Selection of MPEG-7 Image Features for Improving Image Similarity Search on Specific Data Sets

Peter L. Stanchev
Kettering University, Flint, Michigan 48504, USA

Giuseppe Amato, Fabrizio Falchi, Claudio Gennaro, Fausto Rabitti, Pasquale Savino
ISTI-CNR, Pisa, Italy

Abstract

In this paper a technique for evaluating the effectiveness of MPEG-7 image features on specific image data sets is proposed. It is based on well defined statistical characteristics. The aim is to improve the effectiveness of the image retrieval process, based on the similarity computed on these features. This technique is validated with extensive experiments with real users.

1. Introduction

With the continuous increase of production of images in digital format, the problem of retrieving stored images by content from large image archives is becoming more and more relevant. A very important direction towards the support of *content-based image retrieval* is *feature based similarity access*. A feature (or *content-representative metadata*) is a set of characteristics of the image, such as color, texture, and shapes. Similarity based access means that the user specifies some characteristics of the wanted information, usually by an example image (e.g., find images similar to this given image, represents the *query*). The system retrieves the most relevant objects with respect to the given characteristics, i.e., the objects *most similar* to the query. Such approach assumes the ability to measure the distance (with some kind of metric) between the query and the data set images. Another advantage of this approach is that the returned images can be ranked by decreasing order of similarity with the query, presenting to the user the most similar images first. A very important contribution to the practical use of this approach has been the standardization effort represented by MPEG-7, intending to provide a normative framework for multimedia content description. In MPEG-7, several features have been specified for images as visual descriptors.

A lot of research effort has been devoted to the image retrieval problem, adopting the similarity based paradigm, in the last 20 years [1]. Industrial systems, such as QBIC (IBM Query by Image Content) [2], VisualSEEk [3], Virage's VIR Image Engine [4], and Excalibur's Image RetrievalWare [5] are available today. The results achieved with this generalized approach are often unsatisfactory for the user. These systems are limited by the fact that they can operate only at the primitive feature level while the user operates at a higher semantic level. None of them can search effectively for, say, a photo of a dog [6]. This mismatch is often called the *semantic gap* in the image retrieval. Although it is not possible to fill this gap in general terms there is evidence that combining primitive image features with text keywords or hyperlinks can overcome some of these problems, though little is known about how such features can best be combined for retrieval [6].

There is evidence that image features work with different levels of effectiveness depending on the characteristics of the specific image data set. Eidenberger [7] analyses descriptions based on MPEG-7 image features from the statistical point of view on 3 image data sets. He founds, as everybody would expect, that Color Layout, like Color Structure, perform badly on monochrome images. Dominant Color performs equally well on the 3 data sets, etc. This study demonstrates that, even if it not possible, in general, to overcome the semantic gap in image retrieval by feature similarity, it is still possible to increase the retrieval effectiveness by a proper choice of the image features, among those in the MPEG-7 standard, depending on the characteristics of the various image data sets (obviously, more homogeneous the data set is, better results can be obtained).

In this paper we generalize this result. We propose a technique for evaluating the effectiveness of MPEG-7 image features on specific image data sets, based on well defined statistical characteristics of the data set. The aim is to improve the effectiveness of the image retrieval process based on the computed similarity on these features. We also validate this method with extensive experiments with real users.

The layout of the paper is as follows. In section 2 we explain the proposed technique to image feature selection. In section 3 we describe the testing environment. In

section 4 we analyze the results, and finally in section 5 the conclusions and future work are presented.

2 The proposed approach to image feature selection

The major aim of this paper is to develop a technique that allows determining the image features that provide the best retrieval effectiveness for a specific application domain or for a specific data set. Due to the availability of specific image features used in the MPEG-7 standard [8], we base our evaluation on them. The results of the work presented in this paper are more general and can be applied to any feature set, used to support image similarity retrieval.

