



2006-01-01

# Using Sketches and Knowledge Bases for Geo-Spatial Image Retrieval

Michela Bertolotto  
*University College Dublin*

James Carswell  
*Dublin Institute of Technology, jcarswell@dit.ie*

E McLoughlin  
*University College Dublin*

D O'Sullivan  
*University College Dublin*

C Wilson  
*University of North Carolina at Charlotte*

Follow this and additional works at: <http://arrow.dit.ie/dmcart>

 Part of the [Computer Sciences Commons](#)

## Recommended Citation

Bertolotto, M et al. (2006) Using Sketches and Knowledge Bases for Geo-Spatial Image Retrieval. *Elsevier Computers, Environment and Urban Systems* 30, 29–53pp. doi:10.1145/945645.945659

This Article is brought to you for free and open access by the Digital Media Centre at ARROW@DIT. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@DIT. For more information, please contact [yvonne.desmond@dit.ie](mailto:yvonne.desmond@dit.ie), [arrow.admin@dit.ie](mailto:arrow.admin@dit.ie).



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](#)





## Using sketches and knowledge bases for geo-spatial image retrieval

M. Bertolotto <sup>a,\*</sup>, J.D. Carswell <sup>b,1</sup>, E. McLoughlin <sup>a</sup>,  
D. O'Sullivan <sup>a</sup>, D. Wilson <sup>c,2</sup>

<sup>a</sup> Department of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland

<sup>b</sup> Digital Media Centre, Dublin Institute of Technology, Dublin 2, Ireland

<sup>c</sup> Department of Software and Information Systems, University of North Carolina at Charlotte,  
Charlotte, NC 28223, USA

Received 30 March 2003; accepted in revised form 17 March 2004

---

### Abstract

This paper presents research in the field of knowledge management for geo-spatial imagery including scanned aerial photos and satellite images. We have developed a web-based system that allows users to query a database of images not only using metadata, but also drawing sketches of configurations of objects they are interested in as well as inputting textual descriptions of their intended task. Our system integrates case-based reasoning techniques to form a knowledge base from previously issued queries that can be exploited to improve future query processing and to build organizational memory through experience capture. The effective design and implementation of a user-friendly graphic user interface plays an important role for the system to provide improved human–computer interaction and decision support.

© 2004 Elsevier Ltd. All rights reserved.

*Keywords:* Geo-spatial imagery; Knowledge base; Sketch-based queries

---

\* Corresponding author. Tel.: +353 1 716 2913; fax: +353 1 269 7262.

*E-mail addresses:* [michela.bertolotto@ucd.ie](mailto:michela.bertolotto@ucd.ie) (M. Bertolotto), [jcarswell@dit.ie](mailto:jcarswell@dit.ie) (J.D. Carswell), [eoin.a.mcloughlin@ucd.ie](mailto:eoin.a.mcloughlin@ucd.ie) (E. McLoughlin), [dymphna.osullivan@ucd.ie](mailto:dymphna.osullivan@ucd.ie) (D. O'Sullivan), [davils@uncc.edu](mailto:davils@uncc.edu) (D. Wilson).

<sup>1</sup> Tel.: +353 1 4023264.

<sup>2</sup> Tel.: +1 704 687 5498.

## 1. Introduction

The purpose of our research is the development of a knowledge management system called MaGIK (Managing Geo-Spatial Imagery and Knowledge) for the retrieval of geo-spatial imagery matching user-sketched configurations of objects, metadata and task descriptions. The functionality that users require from such a system is twofold:

- Be able to retrieve the images and geo-spatial data they need, and
- Be able to use, augment and manipulate the retrieved images and data to highlight task-relevant information, as well as to capture the results of their interactions in a knowledge base for future support.

For example, a company that uses geo-spatial data for urban planning projects may employ such a system to assist in selecting the location for new civil developments. From a task-based standpoint, the most relevant work product lies not merely in the applicable visual data, but in descriptions of why and how the information has been collected and to what ends it has been successfully (or unsuccessfully) employed. A clear advantage is provided by capturing and leveraging essential underlying information. It is also advantageous to measure and record the human expertise involved in applying such information as part of the organisational task.

The MaGIK system that we are currently developing extends and augments our previous work on sketch-based image retrieval (Agouris, Bertolotto, Carswell, & Stefanidis, 2000, 2002; Carswell, 2000) by providing knowledge management support as well as a user-friendly graphical user interface for improved human–computer interaction and decision support. Classical retrieval systems would require the user to formulate his/her query in a given database language, possibly with the use of a graphical interface where he/she can type in attribute characteristics of the objects. However, these systems would not allow the user to formulate the query in a visual way (e.g., by means of a sketch) and they would not take the shape of objects into consideration. A sketch is a very effective and intuitive way of expressing what the user has in mind. Additionally, sketch-based queries are important in applications where shape is an essential characteristic and they facilitate human–computer interaction by allowing the users to visualise the spatial configuration they want to retrieve. An example of a sketch-based query is shown in Fig. 1 below where the user is interested in retrieving images containing a road intersection, groups of houses and a stadium.

The sketch-based query facility is an innovative aspect of our system and provides users with an option of formulating their queries by drawing sketches (e.g., with the aid of a pen-based input device) of the spatial configuration of objects they intend to retrieve.

Our current system, does not rely solely on the sketch-based image matching algorithms developed in (Carswell, 2000), it may be integrated with any sketch-based retrieval system or algorithms.

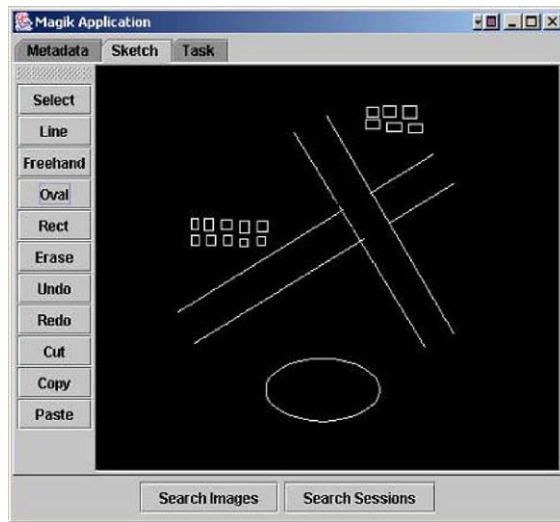


Fig. 1. Query by sketch.

Therefore besides a sketch of object configurations additional metadata regarding the geographic area (e.g., Boston, Massachusetts, USA) and scale (e.g., 2 m per pixel) associated with the query, and task description outlining the specific purpose of the search can also be input to the MaGIK system by the user. Fig. 2 illustrates an example of a task description a user may enter with regard to the development of a shopping center.

