The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1468-4527.htm

Image retrieval from the web using multiple features

A. Vadivel
Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India
Shamik Sural
School of Information Technology, Indian Institute of Technology, Kharagpur, India, and
A.K. Majumdar
Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur, India

Refereed article received
12 March 2009
Approved for publication
16 July 2009

Abstract

Purpose – The main obstacle in realising semantic-based image retrieval from the web is that it is difficult to capture semantic description of an image in low-level features. Text-based keywords can be generated from web documents to capture semantic information for narrowing down the search space. The combination of keywords and various low-level features effectively increases the retrieval precision. The purpose of this paper is to propose a dynamic approach for integrating keywords and low-level features to take advantage of their complementary strengths.

Design/methodology/approach – Image semantics are described using both low-level features and keywords. The keywords are constructed from the text located in the vicinity of images embedded in HTML documents. Various low-level features such as colour histograms, texture and composite colour-texture features are extracted for supplementing keywords.

Findings – The retrieval performance is better than that of various recently proposed techniques. The experimental results show that the integrated approach has better retrieval performance than both the text-based and the content-based techniques.

Research limitations/implications – The features of images used for capturing the semantics may not always describe the content.

Practical implications – The indexing mechanism for dynamically growing features is challenging while practically implementing the system.

Originality/value – A survey of image retrieval systems for searching images available on the internet found that no internet search engine can handle both low-level features and keywords as queries for retrieving images from WWW so this is the first of its kind.

Keywords Internet, Worldwide web, Search engines

Paper type Research paper

Introduction

It has been observed that low-level features such as colour, texture and shape can be efficiently used to perform image retrieval in domain-specific applications. Colour histograms such as the Human Colour Perception Histogram (HCPH) (Vadivel *et al.*,

This work is supported by a research grant from the Department of Science and Technology, Government of India, under Grant SR/FTP/ETA-46/07 dated 25th October 2007.

Online Information Review
Vol. 33 No. 6, 2009
pp. 1169-1188
© Emerald Group Publishing Limited
1468-4527
DOI 10.1108/14684520911011061

2008; Gevers and Stokman, 2004) as well as colour-texture features such as the Integrated Colour and Intensity Co-occurrence Matrix (ICICM) (Vadivel *et al.*, 2007; Palm, 2004) show high retrieval precision in such applications. However, in generic applications such as the retrieval of relevant images from the worldwide web, low-level features alone may not adequately represent the semantic content of images. As a result the retrieval precision tends to drop considerably. This situation can be effectively handled by using keywords to restrict the search space in such applications. For example, while searching for images on the web, use of keywords may be quite helpful in filtering out pages that are not relevant. In the early years, textual keywords were assigned by domain experts to general purpose images, which made the textual keywords highly subjective. These assigned annotations could be combined with low-level features of images for retrieval. However, for annotating images that are embedded into the HTML pages, it is felt that the textual content of the document can be used without any need for manual intervention. Often it is found that the text in HTML documents describes images, and thus is situational and more relevant. Therefore keywords extracted by taking into account their frequency of occurrence in HTML documents can be used for labelling images and combined with their low-level features for improving retrieval precision.

A survey of image retrieval systems for searching images available on the internet may be found in JISC (2008). Out of these image search engines, Google (www.google.com) consists of simple search options with a good advanced search facility. It provides quick and reasonably good search results. Similar to Google, the search options at Yahoo (www.yahoo.com) are also simple. In addition, the retrieval process is quick and retrieved images are relevant, without dead links and duplicates. However, the size of the search scope is unspecified. A few other image search engines have been developed during the last few years. However, none of these search engines can dynamically combine text keyword and image features at the time of retrieval.

Apart from search engine related applications, integration of keywords has been used by researchers in many other applications. Zhou and Huang (2002) proposed a scheme for unifying keywords and visual features of images. They assumed that some of the images in the database have been already annotated in terms of short phrases or keywords. These annotations are assigned either using text surrounding the images in HTML pages or by speech recognition or manual annotation. During retrieval the user's feedback is obtained for semantically grouping keywords with images. While colour moments and colour histograms are used to represent colour content, Tamura's features, co-occurrence matrix features and wavelet moments are extracted for representing texture content. Using a water filling feature and salient points, structural features are represented (Zhou *et al.*, 1999).

An attempt has been made to improve the semantic correlation between keywords in the document title of HTML documents and image features for improving retrieval in news documents (Zhao and Grosky, 2002). This method was used on a collection of 20 documents from a news site. From the titles of the documents, 43 keywords were extracted and HSV-based colour histograms were constructed. While constructing the histograms, saturation and hue axes were quantised into ten levels to obtain a $H \times S$ histogram with 100 bins. The results showed that the proposed technique performed well for a small number of webpages and images.

In general, image search results returned by an image search engine contain multiple topics. Organising the results into different semantic clusters helps users. Cai

et al. (2004) developed a way to analyse the retrieval results of a web search engine. This method used Vision-based Page Segmentation (VIPS) (Cai *et al.*, 2003) to extract the semantic structure of a webpage based on visual presentation. The semantic structure is represented as a tree with nodes, where every node represents the degree of coherence to estimate visual perception.

Tollari *et al.* (2005) designed a technique to map textual keywords with visual features of images. In order to map textual and visual information, semantic classes containing a few image samples are obtained. Each class is initialised with one image and then two classes are merged based on a threshold value on distance between them. The merging process stops when the distances between all classes are higher than a second threshold. Finally the semantic classes are divided into two partitions as reference set and test set. The reference partition provides example documents of each class, which are used to estimate the class of any image of test partition, either using textual or visual information or a combination of both. Visual features such as colour, shape and orientation histograms of the images are extracted from these classes. The performance of this proposed method is high with fewer images having corresponding textual annotations.

Han *et al.* (2005) suggested a memory learning framework for image retrieval applications. Their method uses a knowledge memory model for storing the semantic information by accumulating user-provided data during query interactions. The semantic relationship is predicted by applying a learning strategy among images according to the memorised knowledge. Image queries are finally performed based on a seamless combination of low-level features and learned semantics.

Similarly a unified framework has also been proposed, which uses textual keywords and visual features for image retrieval applications (Jing *et al.*, 2005). The framework builds a set of statistical models using visual features of a small set of manually labelled images that represent semantic concepts. This semantic concept has been used for assigning keywords to unlabelled images. These models are updated regularly when more images implicitly labelled by users become available through relevance feedback. This process accumulates and memorises knowledge learned through relevance feedback from users. The efficiency of relevance feedback for keyword queries is improved by an entropy-based active learning strategy.

An integrated patch model was developed by Xu and Zhang (2007), which is a generative model for image categorisation based on feature selection. Feature selection strategy is categorised into three steps for extracting representative features and also to remove the noise feature. Initially, salient patches present in images are detected and clustered and keyword vocabulary is constructed. Later the region of dominance and the salient entropy measure are calculated for reducing the uncommon noises of salient patches. Using visual keywords, categories of the images are described by an integrated patch model.

The textual keywords appearing on webpages have recently been used for identifying unsuitable, offensive and pornographic websites (Hu *et al.*, 2007). In this framework the webpages are categorised as continuous text pages, discrete text pages and image pages using a decision tree with respect to their contents. These pages are handled by respective classifiers. Statistical and semantic features are used for identifying the pornographic nature of continuous text webpages. Similarly, Bayesian classifiers and image object contours are being used to identify the pornographic content of discrete and image webpages respectively.

Another method has recently been proposed for reducing the gap between the extracted features of the systems and the user's query (Yang and Lee, 2008). The image semantics are discovered from webpages for semantic-based image retrieval using self-organising maps. The semantics are described based on the environmental text, that is, that surrounded by the images. A text mining process has been applied by adopting a self-organising map learning algorithm as a kernel. Some of the implicit semantic information is also discovered after the text mining process. A semantic relevance measure is used for retrieval.