We used six different features (visual descriptors) defined in MPEG-7 for the indexing of images [9]: Scalable Color (***SC***), Dominant Color (***DC***), Color Layout (***CL***), Color Structure (***CS***), Edge Histogram (***EH***) and Homogeneous Texture (***HT***).

In order to pursue our main objective, we perform an extensive user evaluation of the effectiveness of the different image features. Given a specific data set, users should make their relevance assessment by ranking the objects in the data set for a given query. For the same query and by using a specific image feature, we also develop a system that ranks the images in the data set. Our aim is to determine, if exists, for each image feature a characteristic of the data set that allows one to predict (or to emulate) user's behavior. It is possible to reuse the same measure for other data sets without the need of any further validation made by users. Users are involved only during this phase needed to validate the proposed measure. The results reported in this paper are preliminary, since they are based on a single data set and a single relevance measure. We are continuing the experiments by using other data sets, and other relevance measures. It must be observed that the same user assessments can be used to study the behaviour of different relevance measures.

Let us consider a data set composed of N images $(I_1,\ldots,I_N)$, and let us indicate the query as Q. For a specific visual descriptor *vd* the distance between image I_i and the query Q is defined as $d_{vd}(Q,I_i)$. This distance function is an evaluation of the dissimilarity between images. The similarity function can be obtained in different ways from a distance function (e.g. *s=1-d* if *d* is in the range [0,1]).

All the images in the data set can be ranked according to the distance measure d_{vd}. We obtain an ordered list of couples $\left((I_1^{'}, d_{vd}(Q,I_1^{'})),\ldots,(I_N^{'}, d_{vd}(Q,I_N^{'}))\right)$ where $d_{vd}(Q,I_i^{'}) \le d_{vd}(Q,I_j^{'})$ if $I_i^{'}$ proceeds $I_j^{'}$ in the list. Let us consider that a generic query returns to the user k images, ordered in increasing distance $d_{vd}(Q,I)$ (decreasing similarity) with respect to Q. In this paper we evaluate if the following measure, using as queries all the images in the data sets, is appropriate to predict the retrieval effectiveness for a given visual descriptor *vd*:

$$R_k = \frac{avg_Q(Q,I_{Q,k}) - avg_Q(Q,I_{Q,1})}{D}$$

where $avg_Q(Q,I_{Q,1})$ is the average distance measure between the queries and the most similar image (not considering the query image itself). Similarly we define $avg_Q(Q,I_{Q,k})$ where $I_{Q,k}$ is the image ranked k for the given query image Q. D is the average distance between all images in the data set.

This measure is an estimation of the distances between the first k retrieved images in the data set, for the image visual descriptor *vd*. Higher values of R are expected to provide a good "distinction", so that the visual descriptor *vd* is expected to provide good retrieval effectiveness. The intuition suggests that if the k images retrieved are very close one to each other, the users will find difficulties to distinguish between good and bad results.

3 The testing environment

An essential step to validate the usability of this measure requires the evaluation of user's retrieval assessment for a given data set. The user relevance assessments are usually difficult to perform and may require an extensive effort. The standard information retrieval method, based on precision and recall [10], would require providing a user with a query and asking him to go in the entire data set and select the most appropriate images, matching the query image. This technique cannot be adopted if the size of the data set is larger than few hundreds of images. Furthermore, in order to emulate a real world environment, the size of the data set must be larger than several thousands of images.

