Image retrieval based on indexing and relevance feedback

Sanjoy K. Saha [a,*], Amit K. Das [b], Bhabatosh Chanda [c]

[a] CSE Department, Jadavpur University, Calcutta 700 032, India
[b] CST Department, B.E. College (DU), Howrah 711 103, India
[c] ECS Unit, Indian Statistical Institute, Calcutta 700 035, India

Available online 22 May 2006

Abstract

In content based image retrieval (CBIR) system, search engine retrieves the images similar to the query image according to a similarity measure. It should be fast enough and must have a high precision of retrieval. Indexing scheme is used to achieve a fast response and relevance feedback helps in improving the retrieval precision. In this paper, a human perception based similarity measure is presented and based on it a simple yet novel indexing scheme with relevance feedback is discussed. The indexing scheme is designed based on the primary and secondary keys which are selected by analysing the entropy of features. A relevance feedback method is proposed based on Mann–Whitney test. The test is used to identify the discriminating features from the relevant and irrelevant images in a retrieved set. Then emphasis of the discriminating features are updated to improve the retrieval performance. The relevance feedback scheme is implemented for two different similarity measure (Euclidean distance based and human perception based). The experiment justifies the effectiveness of the proposed methodologies. Finally, the indexing scheme and relevance feedback mechanism are combined to build up the search engine.

Keywords: CBIR; Similarity measure; Human perception based similarity measure; Multidimensional indexing; Relevance feedback; Mann–Whitney test

1. Introduction

In a CBIR system, the feature extraction module computes various types of low-level features like shape, texture and color from an image. The search module retrieves the images similar to the query image from the database using a similarity measure based on the features. It should be fast and precision of retrieval should be high.

Apart from the feature set being used by the system, precision of retrieval depends on the similarity measure adopted by the search module. It is evident from the literature that various distance/similarity measures have been adopted by the CBIR systems. Mukherjee et al. (1999) have used template matching for shape based retrieval. A number of systems (Niblack et al., 1993; Srihari et al., 2000) have used Euclidean distance (weighted or unweighted) for matching. Other schemes include Minkowski metric (Fournier et al., 2001), self organising map (Laaksonen et al., 2000), proportional transportation distance (Vleugels and Veltkamp, 2002), etc. For matching multivalued features like color histogram or texture matrix, a variety of distance measures are deployed by different systems. It includes schemes like quadratic form distance (Niblack et al., 1993), Jaccard's co-efficient (Lai, 2000), histogram intersection (Gevers and Smeulders, 2000), etc. The details on combining the distance of various type of features is not available. But, it is clear that Euclidean distance is the most widely used similarity measure.

The simplest approach to search nearest neighbours is the linear search requiring $O(n)$ distance (dissimilarity) computations where n is the number of images in the database. Obviously, it is prohibitive for large value of n. Thus, to obtain a fast response from the retrieval module, an indexing scheme is required. Lots of research work have been carried on in this direction which can be classified

* Corresponding author.
E-mail addresses: sks@becs.ac.in (S.K. Saha), amit@becs.ac.in (A.K. Das), chanda@isical.ac.in (B. Chanda).

as: (a) spatial access methods and (b) distance based indexing methods. In spatial access methods, an image is represented by a finite set of features. Distance between two images is the Euclidean distance between their feature vectors. The Euclidean distance space is used to divide the database into spatial clusters. This is the basic principle in KD tree (Bentley and Friedman, 1979), quad tree (Samet, 1984) and R-tree family (Guttman, 1984; Leutenegger et al., 1997). While searching, some of the clusters are pruned to achieve faster response (Faloutsos et al., 1994). The spatial access methods suffer from curse of dimensionality. As the number of features (dimension) increases, computational cost also increases exponentially (Sebastian and Kimia, 2002). Actually it becomes impractical if dimension exceeds twenty (Weber, 1998). In distance based indexing methods, the database is organised on the basis of the distance of the database elements with respect to one or more key/pivot elements. Elements with similar distance from the key elements are grouped into a cluster and when querying the database, triangle inequality is used to discard some clusters (Berman and Shapiro, 1999). This method reduces computational cost but has a poor indexing efficiency. Moreover, proper selection of key is crucial. A k nearest neighbour graph (knn graph) based scheme has been proposed by Sebastian and Kimia (2002) where each node represents a database element and it is connected to its k nearest neighbour. This method is not guaranteed to provide the nearest neighbour of the query image in spite of the extended neighbourhood concept. Moreover, for a large database it is not practical.

The precision of the set of retrieved images can be improved through a relevance feedback mechanism. As the importance of the features vary for different queries and applications, to achieve better performance, different emphases have to be given to different features and the concept of relevance feedback (RF) comes into picture. Relevance feedback is a learning mechanism to improve the effectiveness of information retrieval systems. For a given query, the CBIR system retrieves a set of images according to a predefined similarity measure. Then, user provides a feedback by marking the retrieved images as relevant to the query or not. Based on the feedback, the system updates the emphasis of individual feature and retrieves a new set. The classical RF schemes can be classified into two categories: query point movement (query refinement) and re-weighting (similarity measure refinement) (Rocchio, 1971). In the query point movement method, the goal is to improve the estimate of the ideal query point by moving it towards the relevant examples and away from bad ones. Rocchio's formula (Rocchio, 1971) is frequently used to improve the estimation iteratively. In (Huang et al., 1997), a composite query is created based on relevant and irrelevant images. Various systems like Quicklook (Ciocca et al., 2001), Drawsearch (Sciascio et al., 1999) have adopted the query refinement principle. In the re-weighting method, the weight of the feature that helps in retrieving the relevant images is increased and importance of the feature that hinders this process is reduced. Rui et al. (1998) have proposed weight adjustment technique based on the variance of the feature values. Systems like ImageRover (Sclaroff et al., 1997), RETIN (Fournier et al., 2001) use re-weighting technique. A close study of past work indicates that re-weighting technique is widely used.

