Towards Query by Sketch*

Michael Springmann[1], Adriana Ispas[2], Heiko Schuldt[1], Moira Norrie[2], Beat Signer[2]

[1] Database and Information Systems Group, University of Basel, Switzerland
{michael.springmann, heiko.schuldt}@unibas.ch

[2] Institute for Information Systems, ETH Zurich, CH-8092 Zurich, Switzerland
{ispas, norrie, signer}@inf.ethz.ch

Abstract

Content-based retrieval has become a very popular and also powerful paradigm for searching in multimedia collections, especially in large collections of images. However, such queries require that one or even several reference images are available prior to the start of the search process. These reference images must be close to the final result so that the user can take them to express her information need. If such reference images are not available or if the information need is covered only by parts of the query object, the result usually does not meet the user's expectation. Therefore, more flexible user interfaces are needed that allow users to sketch a query image by hand drawings and to dynamically select regions of interest from a given query image.
In this paper, we present a novel approach to query by sketch where interactive paper and image similarity search are seamlessly combined. It is based on the iPaper/iServer system of ETH Zurich and the ISIS/OSIRIS content-based image retrieval system of the University of Basel. The paper presents the integrated system which has already very successfully been applied to the development of an interactive museum catalogue. Moreover, it reports on ongoing activities that aim at extending the system to support handwritten sketches, gestures and/or dynamic region selection to make the retrieval process more flexible and less dependent from existing query objects.

Categories and Subject Descriptors

H.5 [Information Interfaces and Representation]: H.5.1 Multimedia Information Systems; H.5.2 User Interfaces; H.3 [Information Storage and Retrieval]: H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries;

General Terms

Algorithms, Design, Human Factors

Keywords

Content-based Image Retrieval, Query by Sketch, Interactive paper, Region-based Image Retrieval, Annotations

1 Introduction

1.1 Motivation

"*A picture says more than a thousand words.*" Increasing numbers of large image collections are available in Digital Libraries. These collections significantly impact the way people access and use information, both in their business and private lives. The Web 2.0, for instance, has generated many

so-called social networks such as flickr[1] and Wikimedia Commons[2] to allow people to share images. Users can manually tag uploaded their images with textual meta data describing the content, known as collaborative tagging. However, although social networks are rather new, it is already clear that it is almost impossible to associate an image with a comprehensive set of objective tags based on user perception in order to support all kinds of queries. This leads to rather poor retrieval quality when issuing a keyword search — which, together with unfocused browsing through a collection, is currently the only support for accessing pictures in social networks.

Many application domains also have to deal with significantly increasing amounts of image data. In healthcare applications, for instance, more and more modalities such as CT (Computed Tomography) and MRT (Magnetic Resonance Tomography) produce high resolution images which are stored in a patient's electronic health record. Meta data usually does not contain any information on the image content. This means that from the point of view of Management of and Access to Virtual Electronic Health Records [Springmann *et alii* 2007], only simple queries like *"Give me all MRT and CT images of Patient X"* are supported, but not more advanced queries like *"Find the most similar images (or health records) of patients having the same pathological deviation in their MRT image as patient X"*.

The information retrieval community has addressed the problem of advanced information access for image data beyond textual keyword queries by making use of the image content itself (Content-based Image Retrieval, CBIR). Despite strong technical support, CBIR systems are still not in widespread use. The reason for this is twofold. First, content-based retrieval ("similarity search") requires a query image to start with that is sufficiently close to the final result, i.e., that precisely expresses the user's information need. Without such a query image, it is difficult or nearly impossible to have good retrieval quality, even if powerful relevance feedback mechanisms are available.

Second, in many applications users are particularly interested in certain regions of the picture. Assume, for instance, a CT or MRT image of a patient with a tumour. From the physician's point of view, similarity should be restricted to the region of the image where the tumour is detected, not to the overall image. However, current approaches either consider similarity in global terms or apply conventional segmentation techniques which do not take into account any application semantics.

A general problem in content-based image retrieval is that image data entails a "sensory gap" between the natural human senses' perception and the computer bit stream representation. Image retrieval is faced with the "semantic gap" between associated metadata and automatically derived features used in the retrieval process and the intended semantics of the search [Smeulders *et alii* 2000]. In addition, the retrieval process is driven by the user. Textual metadata of accompanying tags or derived from occurrence context has to account for possible subjectivity when the generator is human [Rui, Huand, and Chang 1999] or errors produced be protocol misuse when it is automatically generated [Güld *et alii* 2002]. Also, humans do not normally use standard query languages, but rather express their searches in vague natural language [Jain 1996].

We consider that the image retrieval process may be considerably improved by shifting the focus more on the human factor. More specifically, given the taxonomy of user interactions proposed in [Vendrig, Worring, and Smeulders 1999], we consider that the mentioned "query by construction", "query by canvas" or "query by sketch" category has not yet been offered the right attention. Thus, we propose an image retrieval system enhanced with a pen and paper interface for query formulation. Such an interface provides for several interaction scenarios otherwise not supported by information retrieval systems. To search for database images users may draw on paper freehand sketches and/or combinations of intuitive symbols, write textual annotations and specify gesture commands. Also, highlighting the regions of interest in the retrieved images as a means of relevance feedback is much more flexible than automatic segmentation-based region detection.

* The work presented in this paper is funded by the Swiss National Science Foundation (SNF) in the context of the project *Query by Sketch* under contract No. 200021-117800 / 1

[1] http://www.flickr.com

1.2 Sample Application Scenario

To highlight the benefits of the proposed approach, consider a next generation museum that offers its visitors novel interfaces to search for additional information on the exhibits. In particular, this includes a museum guide printed on interactive paper. For a visitor not interested in any of the added value of this guide, it can be used and finally taken home like any other museum guide. For visitors who want to make use of the offered functionality, the museum will provide digital pens for allowing input on paper and screens on the wall to display additional visual information.

The guide will start with a short table of contents or summary of exhibits. By pointing at one of entries, the visitor can get the directions to the selected exhibit. In addition, an area of the guide may allow for textual input, to issue a keyword search on the description of exhibits or annotations. It may also allow drawing a sketch of a piece of art (e.g., a painting) he has in mind and wants to visit, in case the names of works or artists are not known.

