# AN EXOGENOUS APPROACH FOR ADDING MULTIPLE IMAGE REPRESENTATIONS TO CONTENT-BASED IMAGE RETRIEVAL SYSTEMS

James C. French James V. S. Watson X. Jin W. N. Martin
Department of Computer Science, University of Virginia
{french|jvw3n|xj3a|martin}@virginia.edu

## ABSTRACT

Content-based image retrieval (CBIR) uses features that can be extracted from the images themselves. In previous work we have shown that 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[4]. In this paper we show that we can also merge the results of multiple CBIR systems to achieve even greater retrieval effectiveness again without changing the underlying CBIR technology. We also present an example of this combined approach and show that it can dramatically improve retrieval effectiveness in content-based image retrieval systems.

## 1. INTRODUCTION

Content-based image retrieval (CBIR) systems [1,6] 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[4].

The central strategy in our approach is to provide a diversity of representations and search strategies to produce several intermediate results that we can merge into a more effective retrieval result. Our intial work considered a diversity of representations; the current paper extends that work to consider additional search strategies. Both techniques result in substantial improvements in retrieval effectiveness and the combination of techniques is even more impressive.

Our work is analogous to the work in text IR on combination-of-evidence strategies that dates back to the early 90's. Two approaches have generally been used. In the first approach a diversity of queries is used to capture an information need more precisely. The several queries can be combined before searching, or issued individually and the results of each query merged afterwards. The work of Belkin et al.[2] adopts this approach.

The second strategy is to use a diversity of representations, that is, to create several indexes over the same corpus of documents. The typical strategy is to index the corpus with the same technology varying indexing parameters, or to index the corpus with different technologies. Queries are processed in each setting with the results being merged afterwards. The work of Fox and Shaw[3] adopts this strategy. The first approach we adopted for extending CBIR systems to combine multiple evidence was to use a single CBIR technology with multiple image representations[4]. The next approach was to consider the use of multiple CBIR technologies[5]. This paper describes our framework more fully and reports on new experiments corroborating our earlier findings.

## 2. MULTIPLE VIEWPOINT SYSTEMS

In [4] we introduced our channel concept as a means of introducing alternative image representations in the CBIR process. We begin with an overview of the channels.

### 2.1 Single Channel CBIR

This is the conventional approach to CBIR. We are given a corpus of images. We extract a set of features from each image. These features typically capture color, shape and texture information, although spatial and other information might also be used. Image features might be computed globally, or they might be associated with individual objects. After feature extraction the features are generally combined into a feature vector thereby implicitly placing the image (image objects) in a high-dimensional feature space.

In the typical query-by-example approach to retrieval, a query image is presented to the system. The query image is processed in the same way as the stored images to produce a compatible representation: the query vector. Subsequent retrieval is done by producing a ranked list of images at increasing distance from the query vector.

Although details among individual CBIR systems vary, the conceptual model is the same: there is a single representation for each image and that representation is consulted when retrieving images. Thus, we have a single channel into the image collection. Figure 1 shows this simplified conceptual model. A query $Q$ is shown entering the CBIR system which in turn produces a ranked list of results $R$. For our purposes, the CBIR system can be regarded as a black box.


**Figure 1. Conventional CBIR**

### 2.2 Multi-channel CBIR

We defined multi-channel CBIR in [4]. Conceptually it is a straightforward extension of the single channel case. We create several different representations of the images and consult some or all of them during the retrieval process. In our approach we transform the images and index the transformed images.

Figure 2 shows a multi-channel CBIR configuration. The dotted line encapsulates the CBIR black box of Figure 1. Four channels are shown. Details of the channel transforms are covered below. For now it suffices to note that we have used a CBIR technology to index the original images and three transforms of those images. This results in a single set of stored images (the original images) together with four indexes comprising the different representations produced by the CBIR technology.

To retrieve in this multi-channel framework, we transform the query image $Q$ to be consistent with each representation, that is, we produce four queries, $Q_1$, $Q_2$, $Q_3$ and $Q_4$, and process them separately using the appropriate index to produce results $R_1$ through $R_4$. At this point we can either: (a) present the top $k$ results of all channels to the user for inspection or (b) merge the top $k$ results from each channel and present the user with $k$ or more merged results.

