Proceedings of the 2013 International Conference on Information, Operations Management and Statistics (ICIOMS2013), Kuala Lumpur, Malaysia, September 1-3, 2013

A Hybrid Approach to Content-based Image Retrieval

Smarajit Bose[a], Amita Pal[a*], Jhimli Mallick[b] , Sunil Kumar[c]

[a]*Applied Statistics Division, Indian Statistical Institute, Kolkata, India*
[b]*TechBLA Solutions, Kolkata, India*
[c]*ETH, Zurich, Switzerland*

Abstract

Content-Based Image Retrieval (CBIR, in short) refers to the extraction of images of interest (that is, images similar to a so-called query image) from large (image) databases, based on the visual and semantic contents of images. To bridge the semantic gap that exists between the representation of an image by low-level features (namely, colour, shape, texture) and its high-level semantic content as perceived by humans, CBIR systems typically make use of the relevance feedback (RF) mechanism, which iteratively incorporates user-given inputs regarding the relevance of retrieved images, to improve retrieval efficiency. To achieve this objective, one approach is to vary the weights of the features dynamically via feature reweighting. In this work, an attempt has been made to improve retrieval accuracy with RF by combining a CBIR system based on color features alone, with a CBIR system which also incorporates shape information implicitly obtained through prior segmentation of the images. A novel feature reweighting scheme for relevance feedback has also been proposed for boosting the performance further. Results of some experiments have been presented to illustrate the effectiveness of the proposed approaches.

Keywords: Content-Based Image Retrieval (CBIR); Features; Image Segmentation; Relevance Feedback; Feature Reweighting.

1. Introduction

As a direct consequence of rapid advances in digital imaging technology, millions of images are being generated everyday by innumerable sources like military and civilian satellites, military reconnaissance and surveillance flights, fingerprinting and facial-image-capturing devices for security and forensic purposes, scientific experiments, biomedical imaging and home entertainment systems. Large repositories of images have become a commonplace reality due to the availability of cheaper digital storage devices and the internet. However, maintaining such repositories is meaningless in the absence of methodologies that can enable a user to extract or retrieve information (in the form of images) of interest as and when required.

Content-based Image Retrieval (CBIR) systems look for images in large databases that are very similar to a supplied query image, where the search is based on the contents of the image rather than metadata. The term *content* in this context might refer to colors, shapes, textures or any other higher-level descriptor(s) that can be derived from the image itself.

[*] Corresponding author. Tel.: +91-33-2575-2817; fax: +91-33-2577-3104.
E-mail address: pamita@isical.ac.in

In a typical CBIR system, features are extracted from each image in the database and stored in the feature database. The same features are extracted from the query image as well. The system computes the distance or similarity measure of the feature vector for the query image from those of each image in the database and retrieves images (usually a fixed number, specified by the user, known as the *scope* of the system) closest to the query image [2], [6].

The low-level features used to represent an image do not necessarily capture adequately the high-level semantics and human perception of that image. This leads to the so-called *semantic gap* in the CBIR context. A solution to this problem is user intervention in the form of *Relevance Feedback* (RF) [7]. For a given query, the system first retrieves a set of images ranked in order of their similarity to the query image, in terms of a similarity metric, which represents the distance between the feature vector of the query image and that of each image in the database. Then the user is asked to identify images that are relevant or irrelevant (or non-relevant) to his/her query. The system extracts information from these samples and uses that information to improve retrieval results. A revised ranked list of images is then presented to the user. This process continues until there is no further improvement in the result or the user is satisfied with the result. One way of attaining this objective is *feature reweighting*, which essentially assigns greater weights to features that discriminate well between relevant and non-relevant images, thus enhancing retrieval, and smaller weights to those features that do not. Another approach is the instance-based approach, which considers the distance of an image in the database from the query as the minimum distance of the image from the set of all relevant images. This is useful to move through the feature space to the regions with clusters of relevant images. In this work, a novel approach to relevance feedback has also been proposed, which applies a combination of feature reweighting and instance-based approaches to compute relevance scores and hence weights of features, and also the use of a modified initial relevant set for RF. Each of these proposed modifications lead to more accurate retrieval.