Once the visitor gets close to the exhibit and has scrolled to the detail page on the exhibit in the guide, the latter can be used to retrieve additional information. Navigation is done through either pointing on particular areas of the paper or using gestures. The visitor may also highlight particular regions of high interest, for instance to learn more about the smile of Leonardo Da Vinci's Mona Lisa. If the visitor draws a circle around the mouth area on the Mona Lisa picture in his guide, not only additional information about scientific work on this aspect can be shown, but also annotations done by other visitors of the museum, and even a search for similar smiles in other paintings can be issued. The search can be restricted to images that are located in the same museum and directions to access them given. Further search options could be to restrict the search to particular periods by either selecting names of art periods or identifying the start and end from a printed time scale by simply pointing to it.

1.3 Prototype Implementation

The above mentioned scenario is just one example on how query by sketch can add value to the interaction between users and Digital Library systems.

The paper presents a novel approach to queries by sketch and/or annotation and the individual selection of regions of interest from query images. From a systems point of view, its seamlessly combines the iPaper/iServer system of ETH Zurich and the ISIS/OSIRIS content-based image retrieval system of the University of Basel. The iPaper/iServer system acts as advanced user interface and has already been used in a similar setting as described in the scenario e.g. at the Edinburgh Fringe Festival in 2005 [Belotti *et alii* 2005]. The ISIS/OSIRIS has been added as the backend, providing support for image similarity search and relevance feedback and integration framework with other Digital Library (web) services. The paper reports on the integration of both systems which has originated in the context of the EU FP6 Network of Excellence DELOS. Moreover, it identifies the challenges for using iPaper as query front-end for sophisticated search and query refinement.

1.4 Structure of this Paper

The paper is organized as follows. Section 2 presents related work from content-based image retrieval and interactive paper. In Section 3, we introduce the existing prototype system. The challenges for advanced paper-based user interactions in CBIR – namely query by sketch, regions of interest, and the consideration of gestures and annotations – are discussed in Section 4. Section 5 concludes.

---

[2] http://commons.wikimedia.org

2 Basic Technologies and Related Work

2.1 Content-based Image Retrieval (CBIR)

A main challenge in content-based retrieval is the definition of an appropriate similarity measure for images. This similarity measure must solely be based on the information included in the digital representation of images. A common technique is to extract a set of so-called visual features. Features of interest commonly fall into one of the following categories [Del Bimbo A. 1999]: colour, texture, or shape. These features need to be extracted from all objects of a collection, thereby transforming all images from the object space to a high dimensional feature space. For each feature, an appropriate function to compute the similarity between objects in the feature space is needed. Using a distance function, similarity search between objects can be provided by a nearest neighbour search in the feature space.

For a CBIR query, the challenge is to identify those images in the database that are more similar to the reference image than all other images. Again, similarity means the distance of the query image to the images of the collection in the feature space. This is based on the implicit assumption that the closer the feature representations of two images are, the more similar their content is. Hence, the definition of an appropriate distance function is crucial for the success of the feature transformation. Some examples for distance metrics are the Euclidean distance [Niblack 1993], the Manhattan distance [Stricker and Orengo 1995], the maximum norm [Stricker and Orengo 1995], the quadratic function [Hafner *et alii* 1995], Earth Mover's Distance [Rubner, Tomasi, and Guibas 2000], or Deformation Models [Keysers *et alii* 2007b]. Using a distance function, the similarity search becomes a nearest neighbour search in the feature space.

Moreover, it is of utmost importance to determine the most similar objects efficiently — especially for large collections of images. This problem is often solved by using some kind of index structure for the content descriptors (feature vectors) of the images [Witten, Moffat, and Bell 1999]. While the similarity metric influences the effectiveness of the retrieval, the index structure biases its efficiency. The efficiency can also improve using algorithmic optimization during query execution [Springmann and Schuldt 2007].

Over the last couple of years, several CBIR systems have been developed. A comprehensive overview of other existing systems is presented in [Veltkamp and Tanase 2000]. [Müller *et alii* 2004] reviews the use of CBIR in healthcare applications. ISIS (Interactive Similarity Search) is a powerful content-based multimedia retrieval system (presented in detail in Section 3.1) that supports several different media types and the combination of any of these media types with text retrieval. ISIS has been chosen as the backend system for the query by sketch approach.

2.2 Regions of Interest in CBIR

Very simple images, usually created in a controlled / standardized setting, may show a single object in the foreground and plain background. In practice, it is much more common that several objects are present in an image, some in the foreground and some in the background. When a query is issued by giving an example image, the user has a clear idea which regions of the image are relevant to his information need. Good retrieval systems need to take only these regions into account when determining the similarity.

Systems like Blobworld [Carson 1999] or MARS [Rui, She, and Huang 1996] index low level features for regions independently rather than for the complete images. They apply automatic segmentation to identify isolated regions that are assumed to correspond to the boundaries of objects, i.e., they try to detect coherent regions covered by the same object(s). During query evaluation, the systems try to identify regions in the images of the searched collection which are similar to the regions found in the query image. This approach uses several assumptions, which are unfortunately not present in many real world images: Firstly, objects are expected to be not covered by other objects, since this may split objects into several unconnected regions. Secondly, automatic

segmentation is not able to take into account application-specific regions of interest that span several homogeneous regions or that might be fully contained into one of them. For instance, the region of highest interest of a bone fracture will be precisely located where two or more segments of similar colour face each other.

Currently there are several approaches that aim at identifying points of interests rather than entire regions. The approach presented in [Gouet and Boujemaa 2001] extracts colour information of a few hundred characteristic points along the edges of objects within an image. [Keysers, Deselaers, and Breuel 2007] use small subimages (image patches) extracted from such points of interest, [Lowe 1999, Bay, Tuytelaars, and Van Gool 2006] generate scale- and rotation-invariant features from local image gradients. All these derived local features are inherently robust against translations of the image and can retrieve interest points of objects even if they are partially covered. On the other hand, this may lead to degradation of retrieval quality, if irrelevant areas (e.g., in the background) contain such characteristic points. It is therefore important that the user is able to individually express which parts of the image contain relevant, i.e., interesting, points. For many applications, user-defined regions of interest are very valuable information that should always be stored with the image. Therefore current standards such as DICOM and MPEG-7 provide support for storing regions of interests (ROI) as Overlay or Curve Activation Layer [DICOM 2004] or as a Region Locator [MPEG-7 2002]. However, region-based CBIR for regions other than those who have automatically been derived during segmentation is not provided yet in a satisfactory way,

2.3 Interactive Paper

While digital technologies are often used for storing, processing and delivering information, paper is still the preferred medium for reading. Paper supports many interactions including flexible navigation, easy mark-up through free-form annotations or hybrid activities such as reading and writing which are difficult to support in digital environments [Sellen and Harper 2002]. Different workplace studies that have been carried out to investigate the role of paper in working environments revealed that there is a need for integrated, hybrid paper-digital work processes, since paper cannot simply be replaced by digital devices.

