Electronic Letters on Computer Vision and Image Analysis 11(1):1-13, 2012

Novel CBIR System Based on Ripplet Transform Using Interactive Neuro-Fuzzy Technique

Manish Chowdhury, Sudeb Das and Malay Kumar Kundu

Machine Intelligence Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata-700108, India
Received 27 June 2011; accepted 24 January 2012

Abstract

Content Based Image Retrieval (CBIR) system is an emerging research area in effective digital data management and retrieval paradigm. In this article, a novel CBIR system based on a new Multiscale Geometric Analysis (MGA)-tool, called Ripplet Transform Type-I (RT) is presented. To improve the retrieval result and to reduce the computational complexity, the proposed scheme utilizes a Neural Network (NN) based classifier for image pre-classification, similarity matching using Manhattan distance measure and relevance feedback mechanism (RFM) using fuzzy entropy based feature evaluation technique. Extensive experiments were carried out to evaluate the effectiveness of the proposed technique. The performance of the proposed CBIR system is evaluated using a 2×5-fold cross validation followed by a statistical analysis. The experimental results suggest that the proposed system based on RT, performs better than many existing CBIR schemes based on other transforms, and the difference is statistically significant.

Key Words: Ripplet Transform, Relevance Feedback, Content Based Image Retrieval, Artificial Neural Network, Multilayer Perceptron, Fuzzy.

1 Introduction

Over the last few decades, the tremendous growth in quantity of images stored in digital form, have raised many demanding issues in digital data management and retrieval paradigm. Recently, Content Based Image Retrieval (CBIR) system has emerged as an effective solution to address some of these issues efficiently. CBIR is the task of retrieving relevant images from a large image database (DB) by measuring similarities between the query image and database images, based on automatically derived features like color, texture, shape etc. The performance of a CBIR system strongly depends both on the availability of suitable features for proper representation of the semantic aspects of the images automatically, and also the effectiveness of the used similarity measure [1]. High retrieval efficiency and less computational complexity are the desired characteristics of an effective CBIR system.

Wavelet transform (WT) based low level features, provide a unique representation of the image, and are highly suitable for characterizing textures of the image [2]. Wang *et al.* used a $2-$step algorithm using subband variances of wavelet coefficients for image retrieval application [3]. WT is inherently non-supportive to directionality and anisotropy. To overcome the limitations of WT, recently a theory called Multi-scale Geometric

Correspondence to: <st.manishc@gmail.com>

Recommended for acceptance by <Umapada Pal>
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

Analysis (MGA) for high-dimensional signals has been developed, and several MGA tools were proposed like Ridgelet, Curvelet, Bandlet and Contourlet etc.[4, 5, 6].

To improve the retrieval accuracy and to decrease the computational complexity, many machine learning techniques have been used in the CBIR paradigm. Yang *et al.* proposed a method based on NN backpropagation model to distinguish young corn plants from weeds by using color feature of the image as inputs [7]. Pourghassem *et al.* applied Multilayer Perceptron (MLP) for hierarchical medical image classification [8]. To handle the high dimensionality of image database, image classification systems usually rely on a pre-processing step to reduce the computational cost with increase in retrieval accuracy.

Relevance Feedback Mechanism (RFM) has been used as an effective tool to provide significant performance boost in CBIR systems through continuous learning and interaction with the end-users. Most of the RFMs, employ two approaches namely, query vector moving technique and feature re-weighting technique to improve the retrieval results [9]. Feature re-weighting technique uses both the relevant and the irrelevant information, to obtain more effective results [10, 11]. But in all these cases, time complexity per iteration are high and accuracy of the relevant images are low.

The purpose of this paper is to investigate the effectiveness of ripplet transform in CBIR domain. To our knowledge this is the first time ripplet transform is used in CBIR application. The contributions of this presented paper are as follows: (1) using ripplet transform for the first time in the CBIR field, (2) cutting down the computational cost using a MLP based NN, and (3) improving the retrieval performance by modeling the ambiguity and uncertainty associated with user response, using fuzzy feature evaluation index. The proposed CBIR system consists of three phases. In the 1^{st} phase, RT is used to extract the low level features of the images in the DB. During the 2^{nd} phase, a MLP based NN is designed to pre-classify the whole DB. The designed NN is used for correct class identification of the query image. After the classification, Manhattan similarity measure is used to retrieve and display the top 20 ranked images from the particular partitioned DB. The results are improved further with fuzzy based RFM in the 3^{rd} phase of the proposed system

The rest of the paper is organized as follows. RT is described in section 2. Section 3, presents the proposed CBIR system. Experimental results and comparisons are given in section 4, and the conclusion is drawn in section 5.

