Fuzzy Sets and Systems 160 (2009) 3323–3341

Content-based image retrieval using visually significant point features

Minakshi Banerjee[a,*], Malay K. Kundu[a,b], Pradipta Maji[a,b]

[a]*Center for Soft Computing Research, Indian Statistical Institute, 203, B. T. Road, Kolkata 700 108, India*
[b]*Machine Intelligence Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700 108, India*

Abstract

This paper presents a new image retrieval scheme using visually significant point features. The clusters of points around significant curvature regions (high, medium, and weak type) are extracted using a fuzzy set theoretic approach. Some invariant color features are computed from these points to evaluate the similarity between images. A set of relevant and non-redundant features is selected using the mutual information based minimum redundancy-maximum relevance framework. The relative importance of each feature is evaluated using a fuzzy entropy based measure, which is computed from the sets of retrieved images marked relevant and irrelevant by the users. The performance of the system is evaluated using different sets of examples from a general purpose image database. The robustness of the system is also shown when the images undergo different transformations.

Keywords: Content-based image retrieval; High curvature points; Fuzzy feature evaluation index; Color; Invariant moments

1. Introduction

Effective image retrieval from a large database is a difficult problem, and is still far from being solved. Hence, the retrieval of relevant images, based on measuring the similarity between automatically derived features (color, texture, shape, etc.) of the query image and that of the images stored in the database, a problem popularly known as content-based image retrieval (CBIR) [1–3], is a highly challenging task.

There are several popular CBIR techniques in the literature [4–7]. In conventional CBIR approaches, an image is usually represented by a set of features, where the feature vector is a point in a multidimensional feature space. Each feature tries to capture only one property of the image, such as color, texture, shape, etc. It is therefore necessary to select the optimal set of features suitable for a particular type of query, in which the component features may also be varied according to their importance.

Although extensive research has been performed on the CBIR over several decades, we have yet to achieve the desired accuracy from a fully automated CBIR system. An image may implicitly require various kinds of visual reasoning about the meaning or the purpose of different objects. The discrepancy between the actual information and its representation

* Corresponding author. Tel.: +91 3325753100; fax: +91 3325773035.
E-mail addresses: minakshi_r@isical.ac.in (M. Banerjee), malay@isical.ac.in (M.K. Kundu), pmaji@isical.ac.in (P. Maji).

using the computed feature values is known as semantic gap. The accuracy of a CBIR system may be improved by an iterative process of refinement of queries and visual features, guided by user's feedback [8–11], which is commonly known as the relevance feedback mechanism.

Feature extraction is one of the most important parts in designing a CBIR system. The extracted features should be well separated in the feature space to produce effective discrimination between images. The importance of each selected feature of the feature set may further be varied by a suitable weighting factor derived from the feature evaluation index (FEI) [12,13] based on the user's impressions from the relevance feedback.

Natural images generally contain edges and corners, which are locally defined by their positions. These features are visually significant [14] and carry a high proportion of the information contained in the image, with a limited number of pixels. The human visual system is highly efficient in selecting such significant features from an image, which may be used to evaluate the similarity between images [15,16]. This idea is used in the present work.

Image retrieval tasks based on visually significant points are reported in [17–19]. A recent work on the optimal use of color points of interest for the CBIR is presented in [20]. The detection of significant points based on the analysis of statistics of color derivatives is proposed in [21]. Scale invariant feature transform (SIFT) [17] is an efficient algorithm for object recognition, which is based on local extrema of differences of Gaussian filters in scale space. Potential points of interest are identified using a difference of Gaussian functions that are invariant to scale and orientation. Local image gradients are measured at the selected scale in the region around each key point. This representation allows significant levels of local shape distortion and change in illumination. These invariant features have good applications for identifying objects among clutter and occlusion.

Mikolajczyk et al. [18] proposed an interest point detector, which is invariant to scale and affine transformations also. From a multiscale representation of the Harris interest point detector, the characteristic scale is selected by selecting distinctive points at which a local measure (the Laplacian) is maximal over scales. The affine shape of a point neighborhood is estimated based on the second moment matrix. Matching between two images is performed by determining point to point correspondence.

An image retrieval system is proposed in [22], in which local features such as color, texture, etc. are computed on a window of regular geometrical shape surrounding the corner points. The corner points are detected by general purpose corner detectors like [23].

To estimate the local features around significant curvature points, it is necessary to select window size and shape properly. A large window may incorporate lot of insignificant information (noise) along with significant information, whereas small ones may leave out a lot of important information. Hence, the estimation of local features on a fixed sized window will have some limitations as the spread of significant information may be different depending on the type of curvature points (sharp, medium or weak). The characteristics of sharp curvature points will be confined within a small region, while for those of medium and weak type, the region will be larger. These facts indicate the usefulness of extracting the possible high curvature region of interest (ROI), whose shape and size vary adaptively according to the nature of curvature type.

In general, segmentation into multiple regions has been a primary step in image analysis, especially for image retrieval applications [24–26]. A similarity measure between regions was introduced in the FIRM [25], where the properties of all regions are integrated by a family of fuzzy features, in order to evaluate similarity between images. Although segmentation is important for the effective characterization of an image, it is a difficult task without prior knowledge about the classes present in an image. Segmentation becomes difficult when there is a large number of fragmented objects in a scene, or no specific objects. Instead of segmentation and detailed region representation, the information derived from the ROI may be subsequently used for overall scene matching applications.

The proposed technique is based on the assumption that two visually similar images will have similar visual characteristics. A representation that captures local characteristics of visually significant portions is very important in generating inferences from portions that draw visual attention with reduced computational cost. However, some global measurements may also be considered to obtain a better representation, as the notion of similarity is dependent on both local (significant curvature points and around) and global (image in total) views. Looking into these aspects, a fuzzy set theoretic approach is proposed to extract clusters around different types of curvature points (sharp, medium and weak), whose centroid almost depicts a true corner. These points are considered as candidate points for the computation of features. The invariant moments of the extracted point sets are used as features for similarity evaluation. An optimal set of relevant and non-redundant features is selected based on the mutual information criterion, and these features are then used to retrieve images in the first pass. A fuzzy entropy based feature evaluation mechanism [27] is

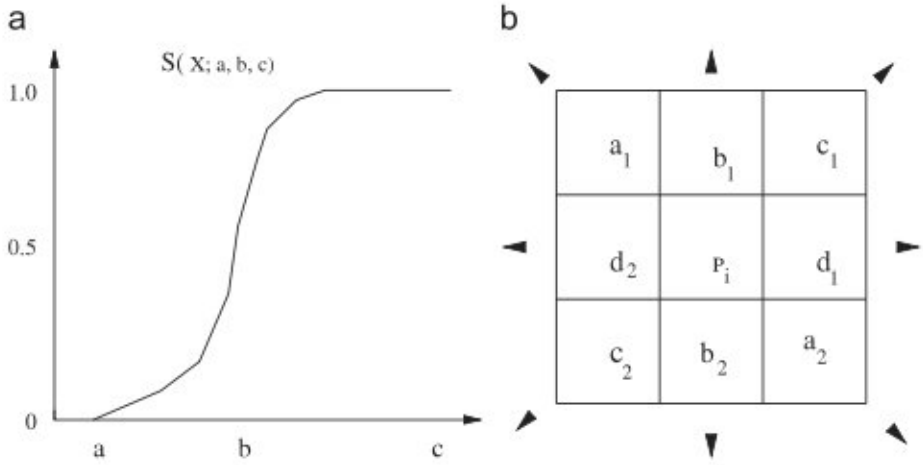

Fig. 1. (a) S type function. (b) 3×3 neighborhood of a pixel.

presented to enhance the accuracy of the system based on user's feedback. The user marks the relevant images and the irrelevant images in the retrieved set. The individual feature weights are updated using a measure known as the fuzzy feature evaluation index (FEI) [27]. This value is computed from the 'intraset ambiguity' and the 'interset ambiguity' as obtained from the relevant and irrelevant set of images. The results of the proposed methodology are compared with those of some well known techniques, namely, the integrated region based approach [28,25] and the color histogram method [29].

The organization of the paper is as follows: In Section 2, the proposed methodology is described in detail, while the experimental steps and corresponding extensive results are demonstrated in Section 3. Concluding remarks are presented in Section 4.

2. Proposed methodology

The feature extraction and feature evaluation methods based on fuzzy methodology are explained in the following subsections. The proposed CBIR method for the retrieval of related images consists of the following phases: 1. Extraction of visually significant points. 2. Computation of color features at significant points. 3. Feature selection. 4. Feature evaluation. Each of the four phases will now be elaborated, one by one.

2.1. Extraction of visually significant points

An efficient CBIR system should be able to model the vagueness that is usually present in the image content, relevance feedback, etc. A fuzzy set theoretic approach may be a good choice for handling the uncertainties that arise at different stages of the processing and analysis of a CBIR system [25,26].

The proposed technique is based on work reported in [30,31]. The potential fuzzy corner regions are extracted using the topographic characteristics of intensity surfaces [32,33] using the terms Plateau Top and Bottom, which are similar to the ridges and valleys of gray level images. The uncertainty arising in locating such points is handled using a fuzzy set theoretic approach. The discontinuities in the intensity surfaces are the possible candidates for curvature points.

The feature computation process involves assigning three membership values to the candidate points. In the first phase, the possible edge candidates P_c are extracted, which lie between the Plateau regions of the Gaussian smoothened image. These points are characterized by gradient membership ($\mu_d(P)$), generated by a S type function shown in Fig. 1(a).