Thus, the search module of a CBIR system has to deal with the number of interrelated issues like similarity measure, indexing and relevance feedback scheme. This is necessary to satisfy the diverging requirements like high precision and fast response. Keeping all these in mind, in this work, we have proposed a human perception based similarity measure and an indexing scheme and finally combined with a novel relevance feedback scheme.

The paper is organised as follows. Section 2 presents a human perception based similarity measure and an indexing scheme is presented in Section 3 which approximates the proposed similarity measure. Mann–Whitney test based relevance feedback scheme is presented in Section 4. Section 5 discusses the experimental system and result. Finally, it is concluded in Section 6.

2. Similarity measure

The collection of features (often referred to as feature vector) convey, to some extent, visual appearance of the image in quantitative terms. Image retrieval engine compares the feature vector of the query image with that of the database images and presents to the users the images of highest similarity (i.e., least distance) in order as the retrieved images. However, it must be noted that the elements in the feature vector carry different kinds of information: shape, texture and color, which are mutually independent. Hence, they should be handled differently as suited to their nature.

The early work shows that most of the schemes deal with Euclidean distance, which has number of disadvantages. One pertinent question is how to combine the distance of multiple features. Berman and Shapiro (1999) proposed following operations to deal with the problem:

$$\text{Addition : distance} = \sum_i d_i \tag{1}$$

where, d_i is the Euclidean distance of ith features of the images being compared. This operation may declare visually similar images as dissimilar due to the mismatch of only a few features. The effect will be further pronounced if the mismatched features are sensitive enough even for a minor dissimilarity. The situation may be improved by using

$$\text{Weighted sum : distance} = \sum_i w_i d_i \tag{2}$$

where, w_i is the weight for the Euclidean distance of ith feature. The problem with this measure is that selection on proper weight is again a difficult proposition. One plausible solution could be taking w_i as some sort of reciprocal of variance of ith feature. Another alternative measure could be

$$\text{Max} : \text{distance} = \text{Max}\ (d_1, d_2, \ldots, d_M) \tag{3}$$

It indicates that similar images will have all their features are lying within a range. Like addition method, it also suffers from similar problem. On the other hand, the following measure

$$\text{Min} : \text{distance} = \text{Min}\ (d_1, d_2, \ldots, d_M) \tag{4}$$

helps in finding the images which have at least one feature within a specified threshold. Effect of all other features are thereby ignored and the measure becomes heavily biased. Hence, it is clear that for high dimensional data, Euclidean distance based neighbour searching can not do justice. This observation motivates us for the development of a new scheme for similarity measure.

A careful investigation on a large group of perceptually similar images reveals that similarity between two images are not usually judged by all possible attributes. This means visually similar images may be dissimilar in terms of some features as shown in Figs. 1–3.

It leads us to propose human perception based similarity measure which states that if K out of M features of two images match then they are considered similar (Saha et al., 2003a). Low value of K will make the measurement too liberal and high value may make the decision very conservative. Depending on the composition of the database the value of K can be tuned.

Distance or range based search basically looks in a region for similar images. In case of Euclidean distance as defined in Eq. (1), the region is a hypersphere. Weighted Euclidean distance as given by Eq. (2) results in a hyper-ellipsoid. Eq. (3) suggests to search in hypercube. While in range search analogous to what suggested by Eq. (3), the region is hyper-cuboid. Our proposed similarity measure, i.e., matching K out of M features leads to a star-shaped region. Fig. 4 shows some examples of such regions.

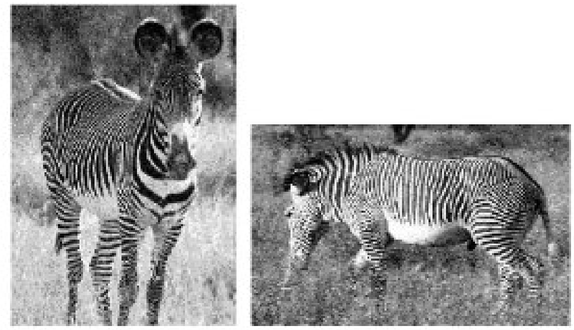


Fig. 2. Similarly textured objects with different shapes.

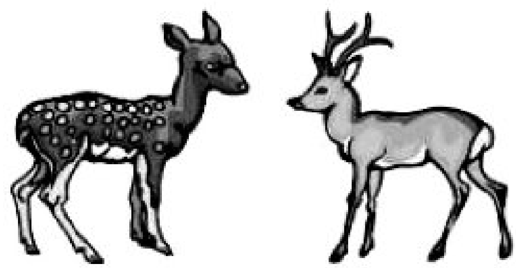
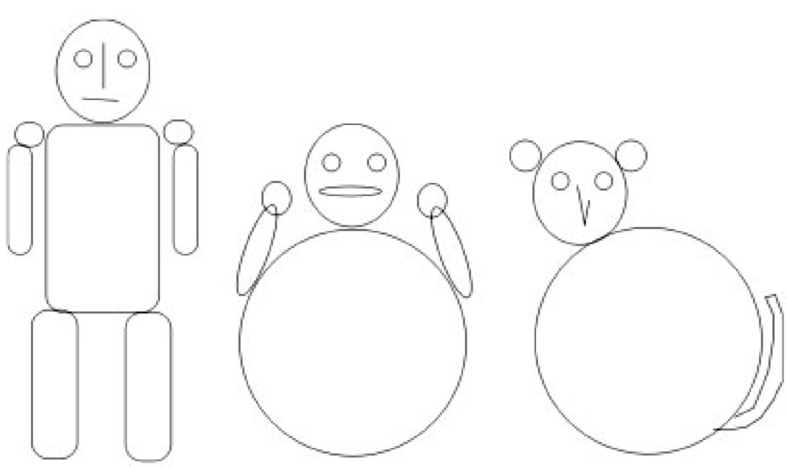


Fig. 3. Similar shapes with different textures.