However, none of the methods discussed above dynamically integrate textual keywords with low-level features for retrieval. It may be noticed that the importance of query content has not been captured by any of these methods. In addition, the advantages of combined queries have not been addressed. In this paper we propose a technique for dynamic integration of keywords with image features for retrieval applications. In the proposed technique we have used a large number of HTML documents containing text and images available on the internet. The HTML documents are fetched using a crawler. The content of the HTML documents is segregated into text, image and HTML tags. Keywords are extracted from the text. These are considered relevant keywords for representing high-level semantics of the images contained in the same HTML document.

In the following section of the paper we discuss image retrieval using keywords and low-level features. Then we explain dynamic extraction from the HTML documents. Following that, the variation in retrieval precision due to weighted features is discussed. Next the architecture is described. Finally the experimental results are given.

Image retrieval using keywords and low-level features

The semantics of an image can be described by a collection of one or more keywords. Let I be a set of images and K be a set of keywords. Assignment of a keyword to an image may be treated as a labelling predicate written as follows:

$$l : K \times I \rightarrow \{True, False\} \tag{1}$$

Thus a keyword $k \in K$can be assigned to an image $i \in I$ if $l(k,i) = True$. Since a keyword can be assigned to multiple images and multiple keywords may be assigned to one image, given a subset of keywords $K' \subseteq K$, K' can be used to define a subset of images from I as follows:

$$C^I_{KD}(K') = \{i | i \in I \; and \; \forall k \in K', l(k,i) = True \} \tag{2}$$

Various Boolean operations like OR, AND and NOT can be used for combining keywords. However, while performing retrieval, the proposed method uses only Boolean ANDing of keywords in the query predicate. If we do not restrict the set of images by keyword, that is, if $K' = \phi$, $C^I_{KD}(\phi) = I$, then all the images in the database form one set. Thus if images could be correctly labelled by a set of K keywords, retrieval based on keywords K would retrieve only relevant images resulting in 100 per cent recall and precision. The main problem with such keyword-based retrieval is that it is not feasible to manually annotate each image with keywords. Another difficulty lies in the fact that it is a subjective process. Different users may describe the same image in different ways. It is also language-dependent. Further, it cannot be ensured that whoever is building the image database will also annotate it. If we consider the

images available on the web that are embedded in HTML documents, authors who have created a webpage may not have correctly updated the image description while publishing the page. Even if it has been done, it could be highly subjective as mentioned above.

In contrast to keyword-based retrieval, content-based retrieval of images searches for images "similar" to a given query image with respect to appropriate low-level features. The commonly used features may be colour, texture, shape or some suitable combination. The "similarity" between a query image and an image stored in a database based on a feature vector f is usually estimated with respect to a suitable distance function. Hence given a query image q, the set of images that are retrieved from the image database I using a feature vector f, a distance function d and a distance range Δd can be denoted by

$$Q(q,f,d,I,\Delta d) = \{i | i \in I \, and \, d(q,i) \leq \Delta d\} \subseteq I \tag{3}$$

Although each image in the set $Q(q,f,d,I,\Delta d)$ is within a distance of Δd from q, it cannot always be guaranteed that the images in $Q(q,f,d,I,\Delta d)$ are semantically close to q. Since retrieval precision considers the set of images that are semantically relevant to the query image, content-based retrieval cannot guarantee high precision except in databases that contain a large number of images, which are both semantically close as well as close in image features to a given query. We can refine content-based retrieval results by only selecting those images that match a set of keywords $K_1 \subseteq K$, if such keyword labelling is available. This leads to a composite content-based image retrieval scheme involving both keywords and image features. The images retrieved from an image database I, with a query involving a set of keywords $K_1 \subseteq K$ and an example image i may be denoted by

$$Q(K_1,q,f,d,I,\Delta d) = \{i \in I | d(q,i) \leq \Delta d \, and \, \forall k \in K_1, \, l(k,i) = True\} \tag{4}$$

From equations (2) to (4), we can write:

$$Q(K_1,i,f,d,I,\Delta d) = C^I_{KD}(K_1) \cap Q(i,f,d,I,\Delta d) \tag{5}$$

Equation (5) represents the set of images that are labelled by each keyword from the set K_1, and are also close to the query image q in terms of the content of the feature f. If we perform a nearest neighbour query, we can choose only those images that are labelled by the keywords of the set K_1 and then rank them in increasing order of their distance from q with respect to feature f and distance function d. For a smaller subset $K_1^{'}$ of keywords such that $K_1^{'} \subseteq K_1$

$$Q(K_1,i,f,d_f,I,\Delta d_f) \subseteq Q\left(K_1^{'},i,f,d_f,I,\Delta d_f\right) \tag{6}$$

As is evident from the above discussion, if a user can specify the keywords that describe the required set of images correctly and if the relevant database images are also labelled with the same set of keywords, we would achieve 100 per cent recall and precision. However, due to the subjective nature of annotations, perfect accuracy is not achieved in practice. Yet since low-level features represent the inherent content of images, there is little subjectivity in specifying them. The main drawback is the inability to capture complete high-level semantics in such low-level features. As a

result, use of low-level features only is also not very accurate. Thus while keywords help in reducing the search space, it would be interesting to find a minimal set of keywords that can be used along with low-level features to achieve a desired level of precision. Let $K_1^{'} \subset K_1 \subseteq K$ be a set of keywords such that

$$C_K^I(K_1) = Q\left(K_1^{'}, i, f, d_f, I, \Delta d_f\right) \tag{7}$$

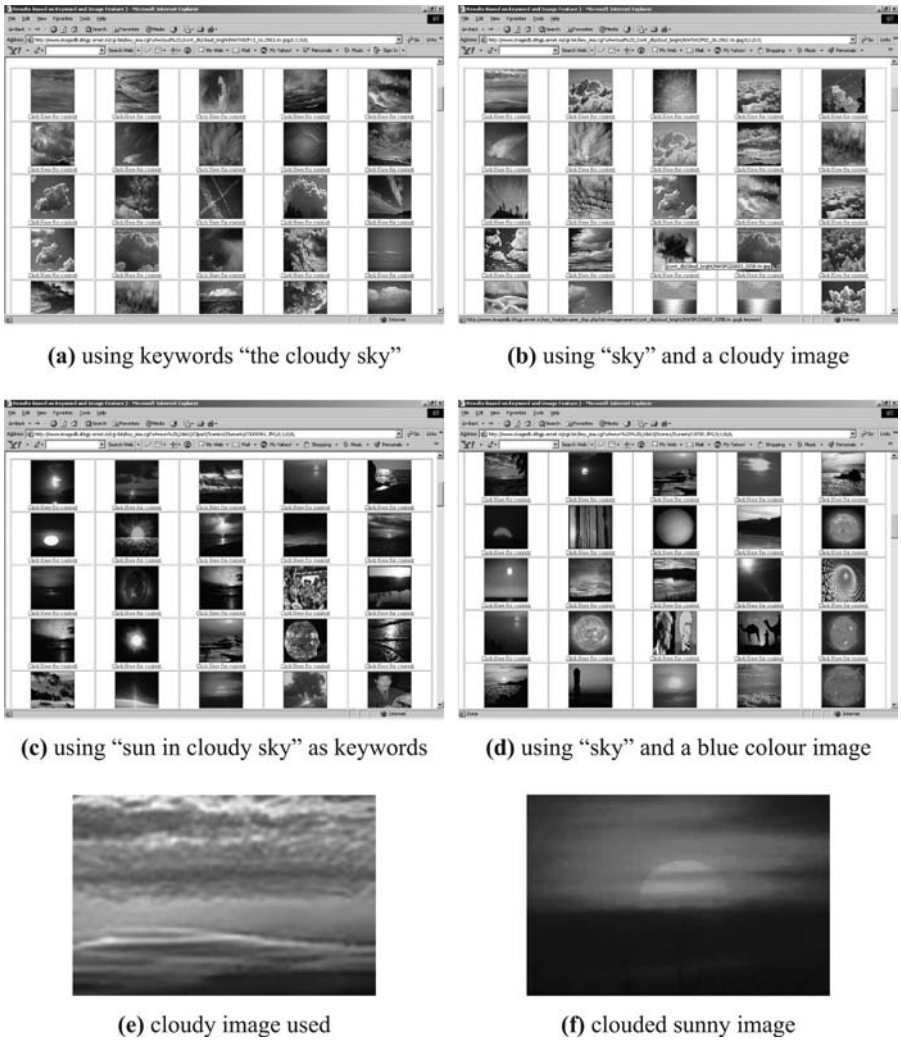
This implies that the set of images retrieved by using only the keywords K_1 is the same as the set of images retrieved using a lower number of keywords $K_1^{'}$ in conjunction with the image feature f. If such a subset $K_1^{'}$ could be found, we may say that the feature f represents the same semantics as that of the set of keywords $\left(K_1 - K_1^{'}\right)$ with respect to the query image q and image database I.