The imagery retrieved by the system in response to such a query consists of a ranked list of images that match the users requirements according to a computed

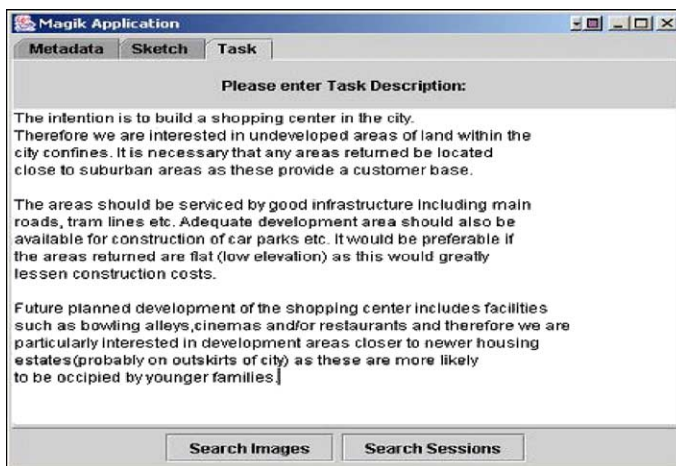


Fig. 2. Query by task description.

similarity metric. Once the relevant information has been retrieved, the user can highlight and elucidate the particular aspects of the imagery that address overarching task-based goals. For example, if a new shopping center is being planned, the user may annotate a retrieved image by highlighting areas of undeveloped land adjacent to areas of housing and areas of good infrastructure, in order to justify a proposed new location. The queries, annotations, and rationale can then be stored as an encapsulated parcel of knowledge in the context of the civil planning task. This parcel can be stored in a knowledge base, thus growing corporate knowledge assets, and can therefore be re-used to support future tasks. For example, if there are records of how the earlier shopping center locations were chosen, those earlier experiences could be used to inform the user of the query process involved in these earlier tasks. This can provide an insight into the rationale behind earlier choices that may apply to the current task. Fig. 3 schematically shows the addition of the knowledge base for experience capture and re-use to a sketch-based image retrieval system.

This paper provides an overview of the system as a whole and then focuses on the latest developments on its knowledge management and task-based querying capabilities.

The remainder of the paper is organised as follows. Section 2 describes related work. Section 3 describes the system architecture design. Section 4 is dedicated to the image retrieval component of our system and the metrics employed in such retrieval. Section 5 discusses how the knowledge management techniques have been integrated into the system to allow for more intelligent image retrieval. Section 6 focuses on how we have implemented the outlined system. Section 7 describes some initial system evaluation and discusses the results of this evaluation. Finally, Section 8 outlines some conclusions and future work.

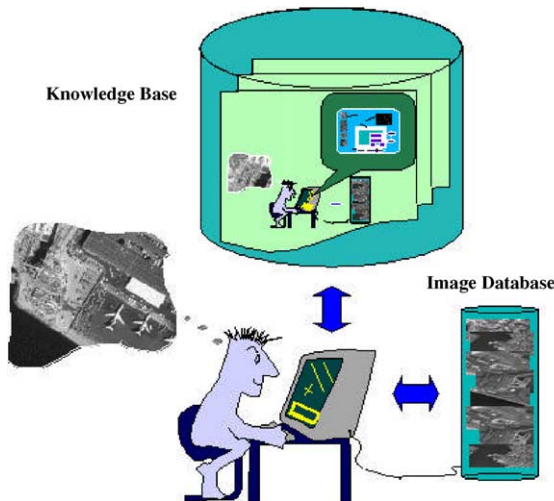


Fig. 3. System overview.

## 2. Related work

The scope of our research ranges from geo-spatial information handling and retrieval to human–computer interaction to knowledge management. Within these areas we focus on specific topics whose related work is described in the following.

### 2.1. Image retrieval by content

Extensive research is being conducted in the field of geographic information systems (GIS) to find efficient methods for storing and retrieving geo-spatial information. Two different formats are used for the representation of geo-spatial data: raster and vector. We concentrate on data in raster format including scanned aerial photos, satellite images, etc. However, additional information can be associated with such data in the form of textual descriptions (metadata) and object attribute information (semantic meaning of the objects contained in a given image).

Since the advent of digital scanners and sensors beginning in the late 1970s, (Chang & Reuss, 1978; Zloof, 1975) image database querying has become a major area of research. Most of the efforts during this time have focused on analysing and comparing the lower-frequency properties of digital imagery. These include: colour, in the form of histogram matching; texture, in the form of image coarseness and contrast matching and composition, where an image is divided into homogeneous regions of colour or texture and the relative positions of these regions analysed (Carson, Belongie, Greenspan, & Malik, 1997; Niblack et al., 1997; Forsyth et al., 1996; Frankel, Swain, & Athitsos, 1996; Gupta, Weymouth, & Jain, 1991; Ogle & Stonebraker, 1995; Pentland, Picard, & Sclaroff, 1996; Sclaroff, Taycher, & La Cascia, 1997).

The expression “image retrieval by content” (Gudivada & Raghavan, 1995) in this paper refers to retrieving images matching to the higher-frequency image characteristics, more specifically, actual shape information (synonymous with outline or edges) of features contained within the imagery, e.g. the outline of a building or buildings. Furthermore, we are interested in doing so completely in the raster/spatial domain.

The majority of work in the area of feature matching has shown some success through matching the image-objects in the vector domain (Blaser, 1998; Chang, 1997; Cohen & Guibas, 1996; Jagadish, 1991; Mehrotra & Gray, 1993). This requires that the raster imagery be converted into scenes of vector objects, often together with attributes and other semantics such as topology before they can be queried. As the process of converting raster imagery into vector scenes of objects is not yet fully automated, this task remains tedious and should not be considered as given.

From a theoretical point of view, to obtain an “optimal” performance in an image information environment, some operations are better or easier done in the raster domain while others are more efficient in the vector domain. Defining spatial relations, for example, the topological (disjoint, touching, overlapping, etc.), directional (above, below, north, south, etc.) and metric (distance) relationships between objects is easier in the vector domain, where properties of the individual objects are known beforehand, (Blaser, 1998; Chang, 1997; Cohen & Guibas, 1996; Jagadish, 1991;

Mehrotra & Gray, 1993), while feature matching is usually done in the raster domain (Agouris & Schenk, 1996; Gonzalez & Woods, 1992; Mehrotra & Gray, 1995). It is well known that the integration of the two domains is still an open problem.

Some cases of image retrieval systems automatically extract keywords (metadata) about the imagery through the analysis of the location (URL) where the image is found on the WWW or from the text in which the image is embedded. Other semantic and/or metadata information in the form of general image colour, texture, dimension, shape, file type, size, and date can also be extracted automatically and indeed are used by some image retrieval systems, e.g. Chabot (Ogle & Stonebraker, 1995), Candid (Kelly, Cannon, & Hush, 1995), Cypress (Carson et al., 1997; Forsyth et al., 1996), ImageRover (Sclaroff et al., 1997), Jacob (Ardizzone & La Cascia, 1997), WebSeer (Frankel et al., 1996), VisualSeek (Smith & Chang, 1996), FIBSSR (Mehrotra & Gray, 1995), Nishida (Nishida, 1999), QBIC (Niblack et al., 1997), PICTION (Srihari, 1995), PhotoBook (Pentland et al., 1996), Virage (Hampapur et al., 1997), Lyco Media Search (<http://www.lycos.com/picturethis/>), Yahoo Image Surfer (<http://ipix.yahoo.com/>) and others (Athitsos, Swain, & Frankel, 1997; Kauppinen, Seppänen, & Pietikainen, 1995; Persoon & Fu, 1977; Smith & Chang, 1997). This definition of image “content” however is not the same as that used in this project. For example, none of the above mentioned systems take into account the actual shape of the objects contained within the imagery. This is understandable of course since the process for generating (extracting) the objects (features) from a raw raster image is not yet straight forward, i.e. not yet fully automated. Table 1 presents a direct comparison of the defining characteristics between the major visual information management systems (VIMS) described in the current literature. The last entry (MaGIK) in Table 1 refers to the image retrieval approach utilized by this research.

First attempts at improved information retrieval employed metadata information to filter the datasets. This approach requires that the metadata associated with available datasets be sufficiently descriptive of the content of those datasets. It has been shown that the use of metadata alone is no longer adequate in terms of intelligent information retrieval.