Image features based on a single attribute like color or shape or texture alone are generally not adequate for satisfactory image retrieval [3]. It has been shown by several researchers in this area that segmentation of the images before matching improves retrieval precision for some image databases. The features derived from each of the clusters obtained by segmentation of the query image are matched with those of the clusters obtained from each image in the database. However, this approach is not uniformly effective for all types of image databases. In this work, features incorporating both color and shape information are used. The color information is obtained from the color co-occurrence matrix (CCM) of the image, while the shape information is extracted through segmentation of the image.

Organization of the paper is as follows. Section 2 provides an overview of the classical approach to CBIR based on relevance feedback. Section 3 presents the proposed approach, together with a couple of new measures of retrieval accuracy. Results are presented in Section 4, while Section 5 gives a brief overview of the contribution made by this work to CBIR.

2. The Classical Approach to CBIR

In a typical CBIR system, the user gives a query image he is interested in and wants to find similar images from a large (image) database $\mathcal{D}$ of size N. A feature extraction algorithm is then used to process each image and extract its features. For an image I, let $\boldsymbol{f}_I = (f_{I1}, f_{I2}, \cdots, f_{Id})'$, a d×1 vector in $\mathbb{R}^d$, be the d features extracted. For a database with N images, the d×N matrix $\boldsymbol{F} = [\boldsymbol{f}_1\ \boldsymbol{f}_2 \ldots \boldsymbol{f}_N]$, whose j-th column is the d×1 feature vector of the j-th image in the database, represents the entire collection of feature vectors that are extracted and stored. The same feature extraction algorithm is used to process the query image Q also and the query feature vector is obtained, say, $\boldsymbol{f}_Q = (f_{Q1}, f_{Q2}, \cdots, f_{Qd})'$. The system consequently calculates the appropriate distance or similarity measure of the feature vector of the query image from those of all the images of the database and retrieves the images (a fixed number, specified by the user, known as the *scope* of the system) closest to the query image.

2.1. Similarity Measure

The similarity between the query image Q and any other image I is inversely proportional to any distance measure between their respective feature vectors. Popular choices of distance measures in CBIR literature are

$$d_1(Q,I) = \sum_{j=1}^{d} w_j \left|f_{Qj} - f_{Ij}\right| \text{ and } d_2(Q,I) = \sqrt{\sum_{j=1}^{d} w_j \left(f_{Qj} - f_{Ij}\right)^2}. \quad (1)$$

based on the L1- and L2-norms, respectively. The usual practice is to initialize the weights as $w_i = 1/d$. In this work, the distance measure $d_2(Q,I)$ has been used throughout, and has been referred to as $d(Q,I)$ for brevity.

2.2. Improvement with Relevance Feedback (RF)

Relevance feedback (RF) is a proven method for reducing the semantic gap through user intervention. The user labels the retrieved images as *relevant* or *non-relevant*. This information is fed back to the system, usually iteratively, leading to improved retrieval accuracy. A popular method for providing this feedback are *feature reweighting* which assigns different weights to different features based, for example, on the Euclidean distance [2], [7]. These weights are modified in each iteration of the relevance feedback. Larger weights are given to those features that discriminate well between relevant and non-relevant images and thus enhance retrieval of relevant images.

A discriminant ratio (as in [7]) can be used to determine the ability of a feature component in separating the relevant images from the non-relevant ones:

$$\delta_i^{(t)} = 1 - \frac{\text{No. of non-relevant images in } D_i^{(t)}}{|\mathcal{R}_t|}, \quad (2)$$

where $\mathcal{R}_t$ denotes the set of relevant images at the t-th RF iteration, $\{F_{rel,j}^{(t)}, j = 1,2,\dots,|\mathcal{R}_t|\}$ denotes the collection of the j-th feature values of all images in $\mathcal{R}_t$ and $D_i^{(t)} = \left[\min_j\left(F_{rel,j}^{(t)}\right), \max_j F_{rel,j}^{(t)}\right]$. The value of $\delta_i^{(t)}$ lies between 0 and 1. It is 0 when all non-relevant images are within the dominant range and thus, no weight should be given for that feature component. On the other hand, when there is not a single non- relevant image lying within the dominant range, maximum weight should be given to that feature component. Based on this, a popular reweighting scheme for feature j is given by

$$w_j^{(t+1)} = \delta_j^{(t)} \frac{\sigma_j^{(t)}}{\sigma_{rel,j}^{(t)}}, \quad (3)$$

where $\sigma_j^{(t)}$ and $\sigma_{rel,j}^{(t)}$ denote respectively the standard deviations of f_j over the sets $\mathcal{R}_t \cup N_t$, N_t being the set non-relevant images at the t-th RF iteration.

