Noname manuscript No.
(will be inserted by the editor)

Interactive Image Retrieval using M-band wavelet, Earth Mover's Distance and Fuzzy Relevance Feedback

Malay K. Kundu · Manish Chowdhury · Minakshi Banerjee

Received: date / Accepted: date

Abstract We propose an interactive content based image retrieval (CBIR) system using M-band wavelet features with Earth Mover's Distance (EMD). A fuzzy relevance feedback (FRF) method is proposed to enhance the retrieval in order to retrieve more images. So that retrieve images are semantically close to the query. $M \times M$ sub-bands coefficient are used as primitive features, on which, for each pixel, energies are computed over a neighborhood and are taken as features for each pixel to characterize its color and texture properties. Based on the energy property, pixels are clustered using Fuzzy C-Means (FCM) algorithm to obtain an image signature. The EMD is used as a distance measure between the signatures for different images of the database. Combining information both from relevant and irrelevant images marked by the user, fuzzy entropy based feature evaluation mechanism is used for automatic computation of revised feature importance and similarity distance at the end of each iteration. The proposed CBIR system performance using M-band wavelets feature are compared to that of MPEG-7 visual features. As MPEG features have almost become a standard benchmarks for both video and image representation and comparison. The proposed FRF technique using EMD is compared with different other similarity measures to test the effectiveness of the proposed system on standard image database.

Malay Kumar Kundu† and Manish Chowdhury*
Machine Intelligence Unit, Indian Statistical Institute
203 Barrackpore Trunk Road
Kolkata-700108, India.
Tel.:+91-33-25753108
Fax: +91-33-25783357
E-mail: †malay@isical.ac.in and *manishchowdhury2005@gmail.com

Minakshi Banerjee
Department of Computer Science and Engineering
RCC Institute of Information Technology
Kolkata-700015, India.

Keywords CBIR · EMD · FRF · FCM · ED · MD · CBD

1 Introduction

Automatic image retrieval is an important research problem considering its usage and ever increasing volume of image data existed in different kinds of databases both on the web as well as in the network computing system. Image retrieval is normally being done using label attachment to each of the image. This may be either a text document generated manually for annotation or feature map extracted from the image automatically. Former methods of manual labelling become irrelevant due to continuous increasing in size of image databases. The image labelling based on feature extracted from inherent image characteristic automatically is known as Content Based Image Retrieval (CBIR)[16–20,34,35], and which is the current practice for image retrieval.

The basic feature of CBIR is used to return a group of images from a very large image database based on the query image given to the system. The accuracy of the system is largely depends upon the quality of visual features used to represent an image information which can capture the overall visual impression of it. To compare the similarity of the images from the database with query image, a different kind of similarity measures may be used. Most popular of them is Euclidean Distance (ED), which is simple and involves low cost computation. It has been shown recently, of the different kinds of distance measures [27] such as Minkowski-form Distance, Histogram Intersection, Kullback-leibler Divergence, Jeffrey Divergence or χ^2 Statistics, recently Earth Mover's Distance (EMD) [22,23] is found to be more accurate in capturing perceptual distance for visual recognition.

Generally, low level features such as color, texture, shape, corner etc., are used to represent as approximate perceptual representation of an image, using which similarity and dissimilarity of the images are computed. But it is found that the perceptual representation of an image in terms of low level features fails to capture entire semantic information of an image and it is often difficult to model accurately. These results lower the accuracy of a CBIR system than expected.

To overcome these difficulties, Relevance Feedback (RF) mechanism is used to enhance the performance of a CBIR system. This idea emerges from the fact that the human observer acquires knowledge about visual subjects, that is gathered over the years, which one learns through different known guided examples. Following the similar approach, a CBIR system may use RF [13–15] to learn more about the similarity and dissimilarity of the images with the query and the human observer as a guide.

There are mainly two types of RF approaches (a) the "weighing approach" where higher weight is given to more distinguishing features in order to reflect the user feedback when calculating similarity and (b) the "probability approach" where the information representing the query image is modified, according to the feedback given by the user. Most existing work in content

based image retrieval (CBIR) uses the weighted approach [31]. RF mostly uses information of positive images only [26]. However, Zin *et al.* [3] have proposed a feature re-weighting technique by using both the relevant and the irrelevant information, to obtain more effective results. Marakakis *et al.* [29] uses Gaussian mixture (GM) models for the image features and query information is updated in a probabilistic manner. This update reflects the preference of the user and is based on the models of both the positive and negative feedback. RF has been considered as a learning and classification process, using classifiers like Bayesian classifiers [4,28], neural network [5], etc. However, trained classifiers become less effective when the training samples are insufficient in number. To overcome such problems, active learning [2] methods are also used.

The computational cost and performance of a CBIR system with RF largely depends on the total numbers of features extracted from different images. Different kinds of derived features are used besides normal features, obtained from spatial data like color, texture, shape, corner etc. The literature on visual features type and its extraction methods is quite rich. People uses frequency domain features like FFT, DCT, Gabor, Wavelet etc., as a tool for features extraction. Among many frequency domain features, wavelets are gaining significant importance in the field of image retrieval [34,37].

MPEG-7 is an ISO/IEC standard developed by MPEG (Moving Picture Expert Group) to facilitate effective uses of audio, visual (color, texture, shape etc.) and motion picture description to address multimedia retrieval. It was formally named as Multimedia Content Description Interface. It is a standard for the multimedia content data that supports interpretation of the information, which can be passed onto, or accessed by a device or a computer code. MPEG-7 is not targeted at any one application in particular; rather it supports a range of applications as possible.

Manjunath *et al.* [10,11] extensively studied the use of MPEG-7 descriptor features based retrieval [9]. MPEG-7 is very extensive and can capture very closely low level visual description through number of descriptors. It induces low level feature extraction algorithms using color, texture, motion, and shape, facilitated for image and video retrieval and benchmarking of newly proposed schemes. MPEG-7 provides Scalable Color Descriptor(SCD), Color Structure Descriptor(CSD), Dominant Color Descriptor (DCD) and Color Layout Descriptor (CLD) for color based retrieval and Texture Browsing Descriptor(TBD), Homogeneous Texture Descriptor(HTD)and Edge Histogram Descriptor(EHD) for texture based retrieval.

The CSD and EHD are generally used as color and texture descriptor to take care of local and global information respectively. But the total numbers of features involved to these descriptors are very large: CSD (256) and EHD (80), which results in high computational cost. So to minimize the computational cost, which is proportional to total number of features used for color and texture description. So, researchers have tried to use wavelet based multiscale color-texture features as an alternative, to reduce the cost of computation with an acceptable level of accuracy for retrieval. Wang *et al.* [8] have used a 2-step algorithm using dyadic wavelet transform for developing a CBIR system for

retrieval of color images. They transformed the image to a color space similar to opponent color space prior to wavelet decomposition. At first, sub-band variances are used as representative features for crude matching, the outputs of which are used for a finer matching.