Indeed, advances in sensor/scanner technology have resulted in the availability of constantly increasing volumes of more and more complex geo-spatial datasets. The increased volume and complexity of data necessitates the development of novel methods to efficiently retrieve information from these large geo-spatial databases. To be effective, these methods have to take into account the actual shape of objects contained within the imagery. The system we are implementing overcomes the limitation of current systems by enabling queries based on the shape of objects sketched directly by the user.

## *2.2. Knowledge management*

As part of an overall effort in intelligent geo-spatial information systems, we are developing case-based knowledge management support for libraries of geo-spatial imagery. Our work exploits existing work in case-based reasoning (Kolodner, 1993; Leake, 1996) for knowledge re-use. Previous case-based reasoning work has

Table 1  
Comparison between “content-based” VIMS

	Color, text and/ or texture queries	Vector shape queries	Raster shape queries	Manual/ semi-auto image preprocessing	Automatic image pre-processing	Aerial/ satellite imagery	Multimedia type imagery
Chabot	X			X			X
Candid	X				X		X
VisualSeek	X				X		X
Cypress	X				X		X
Jacob	X				X		X
ImageRover	X				X		X
Yahoo Image Surfer	X				X		X
Lycos Media Search	X				X		X
WebSeer	X				X		X
WebSeek	X				X		X
FIBSSR		X			X		X
Nishida		X		X			X
Fourier Descriptors			X		X		X
QBIC	X	X		X	X		X
PICTION	X	X			X		X
PhotoBook	X	X		X	X		X
Virage	X	X		X	X		X
MaGIK	X		X		X	X	X



addressed knowledge management (Aarts, 1998; Becerra-Fernandez & Aha, 1999; Klahr, 1997; Leake & Wilson, 2001), textual case-based reasoning (Lenz & Ashley, 1998; Wilson & Bradshaw, 2000), and geo-spatial and image retrieval (Grimnes & Aamodt, 1996; Gross, Zimring, & Do, 1994; Holt & Benwell, 1999; Yeh & Shi, 1999). The challenges consist of integrating and tailoring these methods to address specific needs for geo-spatial image information management, as well as to develop hybrid similarity measures that seamlessly integrate very different types of contextual knowledge afforded by query sketches, result images and metadata, image annotations, textual rationale annotations, and multimedia annotations.

The method of annotating multimedia is related to annotating for the semantic web (e.g., (Champin, Prié, & Mille, 2001)) and multimedia indexing (e.g., Perry & Lewis, 1998; Worring et al., 2002), which focus on developing and leveraging annotated descriptions of the media content. Here, we focus rather on a task-centric view of annotation, providing for and employing annotations about how an image relates to the task at hand, though this will necessarily involve some reference to image content which may later be used to refine indexing. Multimedia database approaches such as QBIC (Niblack et al., 1997) provide for image segmentation and annotation, but also focus on contextualizing individual images, rather than task experiences.

Previous work in CBR has made use of multimedia cases (e.g., Barber et al., 1992; Burke & Kass, 1996), but the case media is not employed directly for retrieval; rather, the case indices are crafted semantic representations of the media content, an overhead we seek to avoid. While we make use of the information as part of task similarity rather than task structure, the system is instrumented to collect information about user interactions in terms of browsing and usage in a manner similar to the usage model in Egyed-Zsigmond, Mille, and Prié (2003).

### 3. System architecture

In this section we describe the architectural components of our system. Such a system relies on a three-tier architecture comprising three main layers, namely the *Client Layer*, the *Application Server Layer*, and the *Image/Knowledge Base Layer* as outlined in Fig. 4.

All communications between the client layer and the image/knowledge base layer are conducted through the application server layer. The application is executed on the client using an applet that runs in a standard web browser's Java Virtual Machine (JVM). The applet communicates with the application server using the existing HTTP networking protocol.

The application server layer contains all executable application programs, including: the image matching and processing algorithms, the database updating and maintenance routines, as well as the knowledge retrieval and management programs.

The image/knowledge base layer contains all image and query data and other information manipulated by the system.

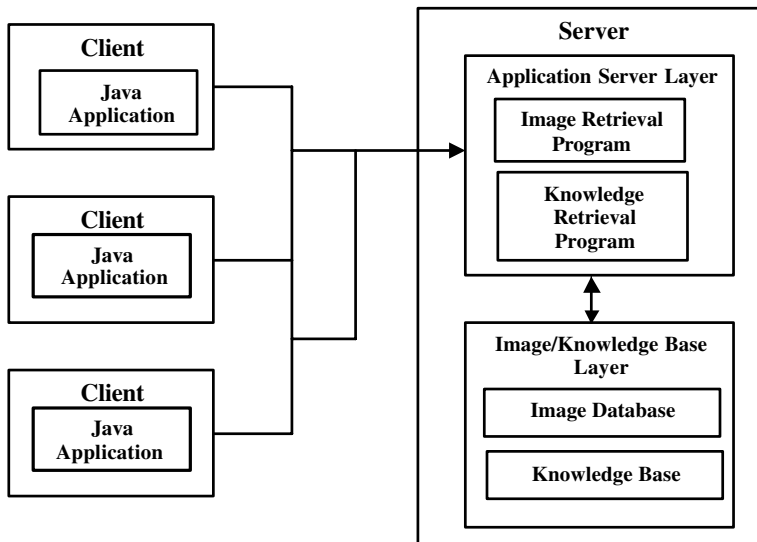


Fig. 4. System architecture.

The functionality of the overall system comprises two main components: image retrieval and knowledge management. In the following sections we describe each of these components separately including the data/information they handle and the functionality they provide.

#### 4. Image retrieval

The image retrieval component of our system includes a comprehensive image database organised into linked digital libraries of raster images, the objects they contain (feature library) and associated textual information (metadata library and associated annotations). Based on these image characteristics, our system provides a baseline functionality for image querying. A typical task-based query to our image repository is a straightforward request to a geo-spatial image database, and it could consist of specified metadata, semantic information, and a sketched configuration of image-objects.

For example, if the user decided to retrieve all images related to their shopping center query, the query process is as follows:

- Process the metadata library and retrieve all images that match to the specified metadata criterion (e.g. scale, location, etc.).
- Follow the links from this subset of images to the semantic library of annotations where it would identify which of these images contained areas of previously undeveloped land close to areas of housing/high population and good infrastructure.

- Use the semantic input criteria to further narrow down the list of images to search within for the required query shapes outlined in the user sketch.
- Use this further subset of imagery to select, through feature linking, a refined subset of features within the feature library to match individual objects in the query sketch against.
- Retrieve all the imagery linked to the best matched features in this refined subset from the feature library.
- Determine which of these images were returned more than once, i.e. which images contain all of the objects in the query configuration.
- Process the spatial relations of the query scene on this final subset of imagery and return a prioritised list of imagery as the query result.

When a feature gets matched to an image, its centroid coordinates within the image are recorded as well as the top left and bottom right coordinates of the query feature's minimum bounding rectangle (MBR), after scaling and rotation have taken place. Spatial relations on the image are determined through the use of the matched query features MBRs instead of the actual image-objects because the image is in raster (non-vectorised) format and therefore no a-priori information is known about any image-object (in particular, their boundaries). Also, it is straightforward to determine the MBR containing the pixels composing the translated/rotated/scaled query object and using their MBRs allows for spatial relations to be built and queried in real-time.

