

IMPROVING IMAGE SIMILARITY SEARCH EFFECTIVENESS IN A MULTIMEDIA CONTENT MANAGEMENT SYSTEM

Giuseppe Amato, Fabrizio Falchi, Claudio Gennaro, Fausto Rabitti, Pasquale Savino
{giuseppe.amato, fabrizio.falchi, claudio.gennaro, fausto.rabitti, pasquale.savino}@isti.cnr.it
ISTI-CNR, Pisa, Italy

Peter L. Stanchev
pstanche@kettering.edu
Kettering University, Flint, Michigan 48504, USA

ABSTRACT

In this paper, a technique for making more effective the similarity search process of images in a Multimedia Content Management System is proposed. The content-based retrieval process integrates the search on different multimedia components, linked in XML structures.

Depending on the specific characteristics of an image data set, some features can be more effective than others when performing similarity search. Starting from this observation, we propose a technique that predicts the effectiveness of MPEG-7 image features based on a statistical analysis of the specific data sets in the Multimedia Content Management System.

This technique is validated through an extensive experimentation with real users.

KEYWORDS

MPEG-7, Image retrieval, Multimedia content management system

1. INTRODUCTION

More and more digital images and video are being captured and stored. In order to use this information, an efficient retrieval technique is required. 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 images with respect to the given characteristics, i.e., the most similar to the query. This

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.

In the last 20 years, a lot of research effort has been devoted to the image retrieval problem, adopting the similarity-based paradigm [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 these generalized approaches 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. 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]. Several semantic image and video models were suggested [7], [8].

There is evidence that different image features work with different levels of effectiveness depending on the characteristics of the specific image data set. Eidenberger [9] analyses descriptions based on MPEG-7 image features from the statistical point of view on three image data sets. For example, he found that Color Layout, like Color Structure, perform badly on monochrome images,

that Dominant Color performs equally well on the three data sets, etc. This study demonstrates that, even if it is 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 define a method that permits to select the most appropriate image features for each image data set, in order 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.

We believe that these results can be practically exploited in the context of Multimedia Content Management Systems and Multimedia Digital Library Systems, where the retrieval process is based on the combination of different techniques for accessing different types of components, like attribute data, text, images and audio/video, etc. This kind of systems is becoming increasingly popular in important application areas like publishing, broadcasting, cultural heritage preservation, healthcare and medicine, biology, e learning, etc. At ISTI-CNR in Pisa we have developed a Multimedia Content Management System, called MILOS [10]. All metadata in the systems, like MPEG-7 for multimedia components, are represented in XML. A key characteristic of MILOS is to support user queries on different multimedia components. The retrieval process is able to integrate the search functions on several multimedia components, linked in XML structures.

In the context of the MILOS project, this research is motivated by the need to improve the effectiveness of similarity-based access to multimedia components by exploiting the statistical characteristic of each multimedia data set. We are now focusing on images represented by MPEG-7 descriptors, and we plan to extend this approach to other media, especially audio and video.

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 our testing environment. In Sections 4 and 5 we analyze the results and their exploitation, and finally in Section 6 the conclusions and future work are presented.

2. IMAGE FEATURE SELECTION APPROACH

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 [11], we base our evaluation on them. As distance functions, we used those suggested by the MPEG-7 group. Moreover, since the datasets used for the experiments are heterogenous, we believed that the results of the work presented in this paper are suitable to be generalized and can be applied to any feature set, used to support image similarity retrieval.

We used six different visual descriptors defined in MPEG-7 for the indexing of images [12]: 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 performed 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 developed a system that ranks the images in the data set. Our aim is to develop an *analytical quality measure*, which we will simply refer to as *measure*, which allows assessing how much a visual descriptor is able to emulate the user perception of image similarity.

Our tests on the MILOS system suggest that 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 same user assessments can be used to study the behavior of different relevance measures (see Section 4 for details).

