Efficient Content-based Image Retrieval Using Weighted Feature Aggregation Scheme

Ivica Dimitrovski[1], Blagojce Jankulovski[1] and Suzana Loskovska[1]

[1] Department of Computer Science and Engineering, Faculty of Electrical Engineering and
Information Technologies,
Skopje, Macedonia
{ivicad, suze}@feit.ukim.edu.mk, blagojcejankulovski@gmail.com

Abstract. This paper presents a content-based image retrieval system for aggregation and combination of different image features. Feature aggregation is important technique in general content-based image retrieval systems that employ multiple visual features to characterize image content. We introduced and evaluated linear combination to fuse different features. The most important step in the feature aggregation is to find suitable weights for the individual features. We have used relevance feedback techniques to determine the salient features and to learn weights for each feature. The weights are used in linear combination scheme that we call weighted feature aggregation. The implemented system has several advantages over the existing content-based image retrieval systems. Several implemented features included in our system allow the user to adapt the system to the query image. The weighted combination of features allows flexible query formulations and helps processing specific queries for which users have no knowledge about any suitable descriptors.

Keywords: Content-based image retrieval, weighted feature aggregation, color features, texture features, shape features, MPEG-7.

1 Introduction

An incommensurable amount of visual information is becoming available in digital form in digital archives, on the World Wide Web, in broadcast data streams, art collections, photograph archives, bio-medical institutions, crime prevention, military, architectural and engineering design, geographical information and remote sensing systems and this amount is rapidly growing. The value of information often depends on how easy it can be found, retrieved, accessed, filtered and managed. Therefore, tools for efficient archiving, browsing and searching images are required.

A straightforward way of using the existing information retrieval tools for visual material, is to annotate records by keywords and then to use the text-based query for retrieval. Several approaches were proposed to use keyword annotations for image indexing and retrieval [1]. These approaches are not adequate, since annotating images by textual keywords is neither desirable nor possible in many cases.

Therefore, new approaches of indexing, browsing and retrieval of images are required.

Rather than relaying on manual indexing and text description for every image, images can be represented by numerical features directly extracted from the image pixels. These features are stored in a database as a signature together with the images and are used to measure similarity between the images in the retrieval process. This approach is known as Content-based Image Retrieval (CBIR).

The aim of CBIR systems is searching and finding similar multimedia items based on their content. Every CBIR system considers offline indexing phase and online content-based retrieval phase. The visual contents of the database images are extracted and described by multidimensional feature vectors in the offline phase. The feature vectors of the database images form the feature database. In the second or online retrieval phase, the query-by-example paradigm is commonly used. The user presents a sample image, and the system computes the feature vector for the sample, compares it to the vectors for the images already stored in the database, and returns all images with similar features vectors. The query provided by the user can be a region, a sketch or group of images.

The quality of response depends on the image features and the distance or similarity measure used to compare features of different images [1]. Regarding the features, different approaches are used but the most common for image content representation are color, shape and texture features. Each extracted feature characterizes certain aspect of the image content. Multiple features are usually employed in many CBIR systems to provide an adequate description of image content. The idea behind these approaches is to employ as many image features as possible, in the hope that at least one will capture the unique property of the target images. It is very challenging problem to measure the image similarity from various individual features because different features are not directly comparable as they are defined in different spaces. Research in feature aggregation is aimed to address this problem [5]. Feature aggregation is a critical technique in content-based image retrieval systems that employ multiple visual features to characterize image content.

In this paper we present efficient content-based image retrieval system which uses weighted feature aggregation scheme. The implemented system has several advantages over the existing content-based image retrieval systems. Several implemented features of our system allow the user to adapt the system to the query image. The weighted combination of features allows flexible query formulations and helps processing specific queries for which the users have no knowledge about any suitable descriptors. For the experiments, public image database is used and the retrieval performance of the aggregation scheme is analyzed in details. The rest of the paper is organized as follows. Section 2 introduces large number of image descriptors that we have included and implemented in our content-based image retrieval system. In Section 3 we describe the weighted feature aggregation scheme. In Section 4 we describe the basic characteristics of the image database used in the experiments and we present the retrieval metrics. Section 5 presents the experimental results, and Section 6 concludes the paper.

2 Features and Associated Distance Measures

In content-based image retrieval systems a set of features is used to find visually similar images to presented query image. The word similar has different meaning for different groups of users. Furthermore, the users usually have different criteria of similarity. To satisfy user's different needs different descriptions are required because different features describe different aspects of the image content.

