JIS

Content-based image retrieval: a comparison between query by example and image browsing map approaches

Christopher C. Yang

Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong

Received 7 January 2004
Revised 18 February 2004

Abstract.

Efficient and effective retrieval techniques of images are desired because of the explosive growth of digital images. Content-based image retrieval is a promising approach because of its automatic indexing and retrieval based on their semantic features and visual appearance. The similarity of images depends on the feature representation and feature dissimilarity function. However, users have difficulties in representing their information needs in queries to content-based image retrieval systems. In this paper, we investigate two approaches, query by example and image browsing map. Activities to support the information seeking behavior are analyzed. The performance of these approaches is measured by a user evaluation. It is found that the image browsing map provides more functionalities and capabilities to support the features of information seeking behavior and produces better performance in searching images.

Correspondence to: Christopher C. Yang, Associate Professor, Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N.T. Hong Kong. E-mail: yang@se.cuhk.edu.hk.

Keywords: digital images; content based image retrieval; search strategies; query by example; image browsing maps; information seeking behaviour; user evaluation; retrieval performance

1. Introduction

Content-based image retrieval (CBIR) is one of the major approaches of image retrieval that has drawn significant attention in the past decade. Low level image features, such as color, texture, shape, and structure are extracted from images. Relevant images are retrieved based on the similarity of their image features. Examples of some of the prominent systems are QBIC, Photobook, and NETRA.

Advances in computer and network technologies coupled with relatively cheap high volume data storage devices have brought tremendous growth in the amount of digital images. There is a need for more effective image retrieval methods. Query by examples (QBE) is the traditional type of query in CBIR. However, the performance of QBE is unsatisfactory when a representative image cannot be identified to be submitted as an example. In this paper, we develop an image browsing map using the Kohonen self-organizing maps for CBIR. The collection of images is trained and mapped onto a two-dimensional map. Users may select the nodes based on the labeled image in the self-organizing map to retrieve similar images or explore the nearby nodes in the map.

We have investigated the differences between query by examples and image browsing maps in terms of the information seeking behavior involved and the performance of these systems. Ten features of information seeking behavior are identified by Ellis and Meho [10, 21]. These features include starting, chaining, assessing, browsing, differentiating, monitoring, extracting, verifying, networking, and information managing. The activities that relate to these features in query by example and image browsing map are discussed and analyzed. A user evaluation was conducted to measure performance.

In this paper, we first present an overview of image retrieval, including the feature extractions, query by example system, and self-organizing image browsing map in Section 2. Section 3 describes the information seeking behavior involve' in query by example and self-organizing image browsing map. Section 4 describes the design of the user evaluation and its result and discussion. Section 5 concludes the contributions and findings of our work.

2. Image retrieval

A recent study of literature in image indexing and retrieval [6] has been conducted based on 100 papers from Web of Science. Two major research approaches, text-based (description-based) and content-based, were identified. It appears that researchers in the information science community focus on the text-based approach while researchers in computer science focus on the content-based approach.

Long before digital images were available, access to image collections was provided by librarians, curators, and archivists through the manual assignment of text descriptors and classification codes [14]. Automatic assignment of text attributes to images was developed by utilizing captions and transcripts later. Text-based image retrieval (TBIR) makes use of the text descriptors to retrieve relevant images. Some recent studies [3, 5] found that text descriptors such as time, location, events, objects, formats, aboutness of image content, and topical terms are most helpful to users. The advantage of this approach was that it enabled widely approved text information retrieval systems to be used for visual retrieval systems. However, manual assignment is time consuming and costly while automatic assignment may not be possible if the image collections do not have accompanied text. Besides, the descriptions of images are often subjective [29]. Different people have different interpretations of an image, such as its important objects, or relationships. Annotators will also face serious difficulties in maintaining the consistency of annotation among images in large databases. Automatic text annotation is an alternative approach in TBIR to reduce the cost and processing time in manual annotation. Chen et al. [4] have developed a prototype system for image retrieval from the Internet using Web mining. Text descriptors are extracted from the Web pages to semantically describe the images. Data mining on the log of user feedback is also utilized. The log mining technique has shown improvement in the performance of retrieval but the associate text on the Web page does not necessarily reflect the content of images. Google is another prominent example of TBIR using automatic text annotation. Other Web search engines have adopted a similar approach [25], for example, WebSeer [13] and AltaVista Photo Finder (http://www.altavista.com).

Content-based image retrieval (CBIR) was started in the 1990s making use of the automatic extraction of lower level image features, such as texture, color, shape, and structure. Extensive research has been done to develop advance techniques to extract image features and measure the similarity between a pair of images based on their feature vectors. In this paper, we focus on the content-based image retrieval techniques. In particular, we compare the performance of query by example and self-organizing image browsing map.

2.1. Content-based image retrieval

In CBIR, the images are indexed by features that are derived directly from the images. The features are always consistent with the image and they are extracted and analyzed automatically by means of computer processing, instead of manual annotation. Due to the difficulty of automatic object recognition, information extracted from images in CBIR is rather low level, such as colors, textures, shapes, structure and combinations of the above.

A number of representative generic CBIR systems have been developed in the last ten years. These systems have been implemented in different environments, some of which are Web-based while some are GUI-based applications. QBIC, Photobook, and Netra are the most prominent examples.

QBIC is developed at the IBM Almaden Research Centre [12, 15, 22]. It is the first commercial CBIR application and plays a vital role in the evolution of CBIR systems. The QBIC system supports low level image features of average color, color histogram, color layout, texture and shape. Additionally, users can provide pictures or draw sketches as example images in query. The visual queries can also be combined with textual keyword predicates.

Photobook [23], developed at the MIT Media Lab, is a set of interactive tools for searching and querying images. It is divided into three specialized systems, namely Appearance Photobook (face images), Texture Photobook, and Shape Photobook, which can also be used in combination. The features are compared by using one of the matching algorithms. These include Euclidean, Mahalanobis, divergence, vector space angle, histogram, Fourier peak, and wavelet tree distances, as well as any linear combination of those previously discussed.

NETRA is a prototype image retrieval system that has been developed at the University of California, Santa Barbara (UCSB) [19, 20]. NETRA supports features of color, texture, shape, and spatial information of segmented image regions to region-based search. Images are segmented to homogenous regions. Using the region as the basic unit, users can submit queries based on features that combine regions of multiple images. For example, a user may compose queries such as 'retrieve all images that contain regions having color of a region of image A, texture of a region of image B, shape of a region of image C'.