With the emergence of new technologies for tracking interactions on paper documents various new interactive paper projects focusing on the integration of paper and digital media have been realised over the last few years. Thereby, the hardware solution most commonly used is the commercial Anoto[3] Digital Pen and Paper technology (see Figure 1). The Anoto solution encodes information on paper using a special printed pattern of tiny dots which are recognised by a camera which is integrated in a digital pen. The ready availability of Anoto-based pens has made it much easier for researchers to investigate innovative ways of bridging the paper-digital divide than previous efforts based on the use of cameras to track user actions such as in the DigitalDesk system [Weller P. 1993].

[3] http://www.anoto.com/

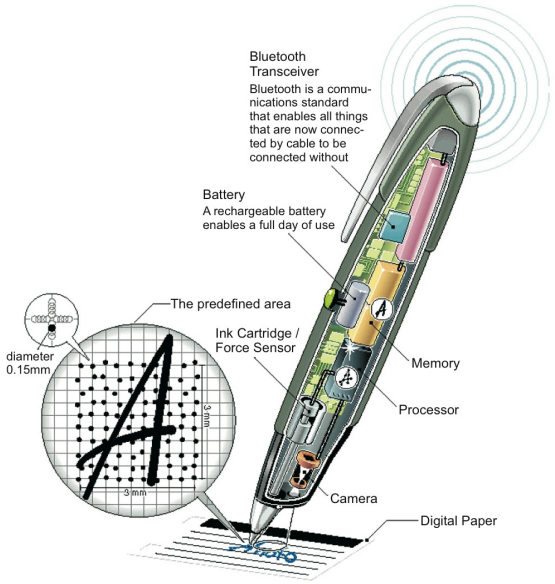

Figure 1: Digital Pen and Paper technology

The ButterflyNet project [Yeh *et alii* 2006] at Stanford is an example of a project to study the potential use of Anoto technologies to support the digital capture of information from scientists' notebooks including notes and their links to digital photographs. In PADD (Paper Augmented Digital Documents) [Guimbretière F. 2003], Anoto technology was used to support paper-based annotation of digital documents. After printing a document together with the Anoto pattern, it can be annotated with a digital Anoto pen and the strokes can later be added as annotations to the digital version of the document. The annotations are embedded as images in the original document. This edit cycle can be repeated multiple times and subsequent printouts of a document will always include changes made to a previous paper version. In the PapierCraft project [Liao, Guimbretière, and Hinckley 2005], the approach has been extended to allow operations such as paper-based copy and paste for text documents.

A restriction of many existing interactive paper solutions is that they only deal with links from paper documents to digital information, whereas linking from digital information to paper is not supported. Furthermore, either the forms of links supported are restricted to specific forms of digital services or the software, and possibly hardware, are tailored to a specific application. However, it is important to provide a general software infrastructure that can support a wide range of applications and technologies. Over the past few years, such a framework for interactive paper, called iPaper [Norrie, Signer, and Weibel 2006] has been developed at ETH Zürich.

2.4 Gestures for Human-Computer Interaction

With the increase of computer applications, which are now the common ground in supporting the users in their very different tasks and areas of activity, a wide variety of input devices have also been developed to provide the appropriate gestures for more natural, easy to use, easy to learn and error-free user interfaces [Buxton 1986a]. Direct manipulation interfaces relying on interaction means provided by a stylus device have been identified as a suitable solution for manipulating computer stored information for better emulating users' mental models [Buxton 1986b]. Some approaches even argue for using the pen for controlling all computer activities as the pen unifies the three input modalities: pointing, data entry and commands [Carr and Shafer 1991]. In [Wolf 1986] and [Wolf and Morrel-Samuels 1987] extensive users studies proving the feasibility of gesture-driven interfaces are presented and in [Wolf, Rhyne, and Ellozy 1989] the benefits and the requirements of gestural interfaces for several applications such as a spreadsheet program, a sketching program, a music program and a mathematical formatter are identified.

3 Combining Interactive Paper and Content-based Image Retrieval

The starting point for query by sketch is the tight integration of an interactive paper interface and a CBIR system. In our previous work, we have combined the iPaper/iServer system as advanced user interface and the ISIS system as underlying multimedia similarity search engine. This integration has been done in the context of the DelosDLMS development in the DELOS FP6 Network of excellence. DelosDLMS is a prototype implementation of a next-generation Digital Library management system [Schek and Schuldt 2006]. Based on the ISIS system and the OSIRIS middleware on top of which it runs, several advanced Digital Library services from different providers have been integrated [Agosti et al 2007]. In DelosDLMS, the iPaper has proven to be a very powerful interface to Digital Libraries (DLs), allowing for novel types of user interactions with a DL.

3.1 ISIS

ISIS (Interactive Similarity Search) is a CBIR system that provides content-based retrieval over different media types including image, audio, and video and also the combination of any of these media types with text retrieval. It also supports relevance feedback which gives the users the option to step-wise refine a query by distinguishing between relevant and non-relevant query results. Image retrieval in ISIS is based on colour moments and Gabor texture moments. The ISIS project has been started in the Database Research Group of ETH Zurich and is now being continued at the University of Basel. ISIS is implemented as a set of Digital Library services on top of OSIRIS [Schuler *et alii* 2004], a peer-to-peer infrastructure for distributed, reliable process execution [Mlivoncic, Schuler, and Tu□rker 2004]. ISIS stores the extracted features and performs queries using the VA-File, a data structure along which is particularly well suited to efficiently implement nearest neighbour search in high-dimensional spaces [Weber, Schek, and Blott 1998], support for region-based image retrieval has been added in [Weber and Mlivoncic 2003]. Figure 2 depicts the query result of a combined keyword and image similarity search in ISIS.

