Visual Micro-clustering Pre-processing for Cross-Language Ad hoc Image Retrieval

Masashi Inoue
National Institute of Informatics, Tokyo, Japan
m-inoue@nii.ac.jp

Abstract

Images are visual representations. However, when one wants to retrieve them semantically, the visual content of the image is less useful than their textual annotations. When multilingual image collections are considered, there is a possibility of using visual features to overcome the gap between languages. In the ImageCLEF 2006 Photo ad hoc task, the effect of clustering-based pre-processing to enhance the matchabilities of textual queries and images was investigated. In our view, if images are nearly visually identical, then they should be regarded as images of similar relevance, even if they have different annotations.

Micro-clustering pre-processing was employed to implement this functionality. We experimentally investigated the effectiveness of the technique on a linguistically heterogeneous image collection that consisted of English and German-annotated images. In current preliminary setting, the use of micro-clustering for pre-processing did not help in the retrieval for either English or German topics.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval;

General Terms

Measurement, Performance, Experimentation

Keywords

Image retrieval, Visual similarity, Micro-clustering, Pre-processing

1 Introduction

We often suffer from insufficient recall in image retrieval as compared with the retrieval of textual documents. The number of digital images is rapidly increasing, but the number of relevant images is smaller than when dealing with text. The reasons for this are threefold: 1) the generation of digital images is less actively performed by people or organizations than is the production of digital text, 2) most images are not organized in a searchable form, and 3) existing image retrieval functionalities are not very powerful. These three factors are interwoven. In this paper, we propose a method to overcome these problems by expanding the search target from mono-lingual collections to multi-lingual collections with the aid of visual features.

In the case of image retrieval based on textual annotations without any translation, the target collection is limited to the images annotated in the query language. This limits a larger set of

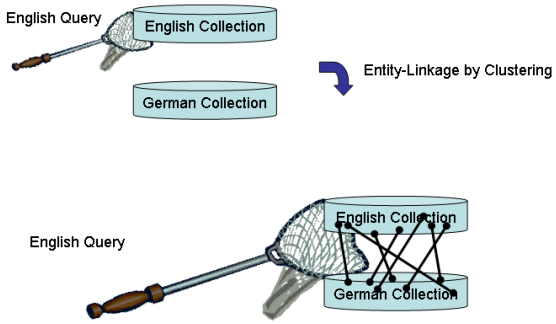

Figure 1: Larger number of images becomes accessible after the micro-clustering pre-processing in cross-language image retrieval.

images because of differences between images on the same topic that are annotated in different languages. For example, when the Web is queried using a search engine with the keyword "sheep" and its Japanese translation, the result will contain some typical pictures of sheep in both search results. In addition to these common images, it may be noticed that the English query retrieves pictures of varieties of breed while the Japanese query finds more pictures related to eating. This suggests that even for the same concept, the available images are different when different languages are used. Cross-language information retrieval (CLIR) techniques may help by expanding the types of images accessible on some topics. Figure 1 illustrates the concept.

The use of visual features is conducted in the framework of a "find similar" task. This procedure can be executed in two ways. First, is finding similar image pairs or groups prior to querying. The second method involves searching for the similar images after the initial result has been retrieved and sometimes makes use of users' feedback. We investigate the first option in this paper. This choice is based on considerations of efficiency. In comparison, similarity calculation between images based on visual features is heavier than that on textual features. Thus, it is usually desirable to conduct such computations off-line. An on-line method of applying the clustering method for image retrieval is by clustering the retrieved results. In the annotation-based image retrieval framework, Chen et al. applied the clustering method but as the post-processing after querying [2].

In the following sections, we first introduce the systems used; particular emphasis is given to the micro-clustering pre-processing. Secondly, we describe the configuration of submitted runs. Thirdly, we show the retrieval results for the submitted runs and additional runs. Furthermore, we analyze these results with discussions. Finally, the paper presents the conclusion.