An important component of the comprehensive digital image database is the feature library that contains a set of distinct features (i.e. image-object shapes) and links to relevant images where such features appear. The role of the feature library is to allow for efficient querying through the optimal organisation of image-object data, in the form of previously sketched queries, and to provide the crucial link between this abridged group of raster features and a library of images. Feature-image linking allows us to avoid matching against the actual images, which can be very time consuming even for a small library of imagery. Thus real-time image querying and retrieval is made possible.

The organisation of the feature library into a tree-like structure (e.g. [Agouris, Carswell, & Stefanidis, 1999](#)) enables it to act like a multi-stage screening mechanism that minimises the risk of wasting considerable time making passes over extensive data that have no chance of selection. For example, the first screening criterion will eliminate as potential matching candidates most of the features within the library. A secondary screening criterion will eliminate the next greatest number of alternatives and so on down through the feature library tree hierarchy.

The feature library is linked many-to-many with the image library. That is, one image could be linked to more than one feature within the feature library and one feature could be linked to more than one image in the image library. Due to the dynamic natures of the image library and query building, the feature library is constantly adding, subtracting and otherwise updating its features, links, and internal organisation. It also therefore needs to be autonomous in that it automatically

maintains its own contents depending on the changing states of these external but integrated components.

## **5. Knowledge management**

Our storage, indexing, and retrieval tools provide a highly effective means for exploiting geo-spatial image information, and they play a natural role in supporting the overarching task-based needs of organisations that rely on such information. Organisations are increasingly concerned about the capture and management of knowledge assets.

Managing the knowledge implicit in using geo-spatial imagery to address particular tasks is crucial for capturing and making the most effective use of organisational knowledge assets. This serves both to facilitate workflow by providing employee access to bestpractice examples, as well as to grow a repository of task-based experience as a resource for support, training, and minimizing organizational knowledge loss as a result of workforce turnover. Our approach addresses task-based knowledge management by providing:

1. A flexible environment to support analysis and elucidation of relevant geo-spatial image information that can easily be integrated as part of existing workflow, and
2. Intelligent tools to support capture and re-use of encapsulated task-based interactions and context.

When interacting with geo-spatial image information in service of a task goal, the user needs to be able to tease out the particular informational aspects that support the task goal. Ideally, two work products emerge – first, the actual information as applied to the task, and second, a record of the information gathering process that allows for incremental development and provides a reference for subsequent justification and refinement if necessary. As the user often needs to make notes and annotations in order to support the former, the latter can be supported in a natural way by integrating annotation tools tailored to the information gathering environment. This also supports efficient interaction, as it minimizes the need to divert attention from the information source.

In order to support the user in constructing the most on-point information kernels, we have designed a task-environment for managing and annotating task-based information. Within this environment we have developed tools for direct image manipulation, such as filters, transformations, highlighting, sketching, and post-it type annotations. These allow the user to, among other things, locate and define regions of interest in the images. The user-defined regions can then be linked to clarifications or rationale, initially in the form of textual annotations, as well as linking relevant associated geo-spatial data. The manipulations and annotations do not alter the underlying images or geo-spatial information, rather they are layered to provide a task-specific view.

Our system provides for the capture and refinement of more general task-based ideas and rationale by allowing for resources to be attached to this experience package. A typical interaction with the system then can capture the sketch and geo-spatial query or queries posed by the user, the results that were found to be useful, as well as the user's annotations of the results. All of the contextual knowledge required to address the task goal can thus be captured as a package or case of experience. These cases are referred to as sessions.

The capture of task-based experience cases is the foundation of the knowledge management process. It enables an increasingly powerful cycle of proactive support that can:

- Make available relevant task knowledge from a user's own previous work.
- Facilitate knowledge sharing by retrieving potentially relevant knowledge from other user experiences.
- Provide training support for novice users.
- Enable automatic proactive support by retrieving potentially relevant knowledge based on the partial task context.

Because the knowledge management system is tightly coupled with the tasks that the user is performing, the system can make proactive recommendations in a natural and unobtrusive manner by monitoring the user's current task context. Based on increments in the geo-spatial image information accessed and annotations provided, the system can correspondingly anticipate and update what previous experiential knowledge would be relevant at that stage, making it available to the user. The knowledge must be provided unobtrusively, so that it need only be accessed when required. Thus the process of knowledge retrieval does not distract from the task at hand, yet makes relevant knowledge available just-in-time.

The development of the task environment leverages tools that are developed for other parts of the system. The most complicated aspects deal with image manipulation and draw on lower-level tools that are used in authoring the geo-spatial image data. The challenge for the interface development in this aspect of the system is to develop a flexible architecture that enables customisation of the user interface, support for multiple types of annotation, and easily afford contextual and user interaction information to the knowledge management tools. In addition, we are investigating human-computer interaction issues in how best to provide knowledge support in a manner that is unobtrusive, so as not to interrupt ongoing work, but also effectively convey the presence and type of knowledge support available.

## **6. Implementation**

Our image retrieval system utilises metadata entry, semantic input and the image matching algorithms developed in (Carswell, 2000) to return relevant imagery. The system, however, is not dependent on the image similarity algorithms described in

(Carswell, 2000), indeed it may be integrated with any such image matching algorithm. We have integrated these components with a graphic user interface (GUI) through which users may interact easily with the system (O'Sullivan, McLoughlin, Bertolotto, & Wilson, 2003).

When a user logs in to the application, they are directed to an interface that enables them to search directly for images that correspond to their current task needs. Queries can be constructed using a combination of metadata, semantic information, and pen-based sketch input. The user can formulate an image library query using any combination of these elements.

The metadata screen allows a user to formulate a query based on the image location, scale, creation date, and semantic keys. For example, a user might be interested in building a shopping center near Boston and wish to view recent images of possible building sites and related areas. As shown in Fig. 5, the user could enter the location as Boston and a small scale value, as they wish to view large areas of land.

Once the user has formulated their query using any combination of the search screens, they can initiate the search.

The resulting matching images are returned to the search screen in a new tab. Matching images are displayed as a list of thumbnails with their associated matching percentage score, as shown in Fig. 6. A subset of the metadata for each image is available as tooltip text when mousing over the image.

The user can browse the images retrieved in the results screen and select any images that are relevant to the task at hand. The selected images are collected in the current user context and made available for manipulation and annotation.

To illustrate the annotation tools, we return to our shopping center example. After retrieving and selecting imagery relevant to Boston, the user can manipulate and/or annotate each image using a substantial set of tools, as shown in Fig. 7.

The tools are a subset of what might typically be found in a fully-fledged image processing suite. We have selected the kinds of image manipulations that would

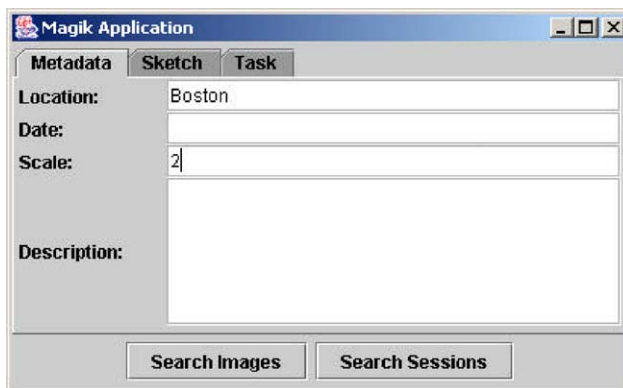


Fig. 5. Query by metadata.