We have performed experiments on a keyword annotated image database to determine whether the same precision can be achieved using a smaller set of keywords and an image feature. We compare retrieval performance of a keyword-only query with a combined keyword and ICICM (Vadivel *et al.*, 2007) feature query. For instance, consider a query with keywords "cloudy sky", that is, $K_1 = \{\text{cloudy, sky}\}$. The images retrieved by this query are shown in Figure 1(a). When the keyword "cloudy" is replaced by an image of a cloudy sky (Figure 1(e)) and the keyword "sky" is retained, that is, $K_1^{'} = \{\text{sky}\}$ and $q =$ an image of a cloudy sky, $f =$ ICICM feature, $d =$ Euclidean distance, the retrieval result from the same image database is shown in Figure 1(b). These two result sets contain images that are not exactly the same, since they have been reordered based on the rank value. This is due to the fact that for $K_1 = \{\text{cloudy, sky}\}$, images labelled with both "cloudy" and "sky" are retrieved. However, for the other query with $K_1^{'} = \{\text{sky}\}$ with an example image, images with label "sky" are retrieved and ranked based on the distance value computed from the features of query image and retrieved images. From the two result sets, some images are found to be common, since $K_1^{'} \in K$ and hence images labelled with both {cloudy, sky} and {sky} are retrieved but with different rank value.

Similar observations can be made from Figures 1(c) and 1(d). Here retrieval based on the keywords "sun in cloudy sky" is compared with that using the keyword "sky" along with the ICICM feature of an image with sun and cloud (Figure 1(f)). Here also the results show considerable similarity. One may therefore conclude that with this image database, ICICM features associated with "cloudy" images capture the semantics of "cloudy" and ICICM features of sun and cloud images correspond to the keyword "sun cloud".

Dynamic extraction of keywords from HTML pages

We have mentioned above that if keywords are correctly assigned to images describing their semantic content, then keyword-based queries should return only the relevant images. However, such labelling of images with keywords cannot always be done in practice. For example, let us consider the large number of images available on the worldwide web, embedded in HTML pages. Most of these images are not annotated properly. However, we argue that images are not embedded arbitrarily in HTML pages. Usually there is a high correlation between the semantic content of an image and the textual description contained in the text of the HTML in which the image is embedded. We therefore propose to use keywords present in text for labelling the images to begin with. While this labelling may not always give correct semantic



(a) using keywords "the cloudy sky"

(b) using "sky" and a cloudy image

(c) using "sun in cloudy sky" as keywords

(d) using "sky" and a blue colour image

(e) cloudy image used

(f) clouded sunny image

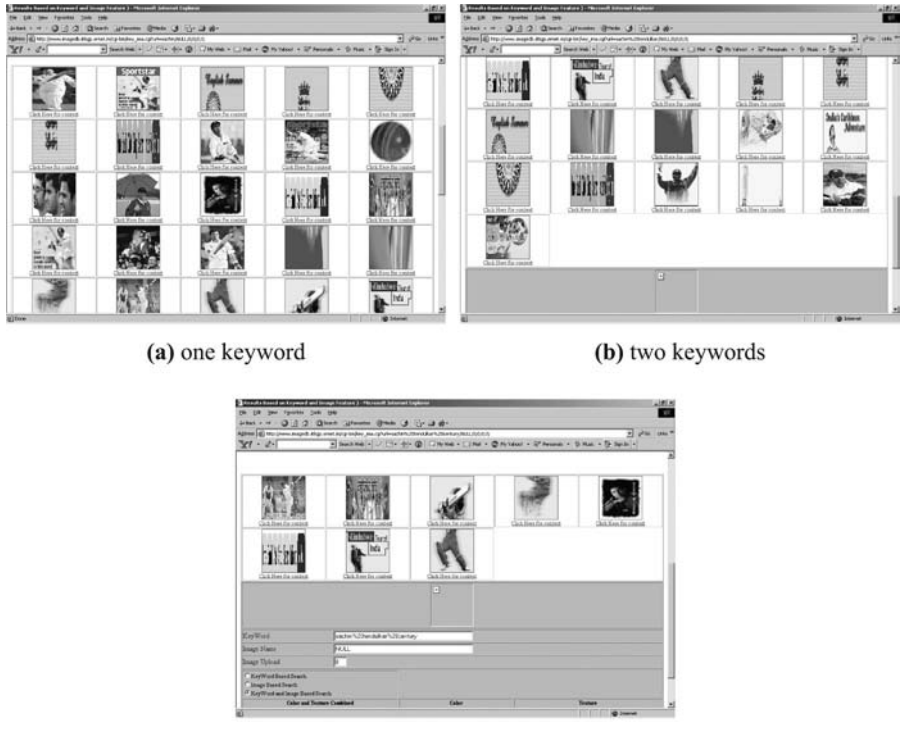
Figure 1.
Retrieval results

representation of the images, in the absence of any other annotation, such keywords are found to be the most suitable for labelling such images. Further, the higher the occurrence of a particular keyword in a page, the higher its probability of describing the content of an image included in that page. This is due to the fact that while developing pages with images, the name of the image and its description and related information are considered to provide more information about the image. For carrying out experiments, the HTML pages along with the images are initially crawled and the keywords are extracted. For keyword extraction, after removing common words and HTML tags, the entire content of the HTML file is processed. The frequency of occurrence of each keyword is calculated.

It is evident that labelling images with keywords from HTML text is not necessarily accurate although it is a good estimate of the semantic content. Thus in this situation the set of images retrieved using a set of keywords K_1 may not contain only the relevant images. As a result we cannot achieve 100 per cent accuracy in all cases. This is in contrast to the situation in which labelling is considered to be 100 per cent accurate. To improve precision however, one can increase the number of keywords. This intuitively should improve precision since pages that contain all such keywords in high frequency are also expected to contain images that are semantically close to the concept expressed by the larger set of keywords.

We performed experiments to see if this indeed is the case. This is illustrated in Figure 2. Figure 2(a) shows a large number of images that were retrieved using a query based on a single keyword. Figures 2(b) and 2(c) show retrieval results as more keywords were associated with the query.