To evaluate the quality of the descriptors we performed two types of experiments:

- **Single descriptor experiments**, where just one descriptor per experiment was used. In these experiments, we took in consideration the rank quality of the retrieved image comparing the descriptor rank with those coming from the users.
- **Compound descriptors experiments**, where the results, for the same query, coming from the different six descriptors were presented together to the user. In this case, we took in consideration the sum of the weights assigned by the users to the images coming from each descriptor.

Let us consider a data set composed of N images (I_1, \dots, 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 the 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 images in the data set can be ranked according to the distance measure d_{vd} with respect the query Q . We obtain an ordered list of pairs $(\langle I'_1, d_{vd}(Q, I'_1) \rangle, \dots, \langle I'_N, d_{vd}(Q, I'_N) \rangle)$, where $d_{vd}(Q, I'_i) \leq d_{vd}(Q, I'_j)$ if I'_i precedes 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 (that can be obtained by using as queries all the images in the data sets) is appropriate to predict the retrieval effectiveness of a given visual descriptor vd :

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

where $avg_Q(Q, I_{Q,1})$ is the average distance between the queries and the most similar image (not considering the query image itself). Similarly we define $avg_Q(Q, I_{Q,k+1})$ where $I_{Q,k+1}$ is the $(k+1)$ -th image ranked for the given query image Q . D is the average distance between all images in the data set. This measure depends on k (the size of the retrieved set), but from the experimental evaluation we will observe that for typical values of k (between 10 and 50) R_k does not varies significantly.

This measure is related to the difference between the average distances of the first retrieved image and of the $(k+1)$ -th nearest image. Higher values of R_k are expected to provide a good “distinction” among the retrieved images, so that the visual descriptor vd is expected to provide good retrieval effectiveness. In fact, the intuition suggests that if the $(k+1)$ -th image retrieved is, on average, not much more distant from the query than the most similar image, then the k images retrieved are more or less at the same distance from the query. Therefore, they should be not very distinguishable.

3. OUR TESTING ENVIRONMENT

An essential step to validate the usability of R_k 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 [13], would require that the users go through the entire data set in order to select the images that had better match the query. This technique cannot be adopted if the size of the data set is larger than few hundreds of images. In order to emulate a real world environment, the size of the data set must be larger than several thousands of images. For this reasons, we adopted a different approach, as described in the following.

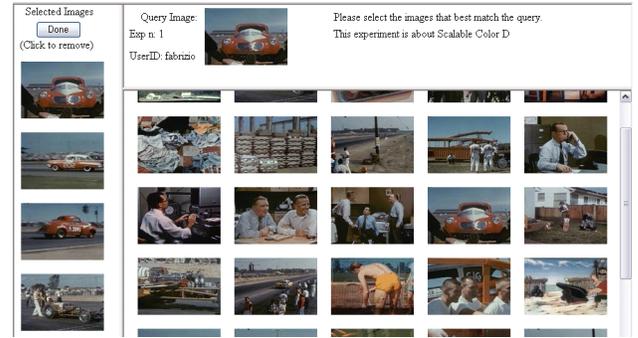


Figure 1 - web experiment interface for image selection.

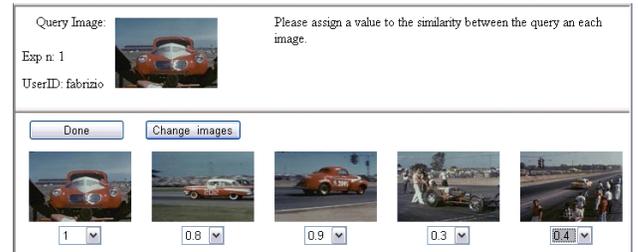


Figure 2 - Web experiment interface for similarity value assignment.

Our testing environment is composed of three main elements:

1. Three image data sets;
2. Six image features (MPEG-7 visual descriptors);
3. 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:

1. 21,980 key frames extracted from the TREC2002 video collection (68.45 hrs MPEG-1);

2. 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);
3. 1,224 photos from the University of Washington (UW), Seattle.

To retrieve images similar to the query we need visual descriptors and a distance functions. MPEG-7 defines some visual descriptors but does not standardize the distance functions. We used the same distance function used in the MPEG-7 Reference Software [14] and suggested in [15].