2 Ripplet Transform Type-I (RT)

Conventional transforms like FT and WT suffer from discontinuities such as edges and contours in images. To address this problem, Jun Xu *et al.* proposed a new MGA-tool called RT [12]. RT is a higher dimensional generalization of the Curvelet Transform (CVT), capable of representing images or $2D$ signals at different scales and different directions. To achieve anisotropic directionality, CVT uses a parabolic scaling law. From the perspective of microlocal analysis, the anisotropic property of CVT guarantees resolving $2D$ singularities along C^2 curves. On the other hand, RT provides a new tight frame with sparse representation for images with discontinuities along C^d curves [12].

There are two questions regarding the scaling law used in CVT: 1) Is the parabolic scaling law optimal for all types of boundaries? and if not, 2) What scaling law will be optimal? To address these questions, Jun Xu *et al.* intended to generalize the scaling law, which resulted in RT. RT generalizes CVT by adding two parameters, i.e., support c and degree d. CVT is just a special case of RT with $c = 1$ and $d = 2$. The anisotropy capability of representing singularities along arbitrarily shaped curves of RT, is due to these new parameters c and d.

2.1 Continuous Ripplet Transform (CRT)

For a 2D integrable function $f(\overrightarrow{x})$, the CRT is defined as the inner product of $f(\overrightarrow{x})$ and ripplets $\rho_{a\,\overrightarrow{b}\,\theta}(\overrightarrow{x})$ as given below [12]

$$R(a, \overrightarrow{b}, \theta) = \langle f, \rho_{a\,\overrightarrow{b}\,\theta} \rangle = \int f(\overrightarrow{x}) \overline{\rho_{a\,\overrightarrow{b}\,\theta}(\overrightarrow{x})} d\overrightarrow{x} \tag{1}$$

where $R(a, \overrightarrow{b}, \theta)$ are the ripplet coefficients and $\overline{(.)}$ denotes the conjugate operator. The ripplet function of the equation (1) is defined as

$$\rho_{a\,\overrightarrow{b}\,\theta}(\overrightarrow{x}) = \rho_{a\,\overrightarrow{0}\,0}(R_\theta(\overrightarrow{x} - \overrightarrow{b})) \tag{2}$$

where $\rho_{a\,\overrightarrow{0}\,0}(\overrightarrow{x})$ is the ripplet element function, $R_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ is the rotation matrix, $\overrightarrow{x}$ and $\overrightarrow{b}$ are $2D$ vectors; b and θ denotes the position parameter and rotation parameter respectively. The element ripplet function is defined in frequency domain as

$$\widehat{\rho}_a(r, \omega) = \frac{1}{\sqrt{c}} a^{\frac{1+d}{2d}} W(a \cdot r) V(\frac{a^{\frac{1}{d}}}{c \cdot a} \cdot \omega) \tag{3}$$

where $\widehat{\rho}_a(r, \omega)$ are the Fourier Transform (FT) of $\rho_{a\,\overrightarrow{0}\,0}(\overrightarrow{x})$ in polar coordinate system, and a is the scale parameter. $W(r)$ is the "*radial-window*"and $V(\omega)$ is the "*angular-window*". These two windows have compact supports on $[\frac{1}{2}, 2]$ and $[-1, 1]$ respectively. They satisfy the following admissibility conditions:

$$\int_{\frac{1}{2}}^{2} W^2(r) \frac{dr}{r} = 1 \tag{4}$$

$$\int_{-1}^{1} V^2(t) dt = 1 \tag{5}$$

These two windows partition the polar frequency domain into 'wedges' as shown in Fig. 1(a).

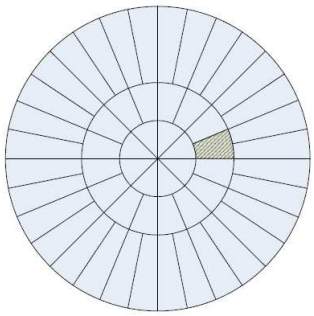

Figure 1: (a) The tiling of polar frequency domain. The shadowed 'wedge' corresponds to the frequency transform of the element function

The ripplets functions decay very fast outside the elliptical effective region, which has the following property for its length and width: $width \approx length^d$. Length and width is the major and minor axis of the ellipse respectively. The customizable effective region tuned by support c and degree d bespeaks the most distinctive property of ripplets − the general scaling. For $c = 1$, $d = 1$, both axis directions are scaled in the same way. So ripplet with $d = 1$ will not have the anisotropic behavior. For $d > 1$, the anisotropic property is reserved for RT. For $d = 2$, ripplets have parabolic scaling; for $d = 3$, ripplets have cubic scaling; and so forth.

