

# Building and Evaluating Systems for Cross-Language Image Retrieval

Paul Clough

University of Sheffield  
Western Bank  
Sheffield, UK

p.d.clough@sheffield.ac.uk

Henning Müller

Medical Informatics Service  
Geneva University and Hospitals,  
Geneva, Switzerland

Henning.mueller@sim.hcuge.ch

## ABSTRACT

In this paper we discuss three systems for image retrieval: two systems (Eurovision and CiQuest) using abstracted features defined by associated text; the other system (medGIFT) providing content-based access to image data by visual features. For text-based access, providing multilingual support can enable users to access images that might otherwise be unavailable to them. Providing information access to image repositories is best supported through a combination of visual and text-based approaches. To facilitate research and development in image retrieval, we have instigated a large-scale evaluation campaign called ImageCLEF, a track of the Cross-Language Evaluation Forum (CLEF).

## 1. INTRODUCTION

Content-based visual information retrieval (CBVIR), or content-based image retrieval (CBIR), has been studied intensively in the field of computer vision in recent years. Images can be retrieved using primitive features based on pixels which form the contents of an image using, for example, a visual exemplar. However, still no breakthrough has been achieved with respect to large and varied databases and one problem is the loss in semantic information when visual features are extracted automatically. An alternative method of access is through associated texts semantically related to the image (e.g. textual captions or meta-data). Many image repositories have this information (e.g. Web image collections, stock photographic databases, medical image repositories and historic photographic collections) enabling the user simple text-based access. Indeed, because of the success of text-based methods techniques to automatically assign keywords or categories to images is also an intense research area. Further approaches attempt to combine both text and visual methods for searching and browsing image collections.

One area of recent interest in the field of information retrieval (IR) is cross-language retrieval (CLIR) that provides multilingual access to document collections. Given that images are “language-independent”, ideally the language used to express the associated texts or textual queries should not affect the success of retrieval, i.e. an image with a caption written in English should be searchable in languages other than English. Providing multilingual access to image repositories offers benefits to both the owners and end-users alike: enabling wider access to images which might otherwise be un-searchable.

We present three image retrieval systems: Eurovision and CiQuest designed to access a collection of historic photographs; medGIFT designed to access medical image databases. Evaluation of such systems is also important and we present ImageCLEF, a large-

scale evaluation held at the Cross Language Evaluation Forum (CLEF<sup>1</sup>).

## 2. EUROVISION

The Eurovision system [1] demonstrates how multilingual access has been provided to a digital library at St Andrews University Library<sup>2</sup> (Scotland) consisting of historic photographs. Like many historic image collections, the photographs are accompanied by structured captions enabling text-based access. The system makes use of online translation tools to translate the user's search request (query translation), the user interface and text associated with results into the user's search language. Although a basic CLIR interface, this demonstrates what can be done with Machine Translation (MT), as well as where further improvements seem possible.

The main challenges of providing access to a historic collection like the St Andrews Photographic Archive are (1) that captions and users' search requests are typically short in length, thereby offering limited context for both retrieval and translation. (2) The images are varied in content and quality; they are mostly black and white, which limits the effectiveness of using colour as a visual feature. (3) Associated captions frequently contain text which is not directly associated with the visual contents of an image (e.g. expressing something in the background) and can cause erroneous or unexplained search results. (4) The St Andrews collection makes use of colloquial and domain-specific language in the captions causing vocabulary mismatch between queries and documents, and problems with translation. (5) Filtering out images which contain query terms but are not judged relevant is important (e.g. the image is too dark or the subject of the query is not clearly visible).

## 3. CIQUEST

Organising a set of documents automatically based upon a set of categories (or concepts) derived from the documents themselves is an obviously appealing goal for IR systems: it requires little or no manual intervention (e.g. deciding on thematic categories) and, like unsupervised classification, depends on natural divisions in the data rather than pre-assigned categories (i.e. requires no training data). Concept hierarchy generation is one such method: it automatically associates terms extracted from a document set and organises them into a hierarchy, each term representing a group of documents.

---

<sup>1</sup> <http://www.clef-campaign.org/>

<sup>2</sup> <http://specialcollections.st-and.ac.uk/photcol.htm>

The CiQuest system is an implementation of concept hierarchies for information access. Previously, concept hierarchies have been studied within the context of retrieving texts and results have shown that automatically generating hierarchies can help users with their search task. In this demonstration, we show the CiQuest system applied to the St Andrews collection of historic images to organize a set of retrieved images by their associated caption text [2]. The version of CiQuest we present is a multilingual version [3] based on using free online translation tools showing how, like Eurovision, a naïve cross-language information system can be easily constructed with minimal technical effort.