2.1.1. Image features. One of the main foci in CBIR is the means for extraction of the features of the images and evaluation of the similarity measurement between the features. Image features refer to the characteristics which describe the contents of an image. In this paper, image features are confined to visual features that are derived from an image directly.

There have been extensive studies of various sorts of visual feature. The simplest form of visual feature is directly based on pixel values of the image. However, these types of visual feature are very sensitive to noise, brightness, hue and saturation changes, and are not invariant to spatial transformations such as translation and rotations. As a result, CBIR systems that are based on pixel values do not generally have satisfactory results. Much of the research in this area has placed the emphasis on computing useful characteristics from images using image processing and computer vision techniques.

Usually, general purpose features in CBIR have included color, texture, shape and structure. Other features are specific to the application domains and require some special knowledge and consequently put constraints on the database. For example, facial CBIR systems require techniques widely studied in image processing for face recognition. In this paper, the aim is to concentrate on general purpose features.

The representation of the content of an image I is usually compiled into a d-dimensional feature vector $\mathbf{f}^I$:

$$\mathbf{f}^I = (f_1^I\ f_2^I\ f_3^I\ \cdots\ f_d^I)^T \qquad \text{(Equation 1)}$$

The dimensionality d of the feature vector directly affects the performance of image query. In the simplest form of query processing without indexing, $O((Nd)^2)$ computations are required to compare each element of all vector pairs, where N is the number of the images.

Understandably, the choice of relevant and suitable features is the key issue when designing CBIR systems. A good feature should contain sufficient discriminating power to distinguish between similar and dissimilar images. Moreover, features should be invariant to spatial transformation such as translation, rotation, and minor changes related to the lighting environment where the image is captured.

Based on the dissimilarity function of image features, the definition of Content-Based Image Query (CBIQ) can be formulated as follow [26]:

Definition 1:

Content-Based Image Query: *Given an image database D of N images and a feature dissimilarity function d(I, J), find the* N_{cutoff} *images* $J \in D$ *with the lowest dissimilarity d(I, J) to the query image I.*

Another definition refers to returning images J which have lower dissimilarity to I than a certain threshold.

In this paper, we focus on chromatic and texture features. This is because chromatic and texture features work well in general images, while shape feature works well in images generated by computer graphics and structure feature works well in images with man-made objects, such as buildings, and bridges.

Chromatic features. Color is a psychophysical phenomenon for human vision. Human visual systems are more sensitive to levels of hue than levels of gray. The color characteristics in images are often an important element of the image content. Many common materials and backgrounds have distinct color properties, for example, grass is green, sea is blue, and human skin has a series of distinguishable colors.

A color space is a part of a three dimensional coordinate system where a color is represented as a vector. Selection of an appropriate color space is the starting point for using color features in CBIR systems. Red, green, blue (RGB) and cyan, magenta, yellow (CMY) are the most popular color spaces in computer technology, such as cathode ray tube, computer storage, printers. However, they do not correspond to human visual perception [34]. The human visual perception corresponds to luminance, hue and saturation (LHS) color space [30]. It is derived from the Maxwell triangle in the RGB space. HSV and HLS are alternatives of LHS with advantages of fast transformation from RGB color space. They are hexagon model and double hexagon model, respectively. CIELAB (CIE L*a*b*) and CIELUV (CIE L*u*v*) are two uniform color spaces developed by CIE (Commission Internationale de L'Éclairage) [7]. One of their limitations is that the transformation is computationally more expensive.

Color histogram [32, 35] is the standard representation of color feature in CBIR system, initially investigated by Swain and Ballard [24]. The histograms of intensity values are used to represent the color distribution. This captures the global chromatic information of an image and is invariant under translation and rotation about the view axis. Despite changes in view, change in scale, and occlusion, the histogram changes only slightly. A Color histogram $H(M)$ of image M is a 1-D discrete function representing the probabilities of occurrence of colors in images, which is typically defined as:

$$H(M) = [h_1, h_2, \ldots, h_n]$$
$$h_k = \frac{n_k}{N} \quad k = 1, 2, \ldots, n \qquad \text{(Equation 2)}$$

where N is the number of pixels in image M and n_k is the number of pixels with image value k. The division normalizes the histogram such that:

$$\sum_{k=1}^{n} h_k = 1.0 \qquad \text{(Equation 3)}$$

Texture features. Many texture features have been investigated in the past, including the conventional pyramid-structured wavelet transform (PWT) features, tree-structured wavelet transform (TWT) features, the multi-resolution simultaneous autoregressive model (MR-SAR) features and the Gabor wavelet features [18]. Experiments have been conducted and have found that the Gabor features produce the best performance [2, 20]. The computation of Gabor features is given as follows. A two-dimensional Gabor function can be formulated as:

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) \times \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right] \qquad \text{(Equation 4)}$$

A self-similar filter dictionary can be obtained as a mother Gabor wavelet $G(x, y)$ by appropriate dilations and rotations of Eq. (3) as:

$$G_{mn} = a^{S-m} G(\theta_x, \theta_y)$$

where h = height of image, w = width of image, hside = (h − 1)/2, wside = (w − 1)/2

$$\theta_x = (x - hside)\cos(n\pi/K) + (y - wside)\sin(n\pi/K)$$
$$\theta_y = -(x - hside)\sin(n\pi/K) + (y - wside)\cos(n\pi/K)$$

a > 1, m,n are integers

Given an image with luminance, $I(x, y)$, a Gabor decomposition can be obtained by multiplying the luminance by the magnitude of the Gabor wavelet:

$$|W_{mn}(x, y)| = I(x, y)\sqrt{G_{mn}i^2 + G_{mn}r^2} \qquad \text{(Equation 6)}$$

The mean and standard deviation of the magnitude of the transform coefficient are used to represent the texture feature for classification and retrieval purposes:

$$\mu_{mn} = \frac{\iint |W_{mn}(x, y)dxdy|}{h \cdot w} \qquad \text{(Equation 7)}$$

$$\sigma_{mn} = \sqrt{\iint (|W_{mn}(x, y)| - \mu_{mn}(x, y))^2 dxdy} \qquad \text{(Equation 8)}$$

The Gabor feature vector is constructed by using μ_{mn} and σ_{mn} as feature components:

$$\bar{f} = \left[\mu_{00} \quad \sigma_{00} \quad \mu_{01} \quad \sigma_{01} \quad \ldots \quad \mu_{(S-1)(K-1)} \quad \sigma_{(S-1)(K-1)}\right] \qquad \text{(Equation 9)}$$

where S is the number of scales and K is the number of orientation. In the following experiments, we use $S = 3$ and $K = 4$.