We use the following six MPEG-7 visual descriptors [12]:

1. SC, based on the color histogram in HSV color space encoded by a Haar transform. We used the 64 coefficients form;
2. DC, presents 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;
3. CL, based on spatial distribution of colors. It is obtained applying the DCT transformation. We used 12 coefficients;
4. CS, based on color distribution and local spatial structure of the color. We used the 64 coefficients form;
5. EH, based on spatial distribution of edges (80 fixed coefficients);
6. 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 [14], 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 of 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:

1. An image is randomly selected from the data set and it is used as image query. For the given query image we select the most similar images of the data set according to a given descriptor;
2. We have developed two type of experiments: for the *single descriptor experiments* the 50 most similar images are selected using one descriptor; for the *compound descriptors experiments* the 10 most similar images for each descriptor are selected and presented together to the user;
3. The images are presented to the user (Figure 1) in a random order without any

indication of their relevance to the query (note that one of the retrieved images is the query itself, which is part of the data set);

4. The user selects, among the images, images he/she considers most similar to the query. He/she can choose between 5 to 10 images. In order to determine if the user evaluation is reliable, we verified if he/she selected the image corresponding to the query. When this did not happen the experiment was rejected;
5. 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.

All users repeat this evaluation for all different descriptors. Ninety users have performed the experiments, reported in this paper.

4. ANALYSIS OF RESULTS

The experimentation aims to verify if the *analytical quality measure* (R_k) defined in Section 2, can be used to select, for a given dataset, the most appropriate visual descriptors. We found that the values R_k , for each visual descriptor, are correlated to the user's *experimental quality assessment* for the same descriptor.

Our methodology is based on the following steps:

1. We define two distinct experimental quality assessments, which are able to quantify how much the images selected by users are in agreement with the result produced by a visual descriptor for the same query. For instance, given a query Q , let us suppose that a visual descriptor returns the images I_1, I_2, \dots, I_{10} , while the user considers relevant only the images I_3, I_4, I_7 . The quality measure will measure how much the rank of the user is similar to the rank produced by the visual descriptor. The experimental quality measures used in the paper are described in detail later in this section.
2. In order to validate our analytical quality measure R_k , we use the correlation coefficient between the vector of the values of the experimental assessments for all the descriptor, and the vector of R_k for the same descriptors. The correlation coefficient measures how closely two variables co-vary. It can vary from -1 (perfect negative correlation) through 0 (no correlation) to +1 (perfect positive correlation). For instance, suppose we have three visual descriptors D_1, D_2 , and D_3 . Just as example, suppose the values of the experimental quality assessments, are (0.3, 0.45, 0.8), and the values of analytical quality measures are (1.6, 1.9, 2.8), for the same descriptors. The correlation coefficient between the two vectors is

1, which means that our analytical measure behaves as the experimental assessments.

3. As it is possible to see from the definition (1), our analytical measure is a function of the size of the retrieved set (k). For a given dataset, we study the correlation coefficient tendency as function of k in order to find its optimum.

We defined two experimental quality assessments:

1. **Score Quality:** the capability of retrieving in the first k results the images that received by the users' higher scores.
2. **Rank Quality:** the capability of ranking the result set of a query coherently with the average rank produced by the users.

Score Quality

The score quality (SQ) concerns the quality of the elements retrieved by the visual descriptors. It might happen that, even if the result set is correctly ranked, the retrieved elements are of limited relevance for the user. Another descriptor can return results that are more relevant. SQ is needed because user's assessments are not performed on the entire dataset, but are limited to the top k images retrieved by the system. Since this is not feasible for large datasets, we ask the users to provide their assessment about the quality of the retrieved images (i.e., how much they are similar to the query).

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 , assuming that the score 0 is assigned in case an image is not selected by the user. We define the score quality $SQ^{vd,u,Q}$ by computing the sum of the scores assigned by the user u as:

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

Note that, SQ ranges from 0 to m .