Popularly used dyadic wavelet transform is not very suitable for analysis of high frequency signals with relatively narrow bandwidth. It decomposes the signal channel into logarithmic tiling of time scale plane. Whereas, M-band wavelet transform divides the signal into a mixture of linear and logarithmic tiling of time scale plane. It gives richer parameter space having greater variety of compactly supported components. It has ability to achieve more rapidly a given frequency resolution as a function of the decomposition scale which results in better resolution at high frequencies and overcomes the drawbacks of standard wavelets[7,21]. The M-band wavelet minimizes the computational cost by reducing the number of features for retrieval as compared to popularly used MPEG-7 visual features, like, CSD and EHD [10].

Traditionally, Euclidean Distance (ED) is used to measure the similarity between the feature vector of the query image and the images in the database [1]. However, the main problem using ED [27] is a scale problem because features that have an inherently larger value, due to its scale value it tries to dominate ignoring the small values features. In contrast, Earth Mover Distance (EMD) measure is a variable size descriptions of distributions [23], that can able to overcome this problem. In EMD, ground distance is calculated between the two signature of same features spaces.

In this paper, we present a noble CBIR method, where M-band wavelet transform is used as a tool for feature extraction. The proposed method has low computational complexity as compared to the MPEG-7 visual descriptor having less number of features. We also propose an approach based on fuzzy relevance feedback (FRF) using a weighted EMD distance as a similarity measure. The weights are computed automatically based on a fuzzy feature evaluation mechanism which uses the information from both relevant and irrelevant retrieved images [1]. The weights of the feature component are updated followed by weighted EMD distance is recomputed for generating better results. Using the same set of features with FRF, the performances of the proposed scheme are also tested using other similarity measure namely Euclidean Distance (ED), Manhattan Distance (MD) and Chessboard Distance (CBD). Different retrieval results obtained are compared, where EMD is found to be better than other similarity measures for almost all types of image example. The system performance is also compared using MPEG-7 visual features [1].

The paper is organized as follows: Section 2 describes the detailed theory used in the proposed work. The proposed CBIR system with block diagrams is explained in Section 3. The proposed methodology with detailed description is demonstrated in section 4. Section 5 describes the experimental system and results. Finally, section 6 concludes the article.

2 Theoretical Preliminaries used in the Proposed Work

In this section, we present a brief outline about M-band wavelet transform which is used for image processing application [36–38] and the similarity EMD distance measure used in the proposed CBIR method.

2.1 M-band Wavelet Transform

M-band wavelet are practically implementable and have their ability to achieve more rapidly a given frequency resolution as a function of decomposition scale. In this paper, we have used M-channel filters for decomposing the time-scale space into M×M sub-bands.

An M-band wavelet system [21] forms a tight frame for functions $f(x) \in L^2(\Re)$. As compared to a dyadic representation where we have one wavelet function, an M-band representation has (M-1) wavelets. Thus there are (M-1) unitary wavelet filters. The M-band wavelet expansion [21] is given by

$$f(x) = \sum_{k} c(k)\varphi_k(x) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} \sum_{i=1}^{M-1} M^{j/2} d_{i,j}(k)\psi(M^j x - k) \tag{1}$$

where φ_k is the scaling function and ψ are the wavelet functions respectively and are associated with the analyzing (or synthesizing) filters. The wavelet coefficient is

$$c(k) = \int f(x)\varphi(x-k)dx \tag{2}$$

And the expansion coefficient of coarser signal approximation of $f(x)$ is

$$d_{i,j}(k) = \int f(x) M^{j/2} \psi(M^j x - k)dx \tag{3}$$

Given the scaling and wavelet filters one constructs the scaling function which is the solution to the following two scale difference equation that involves only the scaling filter as shown as

$$\varphi(x) = \sum_{k=0}^{N-1} \sqrt{M} h(k)\varphi(Mx - k) \tag{4}$$

This equation satisfies the recursive equation and is compactly supported in $[0, (\mathrm{N}-1)/\mathrm{M}-1]$, where the sequence $h(n)$ is the scaling vector of length $N = MG$ and is characterized by the constraints:

$$\sum_{n} h(n) = \sqrt{M} \tag{5}$$

$$\sum_{n} h(n + Mm)h(n) = \delta(m) \tag{6}$$

The (M-1) wavelet function [21] can be defined as:

$$\psi_i(x) = \sum_n \sqrt{M} h_i(n) \varphi(Mx - n), for \quad i = 1, 2, ..., M-1 \tag{7}$$

The scaling function and (M-1) wavelet function also define as Multiresolution Analysis (MRA) [21]. A MRA also satisfy the orthonormality condition, and then the subspaces form an orthogonal decomposition of functional space. A MRA is a sequence of approximation spaces for $L^2(R)$. If the space spanned by the translate of $\psi_i(x)$ for fixed j and $k \epsilon \mathbf{Z}$ is defined by $W_{i,j} = Span\{\psi_{i,j,k}\}$, then it can be shown [7] as

$$W_{0,j} = \bigoplus_{i=0}^{M-1} W_{i,j-1} \tag{8}$$

$$\lim_{j \to \infty} W_{0,j} = L^2(R) \tag{9}$$

Thus the $W_{0,j}$ spaces form a multiresolution space for L^2. An important aspect of M-band wavelets is that a given scaling filter h specifies a unique scaling function $\varphi(x)$ and consequently a unique MRA.

2.2 Earth Mover's Distance(EMD)

In an efficient CBIR, the image representation in terms of features as well as distance measure used for visually discriminating between different images, should have characteristics close to human visual representation. Various similarity measures like Minkowski Distance, Histogram Intersection, or χ^2 Statistics are used by the researchers in CBIR. It is observed that EMD follow more closely, the human like visual discriminate capability as distance or similarity measures [22, 23].

The dissimilarity measures are either based on information theoretic aspects or statistics and thus are not able to handle perceptual similarity. A signature defined as $\{cl_i = (feature_i, weight_{feature_i})\}$ characterizes each image independently and efficiently. A complex image will have a larger signature while a simple image will have a small signature, by adapting the number of clusters depending on the complexity of the image [23].

The computation of EMD is based on a solution to Monge-Kantorovitch mass transfer problem. The problem can be represented in terms of flow of goods between suppliers and consumers [22]. Assuming supply of goods from several suppliers each having a capacity to supply to several consumers each having a consumption capacity, the transportation problem is then to find the least expensive flow of goods which satisfies the consumer's demands. Signature matching then becomes analogous to the transportation problem by defining one signature as supplier and the other as the consumer, and the ground distance between an element in the first signature to an element in the second as the cost for a supplier-consumer pair. The EMD [22] thus measures the minimum amount of work required to transform one signature into other. It is formally defined as follows:

Let $P = \{(p_1, w_{p_1}), ..., (p_m, w_{p_m})\}$ be the first signature with m clusters where p_i is a cluster representative and w_{p_i} is the weight of the cluster $Q = \{(q_1, w_{q_1}), ..., (q_n, w_{q_n})\}$ is the second signature with n clusters. Let $D = [d_{ij}]$ the ground distance matrix where $d_{ij} = d(p_i, q_j)$ is the ground distance between clusters p_i and q_j, chosen according to the task at hand.