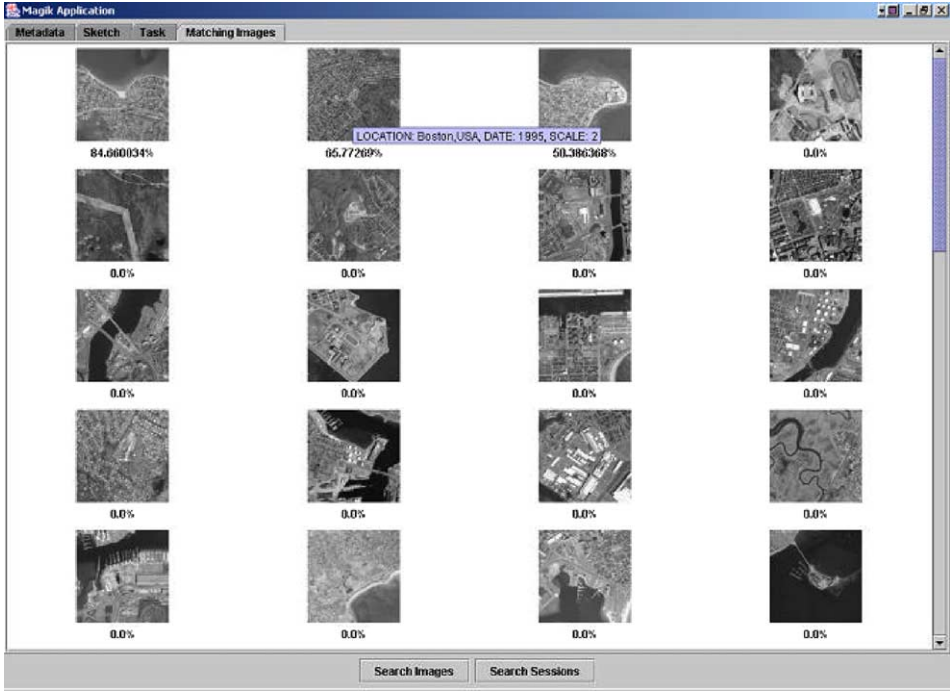


Fig. 6. Matching images.



Fig. 7. Image manipulation screen.

be most useful in helping to analyse and focus on image features (e.g., high-pass filtering). All of the sketching manipulations can be performed in a variety of colours and brush styles. The architecture has also been designed to facilitate the addition of new types of image tools as the need arises.

The user can then go on to add personal media annotations to the image as a whole or to particular highlighted image aspects. Currently, the system supports annotation by text, audio, and video. The system integrates real-time audio and video capture as well as compression. A wide variety of compression formats are supported, including QuickTime, Mpeg and H.263. All textual, audio and video annotations can be previewed before being incorporated as part of the knowledge base, and once recorded they can be saved and uploaded to the image context as a knowledge parcel associated with the task in question. Returning to our shopping center example, in Fig. 8, the user has made use of the transformation and annotation tools in carrying out their task.

They have highlighted some areas of housing by drawing coloured rectangles around them. This indicates that they believe that these areas may provide a good customer base for the development. They have highlighted the infrastructure of the area by means of the white line. They have also added a textual comment to an undeveloped area close to both the populated area and the main road, indicating interest in the feasibility of development in this area.

In Fig. 9, the user has made annotations to another image that indicate that this may also be a suitable site for the construction of the shopping center.



Fig. 8. Annotating an image.





Fig. 9. Annotating an image.

Here the user has uploaded a video file to an area of the image that is highly populated. They may have recorded this video while they were carrying out their task or this may have been an existing file that they had in their possession containing information specific to shopping center construction. The textual and video media annotations are represented by icons, which are painted on the image. If the user mouses over any of these icons the region associated with the annotation is emphasized by a rectangle drawn around the icon. This is shown by the dark rectangle around the camera icon in Fig. 9. The user can click on any of these icons to display a pop-up description as shown in Fig. 10.

The system also supports annotation by cut, copy and paste between a given image and other images in the dataset, as well as images in any application that supports clipboard functionality for the given operating system. A user's entire process of image interaction in the system is stored as an encapsulated session case. Note that a session can be saved and re-opened to continue processing later.

### 6.1. Annotation based retrieval

The capture of task-based experience cases is the foundation of the knowledge management process. In the first phase of the work, we are focusing our annotation-based retrieval on textual annotations, using information retrieval metrics as a basis for similarity. We presume that image retrieval is taking place toward some goal in the context of an overall work flow. Given a textual representation of the

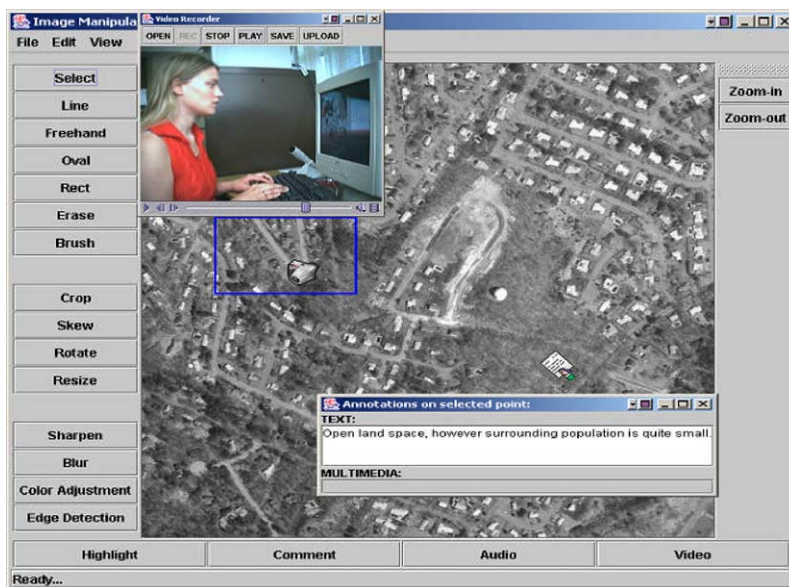


Fig. 10. Viewing annotations.

task context, we can match previously annotated session images to the current context.

This allows both querying by task context alone, as well as providing an additional factor that can be taken into account in conjunction with the original image query components. Moreover, the task descriptions and image query elements can be used to retrieve entire sessions as relevant to supporting the current task. Task descriptions could be provided by the user directly. Since we expect our system to be used in the context of an overall workflow, we have designed the system to link directly with upstream task descriptions, as they are provided to the user. This could also allow multiple users to share the same context for cooperative tasks. We also take relevant parts of any metadata provided as part of the task description context.

The system task-based retrieval employs indexes in three separate spaces:

1. Annotation Index—IR vector space across all textual annotations, where each annotation represents a “document”.
2. Image Index—IR vector space across all images, where the text for each image is composed of the combined text from all annotations and metadata.
3. Session Index—IR vector space across all sessions, where the text for each session is composed of the task description and user metadata query information.

These indices are used in two different types of retrieval: image retrieval and session retrieval.

### 6.1.1. Image retrieval

Image retrieval serves two purposes. First, task-based similarity can be used directly to access annotated images in the image library. Second, it can be integrated with similarities from the other types of query information, such as by image content, to provide a more refined overall metric for retrieval.

In searching for relevant images, similarity is computed as follows. If the image passes the metadata filter:

- Compute similarity in the image index.
- Compute and average similarities for each attached annotation in the annotation index.
- Compute the final image score as the average of overall image and individual annotation similarities.

### 6.1.2. Session retrieval

As the system builds up encapsulated user interactions, another type of retrieval is enabled, retrieving entire previous task-based sessions. This enables a current user to look for the previous image analysis tasks that are most similar to the current task both to find relevant imagery and to examine the decisions and rationale that went into addressing the earlier task. One challenge in retrieving previous sessions has been how to present an entire session to the user in a manner that is compact enough to allow multiple results to be viewed simultaneously while still providing enough information for the user to discriminate potential relevancy. Fig. 11 shows an example of our results for retrieved sessions.