The CRT can only capture the characteristics of high frequency components of $f(\overrightarrow{x})$, since the scale parameter a cannot take the value of infinity. So the 'full' CRT consists of fine scale RT and coarse scale isotropic WT. We can perfectly reconstruct the input function based on its ripplet coefficients.

2.2 Discrete Ripplet Transform (DRT)

As digital image processing needs discrete transform instead of continuous transform, here we describe the discretization of RT [12]. The discretization of CRT is based on the discretization of the parameters of ripplet functions. a is sampled at dyadic intervals. b and θ are sampled at equal-spaced intervals. a_j, $\overrightarrow{b}_k$ and θ_l substitute a, $\overrightarrow{b}$ and θ respectively, and satisfy that $a_j = 2^{-j}$, $\overrightarrow{b}_k = [c \cdot 2^{-j} \cdot k_1, 2^{-j/d} \cdot k_2]^T$ and $\theta_l = \frac{2\Pi}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot l$, where $\overrightarrow{k} = [k_1, k_2]^T$, and j, k_1, k_2, $l \in \mathbb{Z}$. $(\cdot)^T$ denotes the transpose of a vector. $d \in \mathbb{R}$, since any real number can be approximated by rational numbers, so we can represent d with $d = n/m$, $n, m \neq 0 \in \mathbb{Z}$. Usually, we prefer $n, m \in \mathbf{N}$ and n, m are both primes. In the frequency domain, the corresponding frequency response of ripplet function is in the form

$$\widehat{\rho}_j(r,\omega) = \frac{1}{\sqrt{c}} a^{\frac{m+n}{2n}} W(2^{-j} \cdot r) V(\frac{1}{c} \cdot 2^{-\lfloor j \frac{m-n}{n} \rfloor} \cdot \omega - l) \tag{6}$$

where W and V satisfy the following admissibility conditions:

$$\sum_{j=0}^{+\infty} |W(2^{-j} \cdot r)|^2 = 1 \tag{7}$$

$$\sum_{l=-\infty}^{+\infty} |V(\frac{1}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot \omega - l)|^2 = 1 \tag{8}$$

given c,d and j.

The 'wedge' corresponding to the ripplet function in the frequency domain is

$$H_{j,l}(r,\theta) = \{2^j \leq |r| \leq 2^{2j}, |\theta - \frac{\pi}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot l| \leq \frac{\pi}{2} 2^{-j}\} \tag{9}$$

The DRT of an $M \times N$ image $f(n_1, n_2)$ will be in the form of

$$R_{j,\overrightarrow{k},l} = \sum_{n_1=0}^{M-1} \sum_{n_2=0}^{N-1} f(n_1, n_2) \overline{\rho_{j,\overrightarrow{k},l}(n_1, n_2)} \tag{10}$$

where $R_{j,\overrightarrow{k},l}$ are the ripplet coefficients.

The image can be reconstructed through Inverse Discrete Ripplet Transform (IDRT):

$$\widetilde{f}(n_1, n_2) = \sum_j \sum_{\overrightarrow{k}} \sum_l R_{j,\overrightarrow{k},l} \rho_{j,\overrightarrow{k},l}(n_1, n_2) \tag{11}$$

As a generalized version of the existing curvelet transform, RT is not only capable of resolving 2D singularities, but it also has some nice properties:

1. It forms a new tight frame in a function space. Having good capability of localization in both spatial and frequency domain, it provides a more efficient and effective representation for images or 2D signals.

2. It has general scaling with arbitrary degree and support, which can capture 2D singularities along different curves in any directions.

Jun Xu *et al.* have showed that RT can provide a more effective representation for images with singularities along smooth curves [12]. It outperforms DCT and wavelet in nonlinear approximation, when the number of retained coefficients is small. RT can achieve roughly 2 dB higher PSNR on average than JPEG, and provide better visual quality than JPEG2000 at low bit-rates, when applied to image compression. In case of image denoising application, RT performs better than curvelet and DWT. All these experiments show that RT based image coding is suitable for representing texture or edges in images.

To our knowledge this is the first time Ripplet transform has been used for CBIR purpose. We have tried to investigate the effectiveness of the Ripplet transform in CBIR application. From the above mentioned discussion, we expect that RT may provide better result in CBIR domain.

3 Proposed Technique

In the proposed method, 70% data are randomly selected from each class of the images present in the DB as the labeled samples. Then their features based on RT are extracted for the training of a MLP network. The exact configuration of the MLP is selected on trial and error basis, based on the error rate of the output classification. The configuration having minimum error rate is selected as the used pre-classification block.

