Searching and Organizing Images Across Languages

Paul Clough
University of Sheffield
Western Bank Sheffield,
UK +44 114 222 2664
p.d.clough@sheffield.ac.uk

Mark Sanderson
University of Sheffield
Western Bank Sheffield, UK
+44 114 222 2664
m.sanderson@sheffield.ac.uk

Xiao Mang Shou
University of Sheffield
Western Bank Sheffield, UK
+44 114 222 2677
x.m.shou@sheffield.ac.uk

Abstract

With the continual growth of users on the Web from a wide range of countries, supporting such users in their search of cultural heritage collections will grow in importance. In the next few years, the growth areas of Internet users will come from the Indian sub-continent and China. Consequently, if holders of cultural heritage collections wish their content to be viewable by the full range of users coming to the Internet, the range of languages that they need to support will have to grow. This paper will present recent work conducted at the University of Sheffield (and now being implemented in BRICKS) on how to use automatic translation to provide search and organisation facilities for a historical image search engine. The system allows users to search for images in seven different languages, providing means for the user to examine translated image captions and browse retrieved images organised by categories written in their native language.

1. Introduction

The continuous growth of Internet provides great potentials for people from different locations to share information. The need for information organisers to expand their user numbers across language borders also grows. Typically, cultural heritage collection holders will like to increase user numbers to access the data, regardless of languages of the users. To enable this, providing multilingual access to the collections becomes critical.

Currently a great deal of research is being conducted in the field of Cross Language Information Retrieval (CLIR), where documents written in one language are retrieved by a query written in another language. So far query translation, which transforms a user's query into the language of the documents is the dominating approach because this can be made to work successfully with simple translation methods and does not require the overhead of translating collection documents which is often computationally expensive. With the right approach, CLIR systems are able to achieve retrieval effectiveness that is only marginally degraded from the effectiveness achieved had the query been manually translated (Ballesteros & Croft, 1998). Another approach in CLIR is concentrated on means of presenting the retrieved documents in some surrogate form such that users can judge relevance without having access to a full translation (Resnik, 1997; Gonzalo & Oard, 2003).

One area of CLIR research that has received less attention is retrieval from image collections where text is only used to describe the collection objects such as image captions or metadata, and the object's relevance to a query are clear to anyone regardless of their foreign language skills. With the text information, images can then be retrieved using standard IR methods based on textual queries. Here CLIR offers the opportunity of broadly expanding the range of potential searchers to an image archive through multilingual access.

Retrieval from an image collection offers distinct characteristics from one in which the document to be retrieved is natural language text (Armitage & Enser, 1997; Goodrum, 2000). Methods of image retrieval are typically based on visual content (e.g. colour, shape, spatial layout and texture), or by text/metadata associated with the image. Retrieval from such an archive presents a number of

challenges and opportunities. The challenges come from matching queries to the relatively short descriptions associated with each image. Opportunities come from the unusual situation for CLIR systems of users being able to easily judge images for relevance.

This paper is divided as follows: in section 2 we describe some previous work on cross-language image retrieval, section 3 introduce Eurovision, a text-based system for cross-language (CL) image retrieval (Clough & Sanderson, 2006) and section 4 our summary and conclusions.

2. Previous Work

In this section we briefly review two main component research areas: image retrieval by associated text and cross language IR; followed by a description of the occasional attempts at performing image CLIR.

2. 1 Image retrieval via associated text

Retrieval of images by text queries matched against associated text has been long researched, and approaches tend to reflect the nature of the application being addressed and the structure of the image collection.

Chen et al. (1999) presented a method of multi-modal caption retrieval involving both content-based and textual-based modalities, enabling images in Web collections to be browsed. Other approaches have focused on extracting grammatical relations from the captions in order to describe image content. Although in the past it has been shown that Natural Language Processing (NLP) is detrimental to retrieval from documents (Smeaton, 1997), there are some applications which involve short texts, e.g. image captions, where the situation becomes different and NLP can help improve retrieval (Flank, 1998, Elworthy et al. 2001). Given the typically small caption lengths, attention has also been given to expanding the query using lexical resources to reduce the effects of mismatch between query and caption words (Smeaton and Quigley, 1996).