In order to keep session listings small and still provide enough discriminatory information, each session result is summarised to include the following:

- Percent similarity score.
- The most discriminating query information (if more than one) for the session (since we have captured which results were actually used, we know which queries were most fruitful).
- The most important annotations (words, phrases, media buttons that play any audio or video clips uploaded during the session and are deemed to have the high similarity to the current user's context).
- Thumbnail versions of the most important images (images that have been annotated or browsed as part of the similar users context and bear relevance to the current users task and have been ranked accordingly).

In searching for relevant sessions, similarity is computed as follows.

For each previous session above a threshold similarity:

- Compute the preliminary similarity session score in vector space across all sessions, where the text for each session is composed of the task description, annotations and user metadata query information.

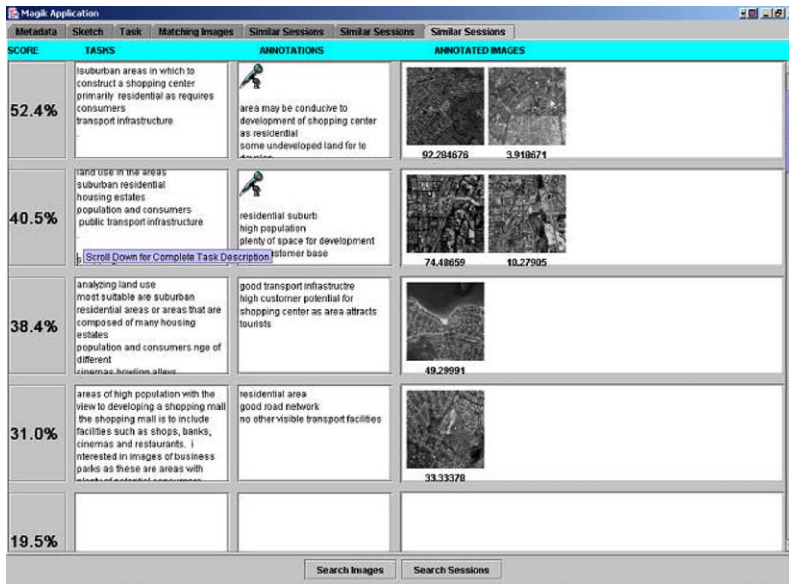


Fig. 11. Matching sessions.

- Compute the number of images annotated and browsed in that session as a fraction of the total number of images returned.
- Compute the final session score as the weighted sum of session similarity, proportion of annotated and browsed images.

The proportions of annotated and browsed images provide a measure of the relative usefulness of a given session, and they are given a parameterized weighting relative (currently lower) to the session index similarity component. The scores for the individual images annotated or browsed in a similar session are calculated by comparing any textual annotations made to those images during that session to the metadata and textual task description entered by the current user. If the user wishes to view the annotations made to an image returned in a similar session, they may do so by clicking on the thumbnail, which brings up the image and all its annotations in the image manipulation screen. Here the user has an opportunity to read a textual comment uploaded by the similar user and also to play audio and video clips uploaded as part of the similar session. The user may further annotate this image if they wish and/or retain the previous users annotations by adding it to their current session image context. The overall similar sessions scores as well as the scores assigned to individual images belonging to these similar sessions vary dynamically based on the interaction of the current user. For example if our user annotates or browses images that another user has already found useful in the context of their task, the scores assigned to these images will increase as the current user proceeds with their own task. Conversely session or image scores will decrease if associated queries, annotations, or images are ignored by our current user.

If a user decides to incorporate the work of another user as part of their own session, this knowledge is implicitly captured by the system and the scores associated with that session as well as the re-used annotations are updated to reflect this recycling process. Once the user saves the desired previous/new annotations, all of these annotations are transferred to the current user's view of the image. The user may perform many different queries and annotate or manipulate many images during the course of a session, and these are all saved as part of the users profile when they exit the application.

## 7. Evaluation

For the initial phase of the implementation, we have conducted testing with 17 different task scenarios using a library of 50 images with novice task-domain users. Sessions were completed each in one pass, without engaging in feedback and refinement. The current experiments are designed more to test whether the system is performing as expected rather than to provide an absolute measure of utility. We plan to carry out more extensive user trials in the near future. The ideal way to test the system, of course, would be to conduct user trials with task-domain experts, and we are actively seeking outlets for such testing.

### 7.1. Image retrieval using annotations

The goal of our evaluation in this instance was to show that with the addition of annotations to images in the library, image retrieval would improve over time. Firstly an empty library of sessions and annotations was constructed. Then our users interacted with the system to create a series of new task sessions. During each session the users added annotations to selected images that they considered relative to their task. Six different categories of task description were outlined, corresponding to civil development in the following areas: airports, hotels, bridges, hospitals, stadiums and shopping centers. An example task description might be: "build a shopping center in the suburbs of the city where infrastructure is good and land is inexpensive."

The users entered task descriptions for each of the outlined categories and initiated searches on the library. The results were returned to them as a set of ranked images in each case. The matching scores were recorded for each image, and the user selected images for annotation that seemed relevant to the task description. The annotations for each image were recorded, but not indexed, in order to provide a baseline retrieval performance for the system with annotations. A total of 20 sessions were added. The experiment was then repeated using the same task descriptions to evaluate how the image scores change with the addition of indexed textual annotations. Fig. 12(a–d) shows the results for the most relevant 4 images (judged by the user) for each session. In each graph there are six separate queries represented along the X-axis where the light/dark columns (e.g. A, A1) give the similarity respectively without and with indexed annotations for that particular query. As expected, there

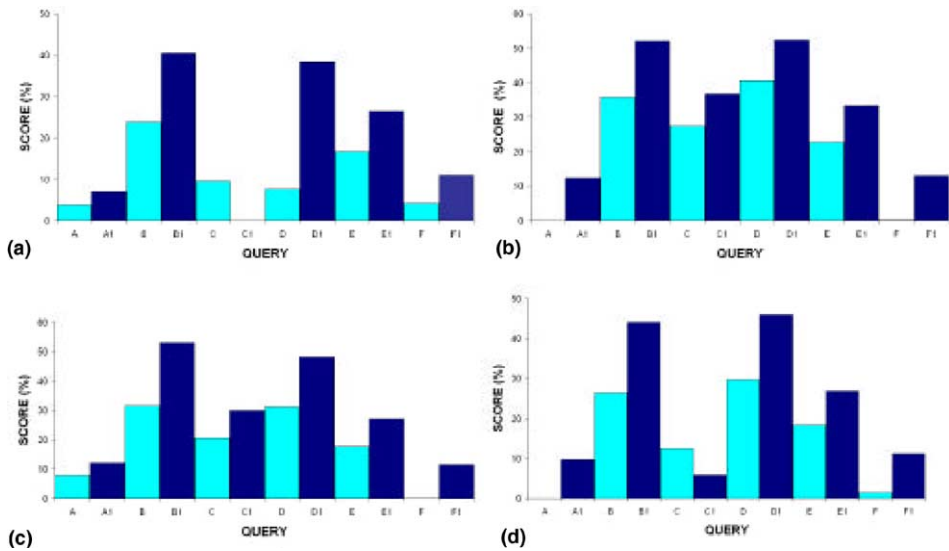


Fig. 12. Evaluation.

was an increase in the image matching scores, demonstrating the usefulness of including textual annotations for retrieval.