Features can be grouped into the following types: color features, texture features, local features, and shape features. The distance function used to compare the features representing an image obviously has a big influence on the performance of the system. In our system the distance functions were selected according to previous research and experiments concerning their influence in the retrieval process [4], [3]. Table 1 gives an overview of the features implemented in our system. The distance functions for each image feature are presented in Table 1.

2.1 Color histogram

The color histogram represents the color content of an image. Color histogram is a global property of an image and it does not consider the spatial information of pixels. To reduce the computation time, we quantized the 256x256x256=16777216 color images into 8x8x8=512 color images in RGB color space. Since R, G and B channels have same distance in its color space the quantization is done into same levels. The resulting histogram has 512 bins. In accordance to [4], we use Jensen-Shannon divergence to compare the color histograms.

2.2 Color moments

Color moments are compact representation of the color [2]. This descriptor is very suitable for images that contain only one object. It has been shown that most of the color distribution information is captured by three low-order moments. The first-order moment captures the mean color, the second-order moment captures the standard deviation, and the third-order moment captures the color skewness. The best results are obtained in combination with HSV color space. We extract the three low-order moments for each color planes. As a result, we obtain only nine parameters to describe the color image.

2.3 Tamura histogram

Tamura [6] proposes six texture features: coarseness, contrast, directionality, line-likeness, regularity, and roughness. Experiments show that the first three features are very important in the context of human perception [6]. So, we use coarseness, contrast, and directionality to create a texture histogram. The histogram consists of 512 coefficients.

2.4 SIFT histogram

Many different techniques for detecting and describing local image regions have been developed [7]. The Scale Invariant Feature Transform (SIFT) was proposed as a method of extracting and describing keypoints which are reasonably invariant to changes in illumination, image noise, rotation, scaling, and small changes in viewpoint [7].

For content based image retrieval good response times are required and this is hard to achieve using the huge amount of data obtained by local features. A typical image of 500x500 pixels will generate approximately 2000 keypoints. The dimension of this feature is extremely high because the size of the keypoint descriptor is 128 dimensional vector.

To reduce the dimensionality we use histograms of local features [4]. With this approach the amount of data is reduced by estimating the distribution of local features for every image. The key-points are extracted from all database images, where a key-point is described with a vector of numerical values. The key-points are then clustered in 2000 clusters. Afterwards, for each key-point we discard all information except the identifier of the most similar cluster center. A histogram of the occurring patch-cluster identifiers is created for each image. This results in a 2000 dimensional histogram per image.

2.5 MPEG 7 visual descriptors

The Moving Picture Experts Group (MPEG) has defined several visual descriptors in their MPEG-7 standard. An extensive overview of these descriptors can be found in [8]. The MPEG-7 standard defines features that are computationally inexpensive to obtain and compare and strongly optimizes the features with respect to the required storage memory. In our research we used the following MPEG 7 descriptors: Color Structure Descriptor, Color Layout Descriptor, Edge Histogram Descriptor, Dominant Color Descriptor and Region Shape Descriptor.

3 Feature Aggregation

Very often, one visual feature is not sufficient to describe different aspects of the image content. A general CBIR system usually requires multiple features to adequately characterize the content of images. Furthermore, it is expected that a proper combination of different visual features could result in improved performances. In CBIR systems using multiple features, the relevant images are ranked according to an aggregated similarity of multiple feature descriptors computed as:

$$\frac{\sum_{i=1}^{n} d_i}{n}$$

where d_i, (i=1, 2, …, n) is i-th feature distance between the query image and an image in the database.

But, before using multiple features it is necessary to understand the impact of the individual features on the retrieval results. To get a higher system performance, methods for multiple features combination are proposed [5]. The basic method uses equal weights assuming that different features have same importance during the search. But in most cases, they don't have the same importance. The general idea is to assign higher weights to a feature that is more important for the query image as it is shown in Fig. 1 where w_i (i=1, 2, …, n) is i-th weight assign for the corresponding feature. Using this scenario the aggregated similarity is computed as:

$$\frac{\sum_{i=1}^{n} w_i d_i}{n}$$

Weights of the features are usually generated from the query image and the images from the user's relevance feedback based on information theory concepts [11].

