

# A collaborative, long-term learning approach to using relevance feedback in content-based image retrieval systems

Vladimir Nedovic and Oge Marques

Department of Computer Science and Engineering, Florida Atlantic University

777 Glades Road, Boca Raton, FL 33431-0991, USA

E-mail: [omarques@fau.edu](mailto:omarques@fau.edu)

**Abstract** - In recent years, an extensive research in the area of Content-Based Image Retrieval (CBIR) has been focused on Relevance Feedback (RF) techniques to improve the retrieval of images. In relevance feedback systems, a search engine dynamically updates the weights of various visual features in the query based on the user's measure of retrieved images' (ir)relevance. In this paper, we propose a novel collaborative relevance feedback approach that relies on the postulation that human-perception subjectivity is narrower than the semantic gap. As relevant results of each query are being remembered by the system, various users over different sessions can contribute to improve the successive query results.

**Keywords** - Content-Based Image Retrieval, Relevance Feedback, Semantic Gap.

## 1. INTRODUCTION

With the increasing use of digital images and the proliferation of the World Wide Web, a vast amount of images has been generated, stored, transmitted, and accessed. In order to make use of such data, efficient methods for searching and retrieving images had to be developed. Initial efforts in image retrieval were built upon existing text-based search engine technologies and directed towards a keyword-based scenario. In this approach, all images are annotated manually before being stored; they can be retrieved later by searching for the corresponding keywords.

To avoid manual annotation and automate the process of image retrieval, Content-Based Image Retrieval (CBIR) techniques based on image's visual contents have been developed in the early 1990s. CBIR systems use low-level features (mostly color, texture, and shape) to represent an image's content and find similar images during the retrieval process. This approach avoids manual annotation and can be made completely automatic because the image's visual contents do not change. CBIR has attracted significant research attention; however, despite all the research efforts, the systems lagged behind the text search engines. There are two main reasons for the inefficiency of such systems:

- The gap between the low level features and high level semantic concepts, i.e. the *semantic gap*

The problem stems from the fact that image primitives (such as color, texture, and shape) as well as the associated similarity measures do not always convey the human notion of an image's semantic meaning or the similarity between two or more images.

- The subjectivity of human perception

Different persons, or even the same person in different circumstances, often have different interpretations of the same visual content.

In an attempt to minimize some of their shortcomings, several CBIR systems have adopted a technique called *Relevance Feedback (RF)*, in which human and computer iteratively interact to refine high level queries to better approach the representations based on low-level image features. The retrieval performance of general CBIR engines without user feedback is usually so poor that even simple RF algorithms (adapted from text-based document retrieval) may significantly improve their accuracy.

In this paper, we propose a novel RF strategy in which users' relevant results are remembered and their habits learned over time. This strategy is strongly based on the premise that the semantic gap is wider than the differences that arise from human perception subjectivity. In other words, if a user explicitly indicates that two images are semantically related, we take the user's word for it and override any contrary hypothesis that may have been formulated on basis of the images' low-level features. This paper is organized as follows: Section 2 gives relevant background information. Section 3 describes our approach in more detail. Finally, our conclusions and suggestions for future work are in Section 4.

## 2. BACKGROUND

Relevance feedback in CBIR is a process of dynamically adjusting an existing query using the information fed-back by the user about the relevance

of previously retrieved images such that the adjusted query is a better approximation to the user's need [1]. The goal of such a process is to capture the user's high level query and perception subjectivity by interacting with him/her and automatically adjusting the weights based on the provided feedback. RF techniques can be divided into groups by various criteria [2], depending on:

- The type of search

Many systems assume that the user is looking for a class of similar items to the given example image (*similarity* or *category* search), whereas others (e.g., [3]) assume that the user is looking for a particular target item (*target search*).

- Feature selection

The task of selecting a limited number of low-level visual features and their variants to represent image's semantics is a critical one considering the issue of a semantic gap. The majority of CBIR systems use standard visual features such as color, shape, texture, and structure. Many systems extract only global features, but there are other approaches that are region-based.

- Online feature weighting

Instead of a user manually specifying the relative importance of particular features in the "ideal query" he/she has in mind, RF algorithms were developed to learn those feature weights automatically from the feedback examples, in a process referred to as *online feature weighting*.

- With vs. without memory

A big disadvantage of several RF implementations is that the acquired knowledge from one session is not stored and cannot be reused to improve the system's performance in successive sessions. In other words, even if the same user repeats the same query (with the same "ideal query" in mind and the same feedback), he/she would have to go through all the iterations of the relevance feedback procedure again. Therefore, the machine learning process is limited to only one session and does not accumulate its "experience" over multiple sessions. To address this issue, some have proposed other approaches in which they attempt to memorize results for some user, assuming that they represent his/her subjectivity characteristics.