Computing EMD thus becomes finding a flow $F = [f_{ij}]$ with f_{ij} the flow between p_i and q_j which minimizes the overall cost. $\mathbf{WORK}(P, Q, F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d(p_i, q_j) f_{ij}$ subject to the constraints:

$$f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n \tag{10}$$

$$\sum_{j=1}^{n} f_{ij} \leq w_{p_i}, 1 \leq i \leq m, \tag{11}$$

$$\sum_{i=1}^{m} f_{ij} \leq w_{q_j}, 1 \leq j \leq n, \tag{12}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = \min(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j}) \tag{13}$$

Constraint Eq.10 ensures movement of goods from suppliers to consumers and not the other way. Constraint Eq.11 defines the upper bound on the capacity of the suppliers while Eq.12 defines the upper bound on the capacity of the consumers. Constraint Eq.13 ensures that maximum possible supplies to be moved from suppliers (P) to consumers (Q), called the total flow. Once the solution to optimal flow is obtained EMD is defined as the work normalized by the total flow:

$$EMD(P, Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d(p_i, q_j) f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \tag{14}$$

EMD by its definition extends to distance between sets or distributions of elements, thereby facilitating partial matches. The cost of moving earth from sand piles to holes, to fill them defines the nearness property properly as compared to histogram, information theoretic or statistics based approach. It can be shown that, EMD is a metric, if the ground distance is a metric and the total weights of two signatures are equal [22].

3 The Proposed Technique

The proposed methodology may be described in terms of three basic functional blocks as follows:

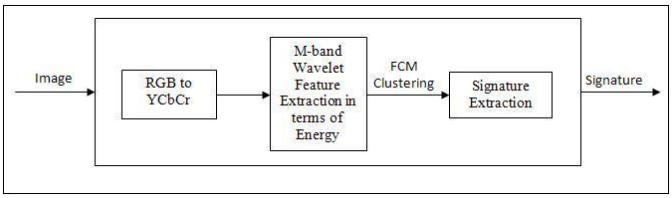


Fig. 1 Block Diagram of Signature Generation

3.1 Signature Generation

The steps of the Signature Generation part as shown in Fig.1 are as follows:

1. RGB color image is converted into YCbCr color plane.
2. YCbCr image is decomposed into $M \times M$ channels without downsampling by using 1D, 16 tap 4 band orthogonal filters as a kernel for wavelet filter and then energy feature at each pixel is computed.
3. FCM clustering algorithm is used in each pixel to obtain the signature of the image.

3.2 Image Retrieval System

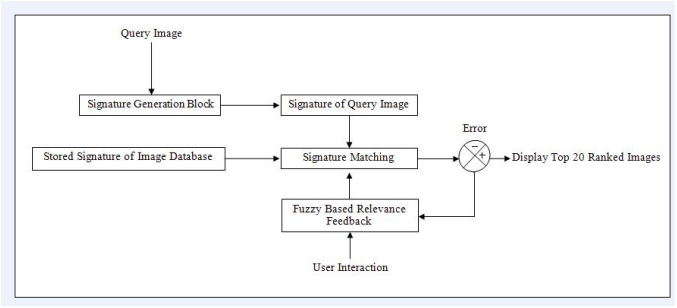


Fig. 2 Block Diagram of Proposed Image Retrieval System

The salient steps of Fig.2 are discussed as follows:

1. The signature of the query image is matched against the stored signature of the images in the database and displayed as the top ranked 20 images.
2. User marks the relevant images with respect to the query image.
3. FRF block uses the information, provided by the user to retrieve better collection of top 20 images for the next iteration.
4. This retrieval process is terminated, when the user is satisfies with the retrieval results.

3.3 Fuzzy Based Relevance Feedback (FRF)

Here, in Fig.3, we list the main steps of the Fuzzy Based Relevance Feedback mechanism.

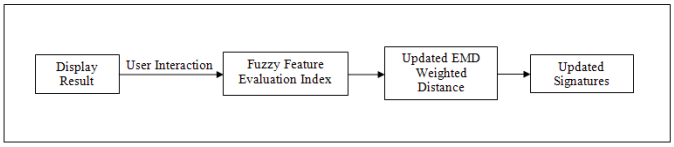


Fig. 3 Block Diagram of Fuzzy Based Relevance Feedback

1. With the user interaction in the display results, relevant and irrelevant images are marked. Fuzzy feature evaluation index computes the relative importance of features.
2. Ranked fuzzy feature is used to recompute the weighted EMD distance.
3. Signature of the images in the database are updated with the modified EMD distance.

4 Implementation Details

The feature extraction process, theoretical details of our proposed FRF using EMD and other similarity measures are discussed in this section.

4.1 M-band Wavelet based Color and Texture Feature Extraction

Human eye shows varying sensitivity response to different spatial frequencies. A Human visual system divides an image into several bands, for complete visualizing the complete image as a whole. This fact motivated us to use the M-band filters which are essentially frequency and direction oriented band pass filters. We use a 1-D, 16 tap 4 band orthogonal filters with linear phase and perfect reconstruction for the multi-resolution analysis. The 1-D, M-band filter transfer functions are denoted by H_i, $1 \leq i \leq 4$. The image, prior to M-band wavelet decomposition, is transformed to Y-Cb-Cr color space. This ensures that the textural characterization of the image is independent of the color characterization. Wavelet decomposition over the intensity plane characterizes the texture information, while the wavelet decomposition over chromaticity planes characterizes color. Wavelet transform is applied to Y, Cb and Cr planes. An over-complete decomposition resulting in the same size of the sub-bands as the image is important. To obtain the features for each pixel of the image, which are subsequently clustered. The 16 sub-bands coefficients obtained are used as the primitive features.

Natural images exhibit spatial variation of the texture. So, texture based retrieval of images assume that textures region may not be homogeneous over very large areas. A localized characterization of textures thus becomes necessary. Hence, the local energy for each pixel in the 16 sub-band images are computed. The Absolute Gaussian energy, for each pixel, is computed over a neighborhood, the size of which is determined using a spectral flatness measure (SFM).

$$energy_{m_1,m_2}(i,j) = \sum_{a=1}^{N}\sum_{b=1}^{N} |Wf_{m_1,m_2}(a,b)|\, G(i-a, j-b) \tag{15}$$

$$1 \le m_1 \le M, 1 \le m_2 \le M$$

where N is the neighborhood size while Wf_{m_1,m_2} is the wavelet transform coefficient obtained by row-wise convolution using the filter H_{m_1} and column-wise convolution with the filter H_{m_2}. The nonlinear transform is succeeded by a Gaussian low-pass (smoothing) filter of the form

$$G(x,y) = \frac{1}{\sqrt{2\Pi\sigma}} e^{-(\frac{1}{2\sigma^2})(x^2+y^2)} \tag{16}$$

where σ defines the spatial support of the averaging filter.