The standard S function of an event (x), which is used in the proposed algorithm, is given by

$$\begin{aligned} S(x; a, b, c) &= 0 && x \leqslant a \\ &= 2 \times \left\{\frac{(x-a)}{(c-a)}\right\}^2 && a \leqslant x \leqslant b \\ &= 1 - 2 \times \left\{\frac{(x-c)}{(c-a)}\right\}^2 && b \leqslant x \leqslant c \\ &= 1 && x \geqslant c. \end{aligned} \tag{1}$$

In $S(x; a, b, c)$ the parameter b is the cross-over point, i.e. $S(b; a, b, c)$= 0.5. The shoulder point is at c, at which $S(x)$=1.0. The point a is the feet point, i.e. $S(a; a, b, c)$= 0.0, as shown in Fig. 1(a).

The assignment of the membership value is based on the local gray level contrast [31], given by

$$\mu_d(P) = S(x; a, b, c), \tag{2}$$

where x is determined from the ratio of contrast between two opposite pixels $(X_m r)$ over a specified window as shown in Fig. 1(b), and

$$X_m r = \min\{X_r\}, \tag{3}$$

where

$$X_r = \left[\frac{1+|a_1-a_2|}{1+|c_1-c_2|}, \frac{1+|c_1-c_2|}{1+|a_1-a_2|}, \frac{1+|b_1-b_2|}{1+|d_1-d_2|}, \frac{1+|d_1-d_2|}{1+|b_1-b_2|}\right]. \tag{4}$$

The parameters a and c are determined from the maximum and minimum value of $X_m r$, which maps the membership between 0.0 to 1.0. By thresholding on $(\mu_d(P))$, the fuzzy edge map, as shown in Fig. 3(b), may be obtained. Two more memberships (μ_f, μ_b) are computed to estimate the strength of connectedness on both sides of the curvature junction. The memberships assigned are related to the value of the curvature subtended at the corresponding points. The flowchart of the proposed algorithm is explained in Fig. 2.

The proposed algorithm assumes that the high curvature points should lie on valid gray level edges. Different sets of curvature points C_{ik} are obtained from the fuzzy edge map by selecting different membership values $\mu_d(P)$ as thresholds and applying fuzzy rules to the computed features (μ_f, μ_b). A cluster of such points is called a corner signature.

The high curvature region of Fig. 3(a) is plotted as (*) on the edge map thresholded at $\mu_d(P) \geqslant 0.6$, and shown in Fig. 3(b). If the threshold is selected at $\mu_d(P) \geqslant 0.5$, both high and medium curvature points will be extracted together. Threshold values $\mu_d(P) \leqslant 0.5$ are not considered because along with the high and medium curvature points, they simultaneously select a lot of spurious curvature points which may reduce the accuracy of the techniques. Experimentally it is found that the proposed algorithm performs better with the values $\mu_d(P) \geqslant 0.5$, typically at (0.6, 0.7 and 0.8).

The feature descriptors, computed from the fuzzy edge map, work successfully in the retrieval of gray level images as shown in Fig. 4, which is reported in [31]. To generate a descriptor rich in structural information, we intend to extract clusters of significant points depicting the change in curvature regions. The clusters of points ROI around the corners carry shape information as well as information about the spatial distribution of those points.

The feature extraction procedure is implemented for color images by converting the (red, green, blue) RGB plane to (hue, saturation, intensity) HSI, and considering only the intensity component to detect the corner signature. The extracted corner signatures for some images are shown in Figs. 5 and 6. The signature does not change significantly under varying imaging conditions, as shown in Fig. 7(a) and (b).

2.2. *Computation of color features at significant points*

The color properties of the extracted candidate points are considered to compute the invariant moments. The invariant moments are computed from the component planes in terms of (c_1, c_2 and c_3), as obtained from the RGB representations.

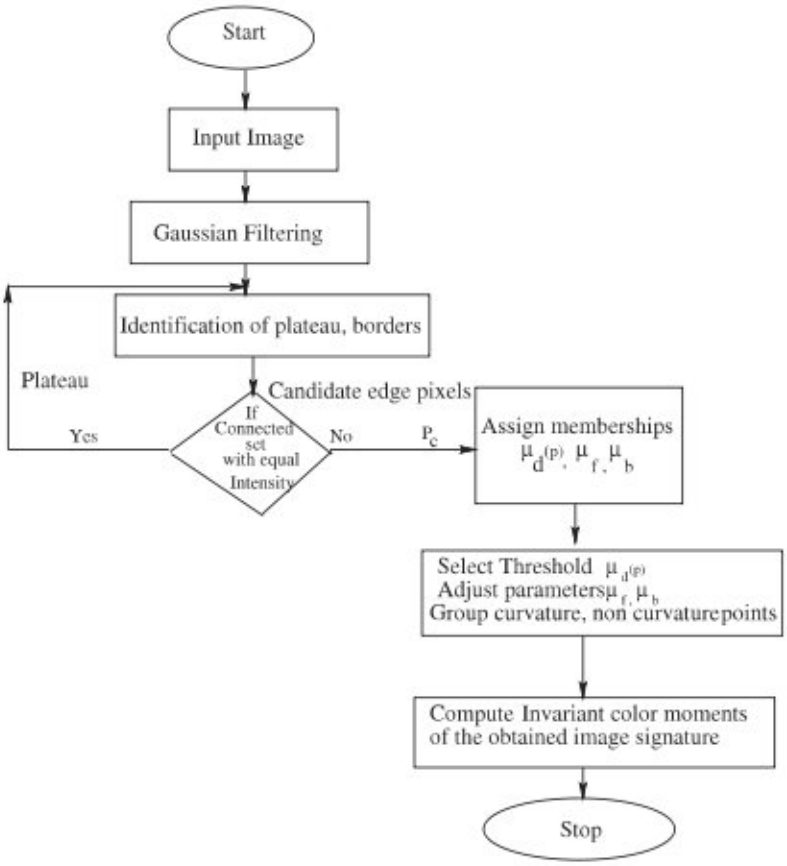

Fig. 2. Building blocks for corner signature extraction.

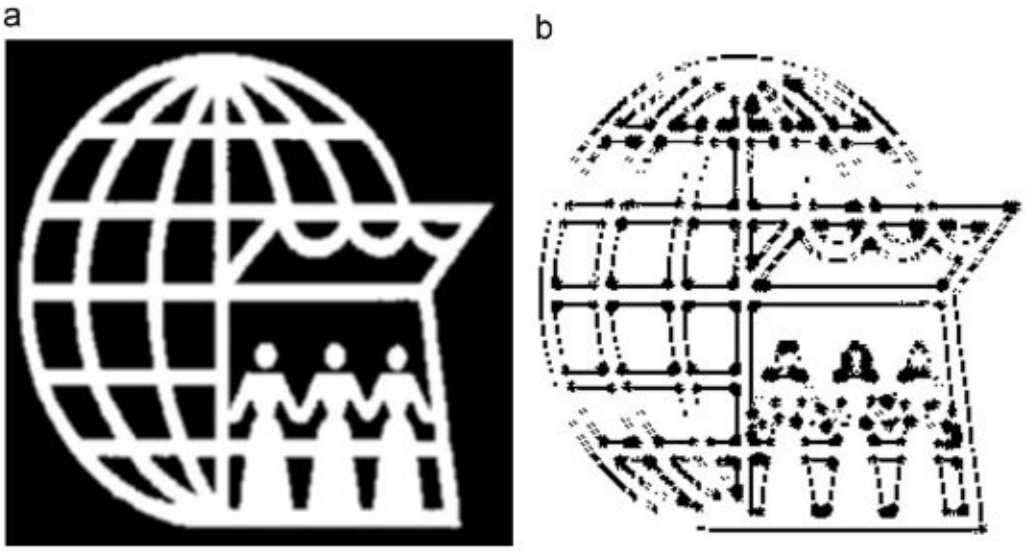

Fig. 3. (a) Original image. (b) The high curvature region, marked as (*) on the edge map, with $\mu_d(P) \geqslant 0.6$.

The invariant feature model in terms of (c_1, c_2 and c_3) as proposed in [34] is chosen here, and is defined as follows:

$$\begin{aligned} c_1 &= \arctan(R/\max(G, B)), \\ c_2 &= \arctan(G/\max(R, B)), \\ c_3 &= \arctan(B/\max(R, G)). \end{aligned} \tag{5}$$

Fig. 4. Retrieved results with feature descriptors from the edge map [31]. The top left image is the query image.

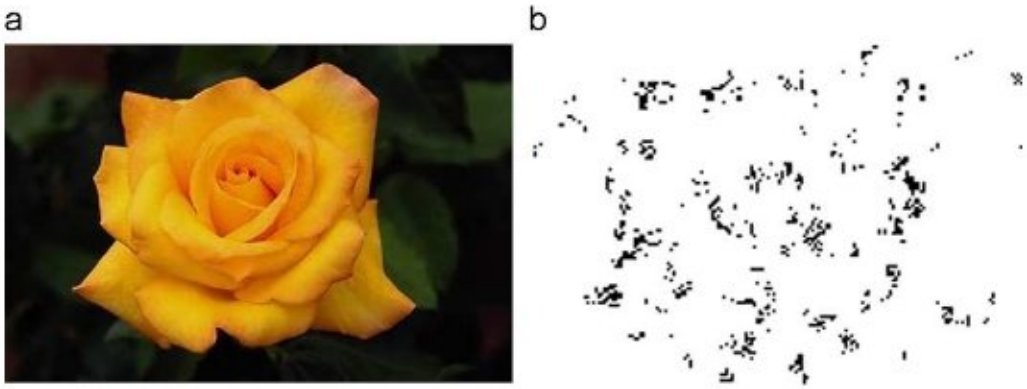

Fig. 5. (a) Flower image. (b) Fuzzy corner signature at $\mu_d(P) \geqslant 0.8$.