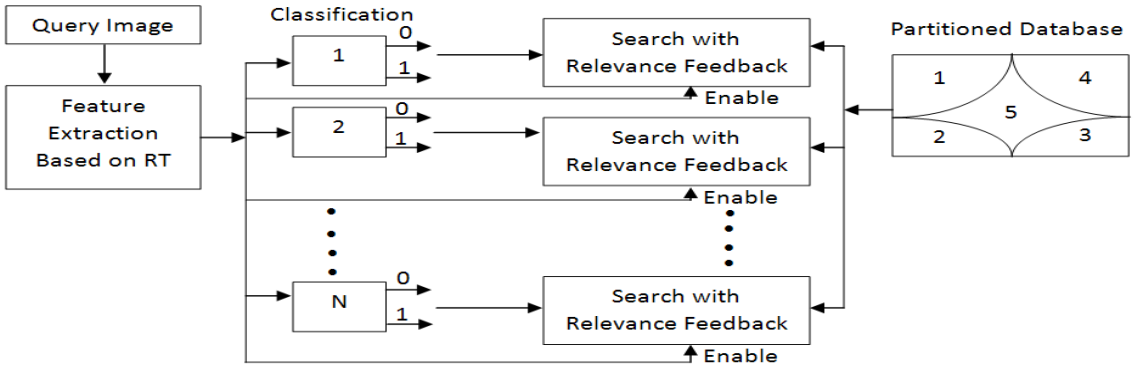

Figure 2: Block Diagram of Image Retrieval System

Block diagram of the proposed CBIR system is shown in Fig. 2. The user enters the query image to the system. RT based features are extracted for this particular query image and by using these features, class identification of the query image is done by the selected MLP based pre-classification block. After classification of the query image, the system computes the similarities between the query image and all the images in the partitioned database by using the Manhattan distance. The system retrieves the top 20 ranked images from the partitioned database and presents them to the user. The user marks the images returned by the search engine as *relevant* or *irrelevant* samples. A fuzzy based relevance feedback algorithm use this feedback information to select a set of better 20 images from the partitioned DB in the next iteration [10]. This retrieval process finishes at a point when the user is satisfied with the retrieved result.

3.1 Feature Extraction using Ripplet Transform Type-I (Phase-I)

Efficient representation of images usually leads to improvements in storage, computational complexity and performance of image processing algorithms. Efficient representation of images can be achieved by transforms. However, conventional transforms such as Fourier transform and wavelet transform suffer from discontinuities such as edges in images. To address these problems recently a new MGA-tool called ripplet transform is introduced [12]. RT is a higher dimensional generalization of the curvelet transform, designed to represent images or two-dimensional signals at different scales and different directions. Specifically, the RT allows arbitrary support c and degree d. These facts motivated us to use RT, which essentially satisfies the multiresolution, good localization, high directionality, general scaling and support, anisotropy and fast coefficient decay properties.

The images in the DB prior to RT based decomposition, are transformed from RGB to YCbCr color space. This ensures that the textural characterizations of the image are independent of the color characterization. RT decomposition over the intensity plane (Y) characterizes the texture information, while the RT decomposition over chromaticity planes (Cb and Cr) characterizes color. Texture and color information are extracted by using RT on each color plane with 4 level (1, 2, 4, 4) decomposition. This decomposition configuration provides 11

$(= 1 + 2 + 4 + 4)$ subbands for each image of the DB for each color plane. As, there are 3 color planes, so altogether we get 33 $(= 3 \times 11)$ subbands for each image of the DB. For each subband S_j, its mean (f^j_{mean}) and standard deviation (f^j_{std}) are computed, and used as the representative features of the image.

$$f^j_{mean} = \frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} |S_j(m,n)| \tag{12}$$

$$f^j_{std} = \sqrt{\frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} (S_j(m,n) - f^j_{mean})^2} \tag{13}$$

where, $S_j(m,n)$ represents a ripplet coefficient at the spatial location (m,n) of the subband S_j of size $M \times N$ and $j = 1, 2, ..., 33$. The final feature vector of an image (I) in the DB is as follows:

$$f^I_{vec} = [f^1_{mean}, f^2_{mean}, ..., f^{33}_{mean}, f^1_{std}, f^2_{std}, ..., f^{33}_{std}]$$

