An Application of Multiple Viewpoints to Content-Based Image Retrieval

James C. French A. C. Chapin Worthy N. Martin
Department of Computer Science
University of Virginia
Charlottesville, VA
01-434-982-2213
{french,acc2a,wnm}@cs.virginia.edu

Abstract

Content-based image retrieval uses features that can be extracted from the images themselves. Using more than one representation of the images in a collection can improve the results presented to a user without changing the underlying feature extraction or search technologies. We present an example of this "multiple viewpoint" approach, multiple image channels, and discuss its advantages for an image-seeking user. This approach has also been shown to dramatically improve retrieval effectiveness in content-based image retrieval systems[3].

1. Introduction

Content-based image retrieval (CBIR) systems[5] search collections of images based on features that can be extracted from the image files themselves without manual descriptive or indexing labor from humans. Identifying such features and methods of extracting them are open areas of research. Using multiple image representations, we have been able to improve the results of existing image retrieval systems without developing any such new methods.

Identifying the most useful distinguishing features of images for searching, especially in cases where the image retrieval task is fairly general and not well specified, is difficult, as is extracting these features from the images in a heterogeneous collection. Frequently a query to an image retrieval system will go unanswered or poorly answered even when there are appropriate images in the collection because while the query seems to the user to communicate the information need, the features actually being used by the system to judge relevance do not quite match the user's expectations.

By performing some simple transformations on query and image collection, we can improve the likelihood of such appropriate but obscured images being found without changing the underlying feature abstraction or retrieval technologies. We have built a system with three additional image representations, or *channels*, beyond the native representation used by the underlying CBIR technology. In this system we perform searches on each channel, and merge the results to obtain an improved result.

2. Context

Multiple channels are an application of the idea of multiple viewpoints[2], an approach to information retrieval that takes advantage of having more than one set of relevance judgments to improve search results, all unified under a common interface. Each viewpoint embodies a set of relevance judgments, possibly using different representations and search technologies, and the results of consulting each viewpoint for documents (e.g. images, text articles, library book records, etc.) are merged to create a single result in which the weaknesses of each viewpoint may be mitigated by the contributions of the others.

We consider CBIR systems that are queried "by example," that is, an image is given as the query and the system's task is to find other images "like" the query. What it means for one image to be "like" another depends on the representations and the method of comparing representations to determine relevance – that is, on the viewpoint. By using several viewpoints, we increase the likelihood that the elements of the query that the user is interested in will be among those considered in choosing results. Our approach is analogous to the work of Belkin et al. [1] and Shaw and Fox [4], and others in text retrieval.

3. Image Channels and Merging

Conventional single channel CBIR systems represent a collection of images by descriptive vectors in a high-dimensional feature space with features such as color, shape, and texture drawn from analysis of the images. We extend this by creating several additional representations of the images (both the query, and those in the collection, and using some or all of these in our multiple channel approach to retrieval. Each single channel uses the same feature extraction techniques and search strategy (we use an existing CBIR technology, treating it as a black box) and the results of all the

Figure 1. First 40 results of Basic CBIR approach.

Figure 2. First 10 results of all four channels.

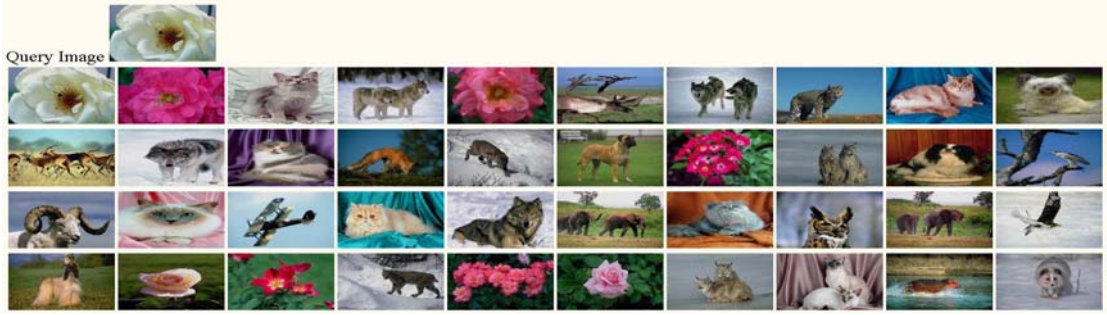

Figure 3. First 40 results of merging grayscale (B+ and B-) channels.

channels are combined to create the multiple channel result.

Whenever more than one search is used together, we must have some way to combine the results sensibly. Merging results can be very difficult, if heterogeneous searches are in use; in our case, since we are using the same CBIR technology for all channels, ranking and similarity measures for the results are comparable. In the results of our experiments in multiple-channel CBIR, we have used a simple similarity-based merge of the top k results of each channel to create a merged result.

However, in presenting the results to a user in an interactive image search, we may choose to present the top k results of each search, in parallel (see Figure 2).