The average score quality of the visual descriptor vd is defined as:

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

Rank Quality

We measure the rank quality by computing the average distance between the rank generated by the visual descriptor and the one produced by the user. We indicate with $\{I_1, \dots, I_m\}$ the set of the images retrieved by the visual descriptor vd processing the query Q , and with $\{r_{I_1}^{vd,Q}, \dots, r_{I_m}^{vd,Q}\}$ the rank assigned to these images by the

visual descriptor; $\{r_{I_1}^{u,Q}, \dots, r_{I_m}^{u,Q}\}$, $n \leq m$, is the rank provided by the user for the same query.

The rank quality $RQ^{vd,u,Q}$ is defined as:

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

Note that, the values of RQ vary from 0 (lowest quality) to 1 (highest quality). We define the average rank quality of a visual descriptor vd , by averaging $RQ^{vd,u,Q}$ for different queries Q and users u as follows:

$$RQ^{vd} = avg_Q (avg_u (RQ^{vd,u,Q}))$$

To compute the relationship between the measure R_k defined in (1) and the experimental results, we use the correlation coefficient. Let X and Y be two vectors of the same number of values n . The correlation coefficient function is:

$$\rho_{X,Y} = \frac{\sum_{i=1}^n (x_i - avg_i(x_i))(y_i - avg_i(y_i))}{\sqrt{\sum_{i=1}^n (x_i - avg_i(x_i))^2 (y_i - avg_i(y_i))^2}} \quad (2)$$

To evaluate the "goodness" of R_k we compute the correlation coefficient between R_k and RQ and between R_k and SQ . The vector X in (2) is composed of six elements (one for each descriptor), with the measure R_k . The vector Y of equation (2) is composed of six elements (one for each visual descriptor) with the values of RQ or SQ . For each dataset, we obtain a different correlation coefficient ρ for RQ and SQ .

We use SQ to estimate the quality of the descriptors for the compound descriptor experiments, and both RQ and SQ for the single descriptors experiments. RQ is appropriate only for this second experiment. Since in the compound descriptor experiments the images coming from different descriptors are presented together, the rank quality is not meaningful.

Figure 3 shows the correlation of SQ with R_k of the six descriptors for the three datasets together with the average correlation value over the three datasets. The average correlation coefficient is maximum for $k = 50$. Therefore, the best k is exactly the size of the m -NN performed. Note that, as it is possible to see, even if the results for the TREC 2002 datasets are not impressive the average correlation for $k=50$ is greater than 0.8.

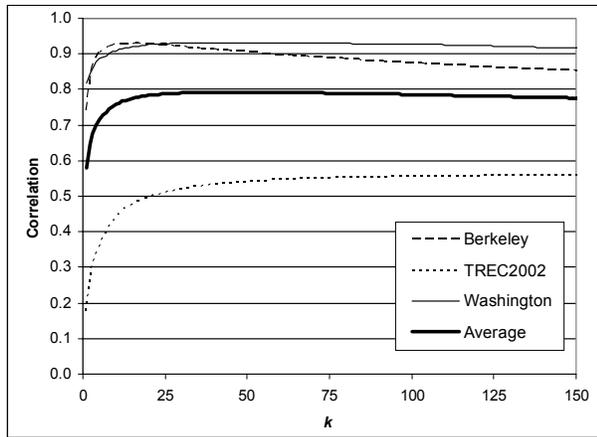


Figure 3 - Correlation coefficient for the single descriptor experiments, between SQ and R_k .

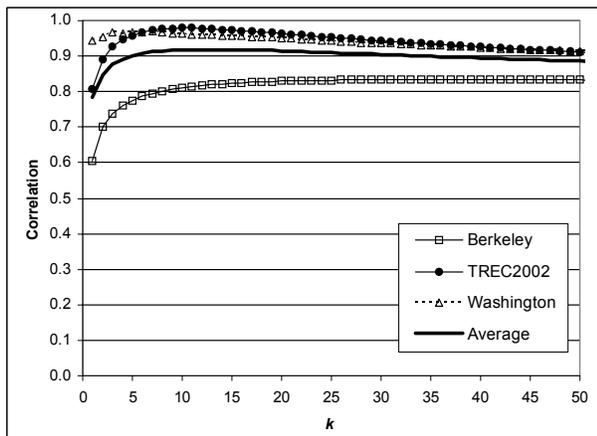


Figure 4 - Correlation coefficient for the compound descriptors experiments, between SQ and R_k .