3.2 MLP-Based Pre-Classification (Phase-II)

To increase the accuracy and to reduce the search space, the proposed CBIR system uses a MLP-Neural Network (NN) based pre-classifier with feedforward backpropagation algorithm for training the network which consists of three layers [13]. The network consist of 66 nodes in the input layer and 10 nodes in the output layer for our used learning network*. By using the rule of thumb, it is assumed that upper bound of the nodes in the hidden layer is

$$Number\ of\ hidden\ nodes = \sqrt{(input\ nodes) * (output\ nodes)} \tag{14}$$

We have tested our network with 10, 15, 20, 25 nodes in the single hidden layer, and it has been found that 26 nodes with single hidden layer produce minimum error rate. We have tested our network with various number of iteration (1000 to 4000) with different learning rate (0.1 to 0.9) to obtain the best possible result. After going through extensive experiment, learning rate (0.7), iteration (4000) and momentum factor (0.95) is selected for classification having minimum error rate. The network with aforesaid configuration is used as a pre-classifier.

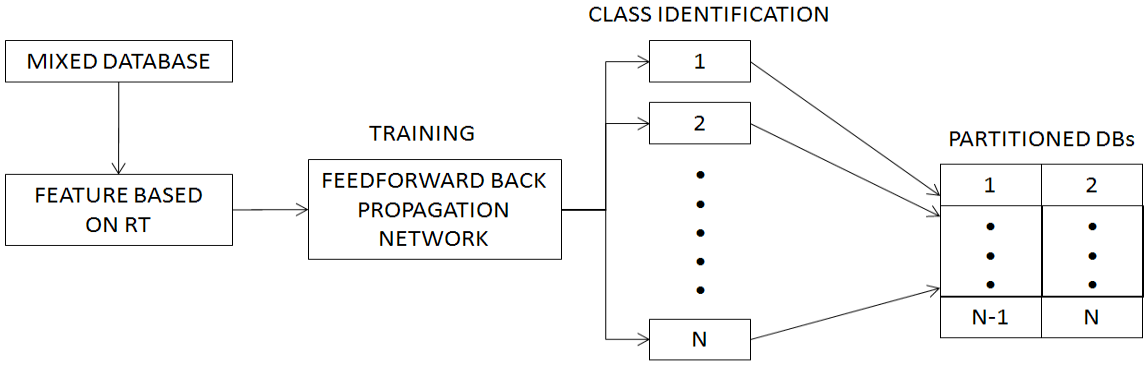

Figure 3: Block Diagram of MLP-NN based pre-Classification

The block diagram of the pre-classification phase is shown on Fig. 3. RT based features are extracted from the mixed DB, from which 70% of the label data are randomly used to train the network. The feedforward

*This configuration is only for the DB used in the experiment. We have used 66 features to represent an image in the DB. So, we have 66 nodes in the input layer. There are 10 different classes of images in the used DB, as a result, we have 10 nodes at the output layer.

backpropagation algorithm is used to train the network. When the training is completed successfully, network is tested on the remaining 30% data randomly.

The used feedforward backpropagation NN is described here. The first layer is the input layer, second layer is the hidden layer and has a log sigmoid activation function, and the third layer (Output layer) has a linear activation function. All the neurons of one layer are fully interconnected with all the neurons of its just preceding and just succeeding layers. Weights measure the degree of correlation between the activity levels of neurons that they connect. The network is initialized with random weights and biases, and was then trained using the Levenberg-Marquardt algorithm (LM) [13]. A momentum term is added to increases the learning rate with stability. The performance of the network is measured by mean squared error (MSE). The smaller the MSE is, the better the network performs.

3.3 Iterative Computation of Weights Based on Feature Relevance (Phase-III)

In the proposed system, fuzzy based relevance feedback (RF) block will be enabled only when the respective class is identified. RF based retrieval systems prompt the user for feedback on retrieval results, and then use this feedback on subsequent retrievals with the goal of increase in the retrieval performance. At each iteration user rates each returned result with respect to how useful the result is for his/her retrieval task at hand. Ratings may be simply *relevant* or *irrelevant*. A fuzzy based RF algorithm uses this feedback information to modify the relative weight of different feature values to select another set of 20 images to retrieve for the user. The retrieval process, for a given query image finishes at a point when the user is satisfied with the retrieved images as shown in Fig. 4.