2.1.2. Query by Examples. Image queries in CBIR systems are traditionally performed by using an example image or series of images. The task of the system is to determine which images are the most similar to the given images. This approach is generally known as Query by Example (QBE). The retrieval interaction begins with an initial selection of reference images. The initial selection can be randomly selected images or some representative images selected by any means. Subsequently, the user can choose one of the images and the system will retrieve those images that are most similar to the reference. One limitation of QBE is that the success of query depends heavily on the initial set of images. In large databases, finding a set of initial images that contains at least one relevant image can be problematic. La Cascia et al. [17] have defined this situation as a page zero problem.

Figure 1 illustrates the CBIR system using QBE [27]. To begin a search, the user has an example image to submit as a query. The query serves as an approximation of the objective image being sought. The CBIR system accesses the images in the database, matches the query against the information in the database, and scores the images in terms of similarity. The matching is based on chromatic and texture features with equal weights. The top k-best images are returned as results. Upon receiving the result, user evaluates if the images in the result are relevant and selects another image from the result or database to refine the query.

2.1.3. Self-organizing image browsing map. Self-organizing maps [16] are a powerful tool for categorization and classification that involve clustering or grouping items of a similar nature. Continuous-valued vectors that represent chromatic and textural features are presented sequentially to the map in time without specifying the desired output.

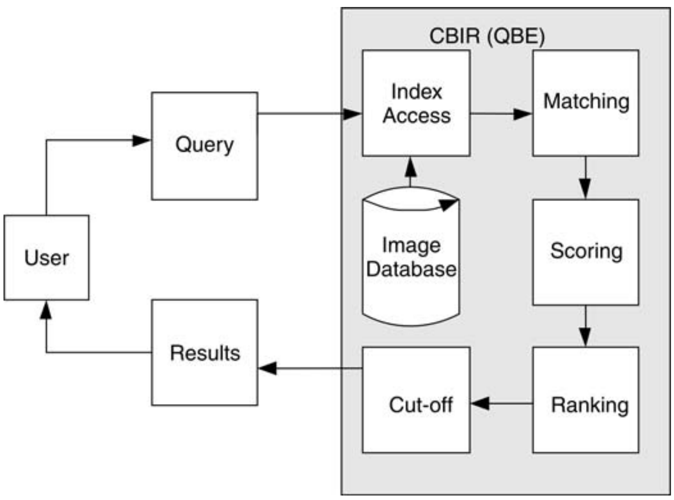

Fig. 1. CBIR system using QBE.

After presenting sufficient input vectors, network connection weights will specify cluster or vector centers that sample the input space such that the point density function of the vector centers tends to approximate the probability density function of the input vectors. Moreover, the connection weights will be organized such that the topologically close nodes are sensitive to inputs that are physically similar. The outline of self-organizing image browsing maps (SIM) [31, 33] is presented below:

Initialize input nodes, output nodes, and connection weights
Use the chromatic feature vector and textural feature vector as the input vector with $N\,(=N_1+N_2)$ elements and create a two-dimensional map (grid) of M output nodes (for example, a 10 by 10 map has 100 nodes). Initialize weight vector of each output node, $w_j(0)$, to small random values.
Present each color image in random order
Represent each image by a vector of its chromatic and textural features and present it to the system in random order.
Compute distances between each input vector and each output node's weight vector
Distance, D_{ij}, between input vector i and output node's weight vector j is

$$D_{ij} = \sum_{k=0}^{N-1} \left(x_{ik}(t) - w_{jk}(t)\right)^2$$

where $x_{ik}(t)$ is the k^{th} element of input vector i at time t and w_{ik}(t) is the k^{th} element of output node's weight vector j at time t.
Updating weights of the winning output node and its neighbors to reduce their distance
The winning output node has the minimum D_{ij}. The winning output node and its neighboring nodes will then be updated as follows:

$$w_{jk}(t+1) = w_{jk}(t) + \eta(t)h_{x_i}(t)\left(x_{ik}(t) - w_{jk}(t)\right)$$

where $\eta(t)$ is the learning factor and $h(t)$ is the neighborhood function.
Label nodes
Train the network through repeated presentation of all color images in the database until it converges. Label the output node, j, by the most similar image, i.
Map the images to the labeled nodes
For each image in the database, map it to the node that has the minimum distance with the corresponding labeled image.

Figure 2 and Figure 3 illustrate the CBIR system using SIM and an example of the result of SIM, respectively. The user selects a SIM node by the labeled image; the images that are grouped into the category are presented

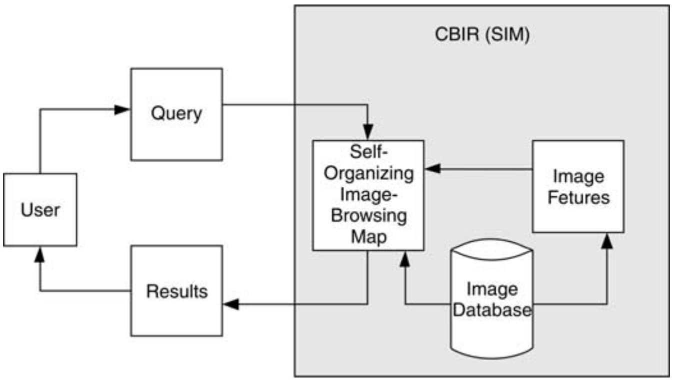

Fig. 2. CBIR system using SIM.

in the right frame as result. Since similar images are automatically grouped into the SIM neighborhood, users may proceed to the nodes in the neighborhood to search for other images based on the observed visual similarity of the labeled images on the SIM nodes.

3. Information seeking behavior

Information seeking behavior focusing on different groups of users has been studied extensively from time to time. There are recent studies of information seeking behavior (ISB) of children, young people, undergraduates, astronomers, chemists, mathematicians, physicists, and social scientists. Cooper [8] presented a study of the ISB of 7-year-old children in a semi-structured situation. Dresang [9] addressed the research needs for the ISB of young people on the Internet. Whitmire [28] presented a study of the ISB of undergraduates using the Biglan model categorizing academic disciplines into three dimensions, hard-soft, pure-applied, and life-nonlife. Brown [1] presented a study of the ISB of astronomers, chemists, mathematicians and physicists.

Ellis [10,11] has developed an information seeking behavior model with six generic features, i.e. *starting, chaining, browsing, differentiating, monitoring,* and *extracting*, based on his study of the ISB of social scientists. The model was later revised by Meho and Tibbo [21] with four additional features, i.e. *accessing, networking, verifying,* and *information management*. The ten features are defined as follows:

- *Starting*: the activities characteristic of the initial search for information, e.g. identifying references that could serve as starting points.
- *Chaining*: identifying new sources of information by following the references in the sources identified during *starting* activities.
- *Accessing*: accessing the sources of information identified and located.
- *Browsing*: looking for information in areas of potential interest.