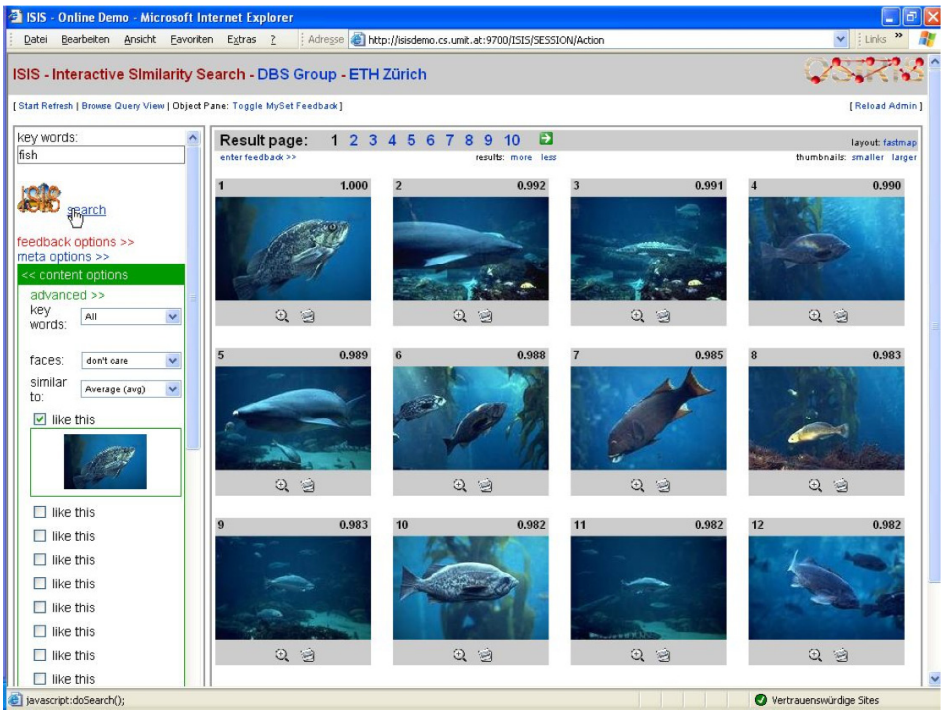

Figure 2: ISIS Search result for keyword "fish" and one reference image

3.2 iPaper

The basic idea of the iPaper architecture shown in Figure 3 is to define active areas on paper documents which then can be linked to digital information or services as well as to other physical resources. Therefore, the iPaper framework has been implemented as a plug-in for our general iServer cross-media platform that allows associations to be defined between arbitrary digital or physical resources [Signer and Norrie 2007b].

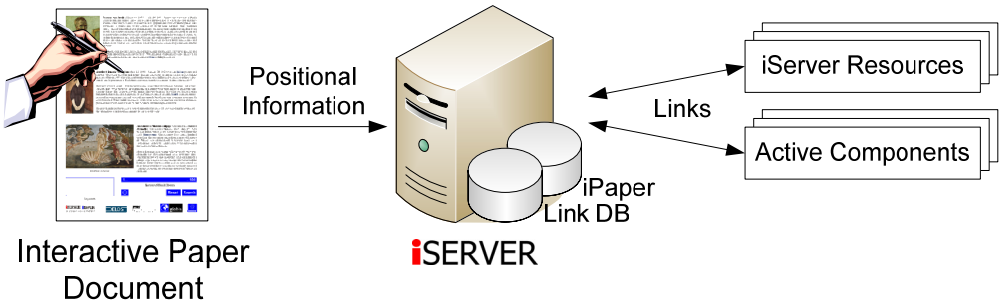

Figure 3: iPaper architecture

After a position within a paper document has been captured by a digital pen, the positional information is sent to the iPaper plug-in that will check if the position lies within any active area. An active area can be linked to other iServer resources or active components. There are currently iServer plug-ins for web pages, movie clips, Flash movies etc. which means that paper-based links can be defined to those types. Of course a paper resource can not only be linked to digital resources but also to paper documents. In this case iServer manages digital associations between different paper documents. While links to other iServer resources return a single piece of information (e.g., paper, a movie or a semantically rich database object), active content is represented by active components which are bound to a piece of Java program code. If a link with an active component target object is selected, a new active component is instantiated and its program code is executed.

The iPaper framework does not only support *enhanced reading* where paper documents are augmented with supplementary information but it also enables the development of *enhanced writing* applications where handwriting information is transformed into digital data. For this, we have developed different active components for capturing and further processing of pen data. For example, handwritten information can be translated into digital text by using commercially available Intelligent Character Recognition (ICR) tools.

While a commercial solution is used for online handwriting recognition within the iPaper framework, we have developed the iGesture recognition framework [Signer, Kurmann and Norrie 2007a] for the processing of online gesture-based commands. iGesture provides some predefined gestures but the iGesture tool component enables users to define and manage their own gesture sets.

The iPaper and iGesture framework by themselves do not support the process of authoring and publishing interactive documents and therefore we have also developed a number of tools to support the authoring and publishing of interactive paper documents, including a tool for the manual authoring of links between existing resources as well as a system to support the large-scale publishing of interactive documents based on a content management system [Weibel, Norrie, and Signer 2007].

3.3 Integration of iPaper and ISIS

As part of the DELOS project, the iPaper/iServer interface and the ISIS CBIR system have been tightly integrated into DelosDLMS, a novel prototype implementation of a next-generation Digital Library management system [Schek and Schuldt 2006]. DelosDLMS provides functionality in a single system that is not available in any known system so far. In particular the interactive paper interface to a Digital Library has been considered as one of the most appreciated components of the DelosDLMS.

As showcase for the iPaper/iServer interface of DelosDLMS and its image retrieval component ISIS, an interactive printed museum guide as a physical user interface has been developed [Agosti *et alii* 2007]. In the interactive museum guide, shown in Figure 4, existing printed text components and images, as well as some new interface components in the form of blue paper "buttons", are associated with specific ISIS queries.

The user starts a search by specifying one or more keywords either by selecting underlined word within the text (similar to the selection of hyperlinks in a web page) or by writing words in a special keyword input area. When the user selects the 'Search' button with the digital pen, a query consisting of the set of keywords is sent to ISIS and a list of images matching these keywords is returned. Another type of query is based on finding image similarities using controls on paper to define factors such as the number of images to be returned. From a technical point of view, the active component concept has been used to develop a paper user interface for the ISIS system.

Figure 4: Interactive Paper Interface to DelosDLMS

The results from ISIS are presented on a computer display to which the pen communicates via Bluetooth. The query can be modified with subsequent input or a new query can be formulated after the reset button is picked. The overall architecture of DelosDLMS is illustrated in Figure 5.