When $K = M$ we arrive at region defined by Eq. (3) and that defined by Eq. (4) if $K = 1$. Hence, our similarity measure is much more generalized and flexible.

Now, the question is how to measure whether a feature of two images matches or not. If the Euclidean distance is considered, then sensitivity of the different features poses a problem. The same distance corresponding to different set of features may not reflect the same quantity of dissimilarity. Moreover, in the beginning of the section we mentioned that the elements of the feature vector carry different kinds of information and are to be treated differ-



(a) (b) (c)

Fig. 1. Similar images: (a, b) are symmetric but differ in circularity; whereas (b, c) are similar in circularity but differ in symmetricity.

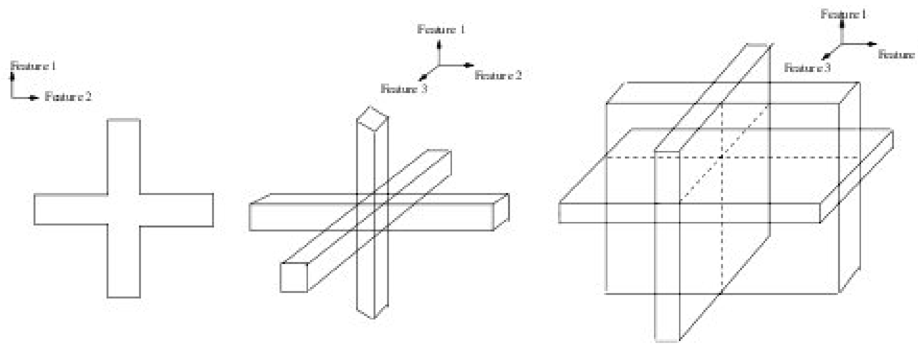


Fig. 4. Search regions for 1 out of 2 (left), 2 out of 3 (middle) and 1 out of 3 (right).

ently suitable to their characteristics. To cope with this problem, the proposed scheme maps real values of feature to character based tag. The mapping algorithm is as follows:

Assume, M is the number of features and τ is the number of buckets into which the range of values of each feature of the images in the database is divided. Thus, each feature value $f_i\,(i = 1, 2, \ldots, M)$ is substituted by a character T_j if f_i falls in jth bucket. Note that, $T_j \in \{A, B, C, \ldots\}$.

The division may be imposed based on absolute value or percentile or some other criterion. Thus the M dimensional feature vector is converted into a tag consisting of M characters. For example, if $M = 8$ and $\tau = 10$, then a tag may look like ADGACBIH.

When we perform query on the database on the basis of Euclidean distance, nearest neighbours are searched within the hypersphere. Basically, for each feature, images within a value range participate. In our scheme, when the characters representing the feature values are compared to check their proximity, it also deals with a range. The differences are (i) there is no floating point operation, and (ii) sensitivity factor of different features is reduced as their ordered grades are considered instead of absolute value. To avoid the boundary problem, at the time of comparing, neighbouring groups may be considered by setting a tolerance range t.

Thus in our scheme, similarity between two images is measured by matching the corresponding features or subset of features based on the criteria suitable to them rather than using a single distance measures considering all the features. A counter, initially set to zero, is increased if a feature is matched and similarity is declared by comparing the count with K. The retrieved images may be ordered based on this count for top order retrieval.

In order to deal with multivalued feature like histograms, some kind of transformation is to be applied on them to obtain fewer and meaningful values. Otherwise, a single independent value may not convey significant information. For example, from color histogram, chromaticity and intensity can be computed and used as features.

It may be noted that, a high value of K will increase the precision but may fail to retrieve sufficient number of images. Hence, the value of K is to be tuned as per requirement.

3. Indexing scheme

In this section, we present an indexing scheme that approximates the human perception based similarity measure (i.e., K out of M feature matching) discussed in the last section.

The database is first partitioned into number of cells on the basis of a set of features called primary key set (PK). Then the records within each cell is ordered according to a set of features called secondary key set (SK). While preparing the cells, overlapping may be allowed to avoid multiple cell searching in case of non-zero tolerance.

The main problem is to select the key sets. Entropy of the individual feature may be a good guideline for such selection. Features with high entropy discriminates images recklessly. As a result, similar images with little variation may be put into different cells. On the other hand, the features with low entropy has low discriminating power. So we pick up feature(s) with median entropy as PK. To avoid exclusion of similar images in the ordering sequence, feature(s) with least entropy is selected as SK.

To perform a query, features of the query image are computed and mapped to character tags. Based on the PK and tolerance value, cells to be searched are decided. Finally, search is carried out in those cells over a range based on SK and tolerance value. Tags of the database images and those of the query image are compared. Images for which at least K tags match are considered as similar. They may be ordered on the basis of the number of the matched features.

Although the scheme is biased towards the matching of PK and SK, still it is a fair approximation of K out of M features similarity search. Suggested scheme of key selection (particularly for PK) tends to compromise between very strict and very liberal search.

Assuming the uniform distribution of the images in the different cells, the size of the search space to execute a query can be estimated as follows. For a single feature, $\frac{N}{\tau}$ number of images fall in the same bucket, where N is the number of images in the database. If tolerance t is considered then $N \times \left(\frac{2t+1}{\tau}\right)$ number of images match with the query image with respect to a single feature. Thus, $N \times \left(\frac{2t+1}{\tau}\right)^l$ images will match with respect to l features.