To quantitatively study the variation in retrieval precision with the number of keywords in the query predicate, we experimented with queries having one, two, three, four and five keywords. We formed a team consisting of graduate students of various levels and research scholars from our department to evaluate results from our web based system by performing queries with keywords. For example, "Rahul Dravid scores century in the second test" was given as "Rahul", "Rahul Dravid", "Rahul Dravid scores", "Rahul Dravid scores century", etc. (Rahul Dravid is a popular Indian cricket player). Each member was given 20 such queries and the average was



(a) one keyword

(b) two keywords

(c) three keywords

Figure 2.
Retrieval result

calculated for plotting the results. For each of these keywords, precision values were estimated for 2, 5, 10, 15 and 20 nearest neighbours. The results are shown in Figure 3. From this figure, it is apparent that the retrieval precision increases with a higher number of keywords as queries. Since the images were crawled from the web, the relevant sets are unknown and hence it is not possible to plot precision versus recall. The precision values were obtained from the user based on the result set.

Weighted combination of keyword and feature-based distances

Let us consider an HTML page that contains N images $i_1, i_2, \ldots i_N$ and M keywords $k_1, k_2, \ldots k_M$. Let the frequency of occurrence of keyword k_j be $f_j, j = 1, \ldots, M$. We consider that all the keywords in one HTML page are equally relevant to all the images in the same HTML page. The association of each keyword $k_j (j = 1, \ldots, M)$ to each image $i_n (n = 1, \ldots, N)$ in an HTML page h can be denoted by:

$$Assoc(k_j, h) = \left(\frac{f_j}{\underset{j=1..M}{Max}(f_j)} \right) \tag{8}$$

Further improvement to such assignment can be made in future by considering locality of reference. In such cases, keywords in the same paragraph as an image can be considered as relevant to that image. In that case the function *Assoc* defined in equation (8) will be a function of image i in the page h also. The font and colour of keywords can also be given different weights. Based on Equation (8), association of a set of keywords K can be denoted by:

$$Assoc(K, h) = \underset{k \in K}{Min}\, Assoc(k, h) \tag{9}$$

The above expression denotes the association of the set of keywords K to each image i_n belonging to the HTML page h. This implies that we consider the frequency of the keyword having maximum frequency in a set of keywords to capture the association of

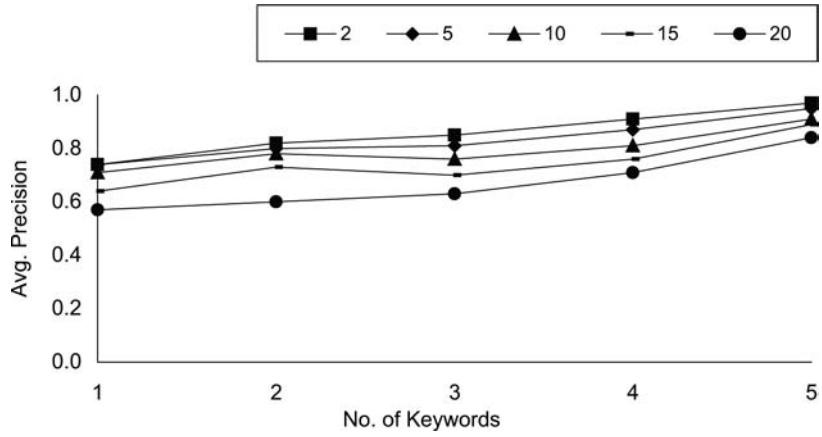


Figure 3.
Average retrieval precision for various nearest neighbours using keywords as queries

a set of keywords to an image in the HTML page. The min operation corresponds to Boolean ANDing of keywords discussed earlier in this paper. We can use the association formula of equation (9) to estimate the likelihood of a query image i specified by a set of keywords K to be present in an HTML page h. This can be denoted by:

$$d(K,h) = \underset{k \in K}{Min}\ Assoc(k,h) \tag{10}$$

It is assumed that each image is identified by its name and occurs only once in a given HTML page. Also, $0 \leq d(k,h) \leq 1$, $d(k,h) = 1$ if the frequencies of all keywords in K are the same, which in turn is equal to the maximum frequency of any keyword in the page h. $d(K,h) = 1$ also if the image name contains all the keywords K. This is likely to be the case because an image name usually has a high semantic correlation with the image content. $d(k,h) = 0$ if even one of the keywords in K is not present in the page h or in the image name. The reason for using keyword-based search before low-level feature-based search is that if the keyword is used before the low-level features, the retrieval set contains images that are associated with the query keyword only. However, if the low-level feature is used first, the result set will have all the images in the database with measured similarity value (i.e. distance) of each image in the retrieval set.

The feature-based distance between a given query image and the database images is estimated based on the Vector Cosine Angle Distance (VCAD). VCAD gives a similarity value in the range [0,1] with 1 denoting an exact match. The VCAD distance between a query image q and a database image i using a feature f is denoted by:

$$d_f^{VCAD}(q,i) = \frac{\overline{F_q}.\overline{F_i}}{||\overline{F_q}||\overline{F_i}||}$$

where $\overline{F_q}$ and $\overline{F_i}$ are the vector representations of the feature f for the images q and i, respectively.

For example, if we consider colour as the feature f, then $\bar{F}$ could be the colour histogram of an image. For a combined keyword and image based query with a set of keywords K and a query image q, the ranking of the database images is done using a weighted sum of the keyword-based distance and feature-based distance. The combined distance of a database image i denoted by $d_{k,q}(i)$ is defined as:

$$d_{K,q}(i) = w_k d(k,h) + w_f d_f^{VCAD}(q,i) \tag{11}$$

where $w_k + w_f = 1$.

In our work we initially assign both $w_k = w_f = 0.5$ and later the weight will be redistributed. The query text matching the image name is assigned 0.25, query text matching the text around the image is assigned 0.125 and the text colour is assigned the remaining 0.125 of the weight. Similarly for the low-level features when ICICM is used, the weights are distributed based on the probability distribution of each sub-matrix of ICICM.

Similar to the result shown in Figure 3, we examined the variation in retrieval precision with number of keywords and image features in the query predicate with queries having one, two, three, four and five keywords and an example image with feature f. The precision value was estimated for each of these keywords with an example

image for 2, 5, 10 and 20 nearest neighbours. The experiments were performed with several query images associated with keywords and the average retrieval precision for different numbers of keywords with a given query image is shown in Figure 4.

In our experiment we randomly selected a number of keywords to form a set of meaningful query strings with a maximum of five keywords. The precision was measured using the first keyword of each query string. Similarly the second keyword of each query string was chosen and concatenated with the respective first keyword and the precision was measured. Finally the average precision was computed. It is evident from Figure 4 that the retrieval precision increased considerably. When five keywords and an example image were combined as a query in the query predicate, precision was about 99 per cent. This is due to the fact that if the keyword is used before the low-level features, the retrieval set contains a set of images, which are associated with the query keyword. Conversely if the low-level features are used first, the result set will have all the images in the database with calculated similarity value (i.e. distance) of all images in the retrieval set, which results in more irrelevant images in the retrieval set. Further increases in the number of keywords, therefore, would not provide any significant improvement in precision.

Architecture of an integrated keyword and feature-based image retrieval system

The proposed retrieval system is broadly divided into two units: a feature extraction unit and a retrieval unit. In the feature extraction unit, HTML pages are crawled using a web crawler, the found HTML documents and images are then pre-processed to extract low-level features and keywords. These features and labels are treated as a secondary index for the underlying image database. While low-level features are inserted in a feature table, high-level features are inserted in a keyword table of the feature database. The architecture of the system is shown in Figures 5(a) and 5(b).

We explain the important blocks in the architecture below.

Web crawler

We use an internet crawler that makes a breadth first search of the World Wide Web starting from a seed webpage. It fetches the images along with HTML documents.