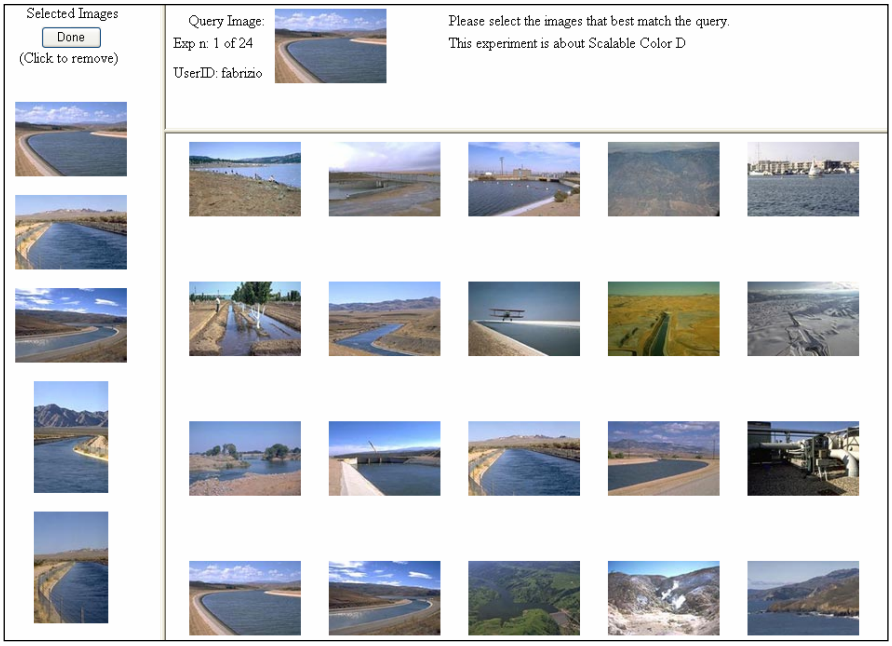

Figure 1. Web experiment interface for image selection

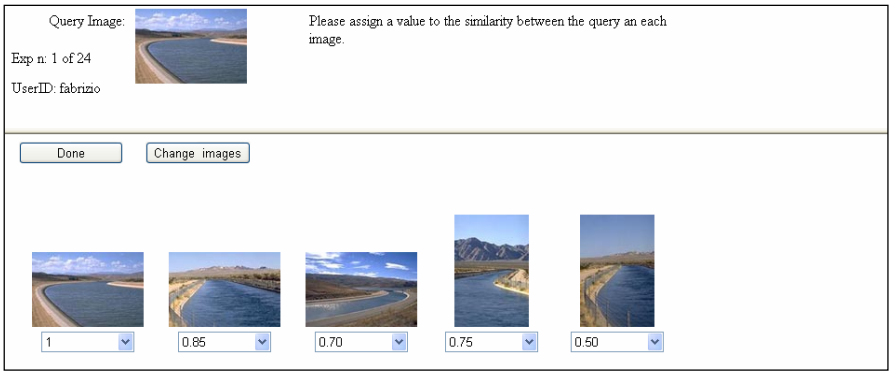

Figure 2. Web experiment interface for similarity value assignment

Our testing environment is composed of three main elements:

- Three image data sets;
- Six image features (MPEG-7 visual descriptors);
- A software module that performs similarity retrieval of images by using different image features and allows users to express their relevance assessment on the retrieved images.

We used the following data sets:

- A subset of the image collection of the *Department of Water Resources* in California. It is available from UC Berkeley (removing B&W and animals we used 11,519 images);
- 21,980 key frames extracted from the TREC2002 video collection (68.45 hrs MPEG1);
- 1,224 photos of the *University of Washington* (UW), Seattle.

In this preliminary evaluation we use only the *Department of Water Resources* collection.

To retrieve images similar to the query we need visual descriptors and a distance functions. MPEG-7 defines some visual descriptors but doesn't standardize the distance functions. We used the same distance function used in the MPEG-7 Reference Software [11] and suggested in [12].

Next table illustrates the characteristics of the six MPEG-7 visual descriptors we used [9].

VD	Description
SC	Based on the color histogram in HSV color space encoded by a Haar transform. We used the 64 coefficients form.
DC	A set of dominant colors taking in considerations their spatial coherency, the percentage and color variance of the color in the image. We used the complete form.
CL	Based on spatial distribution of colors. It is obtained applying the DCT transformation. We used 12 coefficients
CS	Based on color distribution and local spatial structure of the color. We used the 64 coefficients form.
EH	Based on spatial distribution of edges (fixed 80 coefficients).
HT	Based on the mean energy and the energy deviation from a set of frequency channels. We used the complete form.