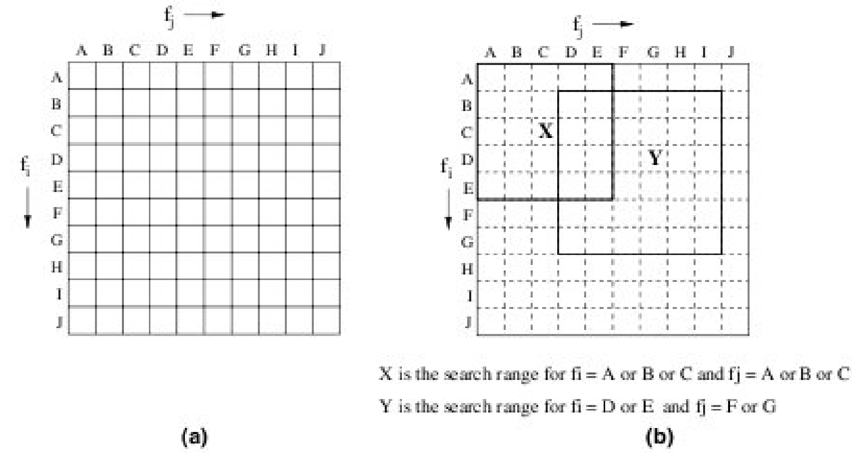


Fig. 5. Partition of the data space (see text for details).

According to the proposed methodology, an image must match in terms of PK and SK in order to be considered as a candidate for retrieval. Thus, to execute a query, in the worst case, the number of records to be searched is $N \times \left(\frac{2t+1}{\tau}\right)^{n_1+n_2}$, where n_1 and n_2 are the number of features in PK and SK, respectively. As the search is restricted to less number of records in comparison to linear search, the response will be faster.

It may be noted that, the number of cells depends on the number of divisions in the feature space and number of features in PK.

At the time of searching, depending on the tolerance range t multiple cells may have to be searched. In order to restrict the search to limited number of cells, the concept of overlapped cell can be adopted at the cost of certain extent of redundancy. For example, $t = 2$, $\tau = 10$ and PK has two elements, f_i and f_j. Then, primary key set can take 100 different values resulting into small cells as shown in Fig. 5(a). Moreover searching process is to be carried in number of cells. The concept of overlapped cells is used to restrict the search in one cell only. As t is 2, for a value, say D, it will consider B–F as matching. For a value of E it will consider C–G as matching. Thus, to restrict the search within a single cell same record has to be maintained in more than one cell. It leads to huge amount of redundancy. To minimise such redundancy, overlapping can be done in such a way that for A, B and C search is restricted to one particular cell. Similarly, search area is one cell for each of the value sets—DE, FG and HIJ. Thus two primary key feature is partition the entire database into 16 overlapping search regions (see Fig. 5(b)).

4. Relevance feedback scheme

The elements of the feature vector represents features of different types. Thus, it is very difficult to find out the correlation hidden among the various features like color and texture features specially. On the other hand, the hidden correlation has a strong implication towards the retrieval of similar images against a query. To cope up with this problem, the concept of relevance feedback (RF) can be used. Once a set of images are retrieved, they may be marked as relevant or irrelevant. This information can be used for refining the similarity measure to improve the performance.

In general, the relevance feedback schemes based on the principle of similarity measure refinement updates the emphasis of each of the features. But in our scheme (Saha et al., 2004a), distance (similarity) measure is refined by updating the emphasis of the useful features only. The term *useful feature* stands for the feature capable of discriminating relevant and irrelevant images within the retrieved set. The most crucial issue is to identify the useful features. Once such features are identified, then their emphasis are adjusted.

4.1. Identification of useful features

A close study of past work indicates that re-weighting technique is widely used for relevance feedback. But, most of the systems address how to update the weight without identifying the good features. In this paper, we present a RF scheme, which first identifies the useful features following a non-parametric statistical approach and then updates their weights.

Useful features are identified using Mann–Whitney test. In a two-sample situation where two samples are taken from two different populations, Mann–Whitney test is used to determine whether the null hypothesis that the two populations are identical can be rejected or not. Specifically, let $X_1, X_2, \ldots, X_n$ be random samples of size n from population 1 and $Y_1, Y_2, \ldots, Y_m$ be the random samples of size m from population 2. Mann–Whitney test (Conover, 1999) determines whether X and Y come from the same population or not. It proceeds as follows. X and Y are combined to form a single ordered sample and ranks 1 to $n + m$ are assigned to the observations from smallest to largest. In case of a tie (i.e., if the sample values are equal), same rank is assigned to the equal samples and the rank is taken as the average of the ranks that would have been assigned to them in case of no tie. Based on the ranks, a test statistic is generated to check the null hypothesis. If the value of the test statistic falls within the critical region then the null hypothesis is rejected. Otherwise, it is accepted.