This model is able to denote the difference between two colors based on their perceptual difference. Among the different color models reported in literature, the normalized RGB representation, illumination and viewing geometry invariant representations are popular, which mostly belong to the HSI family of color models. In addition to these traditional color spaces, new invariant color models (c_1, c_2, c_3) also discount the effects of shading and shadows. Although invariant

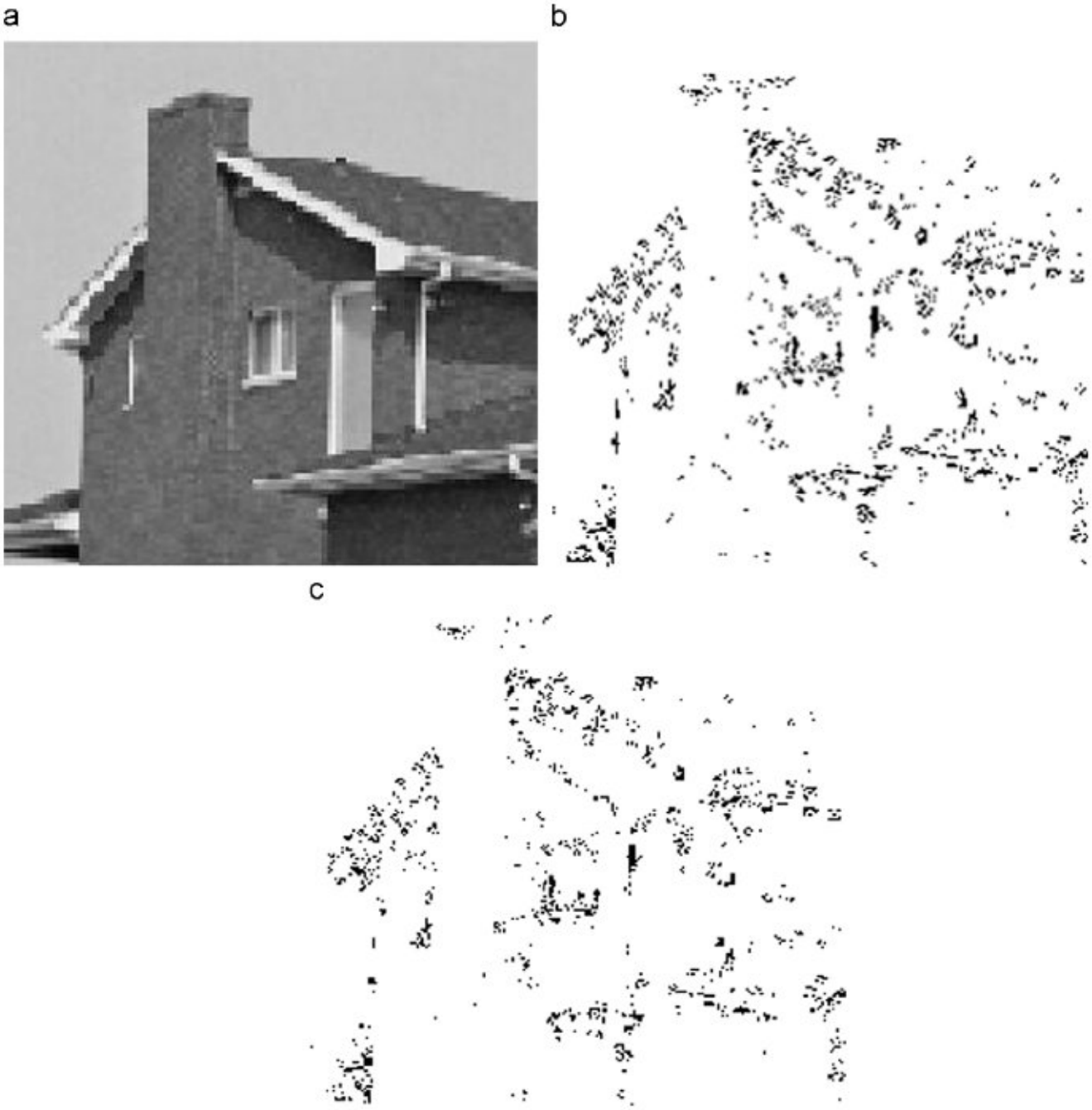

Fig. 6. (a) House image. (b) Fuzzy corner signature at $\mu_d(P) \geqslant 0.7$ and (c) $\mu_d(P) \geqslant 0.8$.

color representations are very popular in CBIR, these models have short comings under certain circumstance due to some loss of discriminatory power among images.

The spatial color moments of the candidates of ROI are extracted (using (5)) from each of the component planes (c_1, c_2, c_3) [35]. The moments are the global descriptors of a shape with invariance properties [35]. The invariant moments are computed as follows:

The moments m_{pq} of order p and q of a function $f(x, y)$ for discrete images are usually approximated as

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y). \tag{6}$$

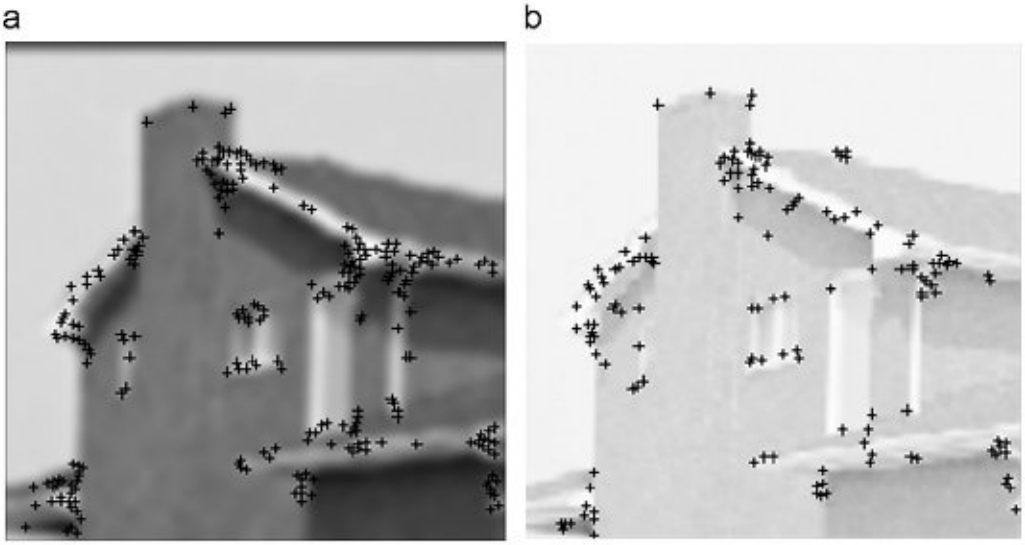

Fig. 7. (a) Representative points (centroids) of the clusters: (a) blurred; (b) illumination change.

The centralized moments are expressed as

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y), \tag{7}$$

where $\bar{x}$=m_{10}/m_{00}, $\bar{y}$=m_{01}/m_{00}.

The centralized moments up to order 2 are expressed as

$$\begin{aligned}
\mu_{00} &= m_{00}, \\
\mu_{20} &= m_{20} - \frac{m_{10}^2}{m_{00}}, \\
\mu_{02} &= m_{02} - \frac{m_{01}^2}{m_{00}}, \\
\mu_{11} &= m_{11} - \frac{m_{10} * m_{01}}{m_{00}}.
\end{aligned} \tag{8}$$

The normalized central moments are computed as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \tag{9}$$

where $\gamma = (p + q)/2 + 1$ for $p + q = 2, 3, \ldots$. A set of seven moments invariant under translation, rotation and scale can be computed from η_{pq}. Among these moments, the invariant moment (ϕ) is considered for the proposed feature extraction, and represented as

$$\phi = \eta_{20} + \eta_{02}. \tag{10}$$

The image is characterized in the following manner. The moment (ϕ) of the extracted significant spatial locations ROI will help to identify the color similarity of the identified regions. Three additional sets of moments are computed, taking all points from each of the (c_1, c_2, c_3) planes. The components of the feature vector F_k are designated as $[f_1, f_2, f_3, \ldots, f_6]$. The features f_1, f_2, f_3 represent the moment values (ϕ) are computed, considering all the points of each of the component planes obtained from (5). These values do not vary with the thresholding levels. The other features f_4, f_5, f_6 represent (ϕ) as obtained from the representative locations ROI of the generated signature (C_{ik}), from each component plane.

The proposed feature vector has six components as

$$f_l = \{\phi_l(S_c)\}_{l=1}^{6}. \tag{11}$$

S_c represents the component planes. ϕ_l are the invariant moments computed using Eqs. (6)–(10). For $l = \{1,2,3\}$, $f(x, y)_{S_c} = \{c_1, c_2, c_3\}$, respectively, of all points $(x, y) \in M \times N$ (the image). For $l = \{4,5,6\}$, $f(x, y)_{S_c} = \{c_1, c_2, c_3\}$, respectively, for $(x, y) \in C_{ik}$ (the clusters), and the default white (1) for other locations, as shown in Fig. 5(b). The features are computed on all images of the database. Each image of the feature database is represented as, $F_k = [f_1, f_2, f_3, f_4, f_5, f_6]$. The representative features $[f_1, f_2, f_3, f_4, f_5, f_6]$ of each image of the database are used for the similarity evaluation, where similarity between two images is measured by the Euclidean distance between the feature vectors F_k.