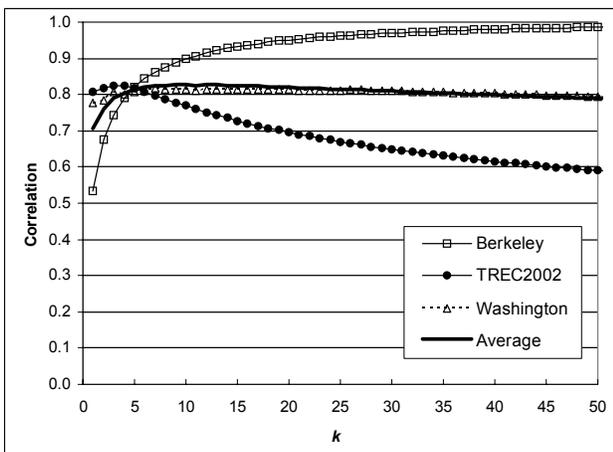


Figure 5 - Correlation coefficient for the single descriptor experiments, between RQ and R_k .

Figure 4 shows the results for the compound descriptors experiments, where all the descriptors were used simultaneously. In this case, also SQ is used. Considering the average correlation between the three datasets, we note that we have an optimum for $k = 10$. In fact, for this type of experiments, we performed a 10-NN for each descriptor, and we evaluated the “goodness” of the results coming from the descriptors. In this case, with SQ , we take in consideration two different aspects: if a certain image, selected by the user, is in the 10-NN of a given descriptor and the value assigned by the user to the selected images. Since the user sees all the 10-NN simultaneously in a random order, we can evaluate which descriptor performs the best 10-NN.

Figure 5 shows the correlation coefficients between R_k and RQ of the six descriptors for the three datasets together with the average correlation value over the three datasets. In this case, we consider the single descriptor experiments. As explained earlier, a correlation close to 1 means that R_k is able to predict the behavior of the experimental quality assessments. For a given dataset, the best correlation is given for different values of k . For example, the correlation is the maximum for $k = 50$ for Berkeley, $k = 3$ for TREC2002, and $k = 7$ for Washington. However, we are interested to choose a value of k , which provides simultaneously a good correlation for all three datasets. In order to accomplish this task we use the average correlation coefficient that reaches a maximum of about 0.82 for $k = 6$. This optimum value of k for R_k , considering the average correlation coefficient, is validated by the fact that in this set of experiments RQ strongly depends on the number of images selected by the users, which ranges between 5 and 10, and it is 5.6 on average. Note that, using RQ we are not considering the general quality of the m -NN performed, but we are evaluating if the images selected by the users, among the ones proposed by the system, have high rank. Let m be the number of images selected by the user. The intuition suggests that the better the m -NN, the higher should be RQ . In fact, in this case, the selected images are probably in the m -NN and then they have a higher rank.

In Figure 6 we report the values of R_k (for $k=11$) and the values of the SQ experimental results (for the compound experiments) for the six descriptors on the three datasets. As we can see, the patterns of the SQ values and R_k for the six descriptors are very similar for the TREC2002 key-frames and for the Washington datasets. The correlation between the two vectors is very close to 1 for these datasets (see Figure 4). Even for the Berkeley datasets, for which the correlation is around 0.8, the prediction is quite good especially considering that we were still able to predict the best and the worst descriptor.