The data set has been indexed by using the six MPEG-7 descriptors. The software module, based on the MPEG-7 Reference Software [11], permits the indexing of images in the data set for all six different descriptors. It supports image similarity retrieval, based on the computation of the distances between the query and the images in the data set. The software can be accessed from a web browser that allows the user, after a login procedure, to perform the following tasks:

- An image is randomly selected from the data set and it is used as the image query. For the given query image we select the most similar images in the data set according to a given descriptor;
- The 50 most similar images are selected and presented to the user together with 10 images randomly selected from the data set (Figure 1). All 60 images are shown to the user without any indication of their relevance to the query and in a random order (note that one of the retrieved images is the query itself, which is part of the data set);
- The user selects, among the 60 presented to him, images he considers most similar to the query. He can choose between 5 to 10 images. In order to determine if the user evaluation is reliable, we verify if he selects the image corresponding to the query. If this does not happen the experiment is rejected;
- The user assigns a relevance judgment to each selected image as a score in the range [0, 1] (Figure 2) with a granularity of 0.05.

This evaluation is repeated for all different descriptors by all users. The experiments reported in this paper have been performed by 90 users.

4. Analysis of results

The analysis of the results, obtained by the interaction of the users with the testing environment, returns information that can be used to judge the capability of each visual descriptor. We have represented the quality of a visual descriptor considering three different aspects:

I. the capability of ranking the result of a query consistently with the average rank produced by the users;
II. the capability of retrieving in the top *k* results images that were scored high also by the users;
III. the capability of retrieving images that are evidently distinguishable from randomly generated result sets.

In order to have an objective evaluation of visual descriptors according to these three criteria, we have defined three measures that can be computed from the data obtained by the testing environment.

Criteria I defines a concept of rank quality. We measure it by computing the average distance between the rank generated by the visual descriptor and that produced by a user. Let $\{ I_1,\dots I_m \}$ be the set of images retrieved by the visual descriptor *vd* processing the query *Q*. Let $\{ r_{I_1}^{vd,Q},\dots r_{I_m}^{vd,Q} \}$ be the rank assigned to these images by the visual descriptor. And let $\{ r_{I_1}^{u,Q},\dots r_{I_n}^{u,Q} \}$, $n \leq m$, be the rank provided by the user for the same query. The rank quality $RQ^{vd,u,Q}$ is defined as:

$$RQ^{vd,u,Q} = \frac{\sum_{i=1}^{n} \left| r_{I_i}^{u} - r_{I_i}^{vd} \right|}{(m-1)\cdot n}$$

We obtain the average rank quality of a visual descriptor *vd*, by computing the average of $RQ^{vd,u,Q}$ by varying the query *Q* and the user *u* as follows:

$$RQ^{vd} = avg_Q\left(avg_u\left(RD^{vd,u,Q}\right)\right)$$

Criteria II concerns the quality of the elements retrieved by the visual descriptor. It might happen that, even if the result set is correctly ranked, the retrieved elements have a low quality, which means that they are judged by the user not to be really relevant. We measure this fact by considering the scores associated by the users to the images retrieved with the visual descriptor. Let $\{ s_{I_1}^{u,Q},\dots s_{I_m}^{u,Q} \}$ be the scores assigned by the user to the images retrieved with the visual descriptor *vd,* supposing that the score 0 is assigned in case an image is not selected by the user. We define the score quality