In CBIR systems, a set of images are retrieved according to a similarity measure. Then, feedback is taken from the user to identify the relevant and irrelevant outcomes. For the time being, let us consider only jth feature and $X_i = \text{dist}(Q_j, f_{ij})$ where, Q_j is the jth feature of the query image and f_{ij} is the jth feature of the ith relevant image retrieved so far. Similarly, $Y_i = \text{dist}(Q_j, f'_{ij})$ where f'_{ij} is the jth feature of ith irrelevant image. Thus, X_is and Y_is form the different random samples. Then, Mann–Whitney test is applied to judge the discriminating power of the jth feature. Let $F_X(x)$ and $G_Y(x)$ be the distribution function corresponding to X and Y, respectively. The null hypothesis, H_0 and alternate hypothesis, H_1 may be stated as follows:

H_0: jth feature cannot discriminate X and Y (X and Y come from same population) i.e.,

$F_X(x) = G_Y(x) \quad \text{for all } x.$

H_1: jth feature can discriminate X and Y (X and Y come from different population) i.e.,

$F_X(x) \neq G_Y(x) \quad \text{for some } x.$

It becomes a two tailed test because, H_0 is rejected for any of the two cases: $F_X(x) < G_Y(x)$ and $F_X(x) > G_Y(x)$. Physically, it can be understood that a useful feature can separate the two sets and X and Y such that X may be followed by Y or Y may be followed by X in the combined ordered list. Thus, if H_0 is rejected then jth feature is taken to be as a useful feature. The steps are as follows:

1. Combine X and Y to form a single sample of size N, where $N = n + m$.
2. Arrange them in the ascending order.
3. Assign rank starting from 1. If required, resolve ties.
4. Compute test statistic T as follows:

$$T = \frac{\sum_{i=1}^{n} R(X_i) - n \times \frac{N+1}{2}}{\sqrt{\frac{nm}{N(N-1)} \sum_{i=1}^{N} R_i^2 - \frac{nm(N+1)^2}{4(N-1)}}}$$

   where $R(X_i)$ denotes rank assigned to X_i and $\sum R_i^2$ denotes sum of the squares of the ranks of all X and Y.
5. If the value of T falls within critical region then H_0 is rejected and the jth feature is considered as a useful one else not.

The critical region depends on the level of significance α which denotes maximum probability of rejecting a true H_0. If T is less than its $\alpha/2$ quantile or greater than its $1 - \alpha/2$ quantile then H_0 is rejected. In our experiment distribution of T is assumed to be normal and α is taken 0.1. If the concerned feature discriminates and places the relevant images at the beginning of the combined ordered list, then T will fall within the lower critical region. On the other hand, if the concerned feature discriminates and places the relevant images at the end of the same list then T falls within the upper critical region.

It may be noted that, the proposed work proceeds only if the retrieved set contains both—relevant and irrelevant images. Otherwise, samples from two different populations will not be available and no feedback mechanism can be adopted.

4.2. Adjustment of the emphasis of features

Adjustment of the emphasis of feature is closely related with the distance/similarity measure adopted by the system. In the current work we have adopted a human perception based similarity measure. However, for easy understanding we first present emphasis adjustment scheme for Euclidean distance. Subsequently we will transfer the idea to the perception based similarity measure.

Weighted Euclidean distance is a widely used metric for CBIR systems. Let an image be described by M features. Then, the distance between two images can be expressed as $\sum_{j=1}^{M} w_j d_j$ where, d_j denotes Euclidean distance between them with respect to jth feature and w_j is the weight assigned to the feature.

In the proposed scheme, w_j is adjusted only if jth feature is a useful one. To explain the strategy for adjustment of weights of the features, let us consider a system that relies on two features only, say, f_1 and f_2. Difference in feature values between the query image and the database image are d_1 and d_2. With $w_1 = w_2$, the search space corresponding to Euclidean distance is a circle (as seen in Fig. 6 with solid line). Now suppose f_1 is a useful feature such that the test statistic of d_1 lies in the lower critical region. That means, f_1 can discriminate between relevant and irrelevant

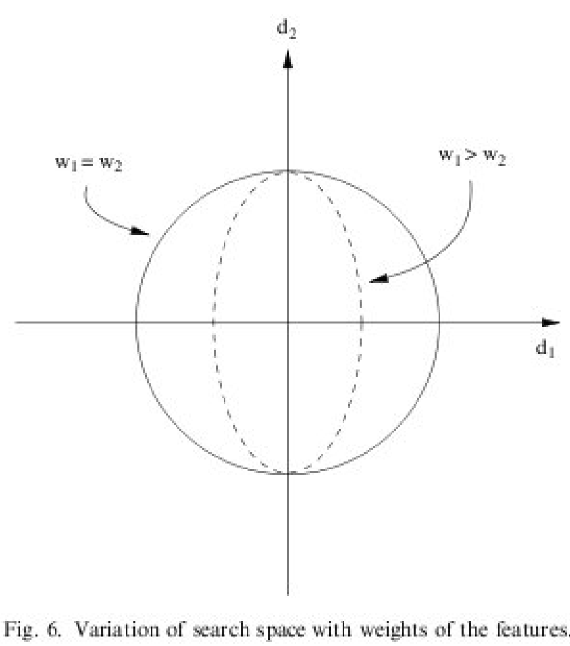


Fig. 6. Variation of search space with weights of the features.

images, and the d_1 of the relevant images are, in general, less than d_1 of the irrelevant images. If we make $w_1 > w_2$, search space is changed to ellipse (as seen in Fig. 6 with dashed line) and thereby we discard irrelevant images as much as possible from the retrieved set. Similarly, if f_1 is a useful feature and the test statistic of d_1 lies in the upper critical region then d_1 of relevant images are, in general, greater than d_1 of irrelevant images. Hence, by making $w_1 < w_2$, more relevant images can be included in the retrieved set. Thus by increasing the weight of the useful feature with lower test statistic, we try to exclude the irrelevant images from the retrieved set. On the other hand, by decreasing the weight of the useful feature with higher test statistic, we try to include the relevant images in the retrieved set.

In essence, the proposed RF scheme works as follows. Once images are retrieved, feedback is taken from the user and useful features are identified. Finally, weights are adjusted as follows:

1. Initialize all w_j to 1.
2. For each jth useful feature where test statistic falls within lower critical region, set w_j as

   $w_j = w_j + \sigma_x^2$

   where, σ_x^2 is the variance of X.
3. For each jth useful feature where test statistic falls within upper critical region, set w_j as

   $w_j = \max\{0, w_j - \sigma_x^2\}$
4. Repeat step 2 and 3 for successive iteration.