Our techniques are exogenous because we are providing multiple representations of the image collection external to the CBIR system.

### 2.3 Simple 4-Channel Model

Color is generally considered to be a three-dimensional attribute. Here we will use red (R), green (G) and blue (B) as the three dimensions so the pixel at location $(x,y)$ is represented by $P(x,y) = (R(x,y),G(x,y),B(x,y))$. To create multiple representations we define a gray-scale operator $g(P(x,y)) = (i, i, i)$ where $i = (R(x,y)+G(x,y)+B(x,y))/3$

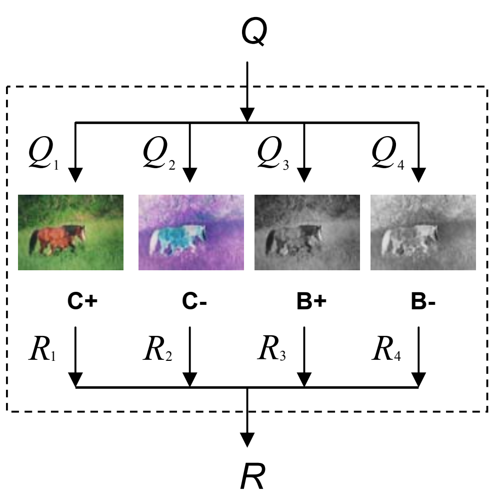



Figure 2. Four-Channel CBIR

and a negative operator $n(v) = 2^b-1-v$ for $b$ bit resolution.

We define our channels in terms of transformations as follows.

$$C+(x,y) = P(x,y)$$
$$C-(x,y) = (n(R(x,y)),n(G(x,y)),n(B(x,y)))$$
$$B+(x,y) = g(P(x,y))$$
$$B-(x,y) = n(g(P(x,y)))$$

Our four channels are, therefore, the color positive (C+) and negative (C-) and the black-and-white positive (B+) and negative (B-) images.

The intuition for including black-and-white channels is to provide channels where shape and texture will not be dominated by color.

Note that channel C+ corresponds to conventional single-channel CBIR systems and is our performance baseline.

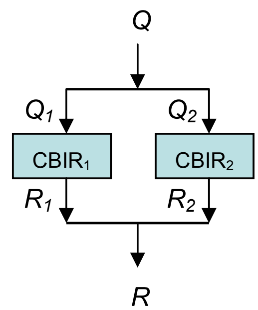



Figure 3. Multiple *k*-channel CBIR

### 2.4 Combining Multiple CBIR Systems

We have also considered combining the results from several *k*-channel CBIR systems[1] into a single retrieval result[5]. Figure 3 depicts two systems, $CBIR_1$ and $CBIR_2$ operating on the same set of images. Each of the CBIR technologies could be single or multi-channel. A query $Q$ is transformed if necessary into $Q_1$ and $Q_2$. These queries are processed as shown in Figure 2 and the results, $R_1$ and $R_2$ can be combined or presented to the user separately.