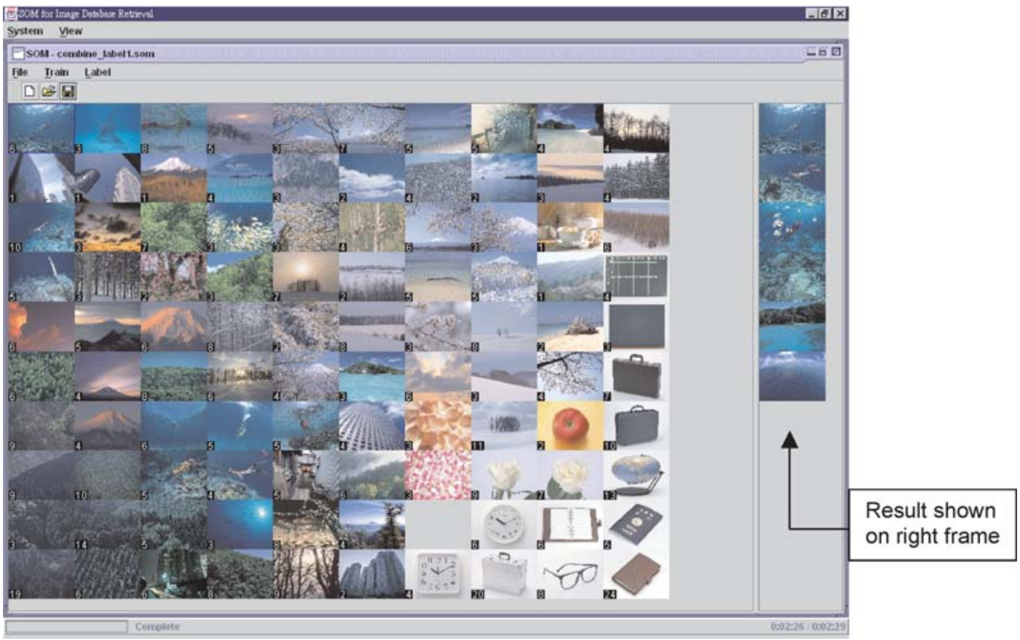

Fig. 3. Example of self-organizing image browsing map.

- *Differentiating*: filtering the amount of information obtained using known differences such as its nature, quality, relative importance, and usefulness.
- *Monitoring*: maintaining awareness of development in the topic of interest by following particular sources.
- *Extracting*: going through a particular source and identifying relevant material selectively.
- *Verifying*: checking the accuracy of the information.
- *Networking*: communicating with a broad range of people.
- *Information Managing*: filing, archiving, and organizing the information collected.

The generic features in Ellis and Meho's ISB model are helpful in developing information retrieval systems since it could increase their usefulness to include capabilities to support activities involved in these ISB features [21]. In the next sections, we investigate how the activities of the ISB features are supported by content-based image retrieval systems using query by examples and self-organizing image browsing maps.

3.1. Information seeking behavior in QBE and SIM

The query processes of QBE and SIM are different although the extraction of image features and similarity measurement are the same. The query processes affect the users' information seeking behavior and hence the success of the search. In this section, we investigate the activities involved in QBE and SIM and map the identified activities onto the features of the ISB model revised by Meho and Tibbo [21].

For QBE, the activities involved in *starting* are selecting an example image from the database as an initial searching query. Based on the initial query, the result of QBE provides the top k-best images that are relevant to the submitted example based on the similarity of their texture and chromatic features (*chaining*). Users may download the relevant images from the image database any time if identified (*accessing*). Users may browse through the top k-best images in the result (*browsing*). Users ignore irrelevant images and take the relevant images and resubmit as an example for query to obtain further results (*differentiating*). Users may investigate the relevant images in detail to extract the features that are representative (*extracting*). For example, blue sky and the scattering of fishes in the ocean are some representative features. Users may examine the image to check the objects in the images are the expected objects (*verifying*). For example, whether the yellow car in the image is a yellow cab or some other yellow car. Currently, the QBE system does not provide functions to support *monitoring, networking* and *information managing*, but they can be added.

For SIM, the activities that map onto the features, *accessing, extracting*, and *verifying*, are the same as those of QBE. SIM does not support *monitoring, networking* and *information managing* either. However, there are differences in the other features. For SIM, users select a node on SIM based on the labeled image as the *starting* point. Users have an overview of the available categories of image in the image database. By selecting a SIM node, the images that are grouped into the node are presented in the right frame for reference (*chaining*). Users may browse through all the labeled images on SIM or browse through the images that are relevant to the selected node in the right frame (*browsing*). The number of images in the right frame is not necessarily k but is determined by the SIM mapping process. Users may ignore the irrelevant images in the right frame and identify the neighborhood nodes that are most relevant to the current node for further exploration (*differentiating*).

4. Experiments

A user evaluation has been conducted to investigate the information seeking behavior and performance of QBE and SIM. In the experiment, 34 undergraduate students of the Department of System Engineering and Engineering Management in the Chinese University of Hong Kong were selected as subjects. They were all Year 3 students with an average age of 21.4. Java applications had been developed as prototypes of SIM and QBE systems.

Image Database. In the experiment, three sets of images were used. The first set (flower) was a collection of 200 images of flowers. The second set (textile) was a collection of 469 images of textile patterns. The third set (comprehensive) was a collection of 2000 images of flowers, forests, sky, underwater scenery, food, buildings etc.

Procedures. Subjects were first given a training session to be familiar with the functionality of the prototype QBE and SIM systems. The training session included a brief introduction to the features of color images and the functionality of QBE and SIM prototypes, and a 30-minute session to practice using the prototypes, using the image database (flower).

Subjects were given two tasks, searching by predetermined textual description and searching for preselected target image. In the first task (Task A), subjects were asked to select the text description using a feature dialog as shown in Figure 4. Features of the images in the database were listed in a tree-structure control. For example, in the flower image database, the first level was the number of flowers in the image and the second level was the major color of the flower(s). Subjects might add their own additional descriptions if they wanted to, for example, rose. Subjects were then given the prototype of QBE or SIM randomly to search for the images that matched their description. Subjects might continue their searching process until they obtained as many images as they wished. In the second task (Task B), subjects were given a target image randomly. The prototypes of QBE or SIM were then assigned randomly to the subject. Subjects submitted and refined their queries until they found the target image. The time taken to complete the task was recorded. Every subject was randomly assigned three tasks of Task A and three tasks of Task B for each of the two image databases (textile and comprehensive). Thus each subject was assigned 6 tasks totally. There was a time limit for Task B to avoid subjects browsing the database endlessly. The time limits for the textile and the comprehensive databases were 60 seconds and 180 seconds, respectively.