2.3. Feature selection method

For natural images (consisting of different objects), the representation obtained from the corner signature, although important, may not be sufficient to differentiate them from other categories. To show this, the extracted features have been ranked using an optimal classifier. The minimum redundancy-maximum relevance based feature selection method, proposed by Battiti [36], is used to extract a set of non-redundant and relevant features. The relevance of a feature with respect to the class and the redundancy between two features are calculated using the mutual information. To calculate the mutual information, the values of each feature thresholded using μ (mean) and σ (standard deviation): any value larger than $\mu + \sigma/2$ is transformed to the state 1; any value between $\mu - \sigma/2$ and $\mu + \sigma/2$ is transformed to the state 0; any value smaller than $\mu - \sigma/2$ is transformed to the state -1 [37]. To evaluate the effectiveness of the selected features, the leave-one-out cross validation is performed using the C4.5 based decision tree algorithm [38]. The C4.5 is a popular decision tree-based classification algorithm. It is used for evaluating the effectiveness of reduced feature set for classification. The selected features are fed to C4.5 for building classification models. The C4.5 is used here because it performs feature selection in the process of training and the classification models it builds are represented in the form of decision trees, which can be further examined. The features $[f_1, f_2, f_3, \ldots, f_6]$ are computed on all images of the (SIMPLIcity) database, which consists of images from 10 different semantic categories. As seen from the classification results, of Tables 1 and 2, the inclusion of the components f_4, f_5, f_6 along with the components of f_1, f_2, f_3 improves the recognition score. This implies that a combination of local properties (information from corner signature) and global moments (considering all points) are important for the effective characterization of an image. However the features f_5, f_6 alone do not play a significant part in improving the recognition score.

The features $[f_1 \cdots f_6]$ are invariant to rotation, translation and scaling by definition, as explained in Section 2.2. To test the robustness of the features against noise, we inject noise at different levels with SNR rations (50, 20, 0 and -10dB) on the query examples of a particular type. We add this class of images into the database images to form a separate noise injected class, and label the class as level 10. The error rate from such class is also shown in Tables 1 and 2. This indicates that the proposed feature is not noise invariant. It has been mentioned that invariant representations can cover a wide range of perceptual variations, but at the same time lacks discriminatory power among images. With the features $f_1 \cdots f_6$ computed directly from the RGB plane, i.e. without using Eq. (5), better recognition has been obtained for some classes, even for the noise injected case, as shown in Tables 3 and 4. We have designated the feature set $[f_1, f_2, f_3, \ldots, f_6]$ as computed from the (c_1, c_2, c_3) model using (5) as set (A), and those computed directly from the RGB plane as set (B). The recognition scores obtained from both cases are shown in Tables 1–4.

2.4. Feature evaluation method

Using either of the feature set (A) or (B), images relevant to a particular query are retrieved from the database. For a given database, different combinations of features (color, texture, shape etc.) may be effective for handling different types of queries [39]. Within a selected feature set, the individual feature weights may be further updated to specify their importance for improvement in retrieval accuracy. If with the selected set of features and total number of images is denoted by (I_s), then k similar images $I_s = \{I_1, I_2, \ldots, I_k\}$, where $I_k \in I_s$, are returned to the user. Let I_r be the set of relevant images and I_{ir} be the set of irrelevant images as marked by the user. $I_r = \{I_j | I_j \text{ relevant, for } I_j \in I_s\}$ and $I_{ir} = \{I_j | I_j \text{ irrelevant, for } I_j \in I_s\}$. The information from I_r and I_{ir} are combined to compute the relative importance of the individual features, from the fuzzy FEI proposed by Pal et al. [27,40] in pattern classification problems.

Table 1
Cost matrix for set (A) considering the three top ranked features.

feature:: 1 rank ::3
feature:: 2 rank ::2
feature:: 3 rank ::1
feature:: 4 rank ::−1
feature:: 5 rank ::−1
feature:: 6 rank ::−1
Errors :: 236(23.2%)

Image	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	
Class 0	77	2	3	2	12	3	1			1		(a)
Class 1	1	75	4	1	3	5			11			(b)
Class 2	7	1	57	8	5	7			13	2		(c)
Class 3	1	3		88			5		2			(d)
Class 4	5	1			94							(e)
Class 5	9	2	4	2	3	74	1	1	1	3		(f)
Class 6	2			9			53	4		32		(g)
Class 7	1	1				1		97				(h)
Class 8	1	16	6	5	1	3		1	66	1		(i)
Class 9	3			3	2	2	4	2		84		(j)
Class 10								10			9	(k)

Table 2
Cost matrix for set (A) considering the four top ranked features.

feature::1 rank ::3
feature::2 rank ::2
feature::3 rank ::1
feature::4 rank ::4
feature::5 rank ::−1
feature::6 rank ::−1
Errors :: 160(15.7%)

Image	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	
Class 0	93		1	2			3			1		(a)
Class 1	3	86		4	1	4			2			(b)
Class 2	5	7	74	5	1	2	1		3	2		(c)
Class 3	1	2		91			3		2	1		(d)
Class 4	2				97			1				(e)
Class 5	5	2	6	2	2	79	2	1		1		(f)
Class 6	1		2	4		2	79	2	1		1	(g)
Class 7	1					1		97			1	(h)
Class 8	2	12	3	3	1	2	1	1	75			(i)
Class 9	3		2	4	2	1	10	2		76		(j)
Class 10							10				9	(k)

For (m) pattern classes $C_1, C_2, \ldots, C_j, \ldots, C_m$ in an N-dimensional $(x_1, x_2, x_q, \ldots, x_N)$ feature space, let the class C_j contains n_j samples. The value of the measure H_{qj} (fuzzy entropy) [33] represents an estimate of 'intraset ambiguity' along the qth coordinate axis in C_j.

The entropy of a fuzzy set having n points is defined as

$$H(A) = \left(\frac{1}{nln2}\right) \sum_i S_n(\mu(x_i)); \quad i = 1, 2, \ldots, n, \tag{12}$$

where the Shannon's function $(S_n\mu(x_i)) = -\mu(x_i)\ln\mu(x_i) - \{1-\mu(x_i)\}\ln\{1-\mu(x_i)\}$.

Entropy is dependent on the absolute values of membership (μ) and satisfies the properties [33], $H_{min} = 0$ for $\mu = 0$ or 1, $H_{max} = 1$ for $\mu = 0.5$.

Table 3
Cost matrix for set (B) considering the three top ranked features.

feature::1 rank ::3
feature::2 rank ::1
feature::3 rank ::−1
feature::4 rank ::2
feature::5 rank ::−1
feature::6 rank ::−1
Errors :: 204(20.0%)

Image	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	
Class 0	82	3	2	2	2	6	1			2		(a)
Class 1	4	69	4	8		7			7	1		(b)
Class 2	9	6	76	3		5	2		1	1		(c)
Class 3	6	9	5	72			4		3	1	1	(d)
Class 4					98	2						(e)
Class 5	4	6	3	2		70		5	4	6		(f)
Class 6	2	1	2	3			84	1		5	2	(g)
Class 7	3		1			6		76		14		(h)
Class 8	3	3	3	5	1	10	2		73			(i)
Class 9	10	2		3	1	3	6	1	1	72	1	(j)
Class 10											18	(k)

Table 4
Cost matrix for set (B) considering the four top ranked features.

feature::1 rank ::3
feature::2 rank ::1
feature::3 rank ::4
feature::4 rank ::2
feature::5 rank ::−1
feature::6 rank ::−1
Errors :: 189(18.6%)

Image	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	
Class 0	82	1	3	2	2	6				4		(a)
Class 1	1	70	5	9		5			8	2		(b)
Class 2	5	4	81	3		2	2		2	1		(c)
Class 3	6	3	7	75			5		3	1		(d)
Class 4					100							(e)
Class 5	4	5	3		1	74		5	3	5		(f)
Class 6	3	4		2			88			3		(g)
Class 7			1			8		76	2	15		(h)
Class 8	1	3	5	6	1	1	1	2	80			(i)
Class 9	7	3	2	4	2	3	6	1	1	70	1	(j)
Class 10			1								17	(k)

To compute H_{qj} of C_j along the qth component, an S-type membership function is considered, with parameters set as

$$\begin{aligned} b &= (x_{qj})av, \\ c &= b + \max\{|(x_{qj})av - (x_{qj})\max|, |(x_{qj})av - (x_{qj})\min|\}, \\ a &= 2b - c, \end{aligned} \tag{13}$$

where $(x_{qj})av$, $(x_{qj})\max$, $(x_{qj})\min$ denote the mean, maximum and minimum feature values, respectively, computed along the qth coordinate axis over all the n_j samples in C_j.