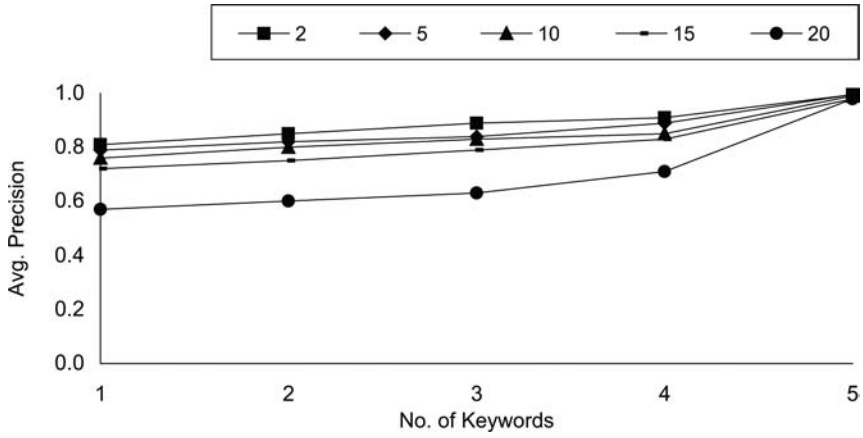


Figure 4.
Average retrieval precision for various nearest neighbours with keywords and image as combined query

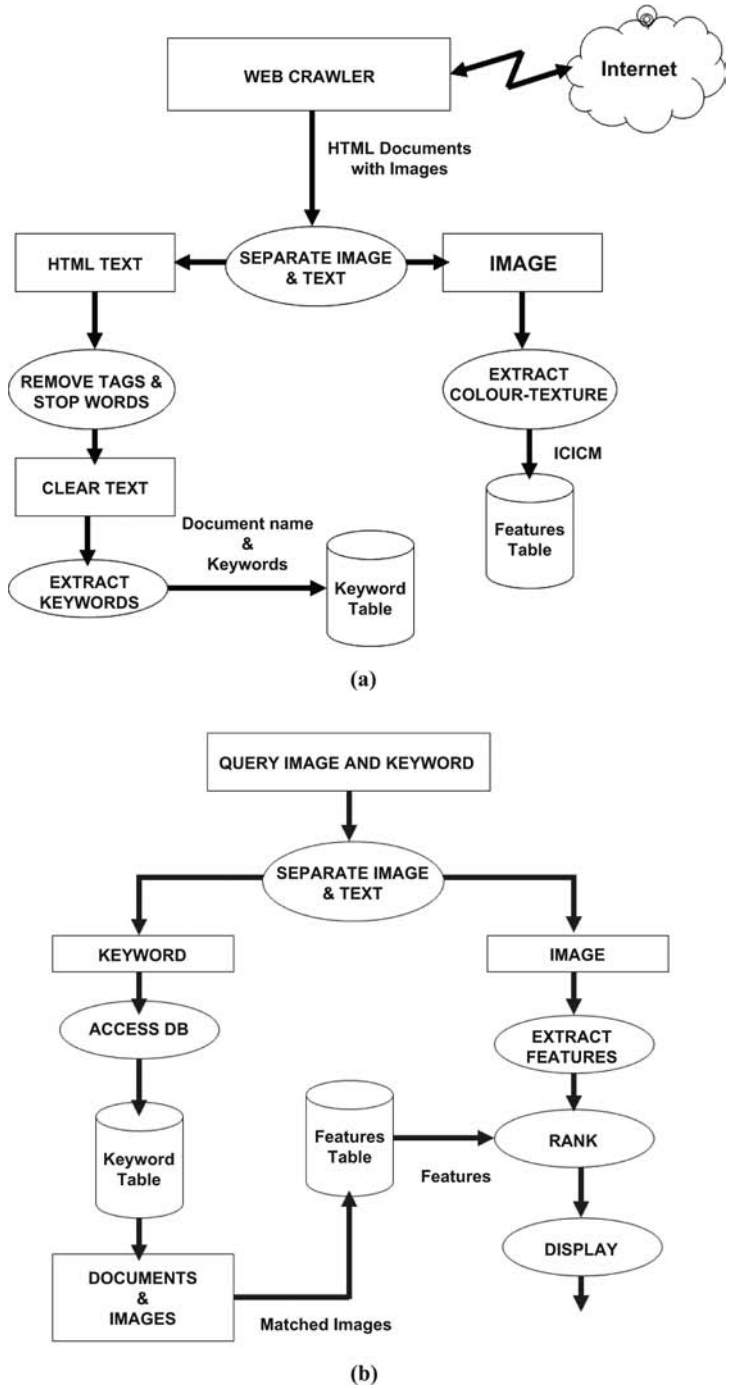


Figure 5.
Architecture of the system: (a) feature extraction unit, (b) retrieval unit

Since the keyword extraction and low-level feature extraction are carried out locally for an entire website, both the HTML and the images are crawled. This is done to avoid computation overhead. The worldwide web may be considered as a graph with the HTML pages forming the nodes and hyperlinks connecting one page to another forming the edges (see Figure 6).

In Figure 6, "d" denotes the depth of a node in the graph. Each node in the graph has three attributes: text, image and URL link. The crawler makes a breadth first search from the root until it reaches a specified depth "d". The content of a node is fetched only