## 7.2. Session retrieval

In order to evaluate the retrieval of similar sessions, three new task descriptions (different from the 17 used for evaluating image retrieval) were then constructed in three of the predefined categories. The top three similar sessions returned by the application were analyzed and their results deemed to be relevant or irrelevant. The first task description corresponded to the “Airport” category. The task description entered by the user outlined that they were interested in viewing images of existing airports as an aid to developing a new airport facility. They sought general information concerning airport locations and orientations with regard to the urban areas which they service, land elevation, infrastructure and the average land space occupied. The top three similar sessions returned for the airport category were then recorded and analyzed. The scores associated with the top three sessions in this instance were 40.68%, 15.14% and 14.82% respectively. The task description of the first similar session outlined a scenario where the user was interested in constructing a new airport. It differed from the task description of the current session, however, in that the user was not interested in retrieving images of existing airports. Rather, they simply wished to view areas of land that would be appropriate for such a new development. Both task descriptions contained text associated with the airport domain such as elevation and land-space, as did some of the annotations uploaded

by the user of the similar session to the images returned in their session. This session was deemed to be useful in fulfilling the current task description.

In the second similar session the user had entered a task requesting the retrieval of images in order to analyze general land usage in the selected cities. Some of the images returned by the similar session depicted airports, and the user had made annotations in this regard. The session score was higher because some more general use terminology, such as land and urban appear in both queries. The session was deemed to be moderately related to the current task. The third most similar session user was interested in constructing a stadium in an urban area. The tasks are similar in that the land sites required for both developments are relatively large when compared to many other development domains and this is reflected in the session score. In both cases it is preferable for the development sites to be located away from the center of urban areas, given cost and the scale of previous development. This session also seems to see gains in similarity from more general use terminology that parallels the domains. We expect that there would be more marked differences allowing for finer distinctions with task-domain experts and more substantial task descriptions. This session was deemed quite similar, but not very relevant.

Similar evaluations were carried out in the development categories of “Bridge”—top three sessions: 35.59% (relevant), 14.65% (very relevant), and 13.79% (not relevant), and “Hotel”—top two sessions: 52.19% (relevant) and 20.14% (relevant). Only two sessions were retrieved in total for the “Hotel” task. While these initial results are only indicative, they do show that the system is performing as expected. We intend to undertake larger scale testing in the near future, and we realize that there are many factors that will need to be accounted for, including a larger range of categories, scaling the number of annotations, and refining vocabulary toward more domain-specific usage.

## **8. Conclusions**

In this paper we presented the geo-spatial image retrieval and knowledge management system we are developing. The main contributions of our work consist of providing an intuitive and effective retrieval environment based on user-drawn sketches and case-base reasoning and knowledge management techniques for developing and annotating task-based information. A knowledge base is built from previous experiences to improve future query processing.

Particular attention has been given to the design of a user-friendly graphic interface that, besides facilitating the query input phase, also displays the results of the query with a clear layout.

Experiments show the initial system implementation performing as expected, and we hope to scale testing as the implementation progresses. As the system matures, we expect to provide knowledge support in a flexible manner that can be easily integrated with existing infrastructure, and we intend to make use of existing standards for communication between system modules and for external communication.

For example, we will incorporate Geography Markup Language (GML) (<http://www.opengis.org>) support as the standard develops. Furthermore, we are investigating optimisation strategies in the form of efficient indexing as well as organisational/storage structures for image databases to obtain the best possible performance.

Once the implementation of the full prototype system nears completion, it will become publicly available over the web. We are currently in the process of transferring the system to the mobile platform and the system will also be available via this medium. The employment of progressive transmission of raster images (e.g., based on wavelets, etc.) will be investigated to facilitate access and delivery of large datasets over slow communication links.

We expect that system capabilities would be enhanced by including more formal representations such as domain ontologies, and we are particularly interested in the possibility of automatically deriving and linking categories based on analysis of task data. We plan to look into adding other resources, such as chaining the supporting experience cases, as well speech-to-text in order to support expanded media retrieval.

In the longer term, we plan to consider the analysis of matching results of images corresponding to the same geographic area in order to detect changes. This technique could be modified to allow for such temporal reasoning on the database. By loosening up object/relation constraints and by analysing matching percentages, we will be able to detect temporal changes in some areas, such as: the elimination of some objects, changes in object shape, and change in location.

Finally, we are planning to integrate personalisation techniques in the feature library organisation: indeed, different users can sketch the same object in very different ways. Each user will then have a personalised version of the feature library.

## Acknowledgements

The funding support of the Research Innovation Fund of Enterprise Ireland is gratefully acknowledged.

## References

- Aarts, A. (1998). A CBR architecture for project knowledge management. In *Proceedings of the fourth European workshop on case-based reasoning* (pp. 414–424). Dublin, Ireland, Berlin: Springer.
- Agouris, P., Bertolotto, M., Carswell, J. D., & Stefanidis, A. (2000). A scene similarity metric for matching configurations of image objects. *XIX ISPRS Congress*, Amsterdam, The Netherlands.
- Agouris, P., Bertolotto, M., Carswell, J. D., & Stefanidis, A. (2002). Scale and orientation-invariant scene similarity metrics for image queries. *International Journal of Geographical Information Science*, vol. 16, Taylor & Francis.
- Agouris, P., Carswell, J. D., & Stefanidis, A. (1999). A feature library approach to on-line image querying and retrieval for topographic applications. *Vision Interface 1999*.