SFM gives a measure of the global frequency content of the image. It is defined as the ratio of arithmetic mean and the geometric mean of the Fourier coefficients of the image. It has been reported in literature that the size of the neighborhood for computation of localized energy range from 11×11 to 31×31, while SFM varies from 1 to 0 [7]. We use a neighborhood size of 11×11 for SFM between 1 and 0.65, 21×21 for SFM between 0.65 and 0.35 while 31×31 for 0 to 0.35. Since images generally are formed by regions, as in from objects and their surroundings, clustering or quantizing the wavelet energy reduces the feature size retaining maximum information. Based on this basic assumption, the energy values, for each sub-band and for each plane of the color image are used as the feature for a pixel and clustered using Fuzzy C-means.

Earth Movers Distance is used as the metric for similarity matching. Earth Movers distance uses a different type of signature over traditional histogram based similarity matching. Rubner *et al.* [23] has successfully shown that the EMD is an efficient metric for content based image retrieval with several advantages over other distance based similarity and dissimilarity measures. The feature vector comprises of the cluster centers location of the energy measurement based cluster over different sub-bands and the numbers of pixels in each cluster form the image signature.

4.2 Proposed Feature Evaluation Mechanism

The information obtained from the set of relevant and irrelevant images as marked by the users are used to automatically specify the weight of the component features. Weight computed are based on a measure defined as Feature

Evaluation Index (FEI). The FEI which automatically estimates the importance of an individual feature can be obtained by considering a pattern classification problem. Let $C_1, C_2, ..., C_m$ are m pattern classes in N dimensional features space where class C_j contains, n_j number of samples.

Let the features values along the q^{th} coordinate along classes C_j are assigned a fuzzy membership function between 0 and 1, using a standard S-type membership function [24]. Entropy(H) of C_j which gives the measure of intraset ambiguity is given by

$$H = (\frac{1}{n_j \ln 2}) \sum_i S_n(\mu_(f_{iqj})); i = 1, 2...n_j \tag{17}$$

where $S_n(\mu(f_{iqj}))$=-$\mu(f_{iqj})$ln$\mu(f_{iqj})$-{1-$\mu(f_{iqj})$}ln{1-$\mu(f_{iqj})$} is the Shannon's function. Entropy [24] is dependent on the absolute values of membership (μ). $H_{min} = 0$ for μ=0 or 1, $H_{max} = 1$ for μ=0.5 .

If the pattern classes C_j with n_j number of samples and C_k with n_k number of samples are combined together and the entropies are computed as follows : H_{qj} is the entropy of class C_j along q^{th} dimension over n_j number of sample and denotes intraset ambiguity. H_{qk} is the entropy of class C_k along q^{th} dimension over n_k number of samples. H_{qjk} then denotes the interset ambiguity along q^{th} dimension between classes C_j and C_k with $(n_j + n_k)$ number of samples. The (FEI_q) for the q^{th} component is defined as

$$(FEI_q) = \frac{H_{qjk}}{H_{qj} + H_{qk}} \tag{18}$$

The criteria of a good feature is that (FEI_q) should be decreasing after combining C_j and C_k as the goodness of the q^{th} features in discriminating pattern classes C_j and C_k increases. The weight w_q is a function of the evaluated (FEI_q) is

$$w_q = F_q(FEI_q) \tag{19}$$

Lower value of FEI_q, indicates better quality of importance of the q^{th} feature in recognizing and discriminating different classes. This approach is used in estimating the importance of each feature component. In conventional CBIR approaches an image I is usually represented by a set of features, $F = \{f_q\}_{q=1}^{N}$, where f_q is the q^{th} features component in the N dimensional feature space. Presently the number of classes are two of which one class constitute the relevant images $I_r = \{I_{r1}, I_{r2}, ..., I_{rm}\}$ and irrelevant images $I_{ir} = \{I_{ir1}$, $I_{ir2}, ..., I_{irm}\}$.

H_{qj} is computed from $I_r^{(q)}$= $\{I_{r1}^{(q)}$, $I_{r2}^{(q)}$, $I_{r3}^{(q)}$,...,$I_{rk}^{(q)}\}$. Similarly, H_{qk} is computed from the set of images where, $I_{ir}^{(q)}$ =$\{I_{ir1}^{(q)}$, $I_{ir2}^{(q)}$, $I_{ir3}^{(q)}$,...,$I_{irk}^{(q)}\}$. H_{qkj} is computed combining both the sets. Images are ranked according to similarity measures. The user marks the relevant and irrelevant set from 20 returned images, for automatic evaluation of FEI.

4.3 Fuzzy Relevance feedback (FRF) using EMD

To compute EMD over M-band wavelet features, Fuzzy C-Means clustering (FCM) is used to cluster the features and obtain the signature. The feature vector comprising of the cluster centers of the energy measurement over sub-bands, with the number of pixels of the image in each cluster comprises the image signature. To keep the computations minimum, FCM was preferred keeping the number of clusters generally ≤ 5. In the present case, it is 3, as it gives results upto the expectation at minimum cost of computation and grossly partition each image of the database into three meaningful clusters. Increasing the number of clusters may include finer details. As a result, the the uncertainties of characterizing the perceptual content may increase.

The results retrieved from the 1^{st} iteration are obtained by measuring EMD between the signature of the query image and the stored images in the database. A better retrieval is then obtained by using a weighted distance from user FRF at successive iterations. In this cases, weight of each component feature of different planes is determined from the feature evaluation mechanism. Perceptual importance as used in the JPEG 2000 is Y : Cb : Cr = 4 : 2 : 1. Here, the weights are chosen heuristically which is based on the convention "Human visual system is less sensitive to chrominance than luminance". However, an automatic scheme which chooses the weights depending on the color-texture complexity of the image will certainly boost the performance of the CBIR system.

In the experiment, an image I is represented in terms of a signature $P = \{(p_1, w_{p_1}), (p_2, w_{p_2}), ..., (p_m, w_{p_m})\} = \{(p_i, w_{p_i})\}_{i=1}^{m}$with m clusters. The cluster centroid p_i constitutes the wavelet features over 16 sub-bands of each Y, Cb and Cr plane and obtained with $p_i = [p_{i_Y}, p_{i_{Cb}}, p_{i_{Cr}}]$ which may also represented as $[f_{1Y}, ..., f_{nY}, f_{1Cb}, ..., f_{nCb}, f_{1Cr}, ..., f_{nCb}]$. Here, p_{i_Y}, $p_{i_{Cb}}$ and $p_{i_{Cr}}$ are the local energy values computed overall sub-bands of each Y, Cb and Cr planes respectively, for e.g $p_{i_Y} = [f_{1Y}, ..., f_{nY}]$ where $p_i \epsilon R^N$ and $N = 3n$ is the feature dimension. Here $n = 16$ and $w_i \geq 0$, for each plane.