### 3. THE PROPOSED SOLUTION

We propose a system in which collaborative contribution of many users over different sessions helps to improve the results of successive queries. This can be very helpful in a scenario where users share not only the same image database, but also the

same interests, educational backgrounds, and presentation styles (i.e. belong to the same "user profile").

Our basic claim is that human perception subjectivity is narrower than the semantic gap and that we should thus rely more on user's judgment than on the low-level feature-extraction engine. A block-diagram of our method is given in Figure 1.

#### 3.1. Feature Extraction

We use only two low-level features for content description, color and shape. The color space of choice was HSV, since it has been designed to model human perception. For color feature, we extract two feature representations: color histograms (for each of the H, S, and V, i.e. three vectors, each with 256 values) and color moments (only the first two moments are extracted for every component, therefore six values); for shape, we extract Fourier descriptors. Since a shape can be extracted from a binary image only, segmentation had to be performed on the original images; Canny edge detection, followed by some morphological operations have been used to extract the foreground object from the image and binarize the result<sup>1</sup>.

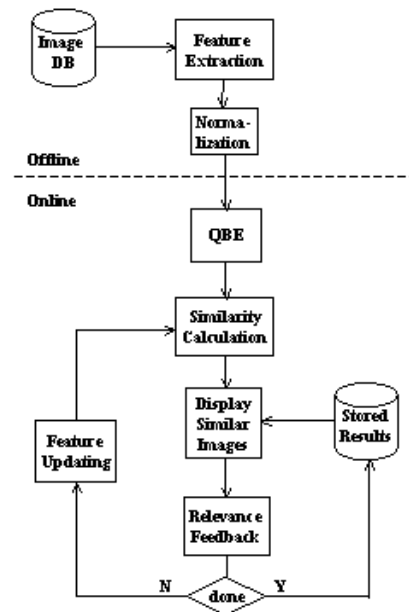


Fig. 1. Simplified block-diagram of our approach

<sup>1</sup> The test database contains images of objects in high contrast with the background in order to facilitate the process of segmentation. In case of patterns or images with more than one object of interest, this part of the process becomes an issue which will be addressed in the future.

### 3.2 Normalization and Similarity Calculation

Once the features have been extracted from all the images in the database, normalization is performed as described in [1]. To ensure equal emphasis of each value within its corresponding feature representation vector, *intra-normalization* is performed.

With the normalized feature-vector values, similarity between the images can be measured. We use histogram intersection for color histogram distance evaluation, simple Euclidean distance for color moments, and a weighted one for translation- and rotation-invariant Fourier descriptors. In order to make Fourier descriptors invariant to rotation, major axis orientation is extracted for each image. When images are compared, the difference in their orientations is found, and one of the images is rotated in order to be positioned appropriately (the phase of its Fourier descriptors is adjusted for the difference in orientations). The similarity of shapes in terms of Fourier descriptors is then evaluated in the frequency domain, so that lower frequencies can be given larger weight.

Similarity measures for each particular feature representation are different and therefore their values can be of different orders of magnitude; in order to ensure that one similarity value does not overshadow the others in the overall similarity process, the *inter-normalization* also has to be performed [1]. Once this is done, the range of similarity values will be [0, 1].

### 3.3 Feature Updating

With all the similarity values calculated, we can present the user with the best results (i.e. those corresponding to the greatest similarity, or smallest difference). The user marks displayed images as being relevant to his query or not in terms of semantics. Once the system has collected his/her judgment on all the presented images, it is able to recalculate the weights in order to refine the query (i.e. the system's understanding of the abstract concept that the user has in mind) and retrieve images that are more similar to the example in the next round. Since the user is presented with options to mark images with different degrees of relevance, the system collects this feedback in a data structure called *Score*, later used in the feature updating process.

For updating the weights, we implemented a *one-class, re-weighting* algorithm. One-class algorithms are those that consider only the positive feedback examples when trying to improve the successive results. Re-weighting is a very intuitive approach used for adjusting the weights given to each individual component (within some feature representation, i.e. *intra-weighting*), which relies on statistical information extracted from the relevant

images: if the variance of positive examples is high for some component, then that component is not important to the user and should receive smaller weight; on the other hand, if it is small, that component is common to all those examples and might be capturing the abstract concept that the user has in mind. The features themselves are updated with the scores obtained from user's feedback (i.e. *inter-weight*). With the new weights, similarity is calculated again, and new results (hopefully more similar to the user's query) are presented to the user in the next round.