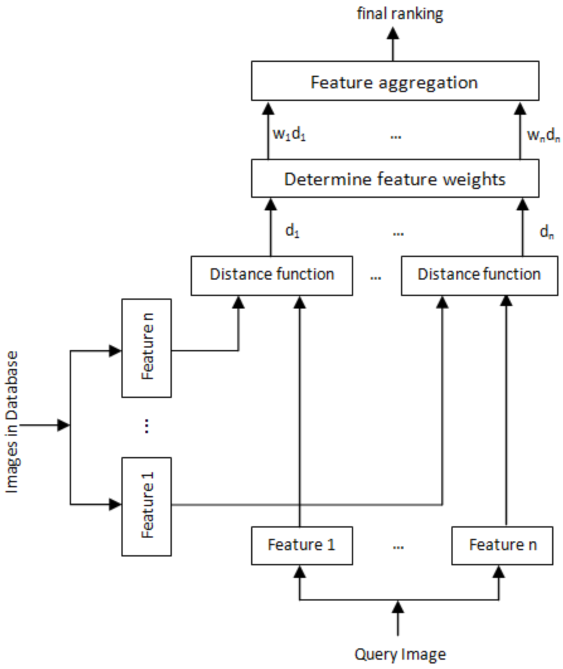

Fig. 1. Feature aggregation in CBIR.

3.1 Normalization of the distances

The distances for each feature vary within a wide range. To ensure equal emphasis of each feature within the feature aggregation scheme we apply normalization of the distances. The distance normalization process produces values in the range [0, 1]. For any pair of images I_i and I_j we compute the distance d_{ij} between them:

$$d_{ij}=distance\_function(F_{Ii},F_{Ij})$$

where F_{Ii} and F_{Ij} are the features of the images I_i and I_j for i, j=1, …, N, where N is the number of images in the database.

For the sequence of distance values we calculate the mean μ and standard deviation σ. We store the values for μ and σ in the database to be used in later normalization of the distances d_i (Fig. 1). After a query image q is presented we compute the distance values between q and the images in database. We normalize the distance values as follows:

$$d_{qI_{norm}} = \frac{1}{2}(1 + \frac{d_{qI} - \mu}{3\sigma})$$

The additional shift of ½ will guarantee that 99% of the distance values are within [0, 1]. For the remaining distances we simply set 1. These images will not affect the retrieval performance because of their dissimilarity with the query image. We convert the distance values into similarity values using $(1 - d_{qI_{norm}})$. At the end of this normalization all similarity values for all features have been normalized to the same range [0, 1] and value 1 means exact match and 0 denotes maximum dissimilarity.

3.2 Weighted feature aggregation

The concept of relevance feedback was introduced into CBIR from text-based information retrieval [13] in the 1990s. With relevance feedback, a user can label a few images as new examples for the retrieval engine if he is not satisfied with the retrieval result. Actually, these new images refine the original query which enables the relevance feedback process to overcome the gap between high-level image semantics and low-level image features.

To learn the weights w_i for the feature aggregation scheme we proceed in similar manner to the algorithms proposed for weighted L_2 distances in [12]. The learning algorithms are derived by minimizing the Leaving-One-Out classification error of the given training set. In our case the labeled images from the relevance feedback were considered as training images for the nearest neighbor system. To improve the performance, we learn weights that minimize the distances among the positively marked images and maximize the distances between the positively marked images and negatively marked images.

4 Benchmark Database for CBIR

We evaluate the methods on public image database called WANG. This database was created by the group of professor Wang from the Pennsylvania State University [10]. The WANG database is a subset of 1000 images of the Corel stock photo database which have been manually selected. The images are divided into 10 classes with 100 images each. Some example images from this database are shown in Fig. 2.

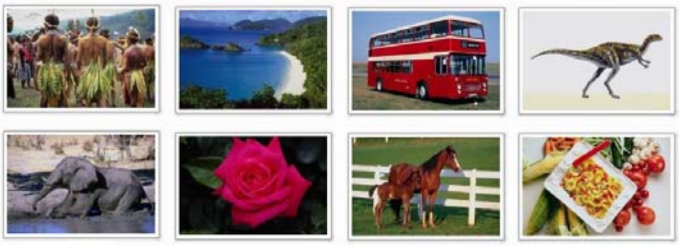

Fig.2. Example images from the WANG database.