From the set of marked images (I_r and I_{ir}), the weight of the features computed over each Y, Cb and Cr are estimated as follows: For each cluster p_i the wavelet based sub-band features of each of Y, Cb and Cr planes are considered. The features considered to compute the feature evaluation index along each component plane e.g. plane Y are $F_Y = \{f_{i_{1Y}}, ..., f_{i_{qY}}\}$, where $i = 1, 2, ..., m$ (m clusters) and $q = 1, 2, ..., n$ (n sub-band features). Similarly features are considered for Cb and Cr plane. The $(FEI)_{qY}$ for the component feature q of Y plane are

$$(FEI)_{qY} = \frac{H_{qtotal}}{H_{qRel} + H_{qIrrel}}. \tag{20}$$

H_{qRel}, H_{qIrrel} and H_{qtotal} are the entropies along the q^{th} dimension of relevant, irrelevant and total returned images respectively. And the $(FEI)_{qCb}$ and $(FEI)_{qCr}$ of other two planes are computed similarly.

The overall weight factor for the Y plane is given by

$$W'_Y = \sum_{i=1}^{m}\sum_{q=1}^{n}(FEI)_{qY} \tag{21}$$

Similarly, the overall weight factor W'_{Cb} and W'_{Cr} for Cb and Cr planes are computed respectively.

A normalization process is used to ensure proper emphasize in each plane even if their features values are of different dynamic ranges. The relative weight factor for the Y plane is as

$$W_Y = \frac{W'_Y}{W'_Y + W'_{Cb} + W'_{Cr}} \tag{22}$$

Using similar formula, W_{Cb} and W_{Cr} are computed. The weights W_Y, W_{Cb} and W_{Cr} reflects user's different importance on the representative feature map for computing overall similarity between images. Multiplying with the weights, actually modify the ground distance $d(p_i, q_j)$ of equation 14, but keep the distribution, *i.e.*, the number of pixels in each cluster unchanged and the total weight remains the same. So, the weighted EMD, *i.e.*, $\text{EMD}_g(P, Q)$, and $g\epsilon(W_Y, W_{Cb}, W_{Cr})$ is computed from the work flow:

$$WORK(F, P^{'}, Q^{'}) = \sum_{i=1}^{m}\sum f_{i,j}d(g(p_i), g(q_j)) \tag{23}$$

where, the centroids p_i and q_j are transformed to $p'_i = [W_Y p_{i_Y}, W_{Cb}p_{i_{Cb}}, W_{Cr}p_{i_{Cr}}]$ and $q'_j = [W_Y q_{j_Y}, W_{Cb}q_{j_{Cb}}, W_{Cr}q_{j_{Cr}}]$ respectively, with the weight updation factor g. The EMD is computed upto k^{th} iteration till it converges, *i.e.*, $W(F^{(K+1)}, P_K^{'(K+1)}, Q_K^{'(K+1)}) \leq W(F^{(K)}, P_K^{'(K)}, Q_K^{'(K)})$. As it is assumed that similar images will have nearly same signature. After multiplying with weights it generate a sequence of work flows $W(F^k, P_k^{'(k)}, Q_k^{'k})$, which is expected to vary in a similar fashion for similar images. As a result, the ranks of the relevant images are not affected much.

In case of irrelevant images, although the signature map may be different, the EMD distance as obtained from the work flows is nearly equal. After FRF, the ground distance (d_{ij}) is modified by the weighting factor W where $W\epsilon[W_Y, W_{Cb}, W_{Cr}]$. The search space that varies over each plane becomes elliptic for unequal weights. As a result, irrelevant images are discarded to a large extent and more relevant images are included due to FRF mechanism because the ground distance varies in accordance to the importance of the component planes.

4.4 Fuzzy Relevance feedback (FRF) using different similarity measures

For relative comparison of performance with different distance measure, results are also provided here using similarity measures namely Euclidean Distance

(ED), Manhattan Distance (MD) and Chessboard Distance (CBD) [39]. To compute these distances using the same set of M-band wavelet features, FCM is used to cluster the features. Thus, M-band wavelet features vector comprising centroids of the query signature ($\sum_{i=1}^{m} p_i$) and the stored images ($\sum_{j=1}^{m} q_j$) are considered.
(a)Euclidean Distance.

$$d_{ED} = \sum_{l=1}^{N} \parallel (p)_l - (q)_l \parallel, \tag{24}$$

(b)Manhattan Distance.

$$d_{MD} = \sum_{l=1}^{N} |(p)_l - (q)_l|, \tag{25}$$

(c)Chessboard Distance.

$$d_{CBD} = \max_{1 \leq l < N} \{|(p)_l - (q)_l|\} \tag{26}$$

where, N is the total number of features by considering energy measurement over sub-band of each component plane as explained in section.4.3.

The results generated from the 1^{st} iteration are obtained by measuring above discussed similarity measures (ED, MD and CBD) between the M-band wavelet features of the query image and the stored images in the database. Top 20 ranked images are displayed. From the returned images of each iteration, the weight specifying the user's importance is computed in a similar fashion as explained in section.4.3. And the weighted similarity measures is
(a)Euclidean Distance.

$$d_{WED} = \sum_{l=1}^{N} \parallel g(p)_l - g(q)_l \parallel, \tag{27}$$

(b)Manhattan Distance.

$$d_{WMD} = \sum_{l=1}^{N} |g(p)_l - g(q)_l|, \tag{28}$$

(c)Chessboard Distance.

$$d_{WCBD} = \max_{1 \leq l < N} \{|g(p)_l - g(q)_l|\} \tag{29}$$

where, g is $g\epsilon(W_Y, W_{Cb}, W_{Cr})$ is used to updated the corresponding features of each components plane.

With each iteration the weight is modified by using the FEI with user feedback and similarity measures (ED, MD, CBD) is calculated on this modified feature vector upto k^{th} iteration till it converges.