In case of human perception based similarity measure (Section 2) also, useful features are identified following the same technique. But, the adjustment of emphasis of a feature is to be addressed in a slightly different manner. In this method, whether or not an image would be retrieved is decided by the count of matched features with the query image. Hence, updation of emphasis of feature has a direct impact on choosing of features for matching. It can be achieved by changing the match tolerance for the useful features. However, the basic principle is similar to that adopted in Euclidean distance based search. When the similar images lie in the close vicinity of the query image in terms of the useful features, i.e., test statistic falls within lower critical region, the tolerance is reduced to restrict the inclusion of irrelevant images. The situation is reverse for the useful features with test statistic falling in the upper critical region. In that case, the similar images are lying in the distant buckets. Thus, to increase the possibility of inclusion of similar images the match tolerance is increased. The steps are as follows:

1. Initialize the tolerance for all features to t.
2. For all jth useful features with test statistic in lower critical region set, $tolerance_j = tolerance_j - 1$
   If $tolerance_j < \mathrm{MIN}$ then $tolerance_j = \mathrm{MIN}$
3. For all jth useful features with test statistic in upper critical region set, $tolerance_j = tolerance_j + 1$
   If $tolerance_j > \mathrm{MAX}$ then $tolerance_j = \mathrm{MAX}$
4. Repeat step 2 and 3 for successive iteration.

MIN and MAX denote minimum and maximum possible tolerance value. In our experiment, we have considered t as 2, MIN as 1 and MAX as $\tau - 1$ where τ is number of buckets in the feature space.

In order to obtain the dual advantages of improved precision and faster response the search strategy should combine the schemes of relevance feedback and indexing. Due to the feedback, tolerance range for useful features may increase or decrease. If it increases for the feature in the primary key set then according to the proposed indexing scheme search may have to carried to additional cells. Thus, the response time may increase. The alternative approach may be to ignore the feedback on features in primary key set. Then, one may have to sacrifice few similar images from the retrieved set which would have been included otherwise. Thus, it may have some detrimental effect on precision. Hence, a careful trade-off is to be imposed depending on the requirement.

5. Experimental results

In our experiment we have used COIL-100 database of Columbia university. It contains 7200 images. Actually 100 different objects are there. For each object 72 different images are generated which correspond to various orientation. Features are computed for the object in the image. Hence, a fast segmentation technique as described in (Saha et al., 2003b) is used to extract the dominant object. Then, various shape, texture and color features are computed. The feature vector is of 48 dimension of which 23 are shape features, 18 features denote texture and remaining 7 represent the color.

Projection method is an interesting technique for extraction of shape information. In our system, petal projection based various shape features are computed (Saha et al., 2003b). In this scheme, an object is divided into a number of petals where a petal is an angular strip area originating from the centre of gravity. Area of the object lying within a petal is taken as the projection along it. Finally, an 8 dimensional petal projection vector is obtained. Based on it, circularity, symmetricity, aspect ratio and concavity are computed. To supplement these features, another set of simple but effective measures of circularity, symmetricity, etc. (Saha et al., 2003b) are used in our system.

We have used a 15×15 texture co-occurrence matrix (Saha et al., 2004b) to describe the texture feature. In order to compute the matrix, the intensity component of the color image is divided into blocks of size 2×2 pixels. Then grey level pattern of the block is converted to a binary pattern by thresholding at the average value of the intensities. Decimal equivalent of the binary string formed from this pattern arranged in raster order gives the texture value.

Table 1
Precision (in %) of retrieval for COIL-100 database

	Euclidean distance based linear search	Proposed similarity measure based linear search	Proposed indexed search
P(10)	82.46	88.52	90.23
P(20)	73.59	79.25	81.45
P(30)	67.31	72.25	74.56

Table 2
Precision (in %) using relevance feedback for Euclidean distance based linear search

	No relevance feedback	Relevance feedback (after iteration 3)
P(10)	82.46	84.74
P(20)	73.59	76.47
P(30)	67.31	70.40

Thus we get the quantized image whose height and width are half of that of the original image and the pixel values range from 0 to 14. Finally, a grey level co-occurrence matrix of size 15×15 is computed from this image. Based on this matrix, features like moments, energy, entropy, homogeneity, etc. are computed.

In order to compute the color feature, a hue histogram is formed based on HSV model. It is then smoothened and normalized. For each of the six major colors (red, yellow, green, …, magenta) and grayness, index of fuzziness is computed as it has been outlined in (Saha et al., 2004b).

To study the performance of perception based similarity measure, each database image is used as the query image and an exhaustive search in the database is carried on once using the Euclidean distance based similarity measure and again using the perception based similarity measure. In the latter case, each feature space is divided into 20 buckets and value of K is taken as 25. Tolerance for matching(t) is taken as 2. As a result, for a value, say D, it will consider B–F as matching. For a value of E it will consider C–G as matching.

Muller et al. (2001) has mentioned that, from the perspective of a user, top order retrievals are of major interest. Secondly, in case of retrieval using perception based similarity measure, as it is quite likely that similar images may spread over multiple buckets, achievement of high recall is quite difficult. Hence, performance is studied based on top order retrievals. It is evident from second and third columns of Table 1 retrieval using human perception based similarity measure is better.