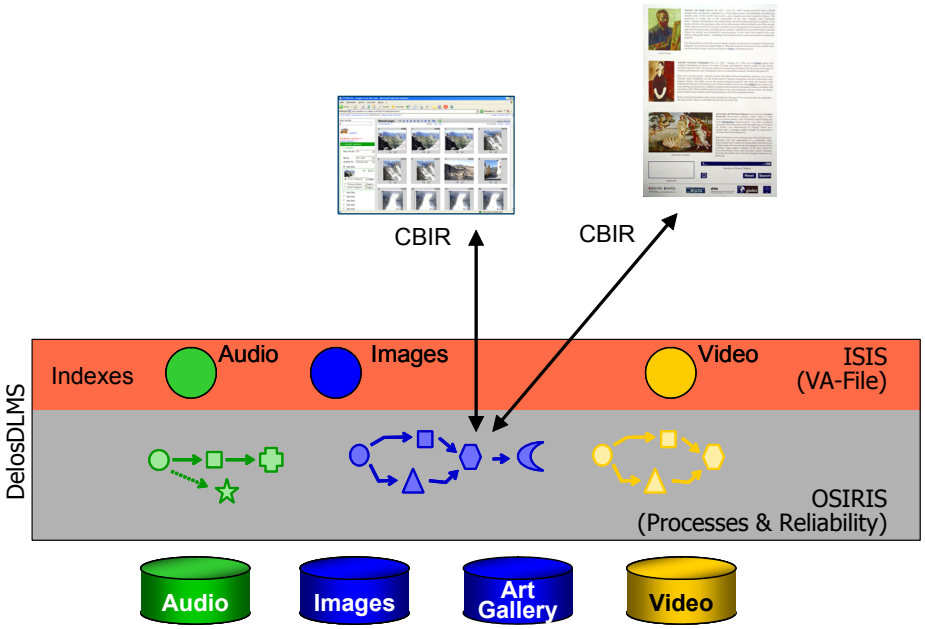

Figure 5: Architecture DelosDLMS and Active Paper

From the paper interface, a user is able to invoke a search process in the OSIRIS part of the DelosDLMS. This process consists of several steps needed to extract the characteristic features of an object (if necessary, i.e., in case this object is not yet in the database), to search for the nearest neighbours using ISIS' VA-file index, and to apply some post processing to the result set. The result of the query initiated and formulated on the iPaper is then sent for display to the ISIS graphical user interface.

4 Requirements and Strategies for Query by Sketch

Having such a powerful and advanced user interface like the iPaper for a content-based image retrieval system like ISIS is a very good starting point for taking query by sketch to the next level: The user can draw sketches on regular paper with an almost regular pen and the image retrieval system can instantly use this input for performing queries. The same is true for the individual selection of regions of interest within an image printed in interactive paper. In order to achieve good query by sketch user experience, several aspects need to be dealt with.

4.1 Sketches as Query Input

Even with the latest developments of digital input technologies, paper remains the preferred medium for high quality and precision designs, where aspects like the thickness of lines, connection of edges, or rectilinearity make a difference. As pointed out in [Veltkamp and Tanase 2000], one of the most significant drawbacks of existing approaches for query by sketch is using the mouse as input device instead of regular pen and paper. With the combination of iPaper and ISIS, this drawback has already been eliminated. In addition to the traditional keyword and example image based queries, our integrated system will finally be able to retrieve images based on similarities with freehand sketches drawn on regular paper. For users with less artistical skills,

searches based on combinations of paper drawn graphic symbols may also be expressed. Further, by using a pen interface, paper hand drawn gesture commands may be interfaced with the retrieval system.

The main challenge in using sketches as query input for searching in image collections is caused by the fact that normally those collections do not contain sketches themselves. This leads to the fact that the information contained in the input differs significantly in type and detail:

- Sketches have the property that out of the three main classes of features used in image retrieval (i.e., colour, texture, and shape), the first two will hardly be present in sufficient detail. If colour information is added, it will be restricted to only very few, homogenously filled colour areas.
- Sketches contain less information, e.g. no details in the image background. An average user may also reduce surfaces and rectangular structures to simple lines.
- Sketches are frequently used to highlight spatial relationships between objects – information that only few CBIR systems can evaluate so far. The reason is that it is very hard to derive such information automatically from photographs or other images due to unsolved problems in object identification/segmentation in the presence of a two-dimensional projection of a three-dimensional scene.

All these aspects need to be taken into account when selecting the image features to perform similarity search. It might be necessary to adopt existing approaches to better account for these particular properties.

For the implementation of the similarity measure, shape features like Elastic template matching [Del Bimbo and Pala 1997] seem most promising. Some pre-processing of the sketch might be needed in order to create connected contour drawing. Another class of features can be deformation models as presented in [Keysers *et alii* 2007b], if the direct pixel values as colour information is taken into account to a much smaller extent and only if present in the sketch, and otherwise only gradient information of detected edges will be used. Common texture features examine as part of their parameters the orientation of objects, often evaluated at various scales e.g., the Gabor texture features used in [Dimai 1999]. This information can still be of high value, thus also these features can provide valuable information in a combination together with others.

4.2 Using Symbols and Gestures as Input for Queries by Sketch

To support searches based on combinations of hand drawn symbols, an important and challenging task is the definition of a relevant collection of symbols. When focusing on specific topics of interest, such as engineering drawings, logic circuit diagrams or maps, the set of symbols may be directly derived from the specific notations employed in the application domain. In the approach proposed in [Garain and Chaudhuri 2004] for the particular case of mathematical expressions, the recognition is performed in two stages. In a first stage, symbols pertaining to a predefined set are recognized. Secondly, groups of symbols bearing a correct mathematical semantics are formed based on spatial relationships and context information. A related approach is the one proposed in [Donaldson and Williamson 2005] for a business process modelling application. Symbols corresponding to single stroke gestures are immediately replaced by the correct diagram component, ambiguous situations being solved by context information.

In the case of a non-specific application domain, which is also the case of our image database, defining a meaningful set of abstractions is not a trivial task. The set of symbols has to be general enough to support specific user queries, but also context inference needs to be enforced. An interesting solution for object categorisation was proposed in [Belongie 2007], where object recognition was further enhanced by using context information provided by Google Sets[4]. The

[4] http://labs.google.com/sets

proposed approach, however, does not accommodate semantics derived from spatial relationships between objects. In the case of a scene comprising a person placed "above" an animal, information about their relative positions could enforce the decision that the animal is actually a horse.