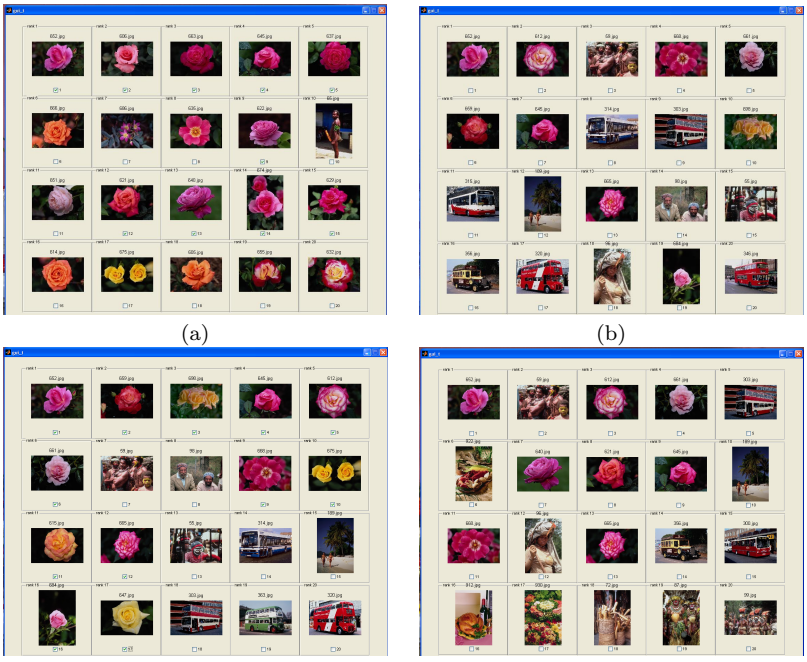
5 Experimental Results

To demonstrate the effectiveness of the proposed CBIR system, two standard image databases were used in the experiments. The first database (SIMPLIcity database (SimDB)) consists of 1000 images of 10 different categories of images, and the second database (Corel database (CorelDB)) consists of 10000 images of 100 different categories image's. All the experiment were done on a Dell machine (DELL PRECISION T7400, 4GB RAM) and programs were written in MATLAB. In all given figures of the experimental results the top left corner images are the query images used in the experiment. The quantitative measures used in the experiments were average precision(%) and average recall(%), and are as follows:

$$Precision(\%) = (\frac{A}{B}) * 100 \tag{30}$$

$$Recall(\%) = (\frac{A}{C}) * 100 \tag{31}$$

where A = Number of relevant images retrieved, B = Total Number of images retrieved and C = Total Number of relevant images in the database.



(a) (b) (c) (d)

Fig. 4 Retrieval results on SimDB using (Top left side image as the query image)(a) M-band wavelet features with EMD (19/20)(b) M-band wavelet features with ED (9/20) (c) M-band wavelet features with MD (12/20)(d) M-band wavelet features with CBD (8/20).

Fig.4(a) shows the effectiveness of the proposed scheme using M-band wavelet feature set and EMD. The performance of the proposed M-band wavelet based CBIR system using other distances (ED, MD and CBD)is shown in Fig.4(b)-(d). The values (e.g.(19/20),(9/20) etc.) given in the caption of the figures indicate the number of correctly retrieved images out of 20 images per frame. The graph of the Fig.5 shows the performance comparison of EMD against other similarity measures (ED, MD, CBD) using M-band wavelet as the features in all the cases. It is clear from the graph of the Fig.5, as well as from the given instances of the retrieval results of Fig.4(a) - (d), that M-band wavelet features with EMD performs better than M-band wavelet features with other distances, even though the CBIR using EMD method is computationally more complex than the CBIR based on other similarity measures and M-band wavelet features.

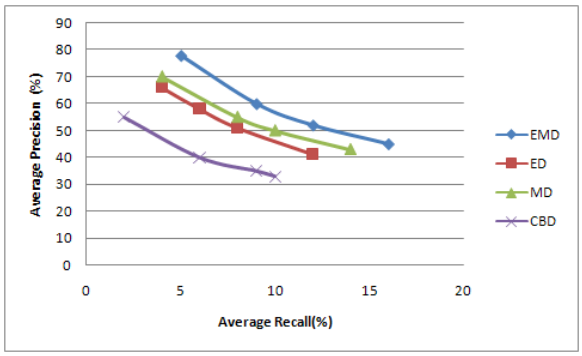
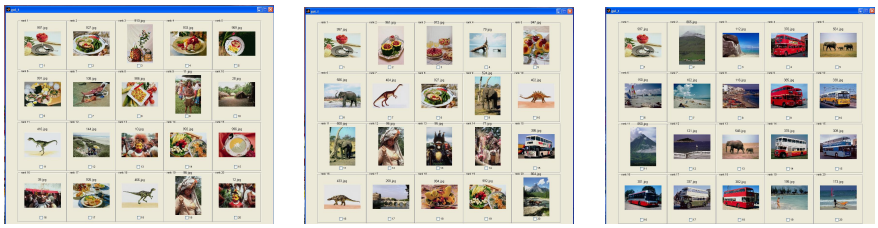


Fig. 5 Performance comparison graph of M-band wavelet features with EMD Versus ED, MD, CBD.

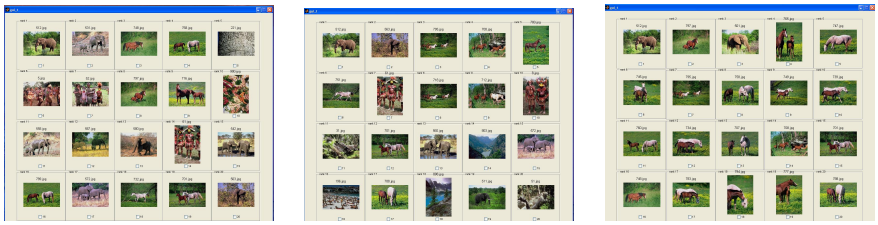
We have compared the performance effectiveness of M-band wavelet feature set with two widely used ISO MPEG-7 visual features, CSD and EHD. CSD represents an image by considering both color distribution (color histogram) and the local spatial structure of the color. An 8×8 structuring block is used for color structure information of the descriptor. EHD captures global spatial distribution of the edges, by dividing the image into 16 sub-images, with fixed number of blocks. Edge information is then calculated for each block in five edge categories: Vertical, Horizontal, 45° diagonal, 135° diagonal and non directional edge. It is expressed as a 5 bin histogram for each image block. CSD and EHD consist of 256 and 80 features respectively. As CSD consists of 256 number of features, which is much larger than M-band(144) and EHD(80), it is expected that CSD should normally performs better than M-band and EHD.

The results of the performance comparisons are shown in Fig.6 and Fig.7, respectively. ED is used as the similarity measure in the comparisons, as it is the most widely used similarity measure by most of the existing CBIR system based on CSD and /or EHD. The results of the Fig.6 describe the performance



(a) (b) (c)

Fig. 6 Retrieval results using ED on SimDB in average cases using (Top left side image as the query image)(a) CSD features (10/20) (b) M-band wavelet features (7/20) (c) EHD features (1/20).