User interfaces of QBE and SIM. In Task A, subjects were first given the dialog box as shown in Figure 4 to describe the images to search for. Once the description was submitted, a user interface as shown in Figure 5 appeared if QBE was assigned. The left frame (feature panel) showed the text description selected by the subject. The center frame (example panel) showed the current example image. The right frame (result panel) showed the search result images. The order of ranking was top to bottom, left to right. The first rank was omitted because it was the example image for the query. The top frame (navigate panel) provided the buttons for subjects to go forward or backward to review previous results.

Fig. 4. Feature dialog.

When the subject started the searching process, an example image was selected from the database. The labeled images available on SIM are provided to be selected as initial examples in order to avoid bias on the SIM. The selected image was presented in the example panel and the initial result was presented in the result panel. If any images in the result panel were found relevant, the subject could select them and place them under the feature panel. The subject might continue to select images from the result panel or the labeled SIM images to refine the query and obtain further results. The subject could exit at any time when he/she thought sufficient images were found.

If the SIM prototype was assigned in Task A, the same dialog box was given to subjects but the user interface as shown in Figure 6 appeared next. The left frame (feature panel) was the same as the feature panel in the QBE prototype. The right frame (SIM panel) presented the self-organizing image browsing map generated for the image database. When subjects clicked on a node in the SIM panel, a pop-up window as shown in Figure 7 appeared. Subjects might select the relevant images in the pop-up window and place them under the feature panel. Similarly, subjects might continue to explore other nodes in the SIM panel until they decided to exit the system.

In Task B, if the QBE prototype was assigned, a user interface as shown in Figure 8 appeared. The user interface was similar to that shown in Figure 5 except that the left frame (target panel) presented the target image to be searched for. The searching process was similar to Task A, but it ended only when the subject found the target image. The time taken was recorded.

If the SIM prototype was assigned in Task B, a user interface as shown in Figure 9 appeared. It was similar to the user interface shown in Figure 6 except that the left frame (target panel) presented the target image to be searched for. Subjects carried out a similar searching process until the target was found, and the time taken was recorded.

Experimental Results. The experimental results of Task A are shown in Table 1. Comparisons are made between QBE and SIM using a small image database

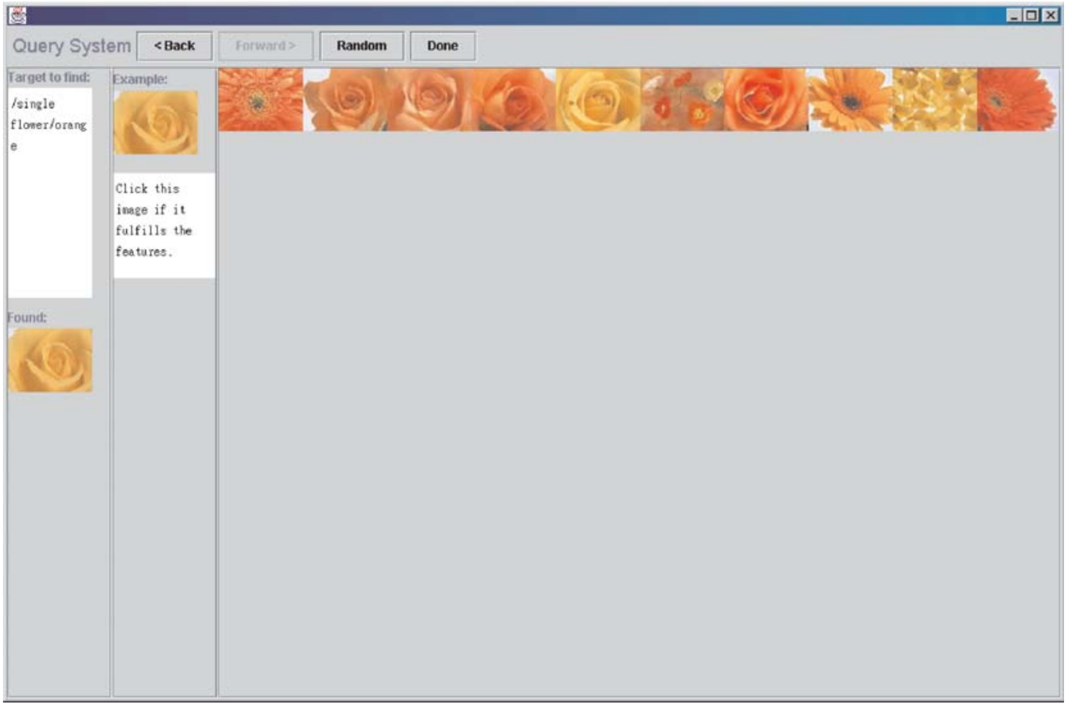

Fig. 5. User interface of QBE prototype for Task A.