For supporting gestures in Query by Sketch, the objective is to extend the iGesture framework with gesture sets corresponding to different sets of meaningful symbols. Sets of rules defining constraints between recognised gestures need to be defined so that complex queries may be expressed. An immediate benefit of using a pen interface is the possibility of employing the temporal order of the sketched gestures to infer relationships between corresponding symbols [Wolf and Morrel-Samuels 1987]. For this, different symbol recognition algorithms [Lladós 2001] need to be evaluated. The flexibility in defining new gesture sets is very helpful when experimenting with new paper-based interfaces for sketch-based queries. For the gesture recognition process iGesture currently offers four different algorithms. However, new algorithms can easily be integrated based on a simple recogniser interface. A test bench enables the manual testing of algorithms as well as the automatic evaluation of multiple algorithms and their configurations in batch mode, followed by an appropriate visualisation of the results.

The benefits of gesture driven user interfaces have been clarified even from very early studies [Wolf and Morrel-Samuels 1987]. Besides the flexible means of interaction, the ease of use and the easy comprehension and assimilation, an important benefit of employing a pen interface is that users may express commands or invoke services offered by the retrieval system without having to switch back and forth between the two information environments, the paper medium and the digital world. After having drawn a sketch or a symbol combination the user may invoke the search service by simply touching a button or other visual representations printed on the paper canvas. Compound queries, comprising combinations of sketches, symbols and query images pre-printed on paper and selected with the pen will also be also supported. Parameters such as the number of results, date or geographic information (cities, regions or even geo-coordinates) associated with an image, or the techniques used for retrieval may be also specified by means of controls printed on paper.

4.3 Region Selection and Region-based CBIR

Image features and similarity measures have to be selected and integrated in the image retrieval service. Existing approaches may have to be extended and combined to exploit the regions selected by the user for best retrieval results. The selection of regions should not be limited to one or more segments identified by automatic segmentation algorithms as provided by most common Region-based CBIR approaches, but rather support more flexible area definitions as defined in the DICOM and MPEG-7 standards to take into account as much domain and context knowledge of the user as possible. Since most CBIR features and similarity measures do not provide support for this on their own, appropriate filtering and weighting schemes have to be introduced to emulate this behaviour.

Region selection may occur in several flavours:

- The simplest case is that the user marks the area of interest on a picture that has been printed on interactive paper. In that case, the covered area can be determined very easily and used to identify segments of interest and adjust weights. For instance, in the case of tourist photographs, one user might be interested in the person standing in front of a famous building to find more images of the trip that they did together. Another user might be interested in architecture and would prefer to find more images of the same building excluding any person in front of it that may hide interesting details about this part of the building.
- Moreover, region selection is also beneficial for sketches. If a user issues a search using a sketch, the system may pay too much attention on minor details of the sketch. Current approaches would offer two possibilities: Either let the user initiate another search which would imply that the sketch has to be re-drawn – or let the user allow to give relevance feedback on each individual result, e.g. judge whether he likes an result image or not. Such

relevance judgements are then used by the system to adjust the query and weighting to improve the result. This means, it tries to show more images for which the user would give positive feedback. The drawback of such a relevance feedback approach is, that the user can only give information how much a result is liked and not why. Region selection could enhance this on both sides: It could be used to modify the query by selecting only the important regions of the sketch and giving less weight on minor details. It also allows the user to highlight not only which images were good results, but which region in the result makes it a good result. Thus, the system gains much more knowledge to improve search than with traditional relevance feedback techniques.

On the technical level, for features that preserve already information on the location in the image, a simple overlay for weighting might be appropriate to adjust weighting based on regions. For instance, [Keysers *et alii* 2007b] use a downscaled version of the image for computing the deformation, to which each pixel contributes. The contribution of a pixel can be adjusted based on the whether it is contained in the region selected by the user. The same applies to techniques using keypoints like [Lowe 1999, Bay, Tuytelaars, and Van Gool 2006], where each descriptor is associated with the location of the keypoint from which it has been extracted. Additionally, such keypoints can be used to recognize objects in the compared image. This information might be used even for other features like the colour and texture moments which have been encoded with weak spatial constraints as described for instance in [Stricker and Dimai 1996].

4.4 Making Use of Annotations for iPaper-based CBIR Queries

Annotations to objects in a collection allow for adding user-defined information either on the complete object or on parts of it. An annotation can be either of textual nature (keywords) of can consist of semantic gestures. Thus, the combination of interactive paper and CBIR is not only beneficial for initiation of similarity queries via paper, but also for enriching objects in a collection, and for exploiting these annotations for query processing. Query by annotation has to be seamlessly integrated with query by sketch and dynamic region selection at the interface level. However, from a functional point of view, it is important to clearly distinguish between sketches (used for query purposes) and annotation (used for additional information to an object). For this, an intelligent character recognition (ICR) extension is needed for the semantic gesture recognition component by integrating ICR functionality such as the one offered by the *MyScript* handwriting recognition engine from *Vision Objects*[5].

For the annotation of existing content, for example in the form of printed images, a semantic mapping process has to be added to the printing process of interactive paper documents which will manage the necessary metadata to be used for later image annotation and the construction of the corresponding queries.

Textual meta data on objects which is added manually by means of annotations with interactive paper will have to be stored together with the reference to the annotated image. Such annotations can refer either to the entire image or only regions of interest, which will be stored and managed together with the annotation. For the textual annotations, stemming will be applied. When a user issues a query that contains both a reference image and a set of keywords, the search component finally needs to merge the objects which are selected due to the nearest neighbour search on the reference image and the images that qualify for their search because of their annotation. For this, algorithms are needed that are able to merge the ranked result lists.

[5] http://www.visionob jects.com/

5 Conclusion and Outlook

In this paper, we have introduced a novel paradigm for querying Digital Libraries. On the basis of interactive paper, a user is able to interact with a system by selecting query objects for content-based image retrieval from a printed paper, to add handwritten keywords, etc. Currently, we are extending the system to allow for even more user-friendly and flexible interactions in content-based image retrieval. This includes handwritten sketches and gestures on interactive paper as query input which, together with relevance feedback, finally allows similarity queries without reference image. In addition, also the dynamic selection of regions of interest will be supported. With this, a user will finally be able to individually extract the relevant parts from either sketches or images.