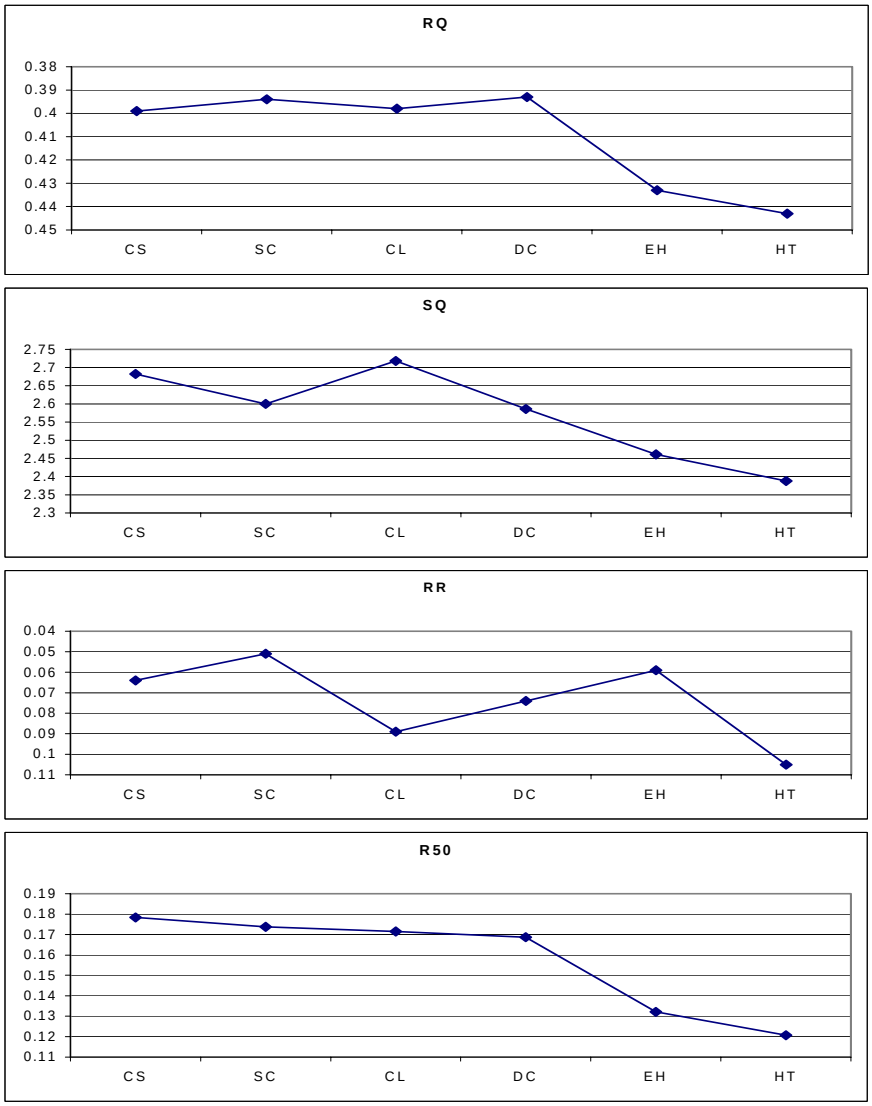

Figure 3. Analysis of results

$SQ^{vd,u,Q}$ by computing the sum of the scores assigned by the user *u* as:

$$SQ^{vd,u,Q} = \sum_{i=1}^{m} s_{I_i}^{u,Q}$$

Also in this case we compute the average score quality of visual descriptor *vd* as:

$$SQ^{vd} = avg_Q\left(avg_u\left(SQ^{vd,u,Q}\right)\right)$$

The measure for assessing criteria III is defined as the ratio between the number of images selected by the user, $rs^{u,Q}$, taken from those randomly generated, and the total number of images selected by the user $ts^{u,Q}$. The random ratio $RR^{vd,u,Q}$ is computed as:

$$RR^{vd,u,Q} = \frac{rs^{u,Q}}{ts^{u,Q}}$$

We compute the average score quality of visual descriptor *vd* as follows:

$$RR^{vd} = avg_Q\left(avg_u\left(RR^{vd,u,Q}\right)\right)$$

Figure 3 shows the results that we obtained using the Berkeley data set. Since the query is part of the data set, one of the retrieved images is the query itself; this image has not been included in the computations reported in