2 System description

2.1 Retrieval Engine

The novelty of our method in the ImageCLEF2006 is solely in the pre-processing of the retrieval. The core ranking process has been conducted by an existing search engine. We used the Lemur Toolkit as the engine[1]. The Lemur Toolkit is an information retrieval toolkit designed with language modeling in mind. We used a unigram language-modeling algorithm for building the document models, and Kullback-Leibler divergence for the ranking. The document models were smoothed using Dirichlet prior [5].

[1] http://www.lemurproject.org/

2.2 Translation

Translation was applied to queries using the Systran machine translation (MT) system[2]. Because of lack of direct translation functionality between German and Japanese in the MT system, English was used as the pivot language when querying German collections using Japanese topics. That is, Japanese queries were first translated into English, and then the English queries were translated into German.

2.3 Visual Clustering

Two types of clustering can be imagined. One is macro-clustering or global partitioning, when the entire feature space is divided into sub-regions. The other is micro-clustering or local pairing, where data points nearby are linked so that they form a small group in a particular small region of the feature space. Figure 2 shows the schematic difference between these two clustering methods. Micro-clustering was used to group images based on their visual similarities.

The use of the micro-clustering technique has been attempted for text processing where terms play central roles for clustering [1]. In this study, the concept of micro-clustering is used but the features and the similarity measure are different. The process of clustering is as follows. First, visual features are extracted from all images. Simple color histograms are used. Since the images are provided in true color JPEG format, the histograms are created for the red (R), green (G), and blue (B) elements of the images. This results in three vectors for each image: $\mathbf{x}_r$, $\mathbf{x}_g$, and $\mathbf{x}_b$. The length of each vector, or the size of the histogram, $i = 256$. These are concatenated and define a single feature matrix for each image: $X = [\mathbf{x}_r, \mathbf{x}_g, \mathbf{x}_b]$. Thus, the size of feature matrix is i by j where $j = 3$.

Further, the similarities between images are calculated using the above feature values. The similarity measure employed was the two-dimensional correlation coefficient r between the matrices. Assuming two matrices A and B, the correlation coefficient is given as

$$r = \frac{\sum_i \sum_j (A_{ij} - \bar{A})(B_{ij} - \bar{B})}{\sqrt{(\sum_i \sum_j (A_{ij} - \bar{A})^2)(\sum_i \sum_j (B_{ij} - \bar{B})^2)}}$$

where $\bar{A}$ and $\bar{B}$ are the mean values of A and B respectively.

Next, a threshold is set that determines which two or more images should belong to the same cluster. In other words, image pairs whose r score is larger than the threshold are considered identical during retrieval. At this stage, the threshold value is determined manually by inspecting the distribution of similarity scores so that relatively small numbers of images constitute clusters. Small clusters containing nearly identical images are preferred since visual similarity does not correspond to semantic similarity; however, visual identity often corresponds to the semantic identity. Finally, re-ranking of ranked lists given by the retrieval engine is conducted using the cluster information. A ranked list is searched from the top and when an image that belongs to a cluster is found, all other members of the cluster will be given the same score as the highly ranked one. This process is continued until the number of images in the list exceeds the pre-specified number, which is 1000 in our study.

3 Description of the runs submitted

The details of the test collection used are given in [3]. There are 20,000 images annotated in English and in German. Instead of viewing the collection as a single bilingual collection, it is regarded as a 20,000 English collection and a 20,000 German collection. Each annotation has seven fields but only the title and description fields were used.

Six runs were submitted for the initial evaluation. The query languages were English, German, and Japanese. The collection languages are English and German. The relationship between query

[2] http://babelfish.altavista.com/

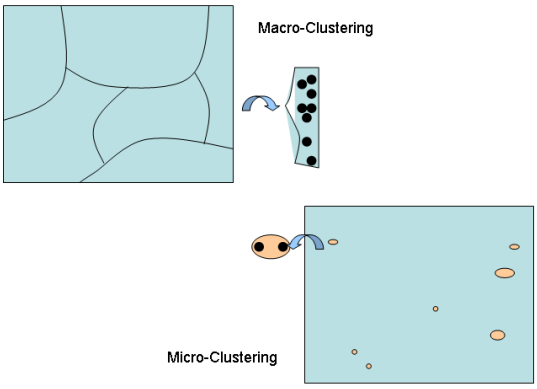

Figure 2: Conceptual difference between Macro- and Micro-clustering. Black dots represent individual data points.