The criteria of a good feature is that it should be invariant to within class variation while emphasizing the differences between patterns of different types [27]. Since $\mu(b) = \mu(x_{qj})av = 0.5$, the value of H_{qj} is 1.0 at $b = (x_{qj})av$ and tends to zero when moving away from b towards either c or a of the S function. Therefore a higher value of H_{qj}

indicates a larger number of samples having $\mu(x)$ equal to 0.5. Therefore, the samples would have a greater tendency to cluster around the mean value, resulting in less internal scatter within the class. After combining the classes C_j and C_k, the mean, maximum and minimum values $(x_{qkj})av$, $(x_{qjk})\max$, $(x_{qjk})\min$, respectively, of the qth dimension over the samples $(n_j + n_k)$ are computed. The value of entropy (H) would therefore decrease as the goodness of the qth feature in discriminating between the pattern classes C_j and C_k increases. The measure denoted as H_{qjk} is called the 'interset ambiguity' along the qth dimension between classes C_j and C_k. Considering the two types of ambiguities, the proposed FEI for the qth feature is shown below:

$$(FEI_q) = \frac{H_{qjk}}{H_{qj} + H_{qk}}. \tag{14}$$

The lower the value of $(FEI)q$, the higher the quality of importance of the qth feature in recognizing and discriminating different classes. The accuracy of retrieval can be improved by emphasizing the weights of the features that help the most in retrieving the relevant images, while reducing the importance of the features that impede the process.

In the proposed algorithm, the number of classes is two. The relevant images constitute the (intraclass) and the irrelevant images constitute the (interclass) image features. To evaluate the importance of the qth feature, the qth component of the retrieved images is considered, i.e. $I^{(q)} = \{I_1^{(q)}, I_2^{(q)}, I_3^{(q)}, \ldots, I_k^{(q)}\}$. H_{qj} is computed from $I_r^{(q)} = \{I_{r1}^{(q)}, I_{r2}^{(q)}, I_{r3}^{(q)}, \ldots, I_{rk}^{(q)}\}$. Similarly, H_{qk} is computed from the set of images $I_{ir}^{(q)} = \{I_{ir1}^{(q)}, I_{ir2}^{(q)}, I_{ir3}^{(q)}, \ldots, I_{irk}^{(q)}\}$. H_{qkj} is computed by combining both sets. The commonly used decision function for measuring the similarity between the query image I_{qr} and other images I, is represented as

$$D_{is}(I, I_{qr}) = \sum_{q=1}^{N} w_q \| f_q(I) - f_q(I_{qr}) \|, \tag{15}$$

where $\| f_q(I) - f_q(I_{qr}) \|$ is the Euclidean distance between the qth component and w_q is the weight assigned to the qth feature component. The images are ranked according to the Euclidean distance. The user marks the relevant and irrelevant sets from 20 returned images, for the automatic evaluation of (FEI). Squire et al. [41] proposed a weight adjustment technique based on the variance of the feature values. Since each feature is weighted by its relative importance, say w_i, the weighted feature vector is now represented as

$$F = \sum_{q=1}^{N} w_q f_q(I), \tag{16}$$

where w_i is the weight associated with the feature component (f_q). The importance of each feature component (f_q) is evaluated as follows. $w_i = 1.0$ indicates the inclusion of the feature component with full importance within the feature vector. For $w_i = 0.0$, no importance is assigned to the component. Initially all components are considered to be equally important, i.e. $w_i = 1.0$, and the candidate images are retrieved using the Euclidean distance metric as the similarity measure. The FEI for each feature component is evaluated from the two classes, relevant (intraclass) and irrelevant (interclass), that are thus obtained. Experimentally better results are obtained if the weights are adjusted based on the FEI values, as $w_i = (FEI_q)^2$. A new retrieved set of the same query is then obtained. This process may be followed over a number of iterations. In the first pass, all features are considered to be equally important. Hence, $w_1 = w_2, \ldots, = w_q = 1$. The feature spaces of the relevant images are then altered by updating the components with w_q. As a result, the ranks of the relevant images are not significantly affected. For irrelevant images, one feature component may be very close to the query, whereas other feature component may be far away from the query feature. But the magnitude of the similarity vector may be close to the relevant ones. After the features of the query and the stored images have been updated with the FEI values, the weighted components are expected to dominate over the feature space, such that the rank of the irrelevant ones are pulled down. Since different components within the feature vector may be of totally different physical quantities, the presence of very large values of some feature components may bias the similarity. One solution is the normalization of the features [42]. The proposed feature set almost represent the same physical quantity (moments). Therefore, normalization is not required in computing the Euclidean distance between the feature vectors.

3. Experiment and comparison

The performance of the image retrieval system is tested upon the following two databases: (a) SIMPLIcity images and (b) COREL 10,000 miscellaneous database downloaded from http://bergman.stanford.edu/cgi-bin/www_wavesearch. The SIMPLIcity database consists of 1000 images from 10 different categories: Africa(0), Beach(1), Buildings(2), Buses(3), Dinosaurs(4), Elephants(5), Flowers(6), Horses(7), Mountains(8) and Food(9). Each category has around 100 images, along with some images undergoing changes due to rotation, translation, scaling, noise injection, illumination, etc.

The main objective here is to design a CBIR system using simple techniques involving a low cost feature extraction mechanism. The moments generate a compact representation with fewer features, compared to other sophisticated features. At the same time, it may yield poor results as the query complexity increases.

The experiments are performed in the following manner. The feature set (A), which is explained in Section 2.3, obtains satisfactory results for almost all categories, except for a few cases. In such cases, better results are obtained by computing the invariant moments (ϕ) directly from the RGB component planes without using (5) i.e. set (B). Such differences in performance can be explained from the fact that a single set of features may not prove to be effective for all types of queries. Moreover, the results also depend upon the availability of similar images in the database. The RGB components are sensitive to varying imaging conditions [6] but have better discriminatory power among images. For an unknown database, feature set (B) becomes a good choice when there is less variation in the imaging conditions or perception. Better recognition scores are obtained in the noise injected class and the class of dinosaurs, as shown in Tables 3 and 4. We start by retrieving queries separately using the sets (A) and (B). The retrieved results are assigned scores by users, as highly relevant (score 3), relevant (score 2) and non-relevant (score 1). The feature set giving better results is chosen for the further feature evaluation scheme.

In the case of moderate results (score 2) for both the cases, the retrieval score is further enhanced by combining both sets of features in a hierarchical fashion. This method generates better results particularly on the 10,000 miscellaneous image database, in which the images do not belong to any specific class. The illumination invariant set (A) is used first to get the short listed candidates (around 100 images) and a second set of retrievals is performed on the short listed candidates using set (B). Each retrieved set can be further subjected to the feature updating scheme to generate still better results.

To evaluate the retrieval performance, standard retrieval benchmarks such as the *recall rate* and the *precision rate* [2] are generally considered. Let n_1 be the number of images retrieved in the top 20 positions that are close to the query. Let n_2 represent the number of images in the database that are similar to the query. The evaluation standard *recall rate* (R) is given by $n_1/n_2 \times 100\%$, and the *precision rate* (P) is given by $n_1/20 \times 100\%$. The performance is evaluated in terms of the precision rate. This may be a better choice for an unknown database, because in order to evaluate the recall rate, the number of relevant images of a particular type has to be known.

The experimental results are shown in Figs. 5–14. The results are explained as follows: The corner signature of the image in Fig. 5(a) is shown in Fig. 5(b) at $\mu_d(P) \geqslant 0.8$. Similarly, for Fig. 6(a), the signatures thresholded at $\mu_d(P) \geqslant 0.7$ and 0.8, respectively, are shown in Fig. 6(b) and (c). The representative corners of Fig. 6(a) under varying imaging conditions (blurred and illumination change) are shown in Fig. 7(a) and (b). The query results of the SIMPLIcity data set are shown in Figs. 8–10. The signature is generated at a threshold value of $\mu_d(P) \geqslant 0.8$. The images are displayed from left to right according to the Euclidean distance, with the top left image as the query image.

The query shown in Fig. 8(a) is of a red flower, which is able to identify similar images that have undergone illumination changes. The image in Fig. 8(b) is from the horse category. The precision obtained is very high in this case. The retrieved images are less dependent on shadows. Fig. 9(a) shows the result when queried with a dinosaur. The precision obtained is very high for such images having distinct objects. The feature is fairly invariant to linear transformations. The further improvement in precision after updating the weights calculated from the FEI values can be seen in Fig. 9(b) from its ability to retrieve blurred, noisy images at the 4th and 6th positions from the left. The results when queried with a flower (yellow), with moments computed from the corner signature only, is shown in Fig. 10(a). The further addition of color moments computed from all points of the component planes generates better results, as shown in Fig. 10(b). The images of this category have objects with some regularity in shape and background. An improvement is also observed in Fig. 11(a) from retrieving images with the literal color properties of the query after adjusting the feature weights. The results when queried with the same flower of Fig. 10(a) with feature set (A) is