## 3. RETRIEVAL EXAMPLE

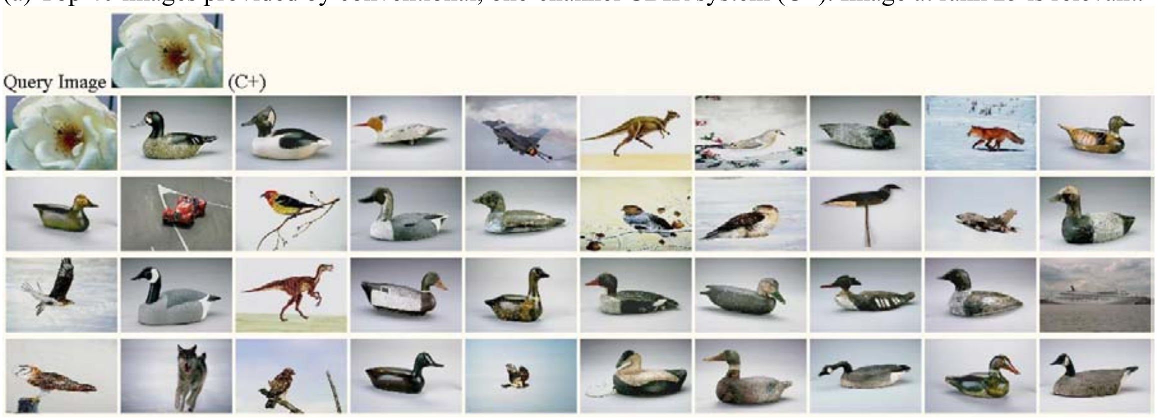
Figure 4 shows one way to deploy our approach, i.e., explicitly expose the channels the the user. We assume that a user is querying an image database for images of roses. For the purposes of this example, we will regard an image as relevant if it contains a rose in the foreground. Figures 4(a) and (b) show the top 40 results produce by the two conventional CBIR technologies. Together they produce a single relevant image. Figure 4(c) shows the top 10 images output on each channel of the CBIR technology used in Figure 4(a). Three relevant images are produced and, interestingly, these are produced on the black-and-white channels. Figure 4(d) shows the top 10 images produced on each channel in response to the second image on the black-and-white channels of Figure 4(c). Eleven relevant images are produced in a single feedback cycle. This example illustrates the potential of the multi-channel approach in interactive retrieval settings.

## 4. EXPERIMENTAL AND RESULTS

Because of space limitation we can only sketch our experimental methodology here. Complete details can be found in [4,5]. Our test data consisted of 3,400 images

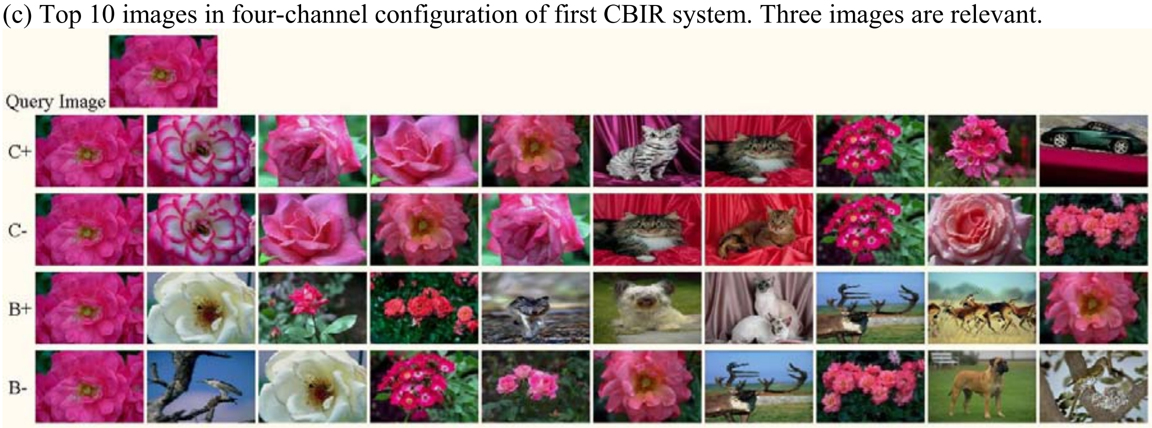
[1] The term *k*-channel CBIR system includes conventional CBIR systems ($k=1$) as well as multi-channel systems ($k>1$).

(a) Top 40 images provided by conventional, one-channel CBIR system (C+). Image at rank 25 is relevant.



(b) Top 40 images from second conventional, one-channel CBIR system (C+). No relevant images found.

(c) Top 10 images in four-channel configuration of first CBIR system. Three images are relevant.



(a) Top 10 images on each channel in response to Image-2 (B+ and B-). Eleven images are relevant.

**Figure 4. Single-channel CBIR Retrieval Results vs. Multi-Channel Approach**

drawn from 34 categories of the COREL image collection. Each category contains 100 images. We used two ground truth definitions for the results reported here: (1) Each of the images in our testbed was labeled as to foreground and background objects, so images were considered relevant if they had a single foreground object in common with the query image; and (2) the COREL class, so images were considered relevant if they were in the same COREL class as the query image.

Two different CBIR technologies were used. Each image in our test data collection was used as a test query in each channel of our multi-system, multi-channel testbed.