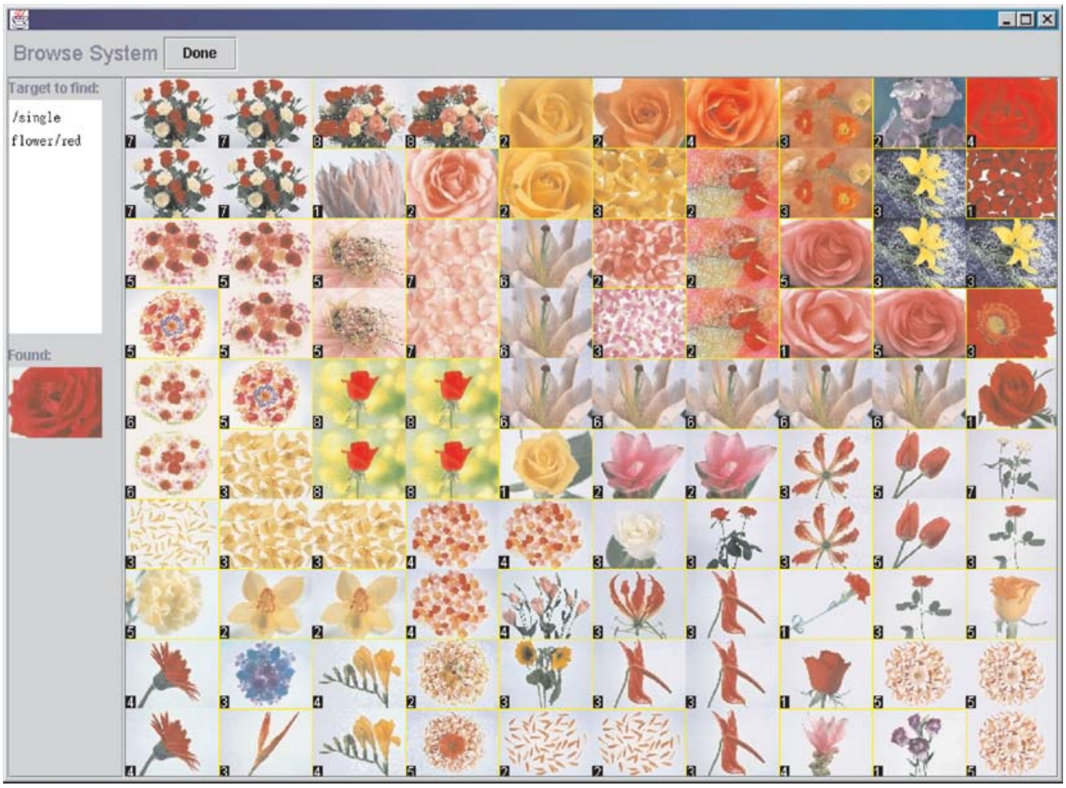

Fig. 6. User interface of SIM prototype for Task A.

Fig. 7. A pop-up window to present images associated with the selected node in SIM.

(textile) and a large image database (comprehensive). T-test is conducted to determine if the difference is significant. It is found that more relevant images are found using the SIM prototype on both small and large image databases. The difference is found to be insignificant for the small image database (textile) but significant at $p \leq 0.01$ for the large image database (comprehensive). It is clear that SIM allows more exploration of relevant images especially for large image databases. The page zero problem in QBE avoids subjects identifying more relevant images because they are not able to acquire better examples to submit as queries after a number of trials. In terms of the chaining feature of the ISB model, the number of references obtained is limited, and therefore, the number of relevant images retrieved is less.

The number of queries and unique queries are also investigated. A query is defined as an example image submitted in QBE and as clicking a node in the SIM panel to open the associated images in the pop-up window in SIM. It is found that subjects submitted significantly more queries and unique queries at $p \leq 0.01$ when using SIM on the small image database. However, there were fewer queries and more unique queries when using SIM on the large image database, and the differences are insignificant. For a small image database, although SIM has already automatically categorized the image database and presented the categories on a two-dimensional map, the categories in the neighborhood are indeed quite similar. As a result, subjects are exploring more nodes in SIM to find the relevant images. For a large image database, the categories in SIM are easier to differentiate in the neighborhood nodes. As a result, the number of queries

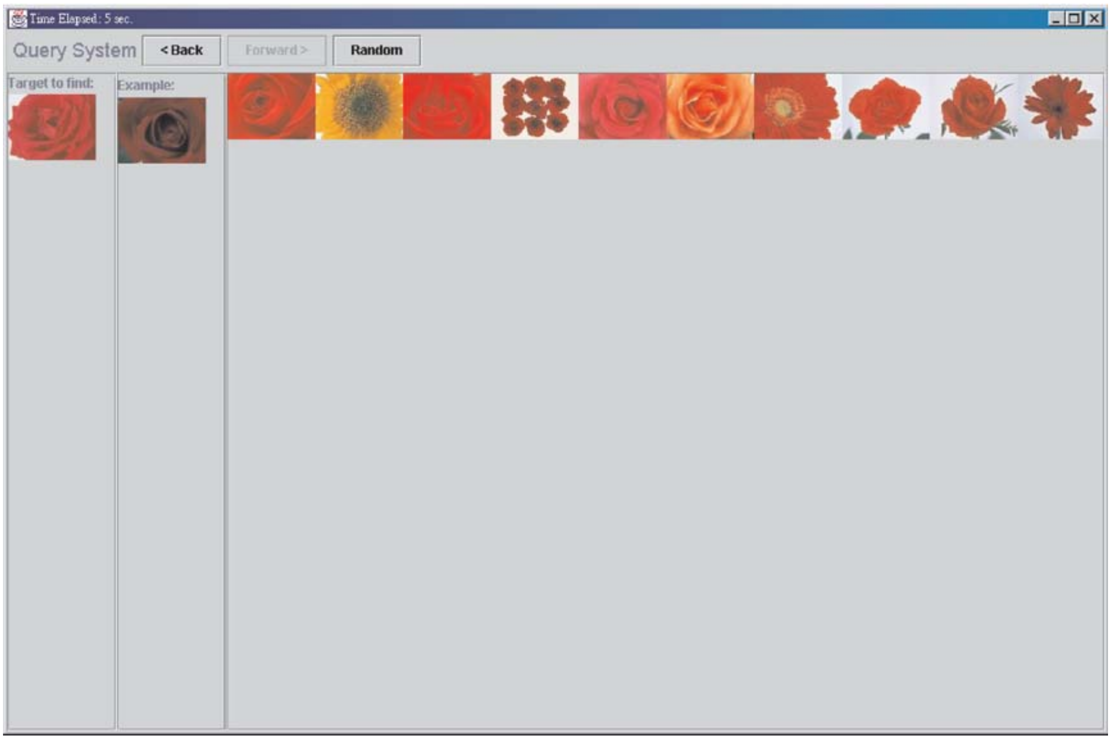

Fig. 8. User interface for QBE prototype for Task B.