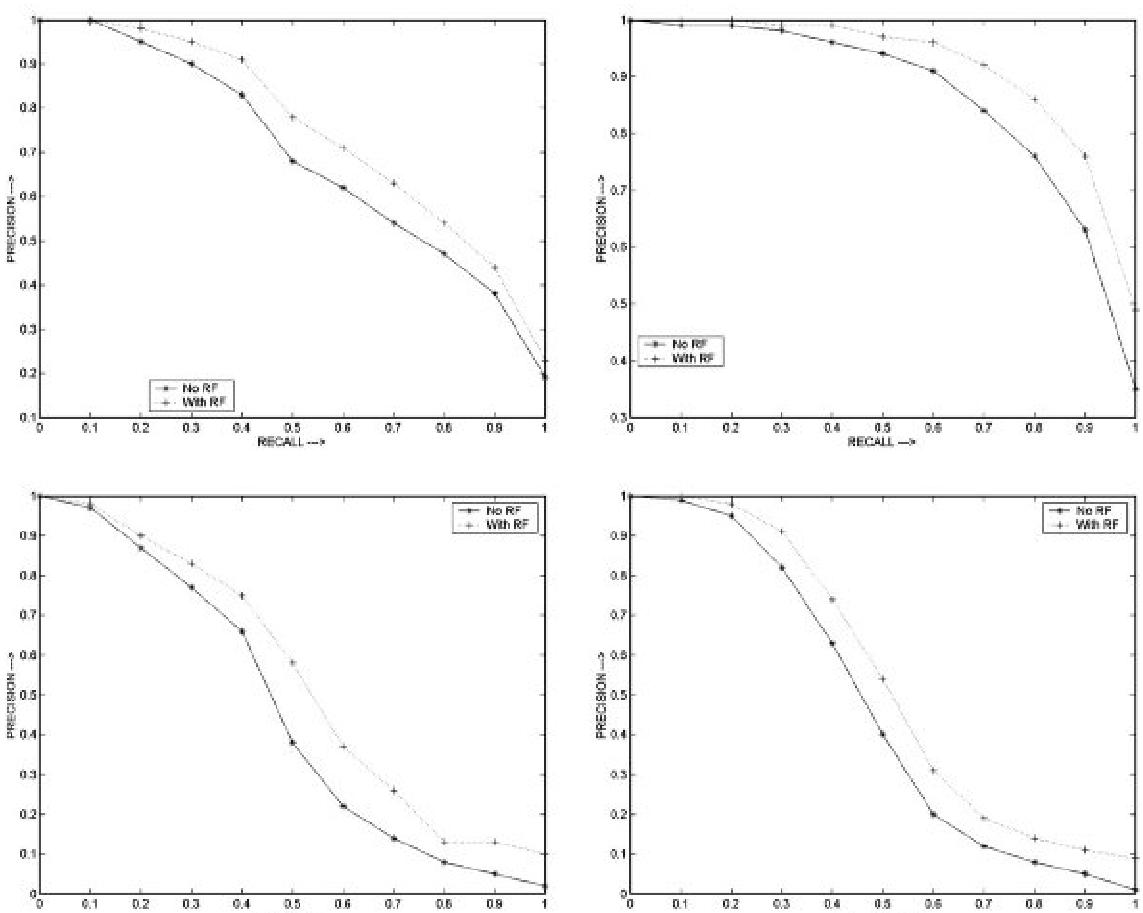


Fig. 7. Recall-precision graph for different object from COIL-100 database; they are (in raster scan order) object 17, 28, 43 and 52.

In order to study the performance of the indexing scheme the database has been partitioned in to number of cells. The primary key set consists of two elements and secondary key contains only one element. As each feature space is divided into 20 buckets, 400 cells of smaller size can be created. Following the concept of overlapped search region as discussed in last section, the database is partitioned into 81 cells. Thus search will be restricted into only one cell. The retrieval performance of indexed search is shown in the third column of Table 1. The basic purpose of indexing is to obtain a faster response and experiment shows that indexed search is around 7 times faster than linear search.

In order to check the capability of proposed relevance feedback scheme, linear search is carried on using each database image as the query image. It is first applied for Euclidean distance based retrieval. Table 2 shows that precision improves for top order retrievals. As the database is already groundtruthed, after retrieving a set of images they are automatically marked as relevant or irrelevant. Thus, feedback is automatically obtained. Hence, to prepare the recall precision graph, all the images retrieved to achieve a particular recall participate in feedback mechanism. Fig. 7 shows the recall precision graphs correspond to a few objects for Euclidean distance based search after third iteration of relevance feedback. It is clear from the graphs that use of proposed scheme improves the performance. Moreover, increase in the time overhead for adopting the relevance feedback mechanism is very low.

As the effectiveness of the proposed relevance feedback scheme is established for Euclidean distance based retrieval, it is applied for the proposed similarity measure based retrieval also. Table 3 shows that precision improves.

Finally, the indexing scheme and relevance feedback scheme are combined. In order to avoid multibucket search, we have not updated the emphasis of features in primary key set. Table 4 shows that better precision is obtained when relevance feedback is combined with indexed search. The response time increases up to some extent due to the additional overhead of relevance feedback scheme.

Table 3
Precision (in %) using relevance feedback for human perception based linear search

	No relevance feedback	Relevance feedback (after iteration 3)
P(10)	88.52	91.07
P(20)	79.25	83.91
P(30)	72.25	79.57

Table 4
Precision (in %) using relevance feedback for indexed search

	No relevance feedback	Relevance feedback (after iteration 1)
P(10)	90.23	92.23
P(20)	81.45	85.55
P(30)	74.56	81.10

6. Conclusion

In this paper we have presented a human perception based similarity measure which considers the matching of only a subset of features and reduces the floating point operation. It also overcomes the curses of dimensionality problem of Euclidean distance measure. The indexing scheme is simple enough and reduces the search domain. As a result a fast response is obtained without sacrificing the precision significantly. Moreover, the precision can be improved by tuning the parameters. As the indexing scheme depends on certain global data like entropy of the features and percentile grouping of the elements, the index can not be updated dynamically. Creation of the index should be an offline process and the index is to be created again when sufficient number of images are added. We have presented a novel relevance feedback scheme based on Mann–Whitney test which can identify the useful features capable of discriminating relevant and irrelevant images within a retrieved set. For weighted Euclidean distance based search, a weight upgradation scheme is suggested based on the variance of the features of the retrieved images. Following the similar principle, a tolerance updation scheme is devised for the proposed similarity search. In both the cases, the schemes are found to be successful and effective. Finally, indexing scheme and relevance feedback are combined to develop a complete retrieval engine for CBIR systems.