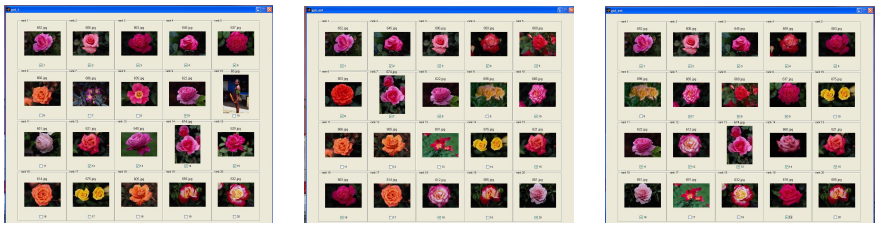


(a) (b) (c)

Fig. 7 Retrieval results using ED on SimDB in case of edge prominent query image using (Top left side image as the query image) (a) EHD features (8/20) (b) M-band wavelet features (5/20) (c) CSD features (2/20).

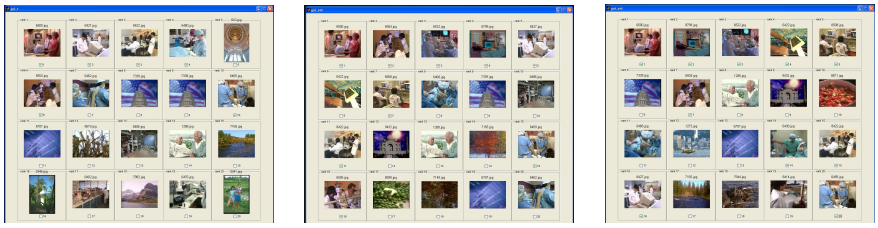
comparisons in average cases, where CSD performs better than M-band and EHD. Similarly, the results of the performance comparisons in case of edge prominency is shown in Fig.7. It is clear from the Fig.6 that, in average case, the M-band wavelet feature set performs better than EHD but poorer than CSD, incase of color prominency. The reverse is true for EHD, where M-band wavelet feature set performs better than CSD but poorer than EHD, incase of edge prominency. M-band wavelet feature set consists of 144 features, which is less than CSD (256) but larger than EHD (80). From the discussion above, it is obvious that even though M-band wavelet feature set is more computationally complex than EHD but less computationally complex than CSD, it has given comparable performances in both the cases.

To improve the performance of the proposed CBIR system, we have used FRF mechanism. Fig.8 shows the improvement in the retrieval result after 1^{st} and 2^{nd} iteration using weighted EMD as the similarity measure on SimDB. It is clear from the Fig.8 that the performance of the proposed CBIR system has improved with every iterations of FRF mechanism, not only in the total number of correctly retrieved images, but also in image ranking. The ranking of the images have improved with every iterations of FRF mechanism as can be seen from Fig.8. It has been observed that normally after 2 to 3 iterations of the FRF mechanism, the retrieval results converge.



(a) (b) (c)

Fig. 8 Results on SimDB using FRF mechanism (Top left side image as the query image) (a) First pass of the retrieval set Using M-band wavelet features and EMD (19/20) (b) First Iteration with weighted EMD (20/20, improved ranking)(c) Second Iteration with weighted EMD (20/20, improved ranking).



(a) (b) (c)

Fig. 9 Results on CorelDB using FRF mechanism (Top left side image as the query image) (a) First pass of the retrieval set Using M-band wavelet features and EMD (6/20)(b) First Iteration with weighted EMD (9/20, improved ranking)(c) Second Iteration with weighted EMD (10/20, ranking of the relevant images as compared to the query image is improved).

The scalability performance effectiveness of the proposed CBIR system based on M-band wavelet feature set using weighted EMD as the similarity measure and with FRF mechanism, is also evaluated using a bigger sized image database, *i.e.*, CorelDB. The results are shown in Fig.9. It is obvious from the results given in Fig.9 that the proposed CBIR system works well on databases containing huge number of images. The CPU-time taken for each iteration is approximately 50 ms for SimDB and 3 sec for CorelDB.

The graph of the Fig.10, shows the effectiveness of the proposed FRF using M-band wavelet feature set with EMD against other similarity measures. It is obvious from the graph that the proposed CBIR system using M-band wavelet feature set with FRF and EMD performs better than the CBIR system using M-band wavelet with FRF and other distances such as ED, MD and CBD.

6 Conclusion

In this paper we have proposed a novel CBIR system based on M-band wavelet feature set using weighted EMD as the similarity measure. To improve the retrieval result we have used an interactive relevance feedback mechanism based on fuzzy feature evaluation procedure. The performance of the proposed CBIR

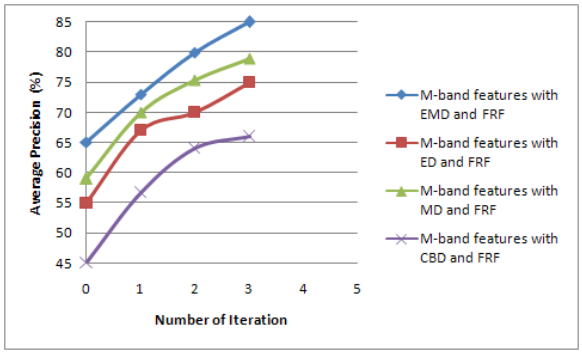


Fig. 10 Performance comparisons on SimDB of FRF mechanisms with M-band wavelet features with EMD similarity measures Versus ED, MD and CBD.

system shows that, due to its simple structure and low computational time requirement it is well suited for real life application paradigm like the internet. In future, we intend to incorporate partial query in our proposed CBIR system and make comparatives studies against MPEG-7 standards. The effectiveness of the proposed feedback mechanism could be tested for retrieval of videos in conjunction with motion information as future scope of research.

References

1. M. K. Kundu and M. Banerjee and P. Bagrecha, An Interactive Image Retrieval in a Fuzzy Framework, in Proceedings 8th International Workshop on fuzzy logic and Application, Lect. Notes in Artificial Intelligence, Vol. 5571, pp. 246–253 (Springer-Verlag, 2009)
2. X. He and O. King and W. Ma and M. Li and H. J.Zhang, Learning a Semantic space from user's Relevance Feedback for Image retrieval, IEEE Trans. on Circuits and Systems for Video technology, 13, pp. 39–48 (2003).
3. Z. Jin and I. King and X. Q. Li, Content-based image retrieval by relevance feedback, in Advances in Visual Information Systems, Lect. Notes in Computer Science., Vol. 1929, pp. 639–648 (Springer, 2000).
4. E.D.Ves and J. Domingo and G. Ayala and P. Zuccarello, A novel Bayesian framework for relevance feedback in image content-based retrieval systems, Pattern Recognition, 39, pp. 1622–1632 (2006).
5. F. Qian and B. Zhang and F. Lin, Constructive learning algorithm-based RBF network for relevance feedback in image retrieval, in Proceedings of the 2nd international conference on Image and video retrieval, Lect. Notes in Computer Science, Vol. 2728, pp. 352–361 (Springer-Verlag, 2003).
6. Q. Cheng and C. Yang and F. Chen and Z. Shao, Application of M-Band Wavelet Theory to Texture Analysis in Content-Based Aerial Image Retrieval International Geoscience and Remote Sensing Symposium, Vol. 3, pp.2163–2165 (2004).
7. M. Acharyya and M. K. Kundu, An adaptive approach to unsupervised texture segmentation using M-band wavelet tranform, Signal Processing, 81, pp. 1337–1356 (2001).
8. J. Z. Wang and J. Li and G. Wiederhold, SIMPLIcity: Semantics-sensitive integrated matching for picture libraries, IEEE Trans. on Pattern Analysis and Machine Intelligence, 23, pp. 947–963 (2001).
9. B. S. Manjunath and P. Salembier and T. Sikora, Introduction to MPEG-7: Multimedia Content Description Interface, John Wiley and Sons Inc, (2002).
10. B. S. Manjunath and J. R. Ohm and V. V. Vasudevan, Color and Texture Descriptors , IEEE Trans. on Circuits and Systems for Video technology, 11, pp. 703–715 (2001).