Table 1: A summary of submitted runs (run names and MAP scores).

	Document Language			
Query Language	English		German	
English	mcp.bl.eng_t.eng_td.skl_dir	0.1193	mcp.bl.eng_t.ger_td.skl_dir	0.0634
German	mcp.bl.ger_t.eng_td.skl_dir	0.1069	mcp.bl.ger_t.ger_td.skl_dir	0.0892
Japanese	mcp.bl.jpn_t.eng_td.skl_dir	0.0919	mcp.bl.jpn_t.ger_td.skl_dir	0.0316

and document languages and submitted runs' names is summarized in the Table 1. In the table, names of runs are assigned according to the following rules. The first element *mcp* comes from the proposed method, micro-clustering pre-processing, and represents our group. The second element *bl* indicates that the baseline method was used. When the micro-clustering pre-processing was applied, this element will be the value of the threshold. For example, *09* denotes the pairs with correlation coefficients greater than 0.9 form a cluster. The next element concerns the query language and the fields of the search topics. Runs using English queries with only title fields are marked by *eng_t*. Similarly, the next element is the collection language and the fields of the annotations. Runs using the German collection with title and description fields are marked by *ger_td*. When half of the English collection and half of the German collection are mixed together, the notation is *half_td*, as shown in Table 2. The last element is the configuration of the retrieval engine. The simple Kullback-Leibler divergence measure was used for ranking (*skl*) and Dirichlet prior used for smoothing (*dir*). In all runs, the same configuration was adopted.

4 Results and discussion

4.1 Language dependency

In baseline runs, the collection language is the determining factor in retrieval performance as shown in the table. Searching in English collection is better in any query language. Furthermore, the translated topics from German to English on the English collection worked better than monolingual German topics on the German collection. The results for Japanese topics were poor, because of low-performing machine translation.

Table 2: A summary of runs against the linguistically heterogeneous collection (run names and MAP scores).

Query Language	Half English Half German Document Collection			
	Without pre-processing		With pre-processing	
English	mcp.bl.eng_t.half_td.skl_dir	0.0838	mcp.09.eng_t.half_td.skl_dir	0.0586
German	mcp.bl.ger_t.half_td.skl_dir	0.0509	mcp.09.ger_t.half_td.skl_dir	0.0374

4.2 Use of Images to Bridge Languages

The advantage of the visual-similarity based pre-clustering becomes clear from the application in the linguistically heterogeneous image collections. Therefore, a linguistically heterogeneous collection was constructed by taking $10,000$ randomly chosen images from the English collection and the remaining $10,000$ images from the German collection. There were no overlapping images. Both English queries and German queries without translation were tested on this single collection with micro-clustering. Table 2 shows the result. It was observed that no improvement was given by the pre-processing.

4.3 Clustering Result

As seen in Figure 3, the generated clusters were small and often of size two: a cluster being formed by a pair of images. We intended this by micro-clustering; we have quite small yet highly restricted clusters. The statistics of the size of cluster are as follows: mean = 12.72, standard deviation= 43.81, minimum= 0, median= 368, and maximum= 0. Some clusters have more than 100 members. Such non-micro clusters are not ideal because when one of their members appears in the list, the cluster dominates the entire ranked list after re-ranking. Thus, clusters bigger than 6 were truncated to size 6.

4.4 Discussion

At this point, improvement could not have been achieved by incorporating visual pre-processing. This failure might be because clusters of irrelevant images were used rather than relevant ones. Because not all of initially retrieved images were relevant, some tactics to select only highly relevant images may be needed. Also, there is a trade-off between the quality of clustering and the degree of search target expansion, and the threshold value used may be conservative to avoid the inclusion of noisy clusters. Additional investigation is needed to clarify the effect of threshold values.