2.3. Instance-Based Methods

As alternatives to feature reweighting schemes based on Euclidean distances, instance based methods have been quite successful in CBIR, for example, as proposed by Zhang *et al.* [8].

Some of the instance-based approaches that have been reported in the literature are described in the next few paragraphs. The purpose of doing so is to lay the groundwork for the proposed combination reweighting scheme developed in Section 3.1.2, which essentially combines the last instance based approach with the reweighting scheme given by Equation (3) to achieve the highest retrieval performance.

2.3.1. Instance-based Cluster Density Method

Let $\mathcal{R}$ and N denote respectively the sets of relevant and non-relevant images with respect to the query image Q. Let $d_R(Q,I) = \min_{I' \in \mathcal{R} \cup Q} d(I',I)$ and $d_N(Q,I) = \min_N d(I',I)$.

Even though a small value of $d_R(q,I)$ means that image I has a high degree of membership to the relevant set, $d_R(q,I)$ alone may not be able to reflect it completely. For example, an image may be very close to the nearest image of the relevant set, but it may be far away from the centre of the set of relevant images if that nearest image is itself an outlier. That is why it is also desirable that the average distance from all the images in the relevant set is small. Thus a modified relevance score of I with respect to Q, involving cluster density, which measures the similarity between Q and I, is given by

$$RS_Q(I) = \left[1 + d_C(Q,I) \times \frac{d_R(Q,I)}{d_N(Q,I)}\right]^{-1}, \tag{4}$$

where $d_c(Q,I) = \frac{1}{|\mathcal{R}|}\sum_{I' \in \mathcal{R}} d(I',I)$, the average distance of Q from all images in its relevant set $\mathcal{R}$.

2.4. Performance evaluation measures

The two most commonly used measures for evaluating the performance of a CBIR method are *Precision* and *Recall*, defined below:

$$Precision = \frac{Number\ of\ relevant\ images\ retrieved}{Number\ of\ retrieved\ images} \tag{5}$$

$$Recall = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ Number\ of\ relevant\ images\ in\ database} \tag{6}$$

Usually the number of images retrieved by the CBIR method (called the *scope* of the method) is a prespecified positive integer. The precision and recall values are calculated for each image in the database, and then these are averaged over all the images (in the database). These averages are then plotted for different values of the scope. Usually, the greater the scope, the larger is the number of relevant images retrieved, leading to higher value of recall. However, the precision typically goes down with increasing scope.

Under relevance feedback, after the user identifies the relevant and non-relevant images in an iteration then, in the following iteration, usually a different set of images will be retrieved due to change in the search criterion. With every RF iteration, the number of relevant images retrieved is expected to be higher. This procedure is repeated a number of times after obtaining the relevance feedback from the user after each step.

There are several issues involved here. For example, it is not desirable to show the same image to the user more than once. Therefore, one can consider retrieving a new set of images at each step. Further if we consider the initial scope S as the number of relevant images the user is looking for, it makes sense to retrieve only $(S - R)$ number of images at every step, where R is the current number of relevant images. In such cases, the total number of images retrieved is changed after every step, and it is expected to be different for different images. It is therefore expected that precision and recall values will increase with each RF iteration, as is evident from each of the precision-recall graphs in Figure 1.

This discussion motivates us to propose two other evaluation measures which have been discussed in Section 3.3.