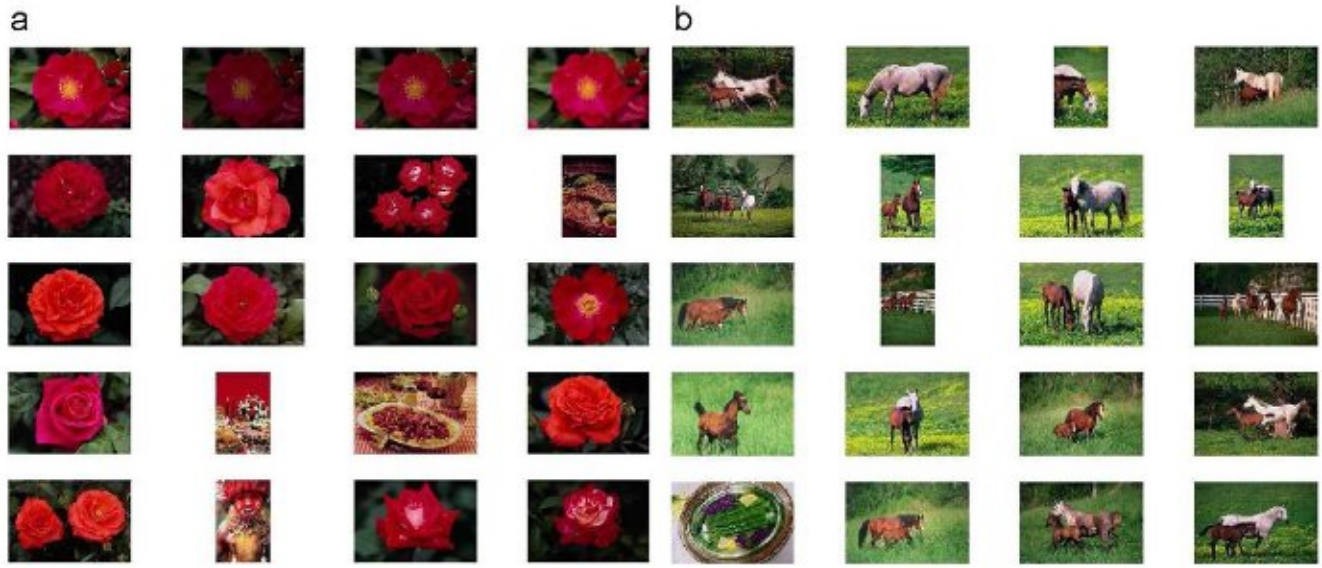

Fig. 8. Retrieved results, with the top left image as the query image. (a) Test for illumination invariance with set (A). (b) Retrieved results using set (A).

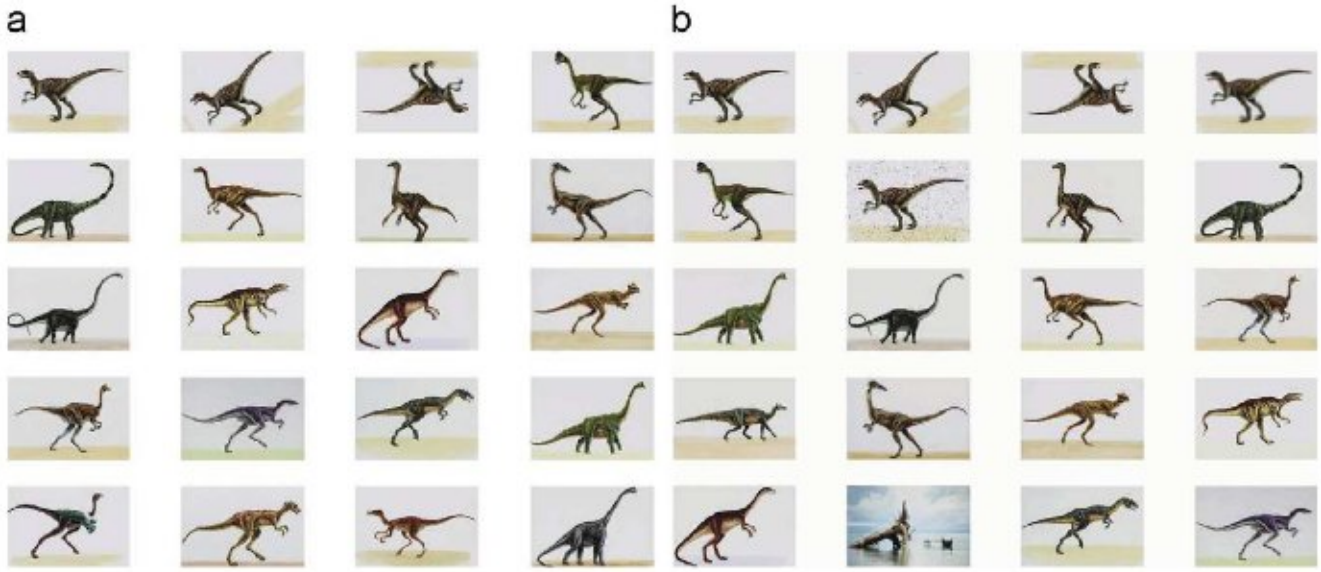

Fig. 9. Retrieved results, with the top left image as the query image. Test for (rot, trans. scale, noise) invariance, using set (B). (b) After feature updating with (FEI), the noisy and blurred images are retrieved.

shown in Fig. 11(b). Similar images under varying illuminations were retrieved, as shown in Fig. 11(b). However, the results show that feature set (B) generates better results in this case.

The results obtained from database (b) of 10,000 miscellaneous images are shown in Fig. 12. The results of Fig. 12(a) show the candidates retrieved by combining the features in a hierarchical fashion. The result after iterative refinement with set (B) on the short listed candidates is shown in Fig. 12(b). As the query image of Fig. 12(a) is a single object, the corner signature should be a more important discriminant in this case. To study this, we have set w_1, w_2, w_3 equal to zero. The result from this observation is shown in Fig. 13(a). The result for a query (scene) is shown in Fig. 13(b). The results obtained prove to be satisfactory for retrieving scenes.

The FEI values for some queries are shown in Table 5. In the case of dinosaurs, i.e. query image of (Fig. 9), it is observed from the FEI values that the importance of the corner signature (f_4, f_5, f_6) is greater than that of the other three components. In case of flowers (Fig. 10(b)), it is seen that all the components have significant contributions to the evaluation of the similarity among the images.

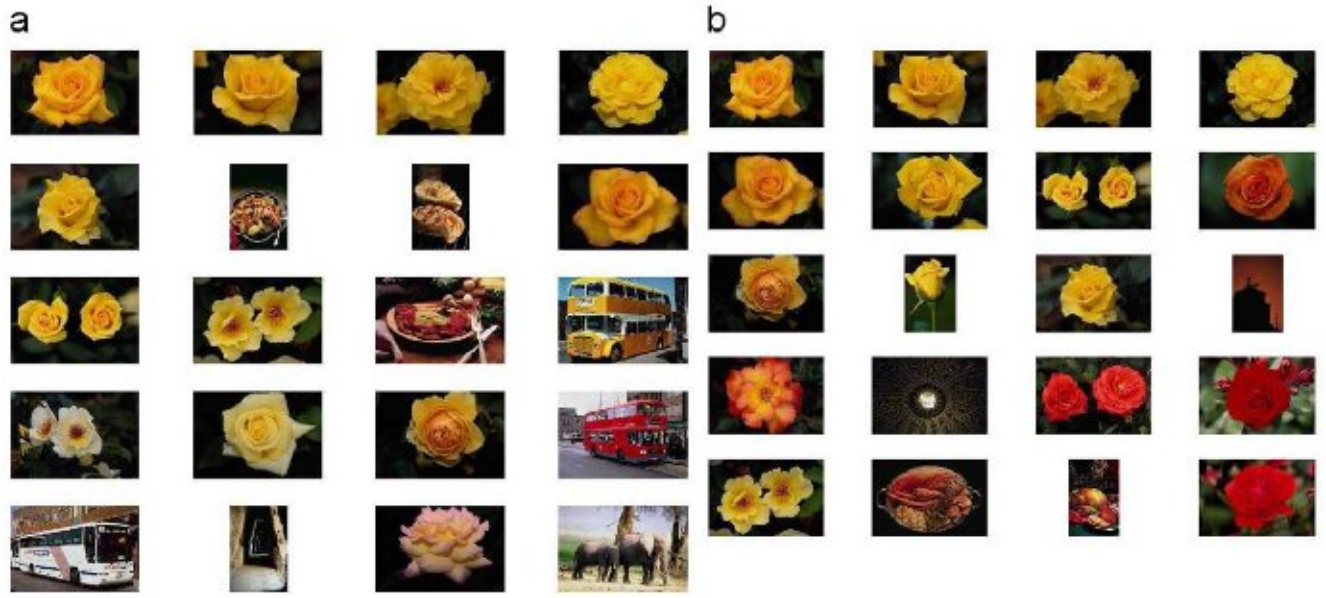

Fig. 10. Retrieved results using set (B), with the top left image as the query image. (a) By taking moments from corner signature only. (b) Also considering moments from all points.

Fig. 11. (a) After feature evaluation with (FEI) on set (B). (b) Using set (A).

Table 5
Feature evaluation index.