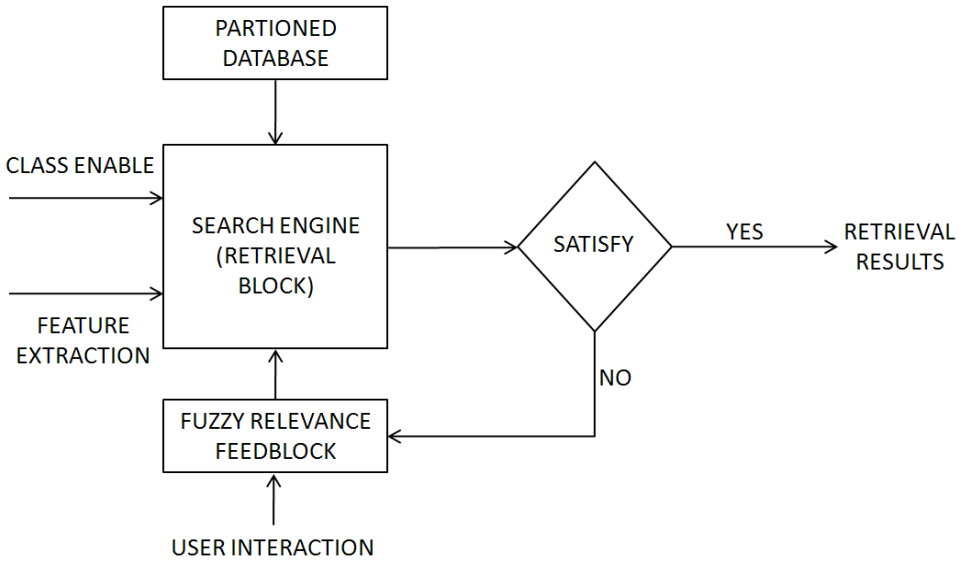

Figure 4: Block Diagram of Relevance Feedback Mechanism

An image I is usually represented by a set of features: $F = \{f_q\}_{q=1}^{N}$, where f_q is the q^{th} feature component in the N dimensional feature space. We have used Manhattan distance with weight w_q for measuring the similarity between query image I_{qr} and other images I in the DB:

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

where $|f_q(I) - f_q(I_{qr})|$ represents the Manhattan distance and w_q denotes the weight assigned to the q^{th} feature component. The weights should be adjusted such that, the features have small variation over the relevant images

and large variation over the irrelevant images. Let k similar images $I_s = \{I_1, I_2, ..., I_k\}$ are returned to the user by the proposed system. From these returned images, the user marks the relevant and irrelevant images. Let $I_C = \{$relevant images (I_r), irrelevant images (I_{ir}), system's returned images $(I_s)\}$, $I_r \cup I_{ir} = I_s$ and $I_r \cap I_{ir} = \phi$. The information from relevant images (I_r), irrelevant images (I_{ir}) and system's returned images (I_s) are combined to compute the relative importance of the individual features by fuzzy feature evaluation index (FEI) [14]:

The entropy of a fuzzy set gives a measure of *ambiguity* associated with a fuzzy feature $f_q^{I_C}$ (feature vector associated with qth feature of I_C) and is computed as follows:

$$H_q^{I_C} = (\frac{1}{M_C \ln 2}) \sum_i^{M_C} S(\mu(f_q^{I_C})) \tag{16}$$

where M_C is the number of images in the class I_C, and the Shannon's function is:

$$S(\mu(f_q^{I_C})) = -\mu(f_q^{I_C}) ln(\mu(f_q^{I_C})) - \{1 - \mu(f_q^{I_C})\} ln\{1 - \mu(f_q^{I_C})\} \tag{17}$$

$\mu(f_q^{I_C})$ is computed using a standard S-type membership function [14]:

$$\begin{aligned} S(f_q^{I_C}; a, b, c) &= 0 && f_q^{I_C} \leq a \\ &= 2 \times \{\tfrac{(f_q^{I_C} - a)}{(c-a)}\}^2 && a \leq f_q^{I_C} \leq b \\ &= 1 - 2 \times \{\tfrac{(f_q^{I_C} - c)}{(c-a)}\}^2 && b \leq f_q^{I_C} \leq c \\ &= 1 && f_q^{I_C} \geq c \end{aligned} \tag{18}$$

where,
$b = avg(f_q^{I_C})$,
$c = b + max\{\ |avg(f_q^{I_C}) - max(f_q^{I_C})|, |avg(f_q^{I_C}) - min(f_q^{I_C})|\ \}$,
$a = 2b - c$,
$avg(f_q^{I_C}), max(f_q^{I_C}), min(f_q^{I_C})$ denote the mean, maximum and minimum values corresponding to the feature vector $f_q^{I_C}$ computed over all the M_C samples in I_C, respectively. Entropy (H) is dependent on the absolute values of membership (μ). $H_{min} = 0$ for $\mu = 0$ or 1, $H_{max} = 1$ for $\mu = 0.5$. Since $\mu(b) = \mu(avg(f_q^{I_C})) = 0.5$, the values of $H_q^{I_C}$ are 1.0 at $b = avg(f_q^{I_C})$ and would tend to zero when moved away from b towards either c or a of the S function. Selecting $b = avg(f_q^{I_C})$ indicates that, the cross over point is near to the query feature component. Higher value of $H_q^{I_C}$, indicates more number of samples having $\mu(avg(f_q^{I_C}))$ equal to 0.5. with a tendency of the samples to cluster around the mean value, resulting in less internal scatter within the class.

