMIR processing



### Notes:

1. CBIR, TBIR and GPS-based search use different
  - Search algorithms & index forms
    - Text-based: actual term roots + synonyms
    - CBIR: color, texture, shape
  - Similarity algorithms
  - Scoring schemes ex.
    - Text-based similarity scores: 0-100 => no to full match:
    - CBIR similarity scores 0-100 => full match to no match, p.g.a. *distance* between signatures
    - Gps-area location
2. Getting 'good' Results through the Merge process is very difficult

### MMIR quality

Quality  $\approx$  Precision as defined by human users

Results – measured by precision scores vary by search technique

### Ex:Image Retrieval w/Oracle

	<i>Precision %</i>	<i>comments</i>
• TBIR:	70+	image descriptors
• CBIR:	<20	based on color match ☹
• CBIR+TBIR	70	assuming input tags
• CBIR+GPS	50+	lack of direction

### Notes:

1. Small studies in # participants, DB size and query set
2. Relevance is very subjective
3. Recall is nearly impossible to measure, p.g.a. #2 above and DB size
4. 2 precision measures used: every 'n'th retrieved and for each relevant image

## Future research & Development



- Query specification:
  - object identification
  - sketch
- Object recognition:
  - Separate multi objects within an image
  - Expand use of image context
  - Record object location in image
- Improve MM merge algorithms
- ...

---

**Project reports:** The CAIM project, <http://caim.uib.no>

- Møller, T. (Sept. 2009). [Bergen By – a Multi-Modal Image Database](#). CAIM-TR-9, Dept. of Information and Media Sciences, Univ. of Bergen.
- Carlson, Christoph. (Sept. 2009). [VISI3 – Context Aware Image Retrieval](#). CAIM-TR 8. Dept. of Information and Media Sciences, Univ. of Bergen.
- Hellevang, Mathias (Sept. 2009). [MMIR3 – Mobile Multimedia Image Retrieval](#). CAIM-TR 7. Dept. of Information and Media Sciences, Univ. of Bergen