3. Proposed Approach

3.1. Segmentation-based similarity

3.1.1. Image Segmentation

A CBIR system searches image databases for images which have content similar to that of the query image. So one can expect that its retrieval efficiency to improve after segmentation ([1], [4]) of the images to identify visually homogeneous subregions or objects within them, which are significant indicators of image content. In this work, the k-means algorithm [4] has been used to segment the images. Before clustering, the following preprocessing is done. Each color image of size $N_1 \times N_2$ in the HSV color space is split into blocks of size $n_1 \times n_2$ (with $n_1 = n_2 = 4$ as suggested in [11]). A total of $b = N_1 N_2 / n_1 n_2$ blocks is thus obtained. If $N_1 = N_2 = 256$, This number is 4096. The d-dimensional feature vector is computed from each block, giving rise to b observations from the image. In the k-means algorithm, k has been taken to be equal to 8, and empty clusters, if any, are discarded.

3.1.2. Proposed Feature Reweighting Strategy

A combination of the reweighting method (Equation (3)) and the instance-based cluster density method (Equation (4)) is used to obtain the weight for each feature. Each of the distances d_R, d_N and d_C in Equation (4) is computed as a weighted Euclidean distance with weights updated in every iteration by the reweighting scheme with weights given by Equation (3).

3.1.3. Proposed Segmentation-based Similarity Measure

The proposed measure of similarity between a query image Q and an image I in the database $\mathcal{D}$, based on the distances among the feature vectors corresponding to their individual segments, is motivated by the following observations:

For the query image Q, let $\boldsymbol{D}_1(Q, I)$ denote the $n_Q \times n_I$ matrix of distances between the feature vectors corresponding to the n_Q segments of Q and those corresponding to the n_I segments of I.

Let $d_{(1)}(Q, I)$ denote the smallest element of $\boldsymbol{D}_1(Q, I)$. Let $\rho_{(1)}(Q, I_j)$, $j = 1,2, \dots, N$, denote the ranks of $d_{(1)}(Q, I_j)$, $j = 1,2, \dots, N$. Suppose that the minimum $d_{(1)}(Q, I)$ is the (p_1, q_1)-th element of $\boldsymbol{D}_1(Q, I)$, and let $\boldsymbol{D}_2(Q, I)$denote the submatrix obtained by deleting from $\boldsymbol{D}_1(Q, I)$ its p_1-th row and q_1-th column.

Likewise, for $i = 2,3, \dots, r$, where r is a prespecified positive integer such that $1 \le r \le \min_{I \in \mathcal{D}} n_I$, let $d_{(i)}(Q, I)$ denote the smallest element of $\boldsymbol{D}_i(Q, I)$, the submatrix obtained by deleting its p_{i-1}-th row and q_{i-1}-th column. Here (p_{i-1}, q_{i-1}) denotes the location of the smallest element $d_{(i-1)}(Q, I)$ of $\boldsymbol{D}_{i-1}(Q, I)$ within the matrix. Let $\rho_{(i)}(Q, I_j)$, $j = 1,2, \dots, N$, denote the ranks of $d_{(i)}(Q, I_j)$, $j = 1,2, \dots, N$.

The more similar Q is to I_j, the higher will be the ranks $\rho_{(i)}(Q, I_j)$, for most of the Q-segments, indexed by i. This motivates a new measure of image similarity in the CBIR context, a segmentation-based distance defined, for a query image Q and an image I in $\mathcal{D}$, as

$$d_{seg}(Q,I) = \sum_{i=1}^{r} \rho_{(i)}(Q,I). \qquad (7)$$

Retrieval accuracy is expected to increase with increase in the value of r within the range specified above. In this work, r has been taken to be less than or equal to 4. The S images I in $\mathcal{D}$ with the lowest values of $d_{seg}(Q,I)$ are retrieved in the 1st iteration of relevance feedback in this proposed segmentation-based CBIR approach (referred to henceforth as the WS approach).

3.2. Selection of Initial Set of Retrieved Images for RF

Henceforth the shorthand notations WOS and WS will be used to denote respectively the conventional CBIR approach without segmentation of images, and the proposed approach based on image segmentation.

To exploit additional information on image content, as captured through segmentation, the following alternative method for specifying the initial set D_{init} of S retrieved images on which relevance feedback is implemented, is proposed. It leads to improved retrieval accuracy with the proposed WS approach relative to the WOS approach, as will be established empirically in Section 4.