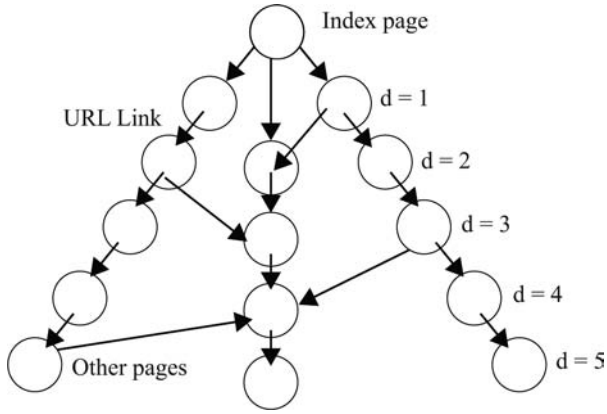


Figure 6.
Graph structure of the WWW

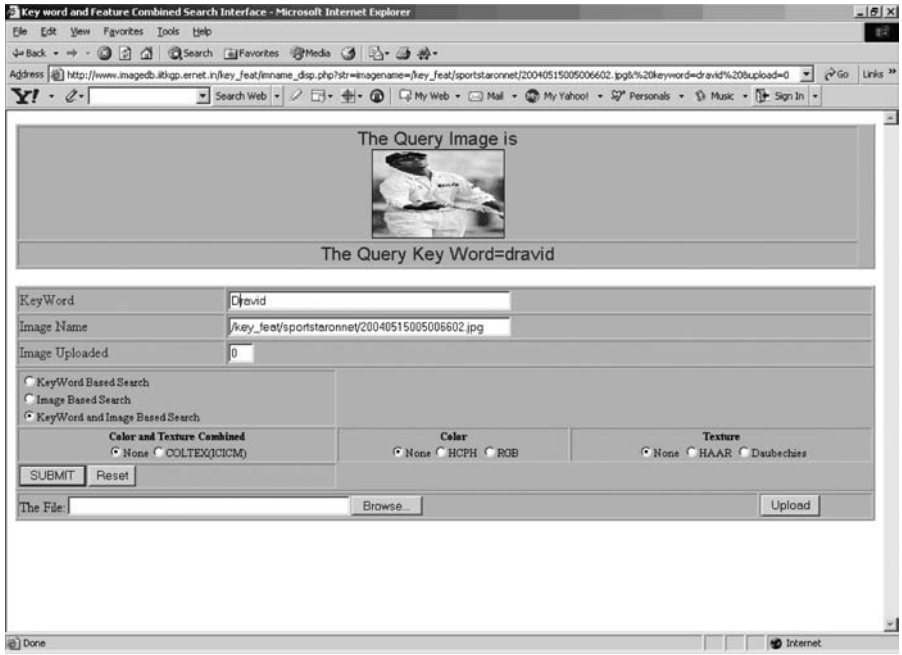
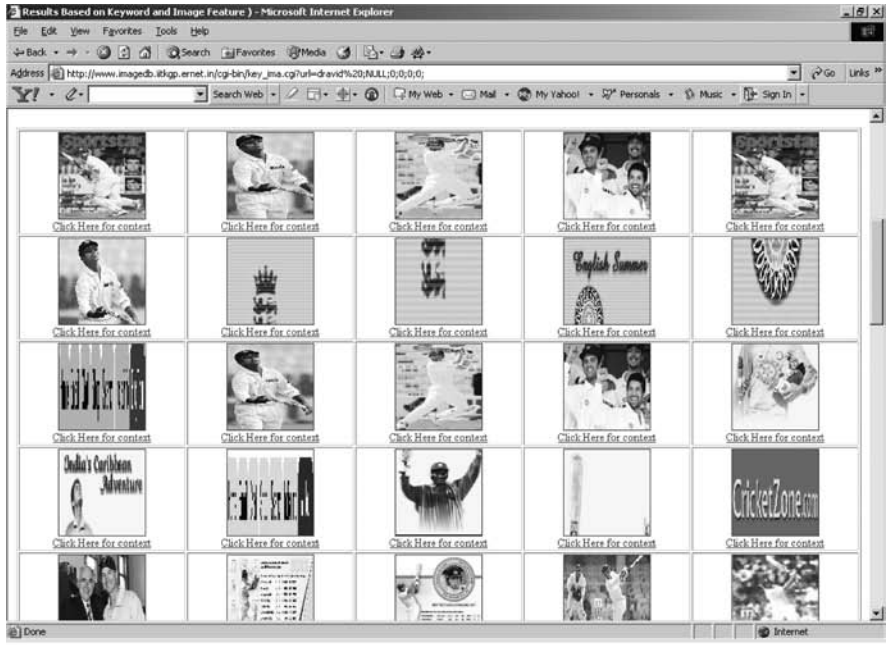
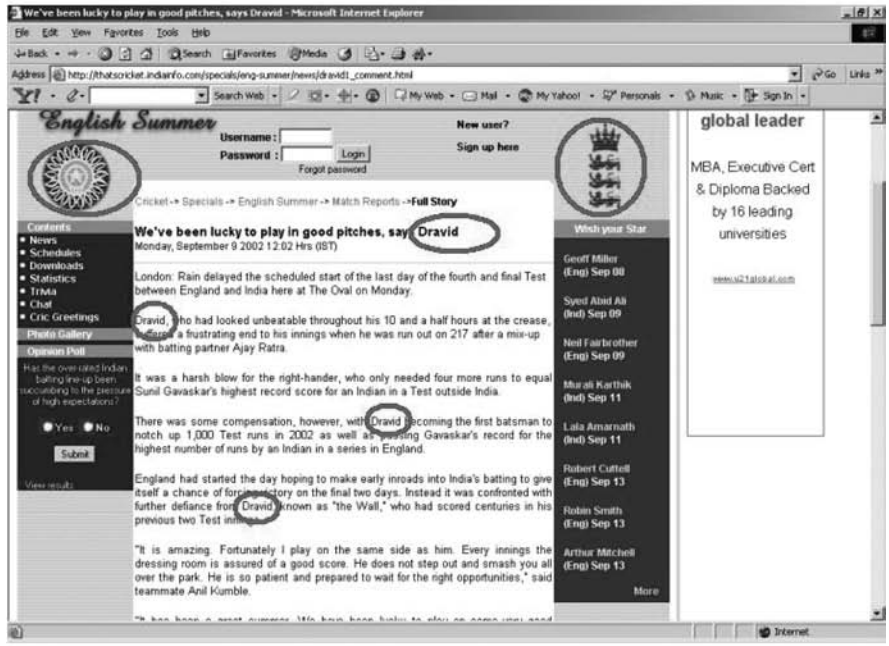


Figure 7.
Query specification interface



(a) retrieval results based on keywords



(b) view of the HTML document of irrelevant images

Figure 8.

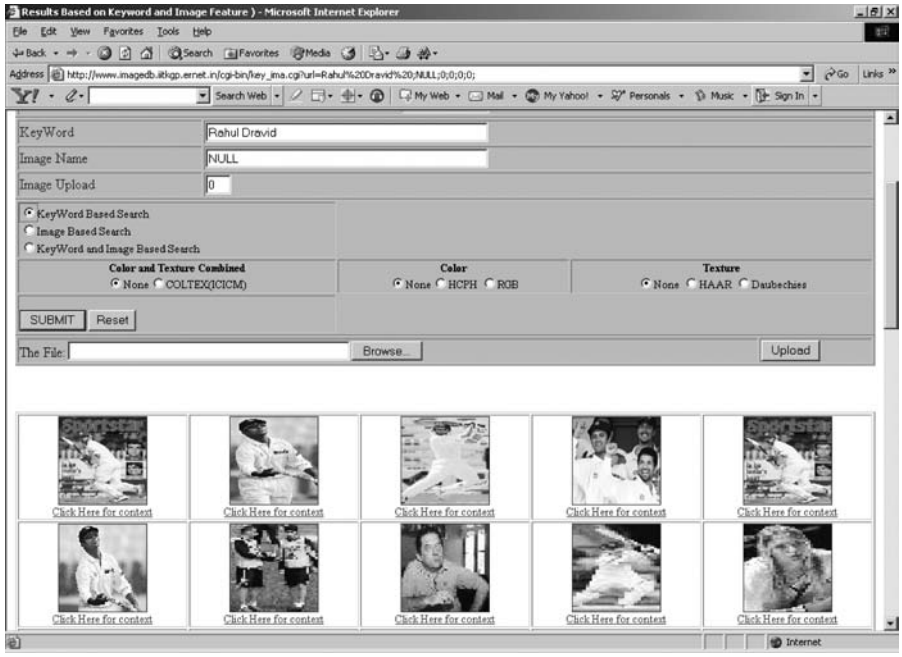


Figure 9.
Retrieval result based on "Rahul Dravid" as a query

if the node contains any images. Further, the link from one node may be referring to the same site or may be referring to another site in the web.

Image retrieval

During retrieval for keyword-based queries the retrieved result is ranked based on the frequency of occurrence of the keywords. As mentioned earlier, locality of reference can be used to further enhance the performance of such queries. However, if the keywords are combined with the images to form a query, the entire image feature in the database is compared with the query image feature to estimate the similarity. The images are assigned weights based on the frequency of occurrence of keywords and the final ranking is determined by the feature and the keyword. A user can also query the database based on image features alone. We used Vector Cosine Angle Distance to measure the similarity between images.

Figure 7 shows the query specification interface. Using this interface, a query can be initiated with keywords, image or both. The interface provides the facility to select various low-level features of an image and its combinations. The system has a provision to let the users upload their own images for performing queries.

For a purely keyword-based retrieval, an input keyword is issued to the system and, based on its frequency of occurrence, the result is ranked and presented to the user. Figure 8(a) shows retrieval results for the keyword "Dravid" as a cricket-related query. As shown in Figure 8(b) some of the images are not images of Dravid – these images are retrieved since both the "Dravid" keyword and the images appear in the HTML document.

