I. INTRODUCTION

We have witnessed the birth of a new industry based on information, with the digital library at the heart of that industry. The Internet, with its' associated search engines and browsers, is the great enabler of this new industry because it provides universal access to digital libraries. The Internet today contains a variety of collections, including patents, transportation schedules, financial data and full texts of professional journals of major societies. Access time to any kind of information has greatly decreased in recent years. The Internet today can be viewed as a collection of over 8 058 000 000 pages (source: www.google.com) and this number is growing rapidly. This trend of increasing the number of pages can be seen by the fact that until the end of 1997 there were only 320 million pages. The Internet also supports access to audio and video clips. Clearly, as images, video and audio become more widely used, the fraction of Internet traffic consisting of text search will decrease. In contrast, image search requests will become more common.

he image-search engines available today are relatively crude compared to the text-search engines. In typical libraries, retrieval tools rely on indices of descriptors for content-based retrieval. Descriptors may be words and phrases. They may be other data that does not appear within the document text such as author names, the publisher or the date of publication. Image descriptors are quite analogous. Image descriptors can be pure text, digital descriptions of the image, and can consist of searchable data that describes the colors, objects, composition, texture or some other feature of the image.

II. GENERAL OVERVIEW OF QBIC

Query By Image Content (QBIC) is a technology that allows queries over collections of images by their content. “Query by content” means that a collection of images can be queried in order to locate images that are similar to the query image, where similarity can be based on color, texture, or other image properties. Querying by content consists of two phases:

- Database creation: A preprocessing step that computes numeric features (also called metadata) for an image or a set of images, and stores these features in a database.
- Database query: The run-time step that uses the computed features to find images similar to the query. During the database creation phase, one or more feature classes are used to compute the features of input images as numeric values. Each feature class creates a feature table in the database where these computed values are stored. During the database query phase, QBIC compares feature data in the query with the computed data in the feature tables. A query can search on one or more features for similarity.

III. EXPERIMENTS AND RESULTS

For experiments the ibm33 digital library was used, containing 33 photos of flowers, also included within QBIC v.3 (demo). The chosen photos are pictures with primal similarity.

We tested the characteristics of QBIC v.3 based on different feature retrievals. The first subsection of the experiment consisted of testing the similarity of pictures from the ibm33 catalog, according to the texture feature. For each of 33 pictures the first 10 most similar candidates were found. The output information is the difference-coefficient according to the texture k. This coefficient varies from...
0.033621 to 3.242527. The most similar pictures according to this criterion are the pictures flower26.jpg and flower12.jpg, $k_t = 0.033621$ (figure 1.b). We also calculated the average difference-coefficient of texture, for every rank up to the 10-th. This coefficient graphically shown in figure 1.a.

Tests for draw feature similarity were made in the second subsection. As in the first subsection, for all 33 pictures, the 10 closest database entries were found. The average difference-coefficient according to draw feature $k_d$ for every rank was calculated. According to this criterion, the most similar pictures are flower08.jpg and flower12.jpg, where $k_d = 0.019531$ (figure 2.b). The dependence of $k_d$ related to the rank of difference is shown in figure 2.a. We observe that $k_d$ tends to increase as a function of the rank.

The third subsection contains tests of the color histogram feature. For all 33 pictures, their 10 closest candidates were found again. The average difference-coefficient according to the color histogram feature was calculated for every rank. The most similar pictures are flower04.jpg and flower18.jpg, with $k_h = 0.005667$. They are shown in figure 3.b. The relation between the difference-rank and $k_h$ is given in figure 3.a.

Fig. 1. a) Dependence of $k_t$ on the rank of difference
b) The most similar pictures by texture feature

Fig. 2. a) Dependence of $k_d$ on the rank of difference
b) The most similar pictures by draw feature

Fig. 3. a) Dependence of $k_h$ on the rank of difference
b) The most similar pictures by color histogram feature

IV. CONCLUSION

This article presents a descriptor-based image search tool. Tests of its characteristics were made over primary similar pictures. These results can be used for further statistical research on a larger scale. Although this kind of image search tool has some anomalies, it is the most common in commercial use. We mention that there are two alternate strategies for image search: image understanding and pixel search. An efficient tool for common use should contain parts of three methods, and must be validated by large scale experimental research.

V. REFERENCES


Abstract – Digital images constitute a large portion of the information available on the Internet. Efficient image searching through digital libraries is becoming an important task. In this article we describe a descriptor-based image search tool and analyze its features experimentally.

Keywords – image search, descriptor, QBIC(Query By Image Content)