The FEI for the q^{th} feature is computed as follows:

$$FEI_q = \frac{H_q^{I_s}}{H_q^{I_r} + H_q^{I_{ir}}} \tag{19}$$

Lower value of FEI_q, indicates better quality of importance of the qth feature in recognizing and discriminating different classes. The precision of retrieval can be improved with these values.

The weight w_q is computed as a function of the evaluated (FEI_q) as shown below:

$$w_q = FEI_q \tag{20}$$

In the first pass, all features are considered to be equally important. Hence, $w_1 = w_2 ... = w_q = ... = w_N = 1$. The feature spaces of the relevant images are therefore altered in a similar fashion after updating the components with w_q. As a result, the ranks of the relevant images are not affected much.

4 Experimental Results and Comparison

To prove the effectiveness of the proposed CBIR system, extensive experiments were performed on 1000 images of SIMPLIcity database [3] with 10 different classes (African, Ocean, Building, Bus, Dinosaurs, Elephant, Flower, Horse, Mountain and Food) of images. The retrieval results obtained using the proposed CBIR system, are compared with some of the existing retrieval systems [20, 17, 18, 19]. The experiments were carried out on a Dell Precision T7400 PC with 4GB RAM and was implemented using MATLAB R2008a.

Two commonly used statistical measures were computed to assess the proposed system's performance, namely precision and recall, which are defined:

$$Precision\ (P) = \frac{N_{RIR}}{N_{RIR} + N_{IRIR}} \tag{21}$$

$$Recall\ (R) = \frac{N_{RIR}}{T_{RID}} \tag{22}$$

where, N_{RIR} is the Number of relevant images retrieved, N_{IRIR} is the Number of irrelevant images retrieved and T_{RID} is the Total number of relevant images in the database.

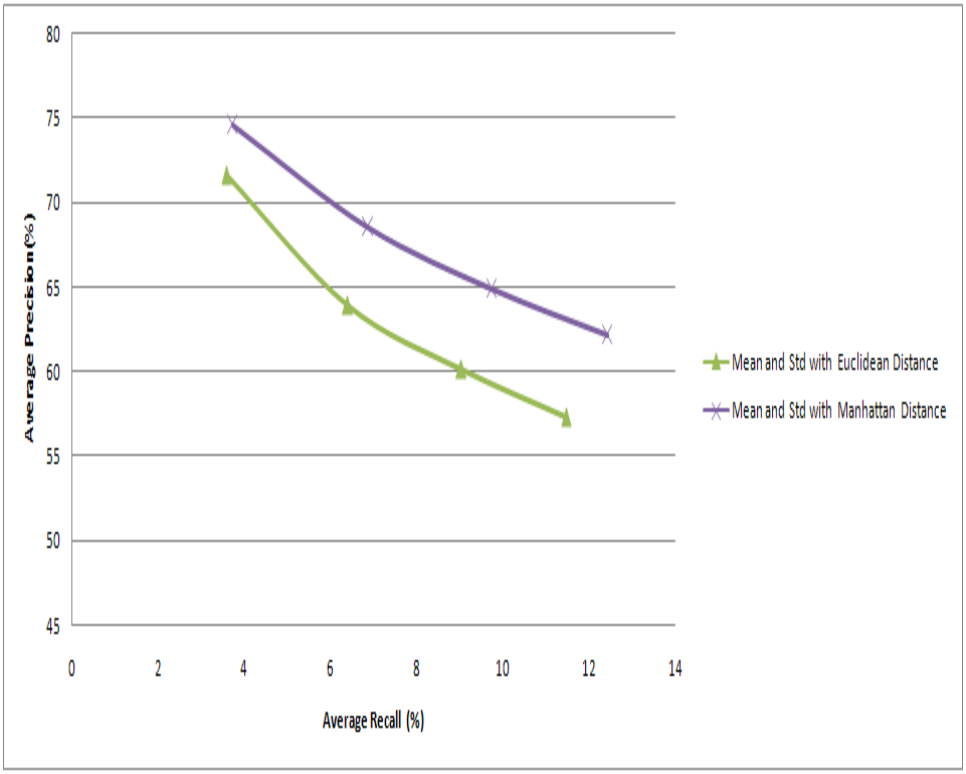

Figure 5: Performance comparison of the proposed system with different similarity measures