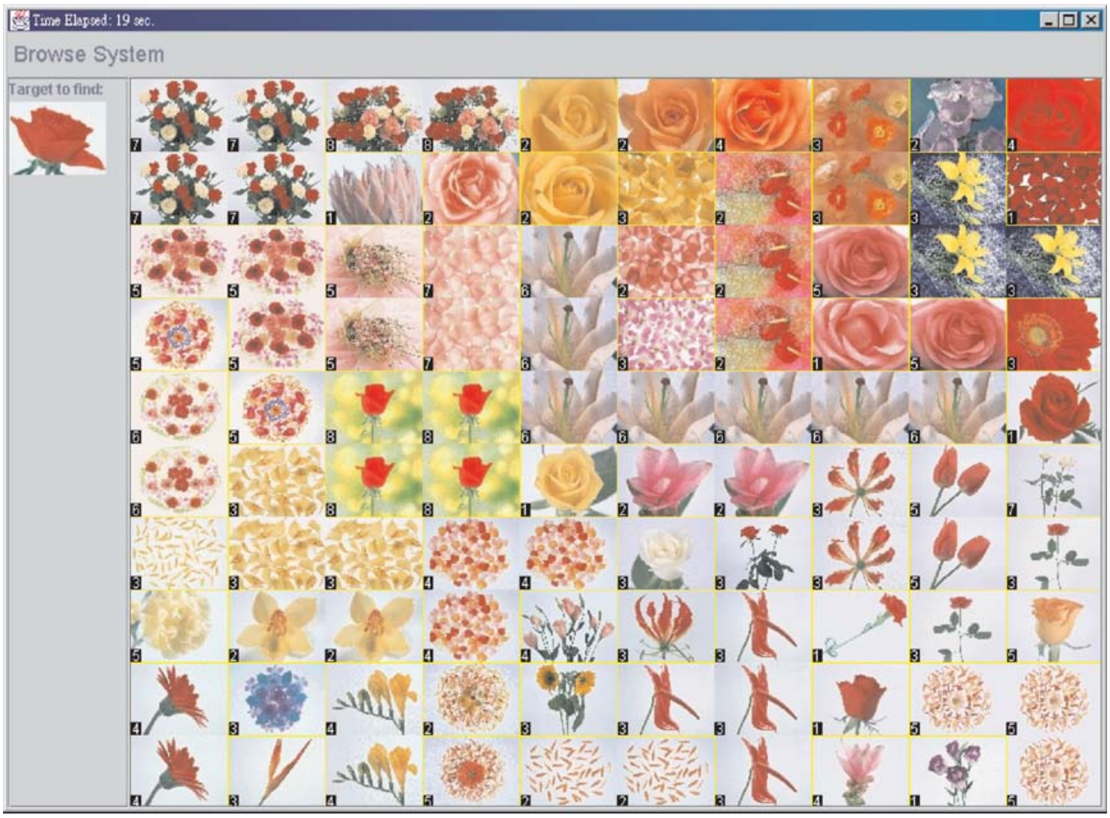

Fig. 9. User interface for SIM prototype for Task B.

in SIM is not significantly more than that in QBE. In terms of the browsing feature in the ISB model, users are able to look for relevant images on the two-dimensional map in SIM and the associated images in the pop-up window using the SIM prototype. Users have more interactions with the system and have more chances of browsing relevant images. It is found that the ratio of the number of unique queries to the number

Table 1
Experimental results of Task A in terms of number of relevant images retrieved, number of queries, number of unique queries, and total time taken and average time taken per query

Database		Textile			Comprehensive		
Prototype		QBE	SIM	*p*-value	QBE	SIM	*p*-value
Number of relevant images retrieved	Mean	4.33	4.76	0.4565	12.05	21.89	0.0028*
	Std	3.44	4.68		12.68	29.97	
Number of queries	Mean	29.05	46.43	0.0003*	48.90	44.10	0.3385
	Std	24.20	40.57		38.01	32.97	
Number of unique queries	Mean	17.82	35.67	0.0000*	28.49	31.71	0.2712
	Std	16.26	29.19		19.97	21.43	
Total time taken (seconds)	Mean	20.61	29.54	0.0681**	20.10	17.43	0.2190
	Std	22.11	43.52		23.08	31.46	
Average time taken per query (seconds)	Mean	33.79	26.49	0.0017*	34.38	35.31	0.3442
	Std	14.79	17.73		13.92	17.15	

* Significant at $p \leq 0.01$; ** Significant at $p \leq 0.07$

of queries is lower for the QBE prototype. This means that users are browsing the same set of images more frequently. In terms of the differentiating feature in the ISB model, differentiating nodes in the SIM neighborhood is relatively easier when the size of the image database increases. When users submit queries by QBE, the number of images in the result is always *k* (the top *k*-best images). The number of irrelevant images in the result increases when the number of relevant images available in the image database is low. The effort in filtering the irrelevant images increases. On the contrary, the number of images associated with a node is based on the mapping process and related to the number of relevant images available. The effort in filtering is comparatively lower.

The total time taken and the average time taken per query are also reported. It is found that the total time taken and the average time taken per query are significantly longer at $p \leq 0.07$ and $p \leq 0.01$, respectively, when SIM is used and the size of the image database is small. However, when the size of the image database is large, the differences in the total time taken and the average time taken per query between using QBE and SIM are insignificant. It is mainly due to a more significant number of interactions with the SIM prototype.

The experimental results of Task B are given in Table 2. It is found that the successful rate is higher when using SIM. The difference in successful rate is bigger when the size of the image database increases. As discussed earlier, SIM provides more functionalities and capabilities for chaining, browsing, and differentiating. It is understandable that users are more able to find a target image using SIM.

There are no significant differences between using QBE and SIM in the other measurements except in the number of queries and the number of unique queries when a large image database is used. There are significantly more queries and unique queries at $p \leq 0.09$ and $p \leq 0.05$, respectively, when SIM is used. This concurs with the result we find in Task A except that the significance occurs when the size of the image database is large because the task is now finding a target image instead of as many relevant images as possible.

We conclude that SIM provides more functionalities and capabilities to support the features of the ISB model and produces better performances in terms of identifying more relevant images and successfully retrieving a target image.

5. Conclusion

In this paper, we have developed prototypes to investigate two approaches of content-based image retrieval, query by examples and self-organizing image browsing maps. These approaches have different supports on three features of information seeking behavior: chaining, browsing, and differentiating. QBE has more limitations than SIM and may suffer page zero problems. On the other hand, SIM has more functionalities and capabilities that automatically map the relevant images to the representative labeled images on the two-dimensional map. Seeking an initial example is no longer a problem but an overall picture of the whole image database is available. The results of a user evaluation show that SIM outperforms QBE.

Table 2
Experimental results of Task B in terms of success rate, number of queries, number of unique queries, and total time taken and average time taken per query.

Database		Textile			Comprehensive		
Prototype		QBE	SIM	p-value	QBE	SIM	p-value
Successful rate		59%	64%		55%	69%	
Number of queries	Mean	7.03	6.77	0.8552	14.07	18.53	0.0857**
	Std	8.50	7.42		12.00	16.51	
Number of unique queries	Mean	4.70	5.74	0.2899	10.35	13.99	0.0425*
	Std	4.80	6.01		7.84	11.79	
Total time taken (seconds)	Mean	37.22	36.26	0.7661	118.34	106.00	0.1708
	Std	22.98	22.77		65.47	62.77	
Average time taken per query (seconds)	Mean	32.54	34.36	0.5674	52.87	47.42	0.2810
	Std	16.73	18.44		31.66	24.08	

* Significant at $p \leq 0.05$; ** Significant at $p \leq 0.09$

Significantly more relevant images are identified when the image database is large. A higher successful rate in retrieving a target image is obtained.