References

Bentley, J.L., Friedman, J.H., 1979. Data structures for range searching. ACM Comput. Surv. 11 (4), 397–409.

Berman, A.P., Shapiro, L.G., 1999. A flexible image database system for content-based retrieval. Comput. Vision Image Understanding 75, 175–195.

Ciocca, G., Gagliardi, I., Schettini, R., 2001. Quicklook2: An integrated multimedia system. Internat. J. Visual Languages Comput. 12, 81–103.

Conover, W.J., 1999. Practical Nonparametric Statistics, third ed. John Wiley and Sons, New York.

Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petrkovic, B., Equitz, W., 1994. Efficient and effective querying by image content. J. Intell. Inform. Systems 3, 231–262.

Fournier, J., Cord, M., Philipp-Foliguet, S., 2001. Retin: A content-based image indexing and retrieval system. Pattern Anal. Appl. 4, 153–173.

Gevers, T., Smeulders, A., 2000. Pictoseek: Combining color and shape invariant features for shape retrieval. IEEE Trans. Image Process. 9 (1), 102–119.

Guttman, A., 1984. R-trees: A dynamic index structure for spatial searching. In: Proc. ACM SIGMOD Conf. on management of data, pp. 47–57.

Huang, J., Kumar, S.R., Mitra, M. , 1997. Combining supervised learning with color correlogram for content-based retrieval. In: Proc. 5th ACM Internat. Multimedia Conference, pp. 325–334.

Laaksonen, J., Koskela, M., Laakso, S., Oja, E., 2000. Picsom—content-based image retrieval with self-organizing maps. Pattern Recognition Lett. 21, 1199–1207.

Lai, T.S., 2000. Chroma: A photographic image retrieval system. Ph.D. thesis, School of computing, Engineering and Technology, University of Sunderland, UK.

Leutenegger, S.T., Lopez, M.A., Edgington, J.M., 1997. A simple and efficient algorithm for r-tree packing. In: Proc. 13th Internat. Conf. on Data Engineering, pp. 497–506.

Mukherjee, S., Hirata, K., Hara, Y., 1999. A world wide web image retrieval engine. WWW J. 2 (3), 115–132.

Muller, H., Muller, W., Marchand-Mailet, S., Pun, T., Squire, D.M., 2001. Automated benchmarking in content-based image retrieval. In: Proc. ICME 2001, Tokyo, Japan, pp. 22–25.

Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Pektovic, D., Yanker, P., Faloutsos, C., Taubin, G., 1993. The qbic project: Querying images by content using color, texture and shape. SPIE 19, 173–187.

Rocchio, J.J., 1971. Relevance feedback in information retrieval. In: Salton, G. (Ed.), The SMART Retrieval System: Experiments in Automatic Document Processing. Prentice Hall, pp. 313–323.

Rui, Y., Haung, T.S., Mehrotra, S., Ortega, M., 1998. Relevance feedback: A power tool in interactive content-based image retrieval. IEEE Trans. Circ. Systems Video Technol. 8 (5), 644–655 (Special issue on Interactive Multimedia Systems for the Internet).

Saha, S.K., Das, A.K., Chanda, B., 2003a. An efficient search technique for cbir systems. In: Proc. 3rd Internat. Workshop on Content Based Multimedia Indexing, Rennes, France, pp. 151–156.

Saha, S.K., Das, A.K., Chanda, B., 2003b. Graytone image retrieval using shape feature based on petal projection. In: Proc. Internat. Conf. on Advances on Pattern Recognition, Kolkata, India, pp. 252–256.

Saha, S.K., Das, A.K., Chanda, B., 2004a. Image retrieval using relevance feedback based on Mann–Whitney test. In: Proc. Indian Conf. on Computer Vision, Graphics and Image Processing, Kolkata, India, pp. 405–410.

Saha, S.K., Das, A.K., Chanda, B., 2004b. CBIR using perception based texture and colour measures. In: Proc. Internat. Conf. on Pattern Recognition, vol. II, Cambridge, UK, pp. 985–988.

Samet, H., 1984. The quadtree and related hierarchical data structures. ACM Comput. Surv. 16 (2), 187–260.

Sciascio, E.D., Mingolla, G., Mongiello, M., 1999. Content-based image retrieval over the web using query by sketch and relevance feedback. In: Proc. Third Internat. Conf. VISUAL '99, Amsterdam, pp. 123–130.

Sclaroff, S., Taycher, L., Cascia, M.L., 1997. Imagerover: A content-based image browser for the world wide web. In: Proc. IEEE Workshop on content-based Access of Image and Video Libraries, San Juan, Puerto Rico, pp. 2–9.

Sebastian, T.B., Kimia, B.B., 2002. Metric-based shape retrieval in large databases. In: Proc. Internat. Conf. on Pattern Recognition, Canada.

Srihari, R., Zhang, Z., Rao, A., 2000. Intelligent indexing and semantic retrieval of multimodal documents. Inform. Retrieval 2 (2), 245–275.

Vleugels, J., Veltkamp, R.C., 2002. Efficient image retrieval through vantage objects. Pattern Recognition 3 (1), 69–80.

Weber, R., 1998. Similarity search in high-dimensional data spaces. Grundalgen von Datenbanken, 138–142.