Our multi-channel merging results were produced using the *combSUM*[3] approach, that is, we summed the similarity values for images across the channels in which the image was included in the response set. We report the multi-system merge results using both *combSUM* and a rank merging algorithm that depends only on the ordinal ranks of the images. Finally, we also report the *perfect* merge — as if the results were merged by an oracle. In this approach, we sorted the merged list by known relevance, that is, we assumed an oracle would place the *r* relevant images in position 1 through *r* while placing the non-relevant images in positions *r+1, r+2* and so on. This merge represents the maximum possible performance achievable by any merge algorithm and gives us an operational upper bound on performance.

All queries retrieved the top 100 images and only those were used in subsequent merging.

The results of our experiments are summarized in Table 1. For brevity we chose to report only the average precision (non-interpolated).[2] Due to space limitations we have had to summarize our results considerably. The table reports the performance gain when the conventional system is considered the baseline (C+).

The baseline performance of the two CBIR technologies is essentially the same. When we run the technologies in a 4-channel configuration the performance of both is boosted substantially. The perfect merge shows what is theoretically achievable. When we merge the technologies, both merging approaches achieve excellent performance gains. The total performance improvement of both techniques combined is an impressive 54-67% better than the baseline performance. Again, we show the perfect merge to establish an operational upper bound on the performance for the aggregate system.

It is clear from these results that the two techniques, multi-channel CBIR and merging multiple technologies, can provide dramatic performance gains.

**Table 1. Average Precision (non-interpolated) of CBIR Technologies**

| | **Foreground Labeling** | | | | | **COREL** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **CBIR-1** | | **Merged** | **CBIR-2** | | **CBIR-1** | | **Merged** | **CBIR-2** | |
| Conventional System (C+) | 0.1049 | | | 0.1089 | | 0.1121 | | | 0.1177 | |
| 4 Channel *combSUM* Merge | 0.1277 | 22% | | 0.1323 | 21% | 0.1300 | 16% | | 0.1405 | 19% |
| 4 Channel Midrank Merge | 0.1333 | 27% | | 0.1367 | 26% | 0.1366 | 22% | | 0.1456 | 24% |
| 4 Channel Perfect Merge | 0.3042 | | | 0.3049 | | 0.3328 | | | 0.3421 | |
| 2 Systems Midrank Merge | | 62% | 0.1695 | | 56% | | 57% | 0.1755 | | 49% |
| 2 Systems *combSUM* Merge | | 67% | 0.1748 | | 61% | | 61% | 0.1808 | | 54% |
| 2 Systems Perfect Merge | | | 0.4250 | | | | | 0.4690 | | |

## 5. CONCLUSIONS

We have described a simple approach for improving the retrieval effectiveness of conventional CBIR systems. Our approach is exogenous, allowing us to treat the CBIR technology as a black box which can be used to provide different channels of retrieval results for subsequent merging or for use in interactive retrieval interfaces. The channels are implemented as additional indexes over simple image transforms. Our approach offers a simple, cost-effective strategy for boosting the performance of CBIR systems.

## REFERENCES


[1] Aslandogan, Y. and Yu, C. Techniques and Systems for Image and Video Retrieval, *IEEE Transactions on Knowledge and Data Engineering*, **11**(1) 1999, 56-63.

[2] Belkin, N., Cool, C., Croft, W., and Callan, J. The Effect of Multiple Query Representations on Information Retrieval System Performance in *Proc. of ACM SIGIR* 1993, 339-346.

[3] Fox, E. and Shaw J. Combination of Multiple Searches in *Proceedings of TREC-2*, March 1994, 243-252.

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

[5] French, J.C., Watson, J.V.S., Jin, X., Martin, W.N. Integrating Multiple Multi-Channel CBIR Systems. Inter. Workshop on Multimedia Information Systems (MIS 2003).

[6] Rui, Y., Huang, T. and Chang, S.-F. Image Retrieval: Past, Present, and Future in *Proceedings of International Symp. on Multimedia Information Processing* (Taiwan, Dec. 1997).


[2] For each query the precision is computed after each relevant image is retrieved. If no relevant images are retrieved, the precision is zero. All the precision values for a query are averaged to yield the query performance. All the query values are averaged to yield the system performance.