2.2 Cross-Language Information Retrieval (CLIR)

Most research around CLIR has concentrated on locating and exploiting translation resources. CLIR is basically a combination of machine translation and traditional monolingual IR and four approaches commonly used for translation include (Gollins, 2001): (1) a controlled vocabulary, (2) machine translation, (3) bilingual parallel corpora, and (4) bilingual dictionaries. Schäuble and Sheridan (1997) suggest that one can translate the query in the source language into the target language, translate each document in the collection into the same language as the query, or translate both queries and documents into an intermediate representation. The effectiveness of retrieval is based on translation and some of the problems arise from (Flank, 2000): (1) the bilingual dictionary may not contain specialised vocabulary or proper names; (2) dictionary terms are ambiguous and can add extraneous terms to the query and (3) the effective translation of multi-word concepts such as phrases. Other problems include lexical ambiguity of words in both the target and source languages. Further information can be found in (Grefenstette, 1998; Ballesteros & Croft, 1998).

2.3 Cross-Language Image Retrieval

There are five typical examples of cross-language image retrieval exist today:

- The IR Game system built at Tampere University (Sormunen, 1998) offers Finish/English cross-language image retrieval from an image archive.
- The European Visual Archive (EVA) (http://www.eva-eu.org) offers English/Dutch/German cross language searching of 17,000 historical photographs.
- Flank (2000) presents a method for accessing 400,000 photographic

images from a commercial image company called PictureQuest (http://www.picturequest.com).

- In 2002, the ImageCLEF track of CLEF (Cross Language Evaluation Forum) was established with the release of one of the first publicly available test collections for cross language image retrieval: approximately 30,000 photographic images from a collection held at St. Andrews University Library and fifty queries (Clough & Sanderson, 2003).
- Sanderson et al., (2004) confirmed the feasibility of an image CLIR system using a test collection study for German and Portuguese.

3. The Eurovison System

The Eurovision system combines existing translation resources to provide Web-access to an image archive provided by St. Andrews University Library. Most image captions contain a number of textual fields which can be exploited during image retrieval. Although our retrieval system is built to search the St. Andrews collection, the information contained in captions for this collection can be found in most annotated pictures, e.g. a description of the image, its author and some kind of categorization. The following sections describe the architecture/interface, and translation.

3.1 The St. Andrews Image Collection

A collection of historic photographs from St. Andrews University Library (Reid, 1999) was used as the dataset for system development and evaluation. This dataset was used because: (1) it represents a real-world collection for which multilingual access can enhance its access, (2) it contains almost 30,000 images with captions of high quality generated by historians, (3) it is being used in ImageCLEF, and (4) the varied quality and content of the collection makes CLIR access a challenge.

All images in the St. Andrews collection are accompanied by a caption consisting of eight distinct fields which can be used individually or collectively to facilitate image retrieval: *Recode ID, Short title, Long title, Location, Description, Data, Photographer, Categories and Notes*. The captions consist of 44,085 terms and 1,348,474 word occurrences; the average caption length is 48 words, with a maximum of 316. All captions are written in British English; they contain colloquial expressions and historical terms. Approximately 81% of captions contain text in all fields, the rest generally without the description field. In most cases the image description is a grammatical sentence of around 15 words and the majority of images (82%) are black and white.

Like many image collections, pictures in the St. Andrews collection have been annotated manually by domain experts (historians). Part of the process is to assign images to one or more pre-defined categories for image storage and management. There are 971 categories in the St. Andrews collection, some more general (e.g. “flowers”, “landscapes”) than others (e.g. names of geographic regions and photographers). Most images are assigned to 3-4 categories.

3.2 Architecture and Interface

The interface is Web-based and generated dynamically using Perl/CGI scripts and JavaScript. A simple modular architecture enables new functionality to be added with relative ease, e.g. changing the translation resource. Users log into the system and can begin by entering search requests in English, French, German, Italian, Spanish, Simplified Chinese or Japanese. Queries are passed our own in-house probabilistic retrieval system, based on the “best match” BM25 weighting operator (Robertson et al., 1998). Images are indexed by a number of caption fields, e.g. title, description and set of manually assigned categories. The default settings of case normalization, removal of stopwords

and word stemming are used and a document ranking scheme used where captions containing all query terms are ranked highest by their BM25 score, and then all other captions containing at least one query term ranked below.