Let D_{WOS} and D_{WS} denote respectively the sets of S relevant images retrieved by WOS and WS. If the number of images in $D_{WOS} \cap D_{WS}$ is c, then D_{init} is taken to be equal to

$$D_{inter} = [D_{WOS} \cap D_{WS}] + {D_{WOS}}^{(1)} + {D_{WS}}^{(1)}, \qquad (8)$$

where ${D_{WOS}}^{(1)}$ and ${D_{WS}}^{(1)}$ are respectively the sets of $S - w_1$ and $S - w_2$ most relevant images in the sets $D_{WOS} - D_{inter}$ and $D_{WS} - D_{inter}$, with $w_1 = w_2 = c/2$ when c is even and, when c is odd, $w_1 = (c-1)/2$, $w_2 = (c+1)/2$.

3.3. Proposed Performance Evaluation Measures

To address the shortcomings of the existing measures, mentioned in Section 2.4, two new measures are proposed:

$$Retrieval\ Efficiency = \frac{Number\ of\ relevant\ images\ among\ the\ retrieved\ images}{Scope} \qquad (9)$$

$$False\ Discovery = \frac{Number\ of\ non\text{-}relevant\ images\ retrieved}{Number\ of\ non\text{-}retrieved\ images\ in\ the\ database} \qquad (10)$$

Retrieval Efficiency is expected to increase with the number of RF iterations and it should converge fast in a few iterations if RF is effective. False discovery, being the ratio of the number of non-relevant images retrieved to the total number of retrieved images, is a measure of erroneous retrieval (that is, the retrieval of non-relevant images), and should be minimized as far as possible.

3.4. Features Used

Feature selection is an extremely crucial aspect of CBIR. As expected, features like colour, shape and texture are key indicators of content. An important representation of the spatial distribution of colour in an image is provided by the colour co-occurrence matrix (CCM) [2], [3]. The $L \times L$ CCM of an image having L colour levels in any one of the dimensions of the HSV (Hue, Saturation, Value) colour space, denoted by $\boldsymbol{P} = \left(\left(p_{ij}\right)\right)$, is such that p_{ij} represents the proportion of pixels with colour level i co-occurring with other pixels with colour level *j*, at a relative position, say, *d*. The diagonal elements of the CCM give the colour distribution in the image, while the non-diagonal elements convey shape information, since colour changes between adjacent pixels indicates the possible existence of an object edge. The feature vector used consists of all L diagonal elements of the CCM as well as a single number to represent the information contained in its non-diagonal elements, defined as

$$ave\_ndiag = \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} (i+j) p_{ij}, \quad (11)$$

where i and j are row and column indices.

It has been observed by researchers that $L_H = 16$ and $L_S = L_V = 3$ are good choices for number of quantization levels of H, S and V for specifying co-occurrence matrices. A co-occurrence distance $d = 1$ has been used in this work and pixel pairs in both vertical and horizontal directions have been considered, leading to symmetric co-occurrence matrices. Thus only upper diagonal elements of the CCMs are considered.

4. Results

To demonstrate the effectiveness of the proposed approach, two image databases were used, which are briefly described below.

Name: DB2000	**Size:** 2000	**No. of categories:** 10	**Size per category:** 200
Name: DB2020	**Size:** 2020	**No. of categories:** 12	**Size per category:** 96-376

The effectiveness of the proposed reweighting scheme for RF (described in Section 3.1.2) as compared to simple reweighting, in the context of the WOS approach, is reported in Table 1, which contains results obtained after 7 iterations of RF.

Table 1: Effectiveness of the Proposed Reweighting Approach

Database	Relative Efficiency		False Discovery	
	With Simple Reweighting	With Proposed Reweighting	With Simple Reweighting	With Proposed Reweighting
DB2000	89.20	94.69	48.61	41.89
DB2020	80.32	86.42	56.67	50.93

A comparison, using all four evaluation measures, between the conventional WOS approach and the proposed WS approach to CBIR, incorporating the proposed feature reweighting scheme and using the initial relevant set proposed in Section 3.2 is reported in Table 2. Marked improvement in retrieval accuracy with the proposed approach is evident in all cases.