References

Agosti M., Beretti S., Brettlecker G., Del Bimbo A., Ferro N., Fuhr N., Daniel Keim D., Klas C.-P., Lidy T., Norrie M.C., Ranaldi P., Rauber A., Schek H.-J., Schreck T., Schuldt H.,Signer B., and Springmann M. DelosDLMS - the integrated delos digital library management system. In C. Thanos and F. Borri (eds.) *DELOS Conference on Digital Libraries Working Notes, Pisa (Italy), 13-14 February 2007*. 71-90

Bay H., Tuytelaars T., and Van Gool L. 2006. Surf: Speeded up robust features. In *Proceedings of the ninth European Conference on Computer Vision, Graz (Austria), 7-13 May 2006.*

Belongie S., Rabinovich A., Vedaldi A., Galleguillos C., and Wiewiora E. 2007, Objects in Context, In *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV 2007), Rio de Janeiro, Brazil, 14-20 October 2007*. Washington: IEEE Computer Society.

Belotti R., Decurtins C., Norrie M.C., Signer B. and Vukelja L. 2005. Experimental Platform for Mobile Information Systems. In T.F. La Porta, C. Lindemann, E.M. Belding-Royer, and S. Lu (eds.), *Proceedings of the 11th Annual International Conference on Mobile Computing and Networking (ACM MobiCom 2005), Cologne (Germany), 28 August - 2 September 2005.* New York : ACM Press. 258-269

Buxton W. 1986a. There's More to Interaction than Meets the Eye: Some Issues in Manual Input, In D.A. Norman and S.W. Draper (eds.), *User Centered System Design: New Perspectives on Human-Computer Interaction*, Hillsdale : Erlbaum. 319-337

Buxton W. 1986b. Chunking and Phrasing and the Design of Human-Computer Dialogues. In H.J. Kugler (ed.) *Proceedings of the 10th IFIP World Computer Congress, Dublin (Ireland), 1-5 September 1986*. 475-480

Carr R.K. and Shafer D. 1991. *The Power of PenPoint*, Boston : Addison-Wesley

Carson C., Thomas M., Belongie S., Hellerstein J.M., and Malik J. 1999. Blobworld: A system for region-based image indexing and retrieval. In . In *Proceedings of the Third International Conference on Visual Information and Information Systems (VISUAL '99), Amsterdam (NL), 2-4 June 1999*. Lecture Notes in Computer Science 1614. Berlin-Berlin-Heidelberg: Springer. 509-516

Del Bimbo A. and Pala P. 1997. Visual image retrieval by elastic matching of user sketches. In *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Washington: IEEE Computer Society. Vol. 19(2) : 121-132

Del Bimbo A. 1999. *Visual information retrieval*. San Francisco : Morgan Kaufmann Publishers Inc.

DICOM. 2004. *Digital imaging and communications in medicine (DICOM). Part 3: information ob ject definitions*. PS 3.3-2004.

Dimai A. 1999. Rotation invariant texture description using general moment invariants and gabor filters. In *Proceedings of the 11th Scandinavian Conference on Image Analysis, Kangerlussuaq (Greenland), 7-11 June 1999*. 391–398

Donaldson A.F. and Williamson A. 2005. Pen-based Input of UML Activity Diagrams for. Business Process Modelling. In *Proceedings of the 1st Workshop on Improving and Assessing Pen-based Input Techniques (HCI 2005 W5), Edinburgh (UK), 5 September 2005.*

Garain U. and Chaudhuri B.B. 2004. Recognition of online handwritten mathematical expressions. In *IEEE Transactions on Systems, Man, and Cybernetics, Part B* Vol. 34(6) : 2366-2376

Gouet V. and Boujemaa N. 2001. Object-based queries using color points of interest. In *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL 2001)*, Washington: IEEE Computer Society. 30-36

Güld M.O, Kohnen K., Keysers D., Schubert H., Wein B.B., Bredno J., Lehmann T.M., 2002, Quality of DICOM header information for image categorization, *In Proceedings of SPIE International Symposium on Medical Imaging, San Diego (USA), 26 February 2002.* SPIE Proceedings Vol. 4685 : 280-287

Guimbretière F. 2003. Paper Augmented Digital Documents. In *Proceedings of 16th Annual ACM Symposium on User Interface Software and Technology (UIST 2003), Vancouver (Canada), 3-5 November 2003*. New York : ACM Press. 51-60

Hafner J., Sawhney H.S., Equitz W., Flickner M., and Niblack W. 1995. Efficient color histogram indexing for quadratic form distance function. In *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Washington: IEEE Computer Society. Vol. 17(7) : 729-736

Jain R. 1996. Infoscopes: Multimedia information system. In B. Furht, (ed.), *Multimedia Systems and Techniques*. Norwell: Kluwer Academic Publishers. 217-253

Keysers D., Deselaers T., and Breuel T.M. 2007a. Optimal Geometric Matching for Patch-Based Object Detection. In *Electronic Letters on Computer Vision and Image Analysis* Vol. 6(1) : 44-54

Keysers D, Deselaers T., Gollan C., and Ney H. 2007b. Deformation Models for Image Recognition. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Washington: IEEE Computer Society. Vol. 29(8) : 1422-1435

Liao C., Guimbretière F., and Hinckley K. 2005. PapierCraft: A Command System for Interactive Paper. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST 2005), Seattle (USA), 23-26 October 2005*. New York : ACM Press. 241-244

Lladós J., Valveny E., Sánchez G. and Martí E. 2001. Symbol recognition: current advances and perspectives. In *GREC '01: selected papers from the fourth international workshop on graphics recognition algorithms and applications, Kingston (Canada), 7-8 September 2001*. Lecture Notes in Computer Science 2390. London : Springer. 104-127

Lowe D.G. 1999. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision (ICCV'99).* Washington: IEEE Computer Society. Vol. 2(2) : 1150

Mlivoncic M., Schuler C., Tu□rker C. 2004. Hyperdatabase Infrastructure for Management and Search of Multimedia Collections. In M. Agosti, H.-J. Schek and C. Tu□rker (eds.), *Digital Library Architectures: Peer-to-Peer, Grid, and Service-Orientation. Proceedings of the 6th Thematic Workshop of the EU Network of Excellence DELOS. Revised Selected Papers. S. Margherita di Pula (Italy), 24-25 June 2004*. Lecture Notes in Computer Science Vol. 3664. Berlin-Heidelberg: Springer.

MPEG-7. 2002. *Multimedia content description interfaces. part 3: Visual*. SO/IEC 15938-3:2002.