Results are presented as a 5 x 4 grid of thumbnails with titles from the captions. Users can browse the results one page at a time, or re-iterate their query. The pre-assigned categories are used to help users navigate the results set and find similar relevant images.

An alternative to the grid display of images is to view returned images by their categories. This provides a summary of the results indicating their contents as described by image categories and can be a more efficient method of searching than viewing images one page at a time. Three methods for viewing these categories are offered to the user: ranked in ascending order by the number of images assigned to each category, alphabetically, and hierarchically. Categories are organised automatically into a hierarchical structure (Clough, Joho & Sanderson 2005) using a co-occurrence relation called *subsumption*, proposed by Sanderson and Croft (1999) for IR. Up to four levels are generated with more frequent/dominant categories ordered at the top of the hierarchy leaving more specific categories nearer the bottom (e.g. "horses and ponies" > "farm implements" > "farming – ploughing"). The user can browse the search results by either navigating through pages of image thumbnails, or viewing the lists of categories.

Similar views of the image categories can also be obtained when browsing, this time generated from the entire collection rather than the results of a search. By listing the categories, the user is able to obtain an overview of the contents of the collection. For example, displaying the categories by the frequency of images assigned to that category shows the most dominant.

3.3 Translation using SYSTRAN

Query, interface and document translation is provided by SYSTRAN (http://www.systransoft.com), one of the oldest and most widely used free on-line Machine Translation (MT) systems (Hutchins & Somers, 1986; Systran, 2002). Cross-language queries are translated into English and passed to the retrieval system in which the English captions have been indexed. Results are displayed as English Web pages and translated dynamically as users interact with the system by calls to SYSTRAN. This method translates the flat and hierarchical category lists, the image captions, and the whole interface in the user's source language. Most un-translated terms are proper names which are either not found in the SYSTRAN dictionaries, or would not commonly be translated from English even manually. The quality of SYSTRAN varies across language due to a range of translation errors (see, e.g. Qu et al. (2000)). For short queries of 2-3 words, SYSTRAN is essentially used for dictionary-lookup as they carry little grammatical structure to help the MT algorithm.

4. Summary and Future Work

In this paper we have discussed an area of CLIR research which to date has received little attention, that of CL image retrieval. We have presented Eurovision, a system for searching image collections by matching user's queries to associated captions. A multilingual search environment is created for the user without any knowledge of a language other than English by using the SYSTRAN MT system to translate user queries, image captions and the interface. We use a collection of historic photographs as a representative dataset and exploit pre-defined categories to enable users to browse images and view the results of a search by category. We also implement a version of concept hierarchies to re-organise the categories into a hierarchical structure to reduce the number of categories users have to view.

For two search tasks conducted on

Eurovision: known-item and category searching, we find that although the absolute success of CL retrieval is poor (43% success), relative to searching in English the system Eurovision operates at 86% monolingual (89% for the known item search and 83% for the ad hoc search). We believe that this high performance figure is a result of the nature of image retrieval: users do not have to view the image caption to judge relevance and they are willing to view many pages of images during retrieval. We also believe that providing browsing through the categories also leads to this degree of success because users have an alternative search method to use when their queries appear to fail. Given retrieval success it would appear that translation of the categories and interface is good enough to enable users to browse with success.

As a CLIR task, image retrieval is one application where even poor translation resources can still achieve good performance. Our experiments have confirmed previous work that image retrieval via associated text is possible, although both English and CL searching could be improved greatly. Plans to improve the Eurovision system include:

- We are planning a larger user evaluation (involving 16-32 users) to compare different user interfaces for cross-language image retrieval.
- We plan to investigate query translation using alternative methods such as bilingual dictionaries and parallel corpora.
- We plan to investigate the use of query expansion through external resources such as a thesaurus, and based on relevance feedback from the user to help reduce vocabulary mismatch.
- We would like to experiment with alternative methods of generating categories for the images automatically (e.g. clustering captions and extracting dominant concepts).
- We are investigating generating the concept hierarchy based on the entire image caption rather than just the categories, as users are frequently uncertain about the categories. Initial results appear promising.

References

Armitage, L. H., & Enser, P. (1997). Analysis of User Need in Image Archives. *Journal of Information Science, 23*(4), 287-299.