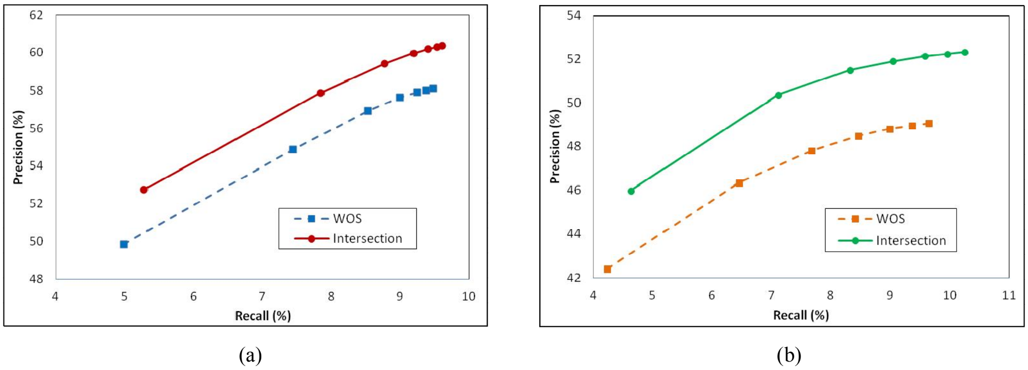

Fig. 1. Precision-Recall graph with proposed method for (a) DB2000; (b) DB2020.

Table 2: Effectiveness of the Proposed Segmentation-based CBIR Approach

Database	Relative Efficiency		False Discovery		Precision		Recall	
	Without Segmentation (WOS)	With Segmentation (WS)+ D_{inter}	Without Segmentation (WOS)	With Segmentation (WS)+D_{inter}	Without Segmentation (WOS)	With Segmentation (WS)+ D_{inter}	Without Segmentation (WOS)	With Segmentation (WS)+D_{inter}
DB2000	94.69	96.10	41.89	39.62	58.11	60.38	9.47	9.61
DB2020	86.42	90.34	50.93	47.67	49.07	52.33	9.64	10.25

5. Summary

In this work, a new hybrid approach for CBIR is proposed where the conventional methods have been combined with a segmentation-based approach. A new relevance feedback mechanism based on a combination of feature reweighting with an instance-based distance is employed. A scheme for combining the two approaches is proposed, and its effectiveness is illustrated with a couple of moderately large databases. The proposed approach was successful in improving the retrieval accuracy significantly.

References

[1] Chen, Y., Wang, J. Z., Krovetz, R., 2005. CLUE: Cluster-based Retrieval of Images by Unsupervised Learning. IEEE Transactions on Image Processing 14, p. 1187.

[2] Das, G., 2007. Reduction of Semantic Gap in Content-based Image Retrieval. PhD thesis, Clayton School of Information Technology, Monash University, Melbourne, Australia.

[3] Hiremath, P.S., Pujari, J. 2007. Content-based Image Retrieval using Color, Texture And Shape Features, in "*Proceedings of 15th International Conference on Advanced Computing and Communications*", p. 780.

[4]Murthy,V.S.V.S., Vamsidhar, E., Swarup Kumar, I. N. V. R., Sankara Rao, P., 2010. Content-based Image Retrieval using Hierarchical and *k*-means Clustering Techniques. International Journal of Engineering Science and Technology, 2, p. 209.

[5] Rui, Y., Huang, T.S., Ortega, M., Mehrotra, S., 1998. Relevance Feedback: A Power Tool for Interactive Content-based Image Retrieval. IEEE Transactions on Circuits and Video Technology 10, p. 644.

[6] Rui, Y., Huang, T.S., Ortega, Chang, S.-F., 1999. Image Retrieval: Current Techniques, Promising Directions and Open Issues. Journal of Visual Communication and Image Presentation 10, p. 39.

[7] Wu, Y., Zhang, A., 2002. A Feature Reweighting Approach for Relevance Feedback in Image Retrieval, in *Proceedings of IEEE International Conference on Image Processing* (*ICIP2002*)*,* Rochester, New York, September 2002.

[8] Zhang, H., 2003. Relevance Feedback in Content-based Image Retrieval, in "*Multimedia Information Retrieval and Management-Technological Fundamentals and Applications*", D. Feng, W. C. Siu, and H. Zhang, Editors. Springer-Verlag, Germany, p. 57.