4. Multiple Channels Applied

Our retrieval example employs a basic CBIR technology, using three color and four texture features, which we treat as a black box. The image collection is 3400 images drawn from 34 categories of the COREL image collection. Our ground truth is based on a labeling of salient foreground and background objects in these images, as described in [3]. We use four channels, the original images, their negatives, their grayscale images, and the negative of the grayscale images; each of these, searched using the basic CBIR technology, constitutes a different viewpoint on the collection.

Here we give a single (but characteristic) example of how multiple channels can improve on single-channel searching. The query image is a rose, and we will call the result images relevant if they contain roses. Figure 1 shows the query image and the first forty results given by the single-channel approach. (Since our query image is drawn from the image collection, the first result of the basic IR system is always the query itself.) Figure 2 shows, instead, the first ten results of each of the four channels, in parallel. The line labeled C+ is the results from the original representation, C- those from the negative channel, B+ those from the grayscale channel, and B- from the negative grayscale channel.

The single-channel approach finds only one rose (other than the query) in its first forty results, and this is at rank 25. A user looking at the first ten or twenty

results of the conventional CBIR approach (Figure 1) might be convinced there are no roses in the image collection at all.

This is due to a mismatch in expectations between the user and the system. The user is trying to communicate that the information need is for roses, but the CBIR technology emphasizes the color, and other elements, rather than those the user intends (a different CBIR technology might emphasize different characteristics, but would be subject to the same mismatching). This is the equivalent of a vocabulary mismatch in text-based information retrieval.

The four channel result, shown in Figure 2, shows the user the first ten results of interpreting the query according to four different viewpoints on the same collection. Here there are three roses, two of which are found by both grayscale (B+) and negative grayscale (B-), one of which is the first result in each of those cases.

In presenting the results to the user in parallel, the system gives the user some idea of how the query image is related to the results in the four viewpoints. Thus, a user could observe that the chosen query image effectively communicates characteristics of the information need (roses) in the grayscale cases, but is not so effective in communicating distinguishing characteristics in the color cases. Also, in cases where there is some overlap between the results of different viewpoints this manner of presenting the results allows the user to see how an image may be judged more relevant under one viewpoint than another.

Suppose that the system also gives the user the option of choosing from which viewpoints to see results. Using only the results from the two grayscale channels produces seven roses in the top 40 images of the merged grayscale channels, B+ and B-, as shown in Figure 3.

Alternatively, the system could keep the multiple channels transparent to the user and instead of the four channel display, provide only a single merged result. The top 40 merged results across all four channels for the query above includes all three roses found in the four-channel top-ten. We have investigated this approach thoroughly in [3] and found that we can improve retrieval effectiveness by 22% on average using the merged output.

The choice between making the viewpoints explicit or simply giving the user their merged result depends on an image retrieval system's intended use. If users are most interested in efficient single-step searching, then the merged result is more appropriate; however if the user is willing to spend more time and thought on choosing which viewpoint(s) to use, then showing results of different viewpoints explicitly may be useful. Further, making the viewpoints explicit is appropriate for a class of users who will work extensively with the system, and are aided by knowing what is available and how to take advantages of the different interpretations of a query available through the channels

Another way in which a user may take advantage of information about the viewpoints in use is to choose a new query based on observations of the relation between the query image and the results in each viewpoint. A user observing how results from the color channels seem related more by broad areas of color than by details of texture might decide that the white rose is not an effective query to the color channel viewpoints because so many other kinds of images have a profusion of white around a dark center; however, few landscapes or wild animals exhibit an abundance of bright pink, so a pink rose may be a better query for the color channels. Choosing a new query may also be done automatically and sent to all channels. Such feedback approaches can enhance search effectiveness.

5. Conclusions

We have described a simple strategy for improving the effectiveness of CBIR systems by simply indexing several transformations of the underlying data. The cost per channel is simply another index and is, therefore, modest. The approach lends itself to a variety of interface approaches from multi-channel browsing interfaces to more efficient merged output streams.

6. Acknowledgement

This work sponsored in part by NASA grant NAG5-12025 and by the National Science Foundation. The opinions expressed here are those of the authors, not the sponsors.

7. References

[1] Belkin, N., Kantor, P., Cool, C. and Quatrain, R. Combining Evidence for Information Retrieval in *Proceedings of Trec-2* (March 1994), 35-44.

[2] French, J. C., Chapin, A. C., Martin, W. N., Multiple Viewpoints as an Approach to Digital Library Interfaces. *Workshop on Document Search Interface Design and Intelligent Access in Large-scale Collections*, July 2002.

[3] French, J. C., Watson, J.V.S., Jin, X and Martin, W. N. Using Multiple Image Representations to Improve the Quality of Content-Based Image Retrieval. Technical Report CS-2003-10, Dept. of Computer Science, Univ. of Virginia, March, 2003.

[4] Shaw, J. and Fox, E. Combination of Multiple Searches in *Proceedings of TREC-3* (April 1995) 105-108.

[5] Yoshitaka, A., and Ichikawa, T. A Survey on Content-Based Retrieval for Multimedia Databases, *IEEE Transactions on Knowledge and Data Engineering*, 11(1), 81-93.