#### 4. MEDGIFT

In the medical field, images (especially digital images) are produced and used for diagnostics and therapy in large quantities. The Radiology Department of the University Hospitals of Geneva alone produced more than 35,000 images per day in 2005. With this quantity of visual data produced, search and navigation using text becomes a problem because of little or incomplete available annotation. Content-based image retrieval methods are one way to overcome such problems in the medical domain (as well as more general domains). The medGIFT project<sup>3</sup> is specialising in applying different approaches to image retrieval in the medical domain. Although the system (at present) is purely visually-based, it demonstrates how visual methods are successful in restricted domains. The system requires an example image or set of images which could be generated from a previous text-based search.

Image browsing is designed for teaching and research purposes and uses very heterogeneous databases (in terms of content and annotation). A variety of techniques are being applied to this problem, including multimodal retrieval (text, visual and structured data) and pre-processing such that the search is concentrated on the most “significant parts” of the images. For a diagnostic aid, systems require much more specialised approaches and medical knowledge must be included into these systems to obtain accurate and useful results. The demonstration system<sup>4</sup> (a web-based interface) shows what can be achieved with purely visual-based retrieval and includes varied techniques for positive and negative relevance feedback.

#### 5. IMAGECLEF

ImageCLEF<sup>5</sup> conducts evaluation of cross-language image retrieval and is run as part of the Cross Language Evaluation Forum (CLEF) campaign. The ImageCLEF retrieval benchmark has previously run in 2003 and 2004 with the aim of evaluating image retrieval from multilingual document collections. ImageCLEF 2005 [4] provided tasks for system-centered evaluation of retrieval systems in two domains: historic photographs and medical images. These domains offer realistic (and different) scenarios in which to test the performance of image retrieval systems and offer different challenges and problems to participants. This benchmark aims to evaluate image retrieval from multilingual document collections and a major goal of ImageCLEF is to investigate the effectiveness of combining

text and image for retrieval. This aims to promote the exchange of ideas which may help improve the performance of future image retrieval systems. ImageCLEF has already seen participation from both academic and commercial research groups worldwide from communities including: Cross-Language Information Retrieval (CLIR), Content-Based Image Retrieval (CBIR), medical information retrieval and user interaction.

Participants are provided with the following: image collections, representative search requests (expressed by both image and text) and relevance judgments indicating which images are relevant to each search request. Campaigns such as CLEF and TREC have proven invaluable in providing standardised resources for comparative evaluation for a range of retrieval tasks and ImageCLEF aims to provide the research community with similar resources for image retrieval. The most recent evaluation campaign ran four tasks: ad-hoc retrieval (find as many relevant images as possible from a static collection given a users’ query) from historic photographs, ad-hoc retrieval from medical images, the automatic annotation of medical images and user-centered retrieval of images, also from a historic photographic collection.

#### 6. CONCLUSIONS

We present three systems for image retrieval that mainly rely on associated textual metadata for searching and browsing image collections (the medGIFT system demonstrating the need for and success of purely visual retrieval in a specialised domain). Given that many images can be associated with text, this enables simple and effective methods for accessing general image collections (although not without problems). Associated text also enables systems to provide multilingual access through the use of translation. The ImageCLEF evaluation campaign provides researchers with a framework in which to evaluate image retrieval systems, particularly text-based systems and multilingual access.

#### 7. REFERENCES

- [1] Clough, P. and Sanderson, M., User Experiments with the Eurovision Cross-Language Image Retrieval System, In Journal of the American Society for Information Science and Technology (JASIST), In Print (expected publication 2006).
- [2] Clough, P., Joho, H. and Sanderson, M. (2005), Automatically Organising Images using Concept Hierarchies, In Proceedings of SIGIR2005, Workshop: Multimedia Information Retrieval, August 15-19, 2005, in Salvador, Brazil.
- [3] Petrelli, D. and Clough, P. (2005) Concept Hierarchy across Languages in Text-Based Image Retrieval: A User Evaluation, In working notes of CLEF workshop, Vienna, Austria, 21-23 September 2005.
- [4] Clough, P., Müller, H., Hersh, W., Deselaers, T., Lehmann, T. and Grubinger, M. (2005), Overview of the 2005 cross-language image retrieval track (ImageCLEF), In working notes of CLEF workshop, Vienna, Austria, 21-23 September 2005.

---

<sup>3</sup> [http://www.sim.hcuge.ch/medgift/w01\\_Presentation\\_EN.htm](http://www.sim.hcuge.ch/medgift/w01_Presentation_EN.htm)

<sup>4</sup> <http://vipier.unige.ch/~muellerh/demoCLEFmed/index.php>

<sup>5</sup> <http://ir.shcf.ac.uk/imageclef/>