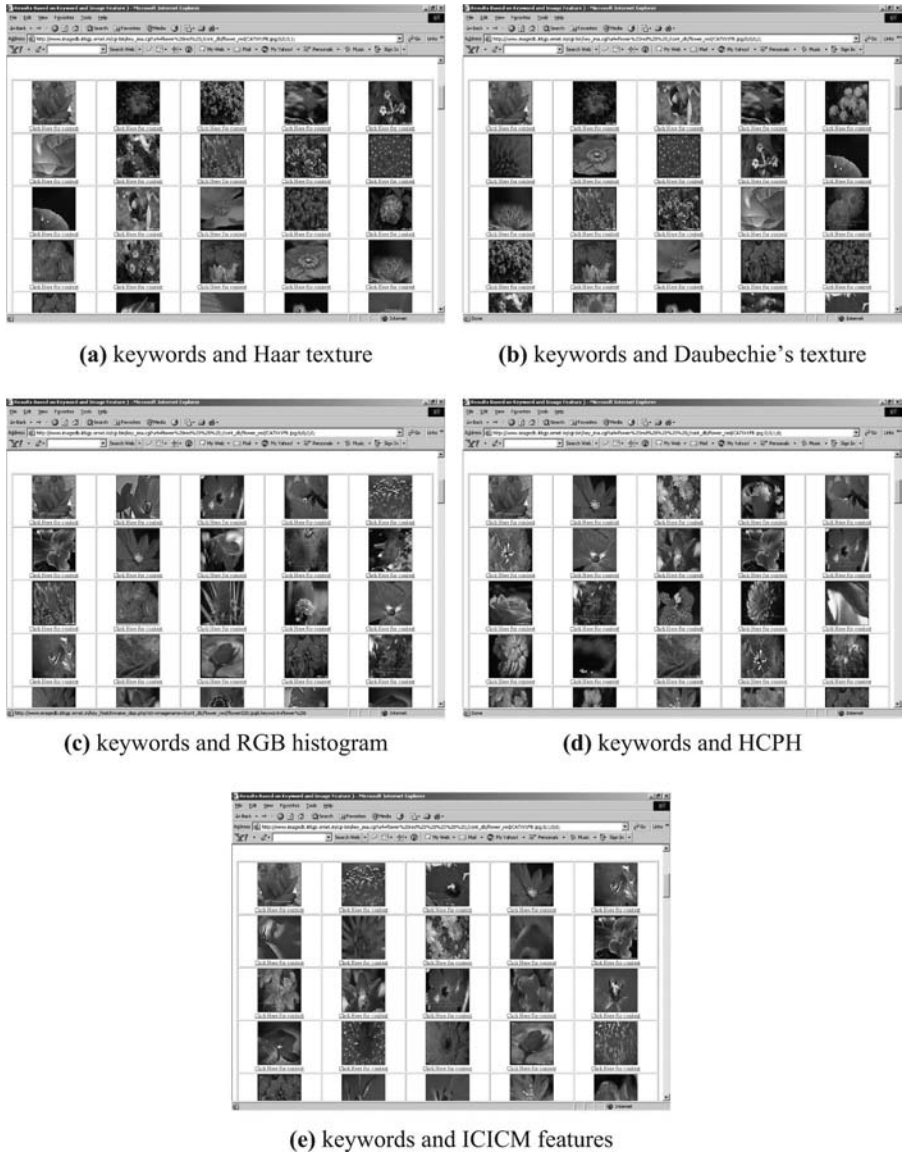
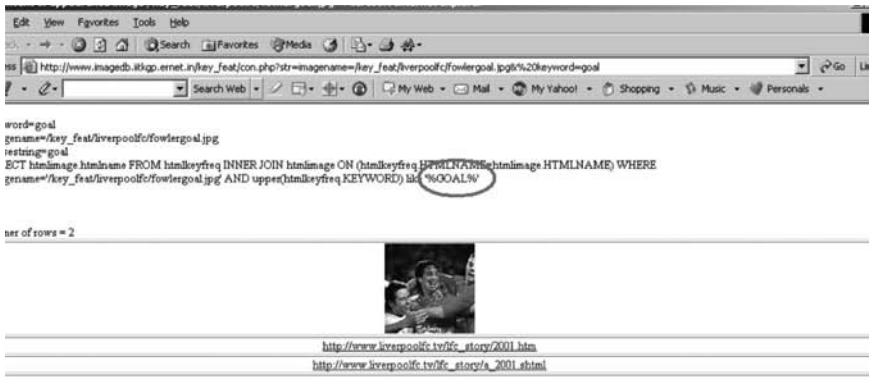


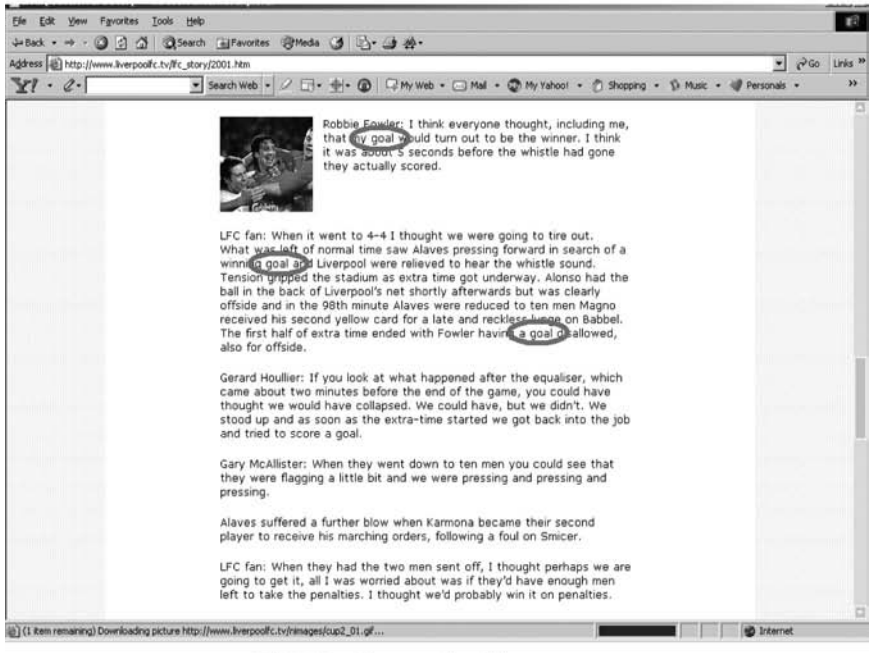
Figure 10.
Retrieval result based on features

Figure 9 shows the retrieval results using more keywords as queries to restrict the search space. Again the query is initiated with the keywords "Rahul Dravid". It is evident from Figure 9 that the number of irrelevant images is considerably reduced in the retrieved result set.

Next we display a retrieval result set using both the keywords "flower rose" and low-level features. Figure 10(a) shows the retrieval results using the Haar wavelet-based texture feature and keywords. Similarly we show the retrieval results



(a) URL of image and keywords



(b) URL of keyword and image context

Figure 11.
Context reference

using keywords along with the following features: Daubechie's wavelet based texture feature, RGB Histogram, HCPH and ICICM in Figures 10(b)-(e) respectively.

In Figure 10 the first image of the retrieved result is the query image. It can be observed that the visual content of the result set closely matches the query images when the keywords are combined with ICICM features. The proposed retrieval system includes the facility to show the original URLs in which the images and documents were found. The user can visit the URL and see the context of each image in the result set. Figure 11(a) shows the results for context reference with "goal" in the context of a soccer query keyword. Figure 11(b) shows the original context of the image and the keywords.

Results

Initially we measured retrieval precision using keywords and ICICM separately. Next, low-level features such as HCPH, ICICM and texture were combined with keywords and the precision was measured. Table I shows the average precision (P) percentage for 10, 20, 50 and 100 nearest neighbours.

We used large numbers of images in different categories and content as query images. We determined the precision for each of these images and computed their average precision. From the results shown in Table I, the performance of the combined method is quite encouraging. The retrieval precision of the keyword and ICICM combination was higher than for other combinations. We also compared the performance of the combination of keyword and ICICM with a number of other techniques, the results of which are shown in Table II. In this experiment we compared Photobook proposed by Pentland *et al.* (1994), Lee and Yoo's (2000) Neural Network Based Image Retrieval system, Zhou and Huang's (2002) Unifying Keywords with Visual Content for Image Retrieval (UKVCIR), and CLUE proposed by Chen *et al.* (2005).