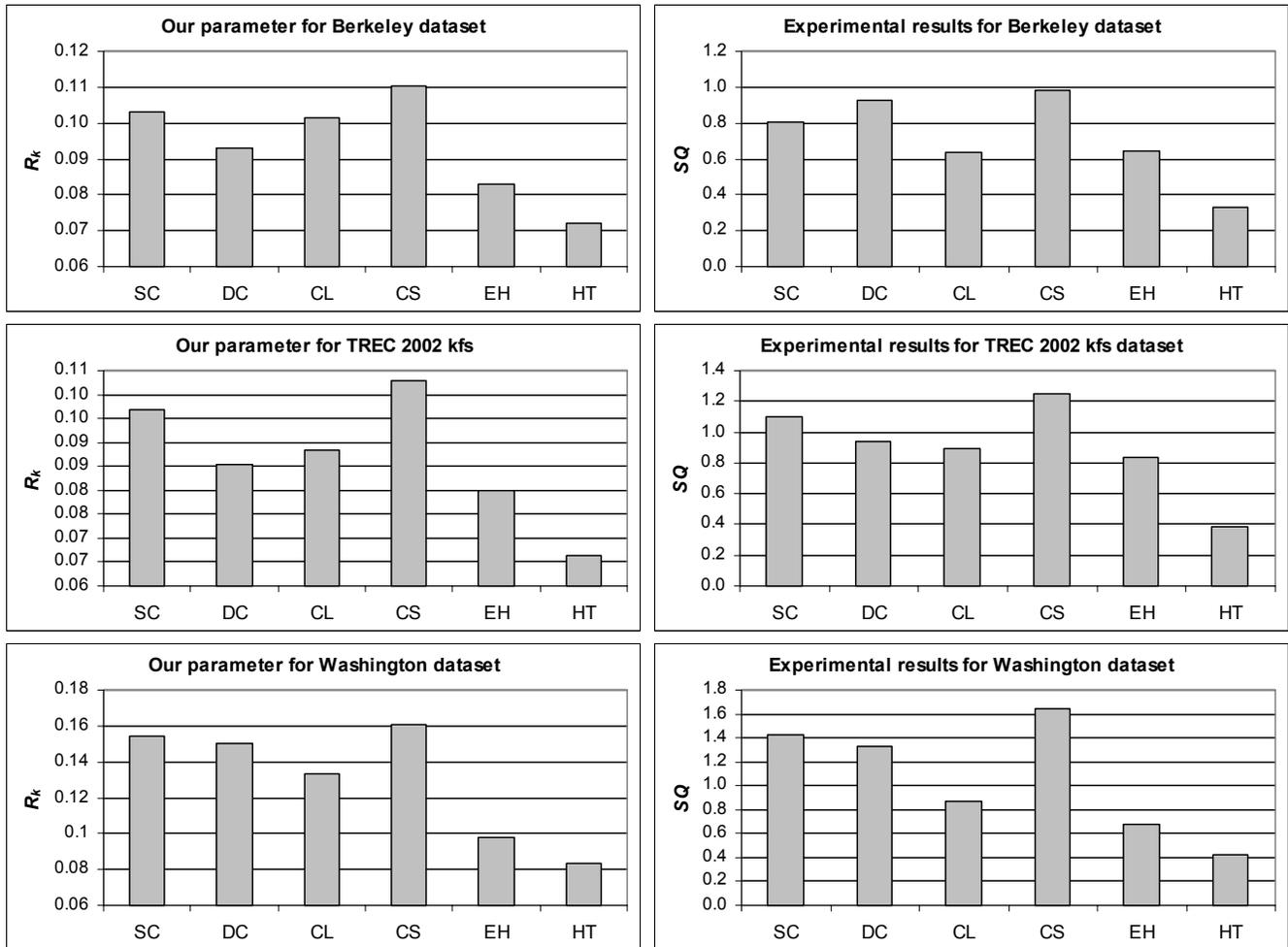


Figure 6 - R_k and SQ (for the compound experiments) for each dataset and $k=10$.

5. EXPLOITING THE EXPERIMENTAL RESULTS

In this section we describe how R_k can be used in practice to select the best visual descriptors for a given dataset (we are actually using this approach in the MILOS Multimedia Content Management System).