4.1 Retrieval metric

Let the database images are denoted by $\{x_1, \ldots, x_i, \ldots, x_N\}$ and each image is represented by a set of features. To retrieve images similar to a presented query image q, the system compares each database image x_i with the query image by an appropriate distance function $d(q, x_i)$. Then, the database images are sorted to fulfill the following distance relation $d(q, x_i) \leq d(q, x_{i+1})$ for each pair of images x_i and x_{i+1} in the sequence $(x_1, \ldots, x_i, \ldots, x_N)$.

Several performance measures based on the precision P and the recall R have been proposed for CBIR systems evaluation [9]. Precision and recall values are usually represented by a precision-recall-graph R -> $P(R)$ summarizing (R, $P(R)$) pairs for varying numbers of retrieved images. The most common approach to summarize this graph into one value is the mean average precision (*MAP*). The average precision *AP* for a single query q is the mean over the precision scores after each retrieved relevant item:

$$AP(q) = \frac{1}{N_R} \sum_{n=1}^{N_R} P_q(R_n)$$

where R_n is the recall after the n-th relevant image was retrieved. N_R is the total number of relevant documents for the query. The mean average precision *MAP* is the mean of the average precision scores over all queries:

$$MAP = \frac{1}{|Q|} \sum_{q \in Q} AP(q)$$

where Q is the set of queries q. An advantage of the mean average precision is that it contains both precision and recall oriented aspects and is sensitive to the entire ranking.

5 Experimental Results

For the evaluation of the different features on the WANG database a leaving-one-out approach has been followed. Every image was used as a query and the remaining 99 images from the same class as the current query image were considered relevant and the images from all other classes were considered irrelevant.

The experimental results for the weighted feature aggregation were performed automatically with simulated user feedback. Each query was performed and all relevant images retrieved among the top 20 results were added to the set of positive samples and all non relevant images among these were added to the set of negative samples. With this approach we have effectively simulated a user who is judging each of the twenty top ranked images regarding its relevance. This procedure was repeated three times.

For the selected features and aggregation methods we reported the mean average precision and complete PR graph.

Table 1 summarizes the result for the MAP on the selected database. Fig. 3 shows the corresponding PR graphs for the selected features and aggregation methods. It shows that different features perform differently on the selected database.

Table 1. Features and their associated distance measures.

Feature	Distance function	MAP (%)
Color histogram	Jensen-Shannon divergence	51.24
Color moments	Euclidean distance	36.76
Tamura histogram	Jensen-Shannon divergence	34.61
SIFT histogram	Jensen-Shannon divergence	48.94
MPEG 7: Edge histogram descriptor	MPEG7-internal distance	40.23
MPEG 7: Color structure descriptor	MPEG7-internal distance	48.37
MPEG 7: Color layout descriptor	MPEG7-internal distance	42.39
MPEG 7: Dominant color descriptor	MPEG7-internal distance	40.21
MPEG 7: Region-based descriptor	MPEG7-internal distance	23.25
Feature aggregation		57.25
Weighted feature aggregation		68.19

The color histogram and color structure descriptor are most suitable for the selected database (mean average precision approximately equals to 50%). In general all color descriptors (color layout, color moments and dominant color descriptor) gave satisfactory results with mean average precision approximately equals to 40. The rest of the features didn’t perform well for the selected database and there is an obvious need for aggregation.