- Agouris, P., & Schenk, T. (1996). Automated aero-triangulation using multiple image multipoint matching. *Photogrammetric Engineering & Remote Sensing*, 62(6), 703–710.
- Ardizzone, E., & La Cascia, M. (1997). Automatic video database indexing and retrieval. *Multimedia Tools and Applications*, 4, 29–56.
- Athitsos, V., Swain, M., & Frankel, C. (1997). Distinguishing photographs and graphics on the World Wide Web. *IEEE Workshop on Content-Based Access of Image and Video Libraries*, Puerto Rico (pp. 10–17).
- Barber, J., Bhatta, S., Goel, A., Jacobsen, M., Pearce, M., Penberthy, L., Shankar, M., & Stroulia, E. (1992). Integrating case-based reasoning and multimedia technologies for interface design support. In J. G. Boston (Ed.), *Artificial Intelligence in Design*, Kluwer Academic Publisher.
- Becerra-Fernandez, I., & Aha, D. (1999). Case-based problem solving for knowledge management systems. In *Proceedings of the twelfth annual florida artificial intelligence research symposium* (pp. 219–223). AAAI.
- Blaser, A. D. (1998). Spatial-Query-by-Sketch: fundamentals for a sketch-based user interface in GIS. Ph.D. Thesis, University of Maine, Orono, ME, USA.
- Burke, R., & Kass, A. (1996). Retrieving stories for case-based teaching. In D. Leake (Ed.), *Case-based reasoning: experiences, lessons, and future directions* (pp. 93–109). Menlo Park, CA: AAAI Press.
- Carson, C., Belongie, S., Greenspan, H., & Malik, J. (1997). Region-based image querying. In *IEEE workshop on content-based access of image and video libraries*, San Juan, Puerto Rico (pp. 42–49).
- Carswell, J. D. (2000). Using raster sketches for digital image retrieval. Ph.D. thesis, Department of Spatial Information Science and Engineering, The University of Maine.
- Champin, P. A., Prié, Y., & Mille, A. (2001). Annotating with uses: a promising way to the semantic web. In *Proceedings K-CAP 2001 workshop on knowledge markup and semantic annotation* (pp. 79–86).
- Chang, N. S., & Reuss, J. (1978). Design considerations of a pictorial database system. *International Journal Policy Analysis Information Systems*, 1, 49–70.
- Chang, S. F. (1997). SaFe/VisualSEEK—Automatic Joint Spatial/Feature Based Image Search System.
- Cohen, S. D., & Guibas, L. J. (1996). Shape-based indexing and retrieval; some first steps. In *Proceedings 1996 ARPA image understanding workshop 2* (pp. 1209–1212).
- Egyed-Zsigmond, E., Mille, A., & Prié, Y. (2003). Club (Trèfle): a use trace model. In *Proceedings of the fifth international conference on case-based reasoning (ICCBR 03)*, Trondheim, Norway, LNAI 2689, Springer-Verlag (pp. 146–160).
- Forsyth, D. A., Malik, J., Fleck, M. M., Greenspan, H., Leung, T., Belongie, S., Carson, C., & Bregler, C. (1996). Finding pictures of objects in large collections of images. In *Proceedings of the 2nd International Workshop on Object Representation in Computer Vision*, Cambridge, United Kingdom (pp. 335–360).
- Frankel, C., Swain, M., & Athitsos, W. (1996). WebSeer: an image search engine for the world wide web. Science Technical Report TR-96-14, Department of Computer Science, University of Chicago.
- Gonzalez, R. C., & Woods, R. E. (1992). *Digital image processing*. Addison-Wesley Publishing Company, Inc, p. 716.
- Grimnes, M., & Aamodt, A. (1996). A two layer case-based reasoning architecture for medical image understanding. In *Proceedings of EWCBR-96*. Springer.
- Gross, M., Zimring, C., & Do, E. (1994). *Using diagrams to access a case library of design, artificial intelligence in design*. Dordrecht, Netherlands: Kluwer, pp. 129–144.
- Gudivada, V. N., & Raghavan, V. V. (1995). Content-based image retrieval systems. *IEEE Computer*, 28(9), 18–22.
- Gupta, A., Weymouth, T., & Jain, R. (1991). Semantic queries with pictures: the VIMSYS model. In *Proceedings of the 17th international conference on VLDB*, Barcelona, Spain (pp. 69–79).
- Hampapur, A., Gupta, A., Horowitz, B., Shu, C.-F., Fuller, C., Bach, J. R., Gorkani, M., & Jain, R. (1997). Virage video engine. In *Proceedings SPIE, vol. 3022*, San Jose, CA, February 1997 (pp. 188–198).
- Holt, A., & Benwell, G. L. (1999). Applying case-based reasoning techniques in GIS. *The International Journal of Geographical Information Science*, 13(1), 9–25.
- Jagadish, H. V. (1991). A retrieval technique for similar shapes. In *Proceedings of ACM SIGMOD international conference on management of data* (pp. 208–217).

- Kauppinen, H., Seppänen, T., & Pietikäinen, M. (1995). An experimental comparison of auto-regressive and Fourier Descriptors in 2D shape classification. *IEEE PAMI*, 17(2), 201–207.
- Kelly, P. M., Cannon, T. M., & Hush, D. R. (1995). Query by image example: the CANDID approach. *SPIE Storage and Retrieval for Image and Video Databases III*, 2420, 238–248.
- Klahr, P. (1997). Knowledge management on a global scale. In *Proceedings of the 1997 spring symposium on artificial intelligence in knowledge management* (pp. 82–85). Stanford, CA: AAAI.
- Kolodner, J. L. (1993). *Case-based reasoning*. San Mateo, CA, US: Morgan Kaufmann.
- Leake, D. B. (1996). *Case-based reasoning: experiences, lessons, & future directions*. AAAI Press/The MIT Press, ISBN 0-262-62110-X.
- Leake, D. B., & Wilson, D. C. (2001). A case-based framework for interactive capture and reuse of design knowledge. *Applied Intelligence*, 14(1), 77–94, Kluwer, Boston.
- Lenz, M., & Ashley, K. D. (1998). Textual case-based reasoning. In *Proceedings of the AAAI-98 Workshop on Textual Case-Based Reasoning, AAAI Papers from the AAAI-98 Workshop, AAAI Technical Report WS-98-12*, AAAI Press: Menlo Park, CA.
- Mehrotra, R., & Gray, J. (1993). Feature-based retrieval of similar shapes. In *Proceedings of ninth international conference on data engineering*, Vienna, Austria (pp. 108–115).
- Mehrotra, R., & Gray, J. (1995). Similar-shape retrieval in shape data management. *IEEE Computer*, 28(9), 57–62.
- Niblack, W., Zhu, X., Hafner, J. L., Breuel, T., Ponceleon, D., Petkovic, D., Flickner, M. D., Upfal, E., Nin, S. I., Sull, S., Dom, B., Yeo, B., Srinivasan, A., Zivkovic, D., & Penner, M. (1997). Updates to the QBIC system. In *Proceedings of SPIE, vol. 3312* (pp. 150–161).
- Nishida, H. (1999). Shape Retrieval from image databases through structural feature indexing. *Vision Interface'99*, Trois Rivieres, Canada (pp. 328–335).
- Ogle, V. E., & Stonebraker, M. (1995). Chabot: Retrieval from a relational database of images. *IEEE Computer*, 28(9), 23–32.
- O'Sullivan, D., McLoughlin, E., Bertolotto, M., & Wilson, D. (2003). Capturing task knowledge for geo-spatial imagery. In *Proceedings of the 2nd international conference on knowledge capture*, Sanibel Island, FL, USA (pp. 78–87).
- Pentland, A., Picard, R. W., & Sclaroff, S. (1996). Photobook: content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3), 233–254.
- Perry, T., & Lewis, S. T. (1998). A novel image viewer providing fast object delineation for content based retrieval and navigation. In *SPIE conference on storage and retrieval for image and video databases VI, vol. 3312* (pp. 436–445).
- Persoon, E., & Fu, K. S. (1977). Shape discrimination using Fourier descriptors. *IEEE Transactions on System Management*, 7(3), 170–179.
- Sclaroff, S., Taycher, L., & La Cascia, M. (1997). ImageRover: a content-based image browser for the World Wide Web. *IEEE workshop on content-based access of image and video libraries*, San Juan, Puerto Rico (pp. 2–9).
- Smith, J. R., & Chang, S.-F. (1996). VisualSEEK: A fully automated content-based image query system. In *ACM multimedia conference*, Boston, MA, USA (pp. 87–98).
- Smith, J. R., & Chang, S.-F. (1997). Visually searching the web for content. *IEEE Multimedia*, 4(3), 12–20.
- Srihari, R. K. (1995). Automatic indexing and content-based retrieval of captioned images. *IEEE Computer*, 28(9), 49–56.
- Wilson, D. C., & Bradshaw, S. (2000). CBR textuality. *Expert Update*, 3(1), 28–37.
- Worring, M., Bagdanov, A., Gemert, J., Geusebroek, J., Minh, H., Schreiber, G., Snoek, C., Vendrig, J., Wielemaker, J., & Smeulders, A. (2002). Interactive indexing and retrieval of multimedia content. In *Proceedings of the 29th Conference on Current Trends in Theory and Practice of Informatics*, Milovy, Czech Republic (pp. 135–148).
- Yeh, A. G. O., & Shi, X. (1999). Applying case-based reasoning to urban planning: a new planning support system tool. *Environment and Planning B: Planning and Design*, 26, 101–116.
- Zloof, M. M. (1975). Query by example. In *Proceedings of AFIPS* (pp. 431–438).