### 3.4 User Interface

We propose a novel user interface in which results of the same queries by previous users (or, even better, of the same user at different times) are displayed to the current one in an attempt to expedite the retrieval process (Figure 2). As in typical RF interfaces, in our case the process also starts with a user presenting a query example to the system<sup>2</sup>. Once the image has been presented to the system, it responds by retrieving images visually similar to the selected one in terms of color and shape (middle column in the figure). If the same example image has already been used in the previous queries, the results of those queries are displayed below it; furthermore, the results of queries in which the retrieved (visually similar) images have been included are also shown in the rows of the rightmost column of the figure. The latter introduces a novelty feature of our user interface: since it is difficult to bridge the semantic gap and describe image semantics by its visual features, we rely on other humans' perception of semantic similarity (claiming that semantic gap is wider than the differences that arise from perception subjectivity). If interested in the particular retrieved image, the user can keep its "satellites" or he/she can delete any number of them. The user is provided with four options in judging the degree of relevance for the retrieved results: he/she can decide that the query example and a particular retrieved image are similar:

- (1) *semantically* (since that is the overall goal of the process, the system "glues" those two images together – the glued images are displayed as "satellites" of each others in the successive rounds of this session or in other sessions)
- (2) *visually* (similar common visual features of both images are given more weight in the query)

---

<sup>2</sup> This can be an image outside of the database (as is typical in a QBE approach), however for testing purposes we decided to use the images already present in the database itself (to verify that the best match will always be the equivalent image).

(3) *both* (do both: glue + adjust feature weights)

(4) *neither* (disregard the retrieved image and its satellites, except those deleted, for the rest of this retrieval session)

The degrees of relevance for the options above are stored as Scores with values 2, 1, 3, and -2, respectively; the scores are used in the process of feature weight updating, which closely followed the algorithm proposed in [1].

A very important feature of our interface is that the collaborative action does not require any additional effort by the user than typical relevance feedback interfaces, since it is incorporated into the usual interface in which images are only marked as relevant to the query or not. The system simply remembers users' "habits" during query sessions by storing relevant results and it uses these results in successive queries to classify and cluster semantically similar images.

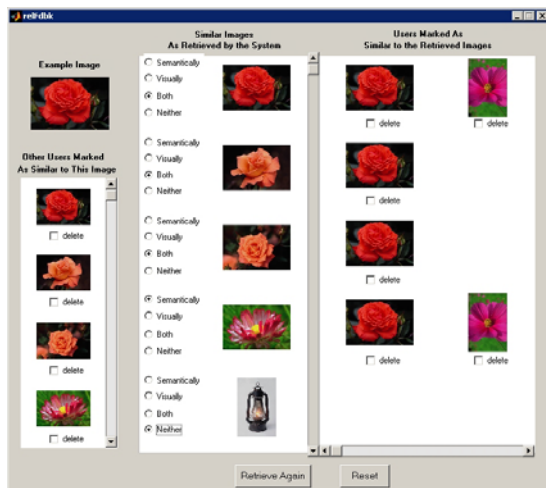


Fig. 2. The system's GUI.

Based on the feedback provided by the user, the system refines its query and retrieves new results; the retrieval process should terminate (and hopefully converge) after three or four such rounds so as not to overwhelm the user. The final results ("glued-pairs" or those images marked by the user as semantically similar or both) are stored to be used in the next session, involving either the same user or someone different.

### 3.5 Implementation and tests

The current prototype has been developed using MATLAB. It has not yet been tested extensively. However, with the images that we are working on (a subset from Corel's collection usually containing objects in high contrast with the background), we achieved the biggest convergence ratio (came "closer" to what we were looking for) after one or two iterations of the algorithm. These results are very desirable, because they mean that the user will be satisfied after a short time and not overwhelmed by repetitious requirements of the relevance feedback interface.

### 4. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel user interface for CBIR systems, in which users can collaboratively mark retrieved images as relevant or not as well as view images marked as semantically similar by previous users and/or sessions.

This is a work in progress and the current prototype can be improved in many ways, such as:

- Add more features (such as texture) and/or more descriptors for the existing features.
- Improve the current RF one-class similarity search algorithm, replacing the option Neither by Neutral and adding a new option, Dissimilar, for which common visual features would be used as negative feedback, in effect transforming the algorithm into a two-class search.
- Test the system extensively and benchmark it against other CBIR systems in the literature.

### REFERENCES

- [1] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra, "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval", IEEE Transactions on Circuits and Video Technology, vol. 8, Sept. 1998.
- [2] X. S. Zhou and T. S. Huang, "Exploring the Nature and Variants of Relevance Feedback", IEEE Workshop on Content-based Access of Image and Video Libraries, Dec. 2001.
- [3] O. Marques and B. Furht, "MUSE: A Content-Based Image Search and Retrieval System Using Relevance Feedback", Multimedia Tools and Applications, Kluwer, Vol. 17, No. 1, pp. 21-50, May 2002.