Müller H., Michoux N., Bandon D., and Geissbuhler A. 2004. A review of content–based image retrieval systems in medical applications - clinical benefits and future directions. In *International Journal of Medical Informatics*, Vol. 73(1) : 1-23

Niblack W., Barber R., Equitz W., Flickner M., Glasman E.H., Petkovic D., Yanker P., Faloutsos C., and Taubin G. 1993. QBIC project: querying images by content, using color, texture, and shape. In W. Niblack (ed.), *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases, San Diego (US), 31 January - 5 February 1993*, SPIE Proceedings Vol. 1908 : 173-187

Norrie M.C., Signer B., and Weibel N. 2006. General Framework for the Rapid Development of Interactive Paper Applications. In *Proceedings of 1st International Workshop on Collaborating over Paper and Digital Documents (CoPADD 2006), Banff (Canada), 4 November 2006*. 9-12

Rubner Y., Tomasi C., and Guibas L.J. 2000. The Earth Mover's Distance as a Metric for Image Retrieval. In *International Journal of Computer Vision* Vol. 40(2). Berlin-Heidelberg: Springer. 99-121

Rui Y., She A., and Huang T. 1996. Automated region segmentation using attraction based grouping in spatial-color-texture space. In *Proceedings of 3rd IEEE International Conference on Image Processing (ICIP 1996), Lausanne (Switzerland), 16-19 September 1996*. 53-56

Rui Y., Huang T. S, and Chang S.F. 1999. Image retrieval: Current techniques, promising directions and open issues. In *Journal of Visual Communication and Image Representation*, 10(1) : 1-23

Schek, H.-J. and Schuldt, H. 2006. DelosDLMS - Infrastructure for the Next Generation of Digital Library Management Systems. In: ERCIM News No. 66, Special Issue on European Digital Library, pages 22-24, July 2006.

Schuler C., Schuldt H., Türker C., Weber R. and Schek, H.-J. 2005. Peer-to-Peer Execution of (Transactional) Processes. In International Journal of Cooperative Information Systems (IJCIS), 14(4): 377-405.

Sellen A.J. and Harper R. 2002. *The Myth of the Paperless Office*. Cambridge : MIT Press

Signer B., Kurmann U., and Norrie M.C. 2007a. iGesture: A General Gesture Recognition Framework. In *Proceedings of 9th International Conference on Document Analysis and Recognition (ICDAR 2007), Curitiba (Brazil), 23-26 September 2007*. Washington: IEEE Computer Society

Signer B. and Norrie M.C. 2007b. As We May Link: A General Metamodel for Hypermedia Systems. In *Proceedings of 26th International Conference on Conceptual Modeling (ER 2007), Auckland (New Zealand), 5-9 November 2007*. Lecture Notes in Computer Science 4801. Springer

Smeulders A.W.M., Worring M., Santini S., Gupta, and Jain R., 2000. Content-based image retrieval at the end of the early years, In *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Washington: IEEE Computer Society. Vol. 22 (12) : 1349-1380.

Stricker M.A. and Orengo M. Similarity of color images. In *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases, San Diego (US), 5-10 February 1996*, SPIE Proceedings Vol. 2670 : 381-392

Stricker M.A. and Dimai A. 1996. Color Indexing with Weak Spatial Constraints, In *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases,, San Diego (US), 28 January - 2 February 1996*, SPIE Proceedings Vol. 2670 : 29-40

Springmann M., Bischofs L., Fischer P., Schek H.-J., Schuldt H., Steffens U., and Vogl R. 2007. Management of and Access to Virtual Electronic Health Records. In C. Thanos and F. Borri (eds.) *DELOS Conference on Digital Libraries Working Notes, Pisa (Italy), 13-14 February 2007.* 431-439

Springmann M. and Schuldt H. 2007. Speeding up IDM without degradation of retrieval quality. In A. Nardi and C. Peters (eds.), *Working Notes of the CLEF Workshop, Budapest (Hungary), 19-21*,

Veltkamp R.C. and Tanase M. 2000. *Content-Based Image Retrieval Systems: A Survey*. Revised and extended version of Technical Report UU-CS-2000-34, October 2000.

Vendrig J., Worring M., and Smeulders A.W.M. 1999. Filter image browsing: Exploiting interaction in image retrieval. In *Proceedings of the Third International Conference on Visual Information and Information Systems (VISUAL '99), Amsterdam (NL), 2-4 June 1999*. Lecture Notes in Computer Science 1614. Berlin-Berlin-Heidelberg: Springer 147–154

Weber R., Schek H.-J., and Blott S. 1998. A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces. In *Proceedings of the 24rd International Conference on Very Large Data Bases*, New York (USA), 24-27 August 1998. San Francisco : Morgan Kaufmann. 194–205

Weber R. and Mlivoncic M. 2003. Efficient region-based image retrieval. In *Proceedings of the twelfth ACM International Conference on Information and Knowledge Management (CIKM '03), New Orleans (USA), 2-8 November 2003.* New York : ACM Press. 69-76

Weibel N., Norrie M.C., and Signer B. 2007. A Model for Mapping between Printed and Digital Document Instances. In Proceedings of ACM Symposium on Document Engineering (DocEng 2007), Winnipeg (Canada), 28-31 August 2007. New York : ACM Press.

Wellner P. 1993. *Interacting with Paper on the DigitalDesk*. In Communications of the ACM Vol. 36(7) : 87-96

Witten I.H., Moffat A., and Bell T.C. 1999. *Managing Gigabytes (2nd edition): Compressing and indexing documents and images*. San Francisco : Morgan Kaufmann Publishers Inc.

Wolf C.G.. 1986. Can People Use Gesture Commands?, In *IBM Research Report (RC 11867)*

Wolf C.G. and Morrel-Samuels P.. 1987. The Use of Hand-drawn Gestures for Text Editing. In *International Journal of Man-Machine Studies* Vol. 27(1) : 91-102

Wolf C.G., Rhyne J.R., and Ellozy H.A. 1989. The Paper-Like Interface. In M.J. Smith and G. Salvendy (eds.), *Proceedings of the Third International Conference on Human-Computer Interaction, Boston (USA), 18-22 September 1989*, Hillsdale : Erlbaum. 494-501

Yeh R.B., Liao C., Klemmer S.R., Guimbretière F., Lee B., Kakaradov B., Stamberger J., and A. Paepcke. 2006. ButterflyNet: A Mobile Capture and Access System for Field Biology Research. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (ACM CHI 2006), Montréal (Canada), 24-27 April 2006*, New York : ACM Press. 571-580