The potential advantages of the approach outlined in this study compared to the usual query translation methods are as follows. First, there is no need to combine rankings given by multiple translated queries. Because the rank aggregation is difficult in IR, trial and error in the design of the merging strategies can be avoided. Second, systems do not have to be concerned about the languages. Even when the language distribution within the collection is unknown, the method can be used.

The limitation of our experimental setting should be noted. The test collection is built upon a random selection from two language collections. Thus, near identical images that might be originally created in a sequential manner could have been split into two languages. However, in reality, similar image pairs may exist only in one language. For example, if one photographer takes photos of an object, it is natural to assume that each of them is annotated in the same language. In future studies, more realistic linguistically heterogeneous collections shall be investigated.

5 Conclusion

In this paper, the experimental runs on ImageCLEF2006 ad hoc task have been presented. The goal was to investigate if images annotated in different languages can be searched beyond the

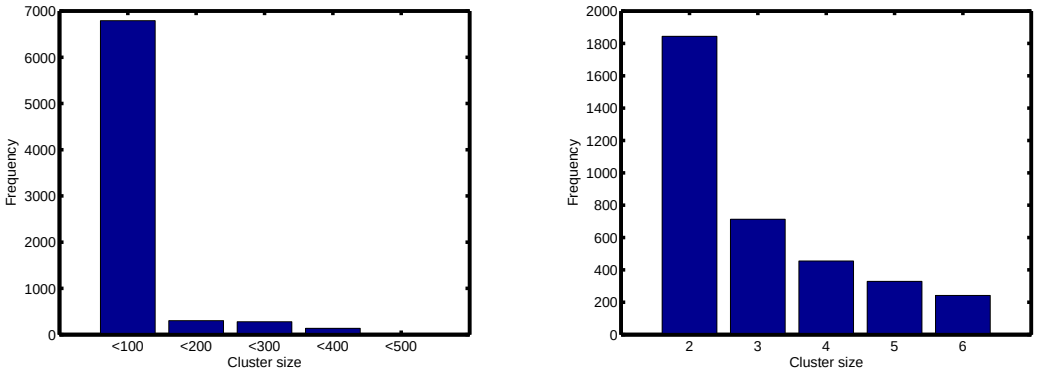

Figure 3: Distribution of cluster sizes. The graph on the right hand side is the close-up of the graph on the left hand side within the rang 2 to 6.

language barriers. A visual feature-based micro-clustering was used for the linkage of near identical images annotated in different languages. After this pre-processing, the retrieval was conducted as a monolingual retrieval. The experiment result does not favor this method.

Cross-language information access technologies have many potential application areas. However, it is not fully understood in what sort of search task they will bring most benefit [4]. Considering the language independent nature of visual representation, cross-language image retrieval can be one such task. Although the work conducted in the present study could not succeed, there could be other possibilities.

References

[1] Akiko Aizawa. A method of cluster-based indexing of textual data. In *Proc. of the 19th Conference on Computational Linguistics (COLING 2002)*, pages 1–7, 2002.

[2] Yixin Chen, James Z. Wang, and Robert Krovetz. CLUE: Cluster-based retrieval of images by unsupervised learning. *IEEE Transactions on Image Processing*, 14(8):1187–1201, Aug. 2005.

[3] Paul Clough, Michael Grubinger, Thomas Deselaers, Allan Hanbury, and Henning Müller. Overview of the ImageCLEF 2006 photo retrieval and object annotation tasks. In *CLEF working notes*, Alicante, Spain, September 2006.

[4] Masashi Inoue. The remarkable search topic-finding task to share success stories of cross-language information retrieval. In *New Directions in Multilingual Information Access: A Workshop at SIGIR 2006*, Seattle, USA, Aug. 2006.

[5] Chengxiang Zhai and John Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Trans. Inf. Syst.*, 22(2):179–214, 2004.