Image	$(FEI1)^2$	$(FEI2)^2$	$(FEI3)^2$	$(FEI4)^2$	$(FEI5)^2$	$(FEI6)^2$
Fig. 9 (iteration 1)	0.64	0.32	0.38	0.02	0.02	0.02
Fig. 9 (iteration 2)	0.62	0.52	0.51	0.02	0.02	0.02
Fig. 10(b) (iteration 1)	0.38	0.12	0.13	0.27	0.29	0.29
Fig. 10(b) (iteration 2)	0.39	0.29	0.11	0.34	0.39	0.39
Fig. 10(b) (iteration 3)	0.35	0.32	0.32	0.38	0.38	0.38

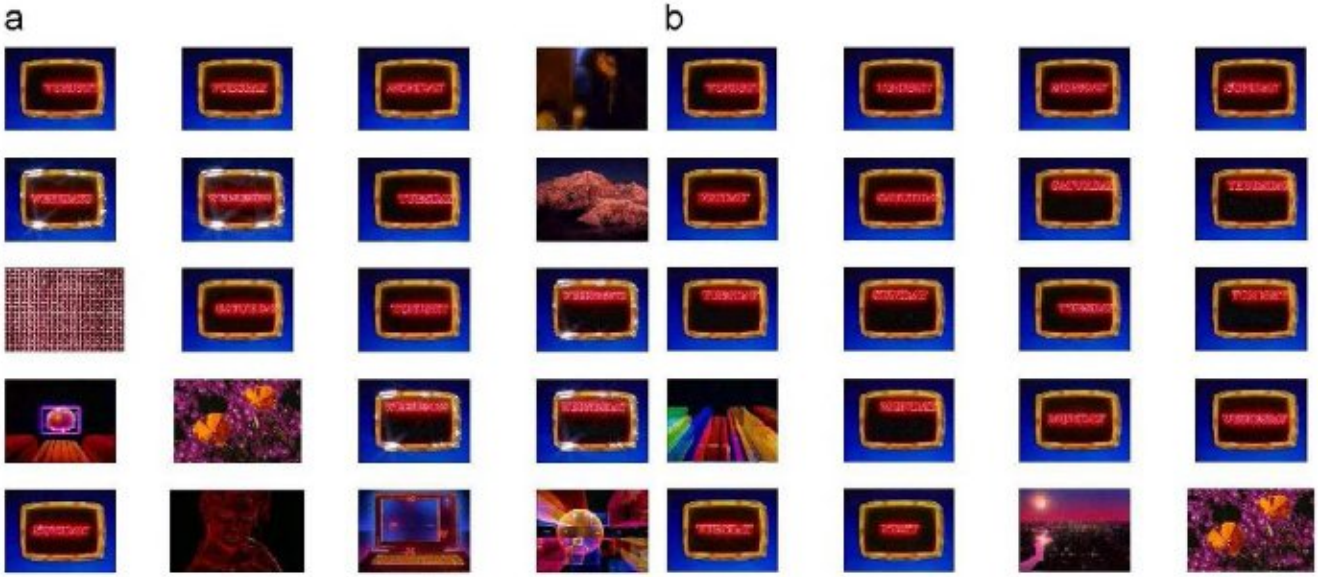

Fig. 12. Retrieved results, with the top left image as the query image. (a) First set of retrieved candidates. (b) After feature updating with (FEI).

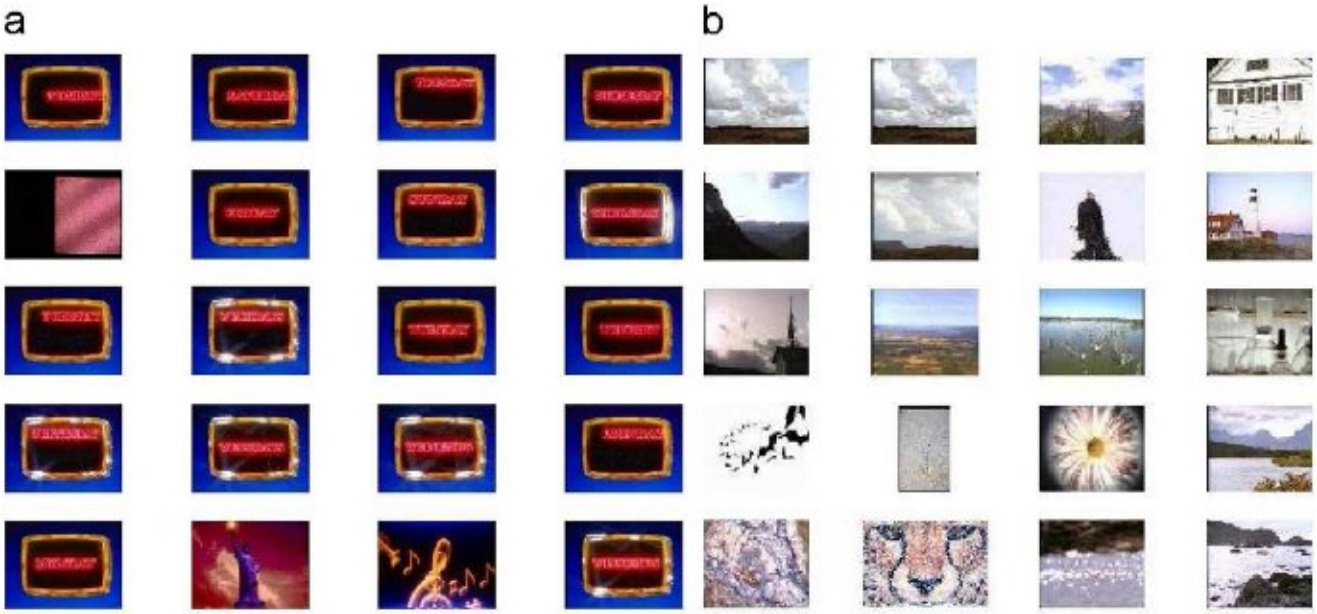

Fig. 13. (a) Taking the contribution from the corner signature only. (b) Results, querying with a natural scene.

To evaluate the performance of the proposed scheme, we have randomly chosen queries from each category of SIMPLIcity data set. The average precision value from each category, after retrieving (10, 20, 40) images as shown in Table 6 show improvement in the results. The performance is evaluated in terms of precision rate given by the fraction of correctly retrieved images within the retrieved set. The overall precision recall curve for the SIMPLIcity test database after the first iteration is shown in Fig. 14.

3.1. Performance comparison

CBIR systems have been reported in the literature, but in most cases the systems rely on user defined parameters. In segmentation dependent techniques the accuracy depends upon the number of classes defined by the user [25,5]. As a result, it is difficult to judge the performance of a CBIR system unless the parameters match exactly. So a pro-

Table 6
Average precision % from our algorithm.

Category	Unweighted features			Weighted features (iteration 1)		
	10 images	20 images	40 images	10 images	20 images	40 images
Africa	68.20	60.25	55.03	70.50	61.00	58.00
Beach	70.00	55.23	50.60	71.48	56.23	52.58
Building	70.26	60.50	54.46	72.40	63.67	56.00
Bus	80.00	70.59	60.67	81.45	72.77	62.67
Dinosaur	100.0	95.0	90.7	100.0	95.0	92.0
Elephant	83.5	75.5	65.8	85.0	77.0	66.0
Flower	90.0	80.5	70.6	92.0	83.0	71.2
Horses	100.0	90.0	80.5	100.0	95.0	83.0
Mountains	70.5	65.8	60.9	72.0	68.0	62.0
Food	60.8	55.8	53.40	62.0	57.0	55.20

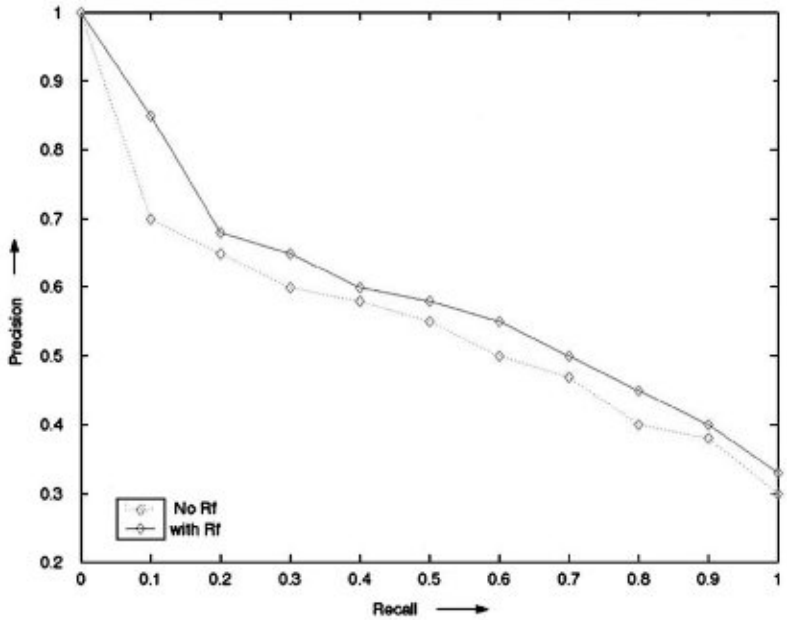

Fig. 14. Precision, recall curves of the SIMPLicity database.

posed scheme can be compared and evaluated best when the results are tested and evaluated over the same database. In order to prove the efficiency of our the technique, we benchmark our results with the well known standard image retrieval algorithms namely [25,28] and color histogram matching [29] using the same data set against the quantitative measure defined as weighted precision in (17). Online demonstrations of their methods are provided at the site (http//:www.wang.ist.psu.edu/IMAGE). The proposed method was compared with the techniques, which have shown useful results on evaluating overall similarity between images. The proposed feature evaluation method was used to study the improvement of results. The results almost converges after two iterations. The proposed scheme is compared with the above-mentioned techniques in order to see, to what extent the simple features (moments) can be used effectively as an alternative to detailed region segmentation in developing a retrieval system. The weighted average of the precision values within the k_1 retrieved images are computed as

$$\bar{p}(i) = 1/(100) \sum_{k_1=1}^{100} n_k/100, \tag{17}$$

where $k_1 = 1, \ldots, 100$ and n_k is the number of matches within the first k_1 retrieved images. The weighted precision as obtained from each category is shown in Table 7.

Table 7
Comparative evaluation of weighted average precision.