Figure 3. In section 2 we described R_k and here we use *k=50* because in the experiments we selected the 50 images most similar to the query. From the graphs it can be seen that the trend of the quality of the descriptors is on average compatible with the trend obtained with our quality predictor R_{50}. In fact we can reliably distinguish between good and bad visual descriptors for the data set. High values of R_{50} correspond on average to high values of the quality estimators rank quality, *RQ,* and score quality, SQ. In fact EH and HT are always distinguishable from CS and SC. The random ratio *RR* is small in percentage, meaning that the probability that the user selects a random image is low. Therefore, even if *RR* has a trend that is not similar to the trend of the predictor, this is not important.

5 Conclusions and future work

Several visual descriptors exist for representing the physical content of images, as for instance color histograms, textures, shapes, regions, etc. Depending on the specific characteristics of a data set, some features can be more effective than others when performing similarity search. For instance, descriptors based on color representation might result not to be effective with a data set containing mainly black and white images. We have proposed a methodology for predicting the effectiveness of a visual descriptor on a target data set. The technique is based on statistical analysis of the data set and queries. Experiments, where we assessed the quality of visual descriptor from the user perspective, have demonstrated the reliability of our approach. In fact, the experiments were conducted with a large number of users to guarantee the soundness of the analysis of results. We are currently testing our approach with other data sets to have additional confirmations of its validity.

As a future work we are seeking for extensions of this technique to a query driven feature selection. The proposed technique is able to choose the most promising query given a target data set. This extension would choose the best feature, taking into consideration the target data sets and the query itself.

We plan to exploit these results, in the context of the design of MILOS, a Multimedia Content Management System under development in ISTI-CNR, Pisa. The rationale is that (as affirmed in [6]) in a system like MILOS, the retrieval process is based on the combination of different types of data (like attribute data and text components) and metadata of different media (typically MPEG-7 for image and audio/video). In this context, we are able to accept the inherent limitations of the similarity-based image retrieval process, as long as it can improve the retrieval process without images. To improve this complex retrieval process, we need to use a method, as proposed in this paper, for selecting MPEG-7 image features to exploit the statistical characteristic of each image data set, managed by MILOS.

References

[1] Yong Rui, Thomas S. Huang, Shih-Fu Chang, "Image Retrieval: Current Techniques, Promising Directions And Open Issues", Journal of Visual Communication and Image Representation, 1999

[2] QBIC™-IBM's Query By Image Content. http://wwwqbic.almaden.ibm.com

[3] J. R. Smith and S.-F. Chang. "VisualSEEk: a fully automated content-based image query system", Proceedings of ACM Multimedia '96, pp. 87-98, 1996

[4] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain, C. Shu, "The Virage Image Search Engine: An open framework for image management", Proc. SPIE, Storage and Retrieval for Still Image and Video Databases, 1996

[5] J. Dowe, "Content-based retrieval in multimedia imaging", Proc. SPIE. Storage and Retrieval for Image and Video Database, 1993

[6] J.P. Eakins, M.E. Graham "Content Based Image Retrieval: A report to the JISC Technology Applications Program", Institute for Image Data Research, Univ. of Northumbria at Newcastle, 1999

[7] H. Eidenberger, "How good are the visual MPEG-7 features?", SPIE & IEEE Visual Communications and Image Processing Conference, Lugano, Switzerland, 2003

[8] MPEG, "MPEG-7 Overview (version 9)", ISO/IEC JTC1/SC29/WG11N5525

[9] MPEG-7, "Multimedia content description interfaces. Part 3: Visual", ISO/IEC 15938-3:2002

[10] G. Salton, M.J. McGill, "An Introduction to Modern Information Retrieval", McGraw-Hill, 1983

[11] MPEG-7, "Multimedia content description interfaces. Part 6: Reference Software", ISO/IEC 15938-6:2003

[12] B.S. Manjunath, P. Salembier, T. Sikora, "Introduction to MPEG-7", Wiley, 2002