11. P. Wu and B. S. Manjunath and S. Newsam and H. D. Shin , A texture descriptor for browsing and similarity retrieval , Signal Processing: Image Communication, 16, pp. 33–43 (2000).
12. Remco C. Veltkamp and Longin Jan Latecki, Properties and Performance of Shape Similarity Measures, in Proceedings of IFCS 2006 Conference: Data Science and Classification, pp. 1–9 (Springer-Verlag, 2006).
13. Peng-Yeng Yin and Bir Bhanu and Kuang-Cheng Chang and Anlei Dong, Integrating Relevance Feedback Techniques for Image Retrieval Using Reinforcement Learning ,IEEE Transaction Pattern Analysis Machine Intelligence, 27, pp. 1536–1551 (2005).
14. J.Han and K.N.Ngan and M.Li and H.J.Zhang, A memory learning framework for effective image retrieval , IEEE Transaction on Image Processing, 14, pp. 521–524 (2005).
15. F. C. Chang and H. M. Hang, A Relevance Feedback Image Retrieval Scheme Using Multi-Instance and Pseudo Image Concepts , IEICE Transaction Information System, E89-D, pp. 521–524 (2006) .
16. Y. Chen and J. Z. Wang and R. Krovet, Cluster based retrieval of images by unsupervised learning , IEEE transactions on Image Processing, 14, pp. 1187–1201 (2005).
17. H.B.Kekre and S. D. Thepade and A. Maloo, Image Retrieval using Fractional Coefficients of Transformed Image using DCT and Walsh Transform , International Journal of Engineering Science and Technology, 2, pp. 362–371 (2010).
18. D. Heesch, A survey of browsing models for content based image retrieval, Multimedia Tools Application, 40, pp. 1380–7501 (2008).
19. A. W. M. Smeulders and M. Worring and S. Santini and A. Gupta and R. Jain, Content based image retrieval at the end of early years , IEEE transactions on Pattern Analaysis and Machine Intelligence, 22, pp. 1349–1380 (2000).
20. W.Xiaoling and M. Hongyan, Enhancing Color histogram for Image Retrieval, in Proceedings of the International Workshop on Information Security and Application, (Academy Publisher, 2009).
21. C. S. Burrus and A. Gopinath and H. Guo, Introduction to Wavelets and Wavelet Transform: A Primer, Prentice Hall International Editions (1998).
22. Y. Rubner and C. Tomasi, Perceptual Metrices for Image Database Navigation, Kluwer Academic Publishers (2001).
23. Y. Rubner and C. Tomasi and L. J. Guibas, The Earth Mover's Distance as a Metric for Image Retrieval , Interational Journal of Computer Vision, 40, pp. 99–121 (2000).
24. S. K. Pal and D. D. Majumder, Fuzzy Mathematical Approach To Pattern Recognition, Willey Eastern Limited (1985).
25. R. Datta, D. Joshi, J.Li and J. Z. Wang,, Image Retrieval: Ideas, Influences, and Trends of the New Age, in ACM Computing Surveys, 40, pp. 1–60 (2008).
26. Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra,Content-based image retrieval with relevance feedback in mars, in Proceedings of IEEE International Conference on Image Processing, 2, pp. 815–818 (1997).
27. S. Santini and R. Jain, Similarity Measures, IEEE transactions on Pattern Analaysis and Machine Intelligence, 21, pp. 871-883 (1999).
28. Z. Shi, Q. He and Z. Shi, Bayes-Based Relevance Feedback Method for CBIR, Advances in Soft Computing, 44, pp.264-271 (2007).
29. A. Marakakis, N. Galatsanos, A. Likas and A. Stafylopatis, Probabilistic relevance feedback approach for content-based image retrieval based on gaussian mixture models, IET Image Processing, 3, pp. 10–15 (2009).
30. T. M. Cover and J. A. Thomas, Elements of Information Theory, Willey Eastern Limited (1991).
31. Woo-Cheol Kim, Ji-Young Song, Seung-Woo Kim, S. Park, Image retrieval model based on weighted visual features determined by relevance feedback, Information Sciences, 178, pp. 4301-4313 (2008).
32. G. Ohashi and Y. Shimodaira, Edge-Based Feature Extraction Method and Its Application to Image Retrieval, Journal of Systemics, Cybernetics and Informatics, 1, pp. 25-28 (2003).
33. Georgy L., G. Farb and Anil K. Jain, On retrieving textured images from an image database, Pattern Recognition, 29, pp. 1461–1483 (1996).

34. R. Ksantini, D. Ziou and F. Dubeau, Image Retrieval based on region separation and multiresolution analysis, International Journal of Wavelets, Multiresolution and Information Processing, 4, pp. 147–175 (2006).
35. M. Banerjee, M. K. Kundu and P. Maji, Content based image retrieval using visually significant point features, Fuzzy Sets and Systems, 160, pp. 3323–3341 (2009).
36. M. K. Kundu and M. Acharyya and , M-band wavelet: Application to texture segmentation for real life image analysis, International Journal of Wavelets, Multiresolution and Information Processing, 1, pp. 115–149 (2003).
37. M. Acharyya and M. K. Kundu, Extraction of noise tolerant, gray-scale transform and rotation invariant features for texture segmentation using wavelet frames, International Journal of Wavelets, Multiresolution and Information Processing, 6, pp. 391–417 (2008).
38. M. Acharyya, R. K. De and M. K. Kundu, Extraction of features using M- band wavelet packet frames and their neuro-fuzzy evaluation for multi-texture segmentation, IEEE Transaction on Pattern Analysis and Machine Intelligence, 25, pp. 1639–1644 (2003).
39. H. Liu, D. Song, S. Rger, R. Hu, and V. Uren, Comparing dissimilarity measures for content-based image retrieval, In: The 4th Asia Information Retrieval Symposium (AIRS2008), pp. 44-50 (Springer-Verlag 2008)