Acknowledgement

This project is supported by an Earmarked Grant for Research from the Hong Kong Research Grants Council, CUHK 7034–98E.

References

[1] C.M. Brown, Information seeking behavior of scientists in the electronic information age: astronomers, chemists, mathematicians, and physicists, *Journal of the American Society for Information Science and Technology* 50(10) (1999) 929–43.

[2] C.C. Chen and C.C. Chen, Filtering methods for texture discrimination, *Pattern Recognition Letters* 20(8) (1999) 783–90.

[3] H. Chen, An analysis of image queries in the field of art history, *Journal of the American Society for Information Science and Technology* 52(3) (2001) 260–273.

[4] Z. Chen, W. Liu, F. Zhang, M. Li and H. Zhang, Web mining for web image retrieval, *Journal of the American Society for Information Science and Technology* 52(10) (2001) 831–839.

[5] Y. Choi and E.M. Rasmussen, Searching for images: the analysis of users' queries for image retrieval in American history, *Journal of the American Society for Information Science and Technology* 54(6) (2003) 498–511.

[6] H. Chu, Research in image indexing and retrieval as reflected in the literature, *Journal of the American Society for Information Science and Technology* 52(12) (2001) 1011–18.

[7] CIE, *Colorimetry*, CIE Publication No. 15.2 (Bureau Central de la CIE, Vienna, 1986).

[8] L.Z. Cooper, A case study of information seeking behavior in 7-year-old children in a semi-structured situation, *Journal of the American Society for Information Science and Technology* 53(11) (2002) 904–22.

[9] E.T. Dresang, More research needed: informal information seeking behavior of youth on the Internet, *Journal of the American Society for Information Science and Technology* 50(12) (1999) 1123–24.

[10] D. Ellis, *The Derivation of a Behavior Model for Information Retrieval System*, (PhD, Dissertation, University of Sheffield, United Kingdom, 1987).

[11] D. Ellis, A behavioral approach to information retrieval system design, *Journal of Documentation* 45(2) (1989) 171–212.

[12] M. Flickner, H. Sawhney and W. Niblack, Query by image and video content: the QBIC system, *IEEE Computer* September (1995) 23–32.

[13] C. Frankel, M. Swain and V. Athitsos, *Webseer: An Image Search Engine for the World Wide Web*, technical report TR-96-14 (CS Department, University of Chicago, 1996).

[14] A.A. Goodrum, M.E. Rorvig, K. Jeong and C. Suresh, An open source agenda for research linking text and image content features, *Journal of the American Society for Information Science and Technology* 52(11) (2001) 948–53.

[15] J. Hafner, H.S. Sawhney, W. Equitz, M. Flicker and W. Niblack, Efficient color histogram indexing for quadratic form distance functions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(7) (1995) 729–36.

[16] T. Kohonen, *Self-Organization and Associative Memory* (Springer, Berlin, 1988).

[17] M. La Cascia, S. Sethi and S. Sclaroff, Combining textual and visual cues for content-based image retrieval on the world wide web. In: *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries, Santa Barbara, California*, (IEEE Computer Society, 1998) 24–28.

[18] M. Lee and C. Pun, Rotation and scale invariant wavelet feature for content-based texture image retrieval, *Journal of the American Society for Information Science and Technology* 54(1) (2003) 68–80.

[19] W.Y. Ma and B.S. Manjunath, NeTra: a toolbox for navigating large image databases, *Multimedia Systems* 7(3) (1999) 184–198.

[20] B.S. Manjunath and W.Y. Ma, Texture features for browsing and retrieval of image data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(18) (1996) 837–42.

[21] L.I. Meho and H.R. Tibbo, Modeling the information seeking behavior of social scientists: Ellis's study revisited, *Journal of the American Society for Information Science and Technology* 54(6) (2003) 570–87.

[22] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos and G. Taubin, The QBIC project: querying images by content using colors, texture and shape. In: W. Niblack (ed.), *SPIE Proceedings Vol. 1908, Storage and Retrieval for Image and Video Databases, 2–3 February 1993, San Jose, California* (SPIE, San Jose, 1993) 3173–87.

[23] A. Pentland, R. Picard and S. Sclaroff, Photobook: content-based manipulation of image databases. Storage and Retrieval for Image and Video Databases II, number 2185, San Jose, CA., February 1994.

[24] M. Swain and D. Ballard, Color indexing, *International Journal of Computer Vision* 7(1) (1991) 11–32.

[25] M. Swain, *Searching for Multimedia on the World Wide Web, technical report CRL 99/1, Cambridge Research Laboratory* (1999). Available at: http://crl.research.compaq.com/publications/techreports/ abstracts/99_1.html.

[26] J.R. Smith, *Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression* (PhD thesis, Graduate School of Arts and Science, Columbia University, 1997).
[27] J.R. Smith, Quantitative assessment of image retrieval effectiveness, *Journal of the American Society for Information Science and Technology*, 52(11) (2001) 969–79.
[28] E. Whitmire, Disciplinary differences and undergraduates' information seeking behavior, *Journal of the American Society for Information Science and Technology* 53(8) (2002) 631–38.
[29] Q. Wu, S.S. Iyengar and M. Zhu, Web image retrieval using self-organizing feature map, *Journal of the American Society for Information Science and Technology* 52(10) (2001) 868–75.
[30] C.C. Yang and S.H. Kwok, Efficient gamut clipping for color image processing using LHS and YIQ, *Optical Engineering*, 42(3) (2003) 701–711.
[31] C.C. Yang, S.H. Kwok and M.K. Yip, Image Browsing of Feature-based Products, *Proceedings of the SPIE International Conference on Electronic Imaging and Multimedia Technology, Photonics Asia, Shanghai, China, October 14–18, 2002*
[32] C.C. Yang and M.K. Yip, Quantization of color histograms using GLA, *Proceedings of the SPIE International Conference on Electronic Imaging and Multimedia Technology, Photonics Asia, Shanghai, China, October 14–18, 2002*.
[33] C.C. Yang and M.K. Yip, Visual thesaurus for color image retrieval using self-organizing map, *Proceedings of the 7th International Conference on Information Systems Analysis and Synthesis, Orlando, July 22–25, 2001*.
[34] J. Zachary, S.S. Iyengar and J. Barhen, Content based image retrieval and information theory: a general approach, *Journal of the American Society for Information Science and Technology* 52(10) (2001) 840–52.
[35] J. Zachary and S.S. Iyengar, Information theoretic similarity measures for content based image retrieval, *Journal of the American Society for Information Science and Technology* 52(10) (2001) 856–67.