Ballesteros, L., & Croft, B. W. (1998). Resolving Ambiguity for Cross-Language Retrieval. *Proceedings of the 21st International Conference on Research and Devlelopment in Information Retrieval*, 64-71.

Chen, F., Gargi, U., Niles, L., & Schütze, H. (1999). Multi-Modal Browsing of Images in Web Documents. *Proceedings of SPIE Document Recognition and Retrieval VI*, 122-133.

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).

Clough, P., Joho, H. and Sanderson, M. (2005), Automatically Organising Images using Concept Hierarchies, *Workshop held at the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Workshop: Multimedia Information Retrieva,* August 15-19, 2005, in Salvador, Brazil.

Clough, P. D., & Sanderson, M. (2003). The CLEF 2003 Cross Language Image Retrieval Track. *Working Notes for the CLEF 2003 Workshop.*

Elworthy, D., Rose, T., Clare, A., & Kotcheff, A. (2001). A Natural Language System for Retrieval of Captioned Images. *Journal of*

Natural Language Engineering, 7(2), 117-142.

Flank, S. (1998). A Layered Approach to NLP-Based Information Retrieval. *Proceedings of 36th ACL and 17th COLING Conferences,* 397-403.

Flank, S. (2000). Cross-Language Multimedia Information Retrieval. *Proceedings of Applied Natural Language Processing and te North American Chapter of the Association for Computational Linguistics.*

Gollins, T. (2000). Dictionary Based Transitive Cross-Language Information Retrieval using Lexical Triangulation. *Masters Dissertation.* Department of Information Studies, University of Sheffield.

Gonzalo, J., & Oard, D. (2003). The CLEF 2002 Interactive Track. In *Springer Lecture Notes in Computer Science (LNCS 2785): Vol. . CLEF 2002* (pp. 372-382). Berlin Heidelberg: Springer-Verlag.

Goodrum, A. A. (2000). Image Information Retrieval. *Informing Science, 3*(2), 63-66.

Hutchins, W. J., & Somers, H. (1986). *An Introduction to Machine Translation.* London, England: Academic Press.

Qu, Y., Eilerman, A. N., Jin, H., & Evans, D. (2000). The Effect of Pseudo Relevance Feedback on MT-Based CLIR. *Proceedings of RIAO 2000.*

Reid, N. (1999). The Photographic Collection in St. Andrews University Library. *Scottish Archives, 5,* 83-90.

Resnik, P. (1997). Evaluating Multilingual Gisting of Web Pages. *Proceedings of AAAI Symposium on Cross-Language Text and Speech Retrieval.*

Robertson, S., Walker, S., & Beaulieu, M. (1998). *Proceedings of TREC-7* (NIST Special Publiation 500-242, pp. 253-264).

Sanderson, M., & Croft, B. W. (1999). Deriving Concept Hierarchies from Text. *Proceedings of the 22nd International Conference on Research and Development in Information Retrieval,* 206-213.

Sanderson, M., Clough, P. D., Paterson, C., & Tung Lo, W. (2004). Measuring a Cross Language Image Retrieval System. *Proceedings of the 26th European Conference on IR Research (ECIR'04),* 353-363.

Schäuble, P., & Sheridan, P. (1997). *Proceedings of the 6th Text Retrieval Conference (TREC-6): NIST Special Publications 500-226. Cross Language Information Retrieval (CLIR) Track Overview.*

Smeaton, A. F. (1997). Information Retrieval: Still Butting Heads with Natural Language Processing? *International Summer School on Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology,* 115-138.

Smeaton, A. F., & Quigley, I. (1996). Experiments on Using Semantic Distances Between Words in Image Caption Retrieval. *Proceedings of the 19th International Conference on Research and Development in Information Retrieval,* 174-180.

Sormunen, E., & Laaksonen, J. (1998). The IR Game - A Tool for Rapid Query Analysis in Cross-Language IR Experiments. *Proceedings of PRICAI'98 Workshop on Cross Language Issues in Artificial Intelligence,* 22-32.

Systran Ltd. (2002). The Systran Linguistics Platform: A Software Solution to Manage Multilingual Corporate Knowledge. *Systran On-line Documentation* Retrieved July 29, 2004, from http://www.systransoft.com/Technology/SLP.pdf