In Figure 7 the graph of R_k for the six MPEG-7 descriptors for the TREC2002-keyframes dataset is presented. Our experiments demonstrated that, for a given k , the greater R_k is, the better the descriptor behaves. In case the system is typically used for nearest neighbor queries with a specific value (say 10-NN) we simply need to select the visual descriptors that has the highest value of R_k for $k=10$ (see Figure 7). It is possible to see that for the most interesting values of k (ranging between 10 and 50), the behavior of R_k is quite stable. In order to decide which visual descriptors to use in a k -NN search of a real Content Management System, we can linearly combine

the distance of the different descriptors giving to them weights related to the values of R_k .

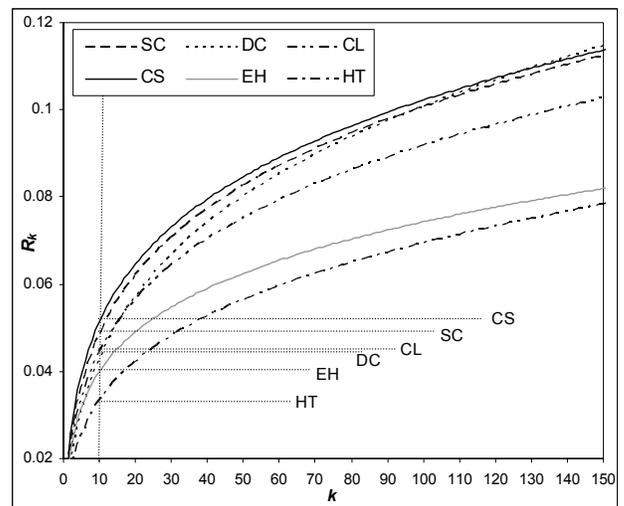


Figure 7 - R_k for the TREC 2002 keyframes dataset

6. 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 can, when performing similarity search. Descriptors based on color representation might result not to be effective with a data set containing mainly black and white images. In this paper, 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 the visual descriptor from a user perspective, have demonstrated the reliability of our approach. The experiments were conducted with a large number of users to guarantee the soundness of the analysis of results.

We have exploited these results in the context of the design of MILOS, a Multimedia Content Management System developed at ISTI-CNR, in Pisa. The motivation 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. To improve the effectiveness of this complex retrieval process, it is essential to use a technique, as proposed in this paper, for selecting MPEG-7 image features to exploit the statistical characteristic of each image data set, managed by MILOS.

As a future work, we are also seeking to extend this technique to support query driven feature selection. This technique should be able to choose the most promising query given a target data set. This extension would help in choosing the best feature, taking into consideration the target data sets and the query itself.

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] Stanchev P., "General Image Database Model", in Visual Information and Information systems, Huijismans, D. Smeulders A., (etd.) Lecture Notes in Computer Science 1614, 1999 (29-36).
- [8] Grosky W., Stanchev P., "Object-Oriented Image Database Model", 16th International Conference on Computers and Their Applications (CATA-2001), March 28-30, 2001, Seattle, Washington (94-97).
- [9] H. Eidenberger, "How good are the visual MPEG-7 features?", SPIE & IEEE Visual Communications and Image Processing Conference, Lugano, Switzerland, 2003
- [10] G. Amato, C. Gennaro, F. Rabitti, P. Savino. "Milos: A Multimedia Content Management System", accepted for publication to the SEBD 2004 Conference.
- [11] MPEG, "MPEG-7 Overview (version 9)", ISO/IEC JTC1/SC29/WG11N5525
- [12] MPEG-7, "Multimedia content description interfaces. Part 3: Visual", ISO/IEC 15938-3:2002
- [13] G. Salton, M.J. McGill, "An Introduction to Modern Information Retrieval", McGraw-Hill, 1983
- [14] MPEG-7, "Multimedia content description interfaces. Part 6: Reference Software", ISO/IEC 15938-6:2003
- [15] B.S. Manjunath, P. Salembier, T. Sikora, "Introduction to MPEG-7", Wiley, 2002