Class	Our method	SIMPLIcity	Histogram based	FIRM
Africa	0.45	0.48	0.30	0.47
Beach	0.35	0.32	0.30	0.35
Building	0.35	0.35	0.25	0.35
Bus	0.60	0.36	0.26	0.60
Dinosaur	0.95	0.95	0.90	0.95
Elephant	0.60	0.38	0.36	0.25
Flower	0.65	0.42	0.40	0.65
Horses	0.70	0.72	0.38	0.65
Mountains	0.40	0.35	0.25	0.30
Food	0.40	0.38	0.20	0.48

As seen from Table 7, the SIMPLIcity method has reported better results than the color histogram method for all categories. It is difficult to obtain satisfactory results for retrieving from all categories using the same set of features. However, the FEI can provide a good measure for improvement in precision. In most of the cases, the average precision for each category can be made better than the SIMPLIcity and FIRM by the feature updating scheme. The SIMPLIcity and FIRM methods are segmentation based. In effect, better results are obtained with these algorithms when each object has different texture properties and the number of objects matches the number of classes. The algorithm is implemented in MATLAB on a SUN Blade system with a 700 MHz Processor. The average CPU time required for computing the features is 10 s.

4. Conclusion and scope of future work

In the current work, major emphasis has been placed on developing significant feature descriptors and selecting an optimal set of features suitable for retrieving a set of images perceptually relevant to the query image. From different query examples and classification results using different training sets on a standard database, it is shown that a single set of features may not be suitable for handling all types of queries. The set of features that may be able to retrieve images with a range of illumination variations may fail to retrieve noisy images. To overcome such limitations, the features are combined and the weights are updated to obtain a trade-off between feature selection and accuracy in the case of retrieving from an unknown database. Although the proposed CBIR system is not capable of handling very complex types of queries, it can identify relevant images that differ visually in some characteristics due to translation, rotation, scaling, blurring, illumination change, etc. The suitability of the proposed features may be increased if the application is extended to specific images like logos or facial images, rather than general purpose images. To achieve these goals we shall try to introduce other sophisticated features and incorporate some other mechanism, like surrounding text for web images, to obtain better applications.

Acknowledgment

Minakshi Banerjee is grateful to the Department of Science and Technology, New Delhi, India, for supporting the work under Grant (SR/WOS-A/ET-111/2003).

References

[1] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1349–1380.
[2] A.D. Bimbo, Visual Information Retrieval, Morgan Kaufmann Publishers, Inc., San Francisco, USA, 2001.
[3] S.C. Cheng, C.T. Kuo, H.J. Chen, Visual object retrieval via block-based visual-pattern matching, Pattern Recognition 40 (4) (2007) 1695–1710.
[4] J.R. Smith, S.F. Chang, VisualSEEK a fully automated content-based image query system, in: Proc. ACM Multimedia, 1996, pp. 87–98.
[5] C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein, J. Malik, Blobworld a system for region-based image indexing and retrieval, in: Proc. of Visual Information Systems, 1999, pp. 509–516.

[6] T. Gevers, A.W.M. Smeulders, Combining color and shape invariant features for image retrieval, Image and Vision Computing 17 (7) (1999) 475–488.
[7] A.K. Jain, Vailaya, image retrieval using color and shape, Pattern Recognition 29 (1996) 1233–1244.
[8] X.S. Zhou, T.S. Huang, Relevance feedback in content based image retrieval: some recent advances, Information Sciences 148 (2002) 129–137.
[9] J. Han, K.N. Ngan, M. Li, H.J. Zhang, A memory learning framework for effective image retrieval, IEEE Transactions on Image Processing 14 (4) (2005) 521–524.
[10] Y. Chen, J.Z. Wang, R. Krovetz, Clue: Cluster-based retrieval of images by unsupervised learning, IEEE Transactions on Image Processing 14 (8) (2005) 1187–1201.
[11] Z. Jin, I. King, X. Q. Li, Content-based retrieval by relevance feedback, in: Lecture Notes in Computer Science, vol. 1929, Springer, Berlin, 2000, pp. 521–529.
[12] M. Acharyya, M.K. Kundu, R.K. De, Segmentation of remotely sensed images using wavelet features and their evaluation in soft computing framework, IEEE Transactions on Geoscience and Remote Sensing 41 (12) (2003) 2900–2905.
[13] M. Acharyya, M.K. Kundu, R.K. De, Extraction of features using m-band wavelet packet frames and their neuro-fuzzy evaluation for multi-texture segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (12) (2003) 1639–1644.
[14] D.G. Lowe, Perceptual Organization and Visual Recognition, Kluwer Academic Publishers, USA, 1985.
[15] S.C. Cheng, Content-based image retrieval using moment-preserving edge detection, Image and Vision Computing 21 (9) (2003) 809–826.
[16] Y.H. Yua, C.C. Chang, A new edge detection approach based on image context analysis, Image and Vision Computing 24 (10) (2006) 1090–1102.
[17] D. Lowe, Distinctive image features from scale invariant keypoints, International Journal of Computer vision 2 (60) (2004) 91–110.
[18] K. Mikolajczyk, C. Schmid, Scale and affine invariant interest point detectors, International Journal of Computer vision 1 (60) (2004) 63–86.
[19] H. Zhang, R. Rahamani, S.R. Cholleti, S.A. Goldman, Local image representations using pruned salient points with applications to CBIR, in: Proc. 14th Annu. ACM Internat. Conf. on Multimedia, USA, 2006, pp. 287–296.
[20] V. Gouet, N. Boujemma, About optimal use of color points of interest for content-based image retrieval, Technical Report, RR-4439, INRIA, France, May 2006.
[21] J. Weijer, T. Gevers, A. Bagdanov, Boosting color saliency in image feature detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (1) (2006) 150–156.
[22] E. Loupias, N. Sebe, Wavelet-based salient points: applications to image retrieval using color and texture features, in: Proc. 4th Internat. Conf. on Advances in Visual Information Systems, VISUAL 2000, 2000, pp. 223–232.
[23] C. Harris, M. Stephens, A combined corner and edge detector, in: 4 th Alvey Vision Conf., 1988, pp. 147–151.
[24] D. Zhang, M. Chen, Y. Lim, An efficient and robust technique for region based shape representation and retrieval, in: Internat. Conf. on Computer and Information Science, ICIS 2007, 2007, pp. 801–806.
[25] Y. Chen, J.Z. Wang, A region-based fuzzy feature approach to content-based image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (9) (2002) 1–16.
[26] J.C. Martnez, J.M. Medina, C.D. Barranco, G. Perales, J.M.S. Hidalgo, Retrieving images in fuzzy object-relational databases using dominant color descriptors, Fuzzy Sets and Systems 158 (3) (2007) 312–324.
[27] S.K. Pal, B. Chakraborty, Intraclass and interclass ambiguities (fuzziness) in feature evaluation, Pattern Recognition Letters 2 (1984) 275–279.
[28] J.Z. Wang, J. Li, G. Wiederhold, Simplicity: semantics-sensitive integrated matching for picture libraries, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (9) (2001) 947–963.
[29] Y. Rubner, L.J. Guibas, C. Tomasi, The earth mover's distance, multi-dimensional scaling and color-based image retrieval, in: Proc. DARPA Image understanding Workshop, 1997, pp. 661–668.
[30] M. Banerjee, M.K. Kundu, Content based image retrieval with multiresolution salient points, in: Fourth Indian Conf. Computer Vision, Graphics and Image Processing, ICVGIP 2004, India, 2004, pp. 399–404.
[31] M. Banerjee, M.K. Kundu, Edge based features for content based image retrieval, Pattern Recognition 36 (11) (2003) 2649–2661.
[32] A. Rosenfeld, Fuzzy digital topology, in: J.C. Bezdek, S.K. Pal (Eds.), Fuzzy Models for Pattern Recognition, IEEE Press, 1991, pp. 331–339.
[33] S.K. Pal, D.D. Majumder, Fuzzy Mathematical Approach to Pattern Recognition, Wiley Eastern Limited, New York, 1985.
[34] T. Zevers, A.W.M. Smeulders, Color-based object recognition, Pattern Recognition 32 (1999) 453–464.
[35] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Wiley, New York, 1985.
[36] R. Battiti, Using mutual information for selecting features in supervised neural net learning, IEEE Transactions on Neural Network 5 (4) (1994) 537–550.
[37] P. Maji, f-information measures for efficient selection of discriminative genes from microarray data, IEEE Transactions on Biomedical Engineering (2008) 1–7.
[38] J. Quinlan, Programs for Machine Learning, Morgan Kaufmann, CA, 1993.
[39] H. Muller, W. Muller, D. Squire, S.M. Maillet, T. Pun, Performance evaluation in content-based image retrieval: overview and proposals, Pattern Recognition Letters 22 (2001) 593–601.
[40] S.K. Pal, B. Chakraborty, Fuzzy set theoretic measures for automatic feature evaluation, IEEE Transactions on Systems, Man and Cybernetics 16 (5) (1986) 754–760.
[41] D.M. Squire, W. Muller, H. Muller, T. Pun, Content based query of image databases: inspirations from text retrieval, Pattern Recognition Letters 21 (2000) 1993–1998.
[42] Y. Rui, T.S. Huang, M. Ortega, S. Mehrotra, Relevance feedback: a power tool for interactive content-based image retrieval, IEEE Transactions on Circuits and Systems for Video technology 8 (5) (1998) 644–655.