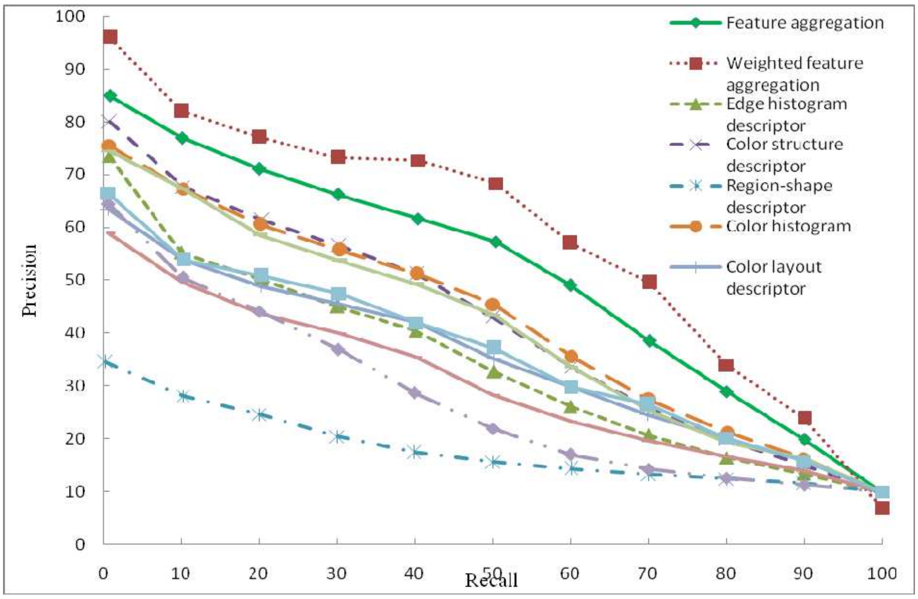

Fig.3. PR graphs for each of the selected features.

The results in Table 1 show that the performance was improved by adding more features in the retrieval process. Furthermore, the best results were obtained using weighted feature aggregation. Compared with the basic feature aggregation method the performance increases by 11%.

6 Conclusions

This paper describes a content-based image retrieval system for aggregation and combination of different image features. For it, we have implemented and tested various feature extraction algorithms. The CBIR system supports query by image retrieval. The query interface supports inclusion of more than one feature in the retrieval process.

Using one feature is not good enough to retrieve target images in all the cases. To get a better retrieval performance feature aggregation is necessary. The feature aggregation with equal weights provides improvement over the individual features, but to improve precision the salient features need to be determined and assigned higher weights.

The relevance feedback technique is used to determine and capture more precisely the query concept presented by the user. Weights for the features can be generated from the images involved in the relevance feedback process. The results obtained with weighted feature aggregation are clearly the best.

The implemented system has several advantages over the existing content-based image retrieval systems. The diversity of the implemented features included in our system allows the user to adapt the system to the query image. The weighted combination of features allows flexible query formulations and helps processing specific queries for which users have no knowledge about any suitable descriptors.

References

1. Datta, R., Joshi, D., Li, J., Wang, J. Z.: Image Retrieval: Ideas, Influences, and Trends of the New Age", ACM Transactions on Computing Surveys, 40(5), (2008).
2. Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., Equitz, W.: Efficient and Effective Querying by Image Content", Journal of Intelligent Information Systems, Vol. 3, No. 3/4, pp. 231–262, (1994).
3. Eidenberger, H.: Distance measures for MPEG-7-based retrieval, Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, Pages: 130 - 137, Berkeley, California, (2003).
4. Deselaers, T., Keysers, D., Ney, H.: Features for Image Retrieval: An Experimental Comparison", Information Retrieval, vol. 11, issue 2, The Netherlands, Springer, pp. 77-107, (2008).
5. Zhang, J., Ye, L.: An Unified Framework Based on p-Norm for Feature Aggregation in Content-Based Image Retrieval, Ninth IEEE International Symposium on Multimedia, page(s): 195-201, Taichung, (2007).
6. Tamura, H., Mori, S., Yamawaki, T.: Textural Features Corresponding to Visual Perception, IEEE Transaction on Systems, Man, and Cybernetics, 8(6), pp. 460–472, (1978).
7. Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 60(2), pp. 91–110, (2004).
8. Martinez., J. M.: Overview of the MPEG-7 Standard, v 6.0, MPEG Requirements Group, ISO/MPEG N4674, (2002).
9. Muller, H., Muller, W., Squire, D. McG., Marchand-Maillet, S., Pun, T.: Performance Evaluation in Content-Based Image Retrieval: Overview and Proposals, Pattern Recognition Letters (Special Issue on Image and Video Indexing), 22(5), pp. 593–601, (2001).
10. Wang, J. Z., Li, J., Wiederhold, G.: SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 23, no.9, pp. 947-963, (2001).
11. Deselaers, T., Weyand, T., Ney, H.: Image Retrieval and Annotation Using Maximum Entropy, CLEF Workshop 2006, Alicante, Spain, (2006).
12. Paredes, R., Vidal, E.: Learning weighted metrics to minimize nearest neighbor classification error. PAMI, 28(7):1100-1110, (2006).
13. Rocchio, J.: Relevance feedback in information retrieval. In The SMART Retrieval System: Experiments in Automatic Document Processing, pp. 313-323. Prentice-Hall, Englewood Cliffs, NJ, USA, (1971).