Table II shows that the Photobook method's retrieval performance is the worst. The neural network based image retrieval method achieves high retrieval precision for ten and 20 nearest neighbours. However, the retrieval performance of ICICM feature with

Table I.
Retrieval precision of the features used in our retrieval system

Features	P(10)%	P(20)%	P(50)%	P (100)%
Keyword	46.00	35.90	25.40	19.98
ICICM	55.00	46.90	35.40	22.98
Harr (texture) and keyword	85.00	76.78	55.09	50.23
Daubechie's (texture) and keyword	85.98	77.00	56.45	51.76
HCPH and keyword	89.65	82.98	60.01	54.14
ICICM and keyword	*92.33*	*87.28*	*64.21*	*51.00*

Table II.
Comparison of retrieval precision

Features	P(10)%	P(20)%	P(50)%	P(100)%
Photobook	40.90	31.08	20.87	10.05
NNIR	85.00	70.76	59.51	41.71
UKVCIR	87.11	79.09	60.71	50.00
CLUE	75.89	69.10	63.91	50.37
ICICM with keyword	*92.33*	*87.28*	*64.21*	*51.00*

keywords is much better. It can also be seen that the precision values of CLUE for 50 and 100 nearest neighbours are comparable to those of ICICM.

Conclusion

We have studied the important role of textual keywords in describing high-level semantics of an image. We felt that HTML documents could be effectively used to define the high-level semantics of images on the worldwide web. A crawler fetches the HTML documents along with their images from the internet. Keywords are extracted from the HTML documents after eliminating HTML tags, stop words and common words. The words' frequency of occurrence in the HTML documents is computed to assign weights to the keywords. During retrieval using only keywords as queries, the final ranking is based on frequency of occurrence. In addition, the image name is also treated as a keyword with high value occurrence frequency to improve the retrieval precision. It is observed from the experiments that using more keywords as queries considerably enhances the precision by leaving out irrelevant images. ICICM has been used as a low-level feature of the images and is combined with keywords. Combining both high-level and low-level features of images achieves high average retrieval precision.

References

Cai, D., Yu, S., Wen, L.R. and Ma, W.Y. (2003), "VIPs: a vision based page segmentation algorithm", Microsoft Technical Report, MSR-TR-2003-79, Microsoft Research Asia, Beijing.

Cai, D., He, X., Ma, W-Y., Wen, J-R. and Zhang, H. (2004), "Organizing WWW images based on the analysis of page layout and web link structure", *Proceedings of International Conference on Multimedia Expo, Beijing, China, IEEE*, pp. 113-6.

Chen, Y., Wang, J.Z. and Krovetz, R. (2005), "CLUE: cluster-based retrieval of images by unsupervised learning", *IEEE Transactions on Image Processing*, Vol. 14 No. 8, pp. 1187-201.

Gevers, T. and Stokman, H.M.G. (2004), "Robust histogram construction from colour invariants for object recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26 No. 1, pp. 113-8.

Han, J., Ngan, K.N., Li, M. and Zhang, H-J. (2005), "A memory learning framework for effective image retrieval", *IEEE Transaction on Image Processing*, Vol. 14 No. 4, pp. 511-24.

Hu, W., Wu, O., Chen, Z., Fu, Z. and Maybank, S. (2007), "Recognition of pornographic webpages by classifying texts and images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29 No. 6, pp. 1019-34.

Jing, F., Li, M., Zhang, H-J. and Zhang, B. (2005), "A unified framework for image retrieval using keyword and visual features", *IEEE Transaction on Image Processing*, Vol. 14 No. 7, pp. 979-89.

JISC (2008), "Review of image search engines", (online), available at: www.jiscdigitalmedia.ac.uk/stillimages/advice/review-of-image-search-engines/ (accessed 17 September 2009).

Lee, H.K. and Yoo, S.I. (2000), "Nonlinear combining of heterogeneous features in content-based image retrieval", in Casasent, D.P. (Ed.), *Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision XIX: Algorithms, Techniques, and Active Vision*, Vol. 4197, pp. 288-96.

Palm, C. (2004), "Colour texture classification by integrative co-occurrence matrices", *Pattern Recognition*, Vol. 37 No. 5, pp. 965-76.

Pentland, A., Picard, R.W. and Sclaroff, S. (1994), "Photobook: tools for content-based manipulation of image databases", *Proceedings of the SPIE Conference on Storage Retrieval for Image and Video Databases, San Jose, CA*, Vol. 2185, pp. 34-47.

Tollari, S., Glotin, H., Le, J. and Maitre, J.L. (2005), "Enhancement of textual images classification using segmented visual contents for image search engine", *Multimedia Tools and Applications, Special Issue on Metadata and Adaptability in Web-based Information Systems*, Vol. 25 No. 3, pp. 405-17.

Vadivel, A., Sural, S. and Majumdar, A.K. (2007), "An integrated colour and intensity co-occurrence matrix", *Pattern Recognition Letters*, Vol. 28 No. 8, pp. 974-83.

Vadivel, A., Sural, S. and Majumdar, A.K. (2008), "Robust histogram generation from the HSV colour space based on visual perception", *International Journal on Signals and Imaging Systems Engineering*, Vol. 1 Nos 3/4, pp. 245-54.

Xu, F. and Zhang, Y-J. (2007), "Integrated patch model: a generative model for image categorization based on feature selection", *Pattern Recognition Letters*, Vol. 28 No. 12, pp. 1581-91.

Yang, H-C. and Lee, C-H. (2008), "Image semantics discovery from webpages for semantic-based image retrieval using self-organizing maps", *Expert Systems with Applications*, Vol. 34 No. 1, pp. 266-79.

Zhao, R. and Grosky, W.I. (2002), "Narrowing the semantic gap – improved text-based web document retrieval using visual features", *IEEE Transactions on Multimedia*, Vol. 4 No. 2, pp. 189-200.

Zhou, X.S. and Huang, T.S. (2002), "Unifying keywords and visual contents in image retrieval", *IEEE MultiMedia*, Vol. 9 No. 2, pp. 23-33.

Zhou, X.S., Rui, Y. and Huang, T.S. (1999), "Water-filling: a novel way for image structural feature extraction", pp. 570-4, *Proceedings of International Conference on Image Processing, Kobe, Japan*, Vol. 2, IEEE.

About the authors

A. Vadivel is an Assistant Professor at the National Institute of Technology, Tiruchirappalli, India. He received his MTech and PhD from the Indian Institute of Technology (IIT), Kharagpur, India, in 2000 and 2006 respectively. His research interest is image and video processing. He was awarded an Indo-US Research Fellow Award in 2008 by the Indo-US Science and Technology Forum, India. A. Vadivel is the corresponding author and can be contacted at: vadi@nitt.edu

Shamik Sural is an Associate Professor in the School of Information Technology, IIT Kharagpur, India. He received his PhD from Jadavpur University in 2000. He has worked in a number of organisations both in India and the USA in various capacities. Dr. Sural has served on the Programme Committee of a number of international conferences. He is a senior member of the IEEE. He has published more than 60 papers in important international journals and conferences. His research interests include database security, data mining and multimedia database systems.

A.K. Majumdar is a Professor in the Department of Computer Science and Engineering, IIT Kharagpur. Professor Majumdar received MTech and PhD degrees from the University of Calcutta in 1968 and 1973 respectively. He also earned a PhD in Electrical Engineering from the University of Florida, Gainesville, in 1976. He has more than 140 research publications in international journals and conferences. Professor Majumdar is a Fellow of the Institute of Engineers (India), a Fellow of the Indian National Academy of Engineering and a Senior Member of the IEEE. His research interests include data and knowledge based systems, design automation and image processing.

To purchase reprints of this article please e-mail: **reprints@emeraldinsight.com**
Or visit our web site for further details: **www.emeraldinsight.com/reprints**