Euclidean distance is extensively used as the similarity measure in CBIR domain. Euclidean distance is always not the best similarity measure, as the distances in each dimension are squared before summation which results in large dissimilarity [15]. A moderate approach is by using the sum of the absolute differences rather than their squares as the overall measure of dissimilarity. This sum of absolute difference distance measure is called as city block distance or the Manhattan distance. In the proposed CBIR system this Manhattan distance is used as the similarity measure. The graph of the Fig.5, supports our choice of Manhattan distance as the similarity measure over Euclidean similarity measure with mean and standard deviation as feature vector. In

addition to the performance boost, Manhattan similarity measure also leads to lower computational complexity [16].

If the user is satisfied with the retrieved results without using RFM then the retrieval process is ended. If, however, the user is not satisfied, he/she can select top query based relevant images as positive feedbacks and the remaining as negative examples for updating the weight parameters and revising the features. Using these feedbacks, Feature Evaluation Index (FEI) as the weights of the features are computed as described in section.3.3. Then, all the images are re-sorted based on the weighted similarity metric. If the user is still not content with the result, he/she can repeat the process.

Fig.6 and Fig.7, shows two examples of the retrieval results obtained by our proposed system, using different query images. In the example of Fig.6, an image from the "*Elephant*" class was used as the query image. Similarly, an image from the "*Flower*" class was used as the query image in the example of Fig.7. In both the cases, we can see all the retrieve images were from their respective classes as of the query images. But, with every iteration of the RFM, the ranking of the retrieve images got changed, and more and more similar images came up as the improved retrieval results.

Figure 6: (a) First pass of the retrieval results using RT and Manhattan Distance (top left side image is the query image) (b) First Iteration (c) Second Iteration (d) Third Iteration.

We have compared the performance of the proposed CBIR system with some of the existing CBIR methods, and the comparison is given in Table 1. The results shown in the Table 1, reflects the performance without using the MLP based pre-classification step and the relevance feedback mechanism. The numbers in the Table 1, represent the average precision computed over the corresponding image class. From Table 1, it can be easily seen, that the proposed CBIR system provide improved retrieval performance over other existing CBIR algorithms namely, SIMPLIcity [17], FIRM [18], using salient points (salient points detected by Harris corner detector (SP by HCD), color salient points (CSP)) [19] and contourlet Harris detector (CHD) [20]. Only for the classes "*Sea*", "*Building*" and "*Buses*" the performance of CHD based method outperforms our method. The **Bold** values indicate the highest retrieval performance.

To test whether the retrieval accuracy of the proposed CBIR system based on RT, is significantly higher than those of other conventional transforms like wavelet (WT), curvelet (CV), contourlet (CN) etc., a paired

(a) (b)

Figure 7: (a) First pass of the retrieval results using RT and Manhattan Distance (top left side image is the query image) (b) First Iteration.

Class	Wang[17]	FIRM[18]	SP by HCD[19]	CSP[19]	CHD[20]	Proposed Method
Africans	0.48	0.47	0.40	0.48	0.44	**0.50**
Sea	0.32	0.35	0.31	0.34	**0.43**	0.42
Building	0.35	0.35	0.32	0.33	**0.49**	0.39
Buses	0.36	0.60	0.44	0.52	**0.64**	**0.64**
Dinosaur	0.95	0.95	0.92	0.95	0.98	**0.99**
Elephant	0.38	0.25	0.28	0.40	0.48	**0.55**
Flower	0.42	0.65	0.58	0.60	0.77	**0.84**
Horse	0.72	0.65	0.68	0.70	0.75	**0.86**
Mountain	0.35	0.30	0.32	0.36	0.31	**0.44**
Food	0.38	0.48	0.44	0.46	0.32	**0.56**
Average	0.47	0.51	0.47	0.51	0.56	**0.62**

Table 1: Comparison with other existing CBIR systems

t-test was performed on the results of the 2×5-fold cross validation. Let μ_w, μ_{cn}, μ_{cv} and μ_{rt} be respectively the average precisions (retrieval accuracies) of wavelet, contourlet, curvelet and ripplet transforms. The null hypothesis is that the difference between the means of the two techniques is zero ($H_0 : \mu_{rt} - \mu_{tr} = 0$); $tr = \{w, cv, cn\}$, and the alternative hypothesis is that the difference is positive ($H_a : \mu_{rt} - \mu_{tr} > 0$). For a trial i of the cross validation, let P^i_{rt} and P^i_{tr} be respectively the results obtained by using RT tr; where $tr = \{w, cv, cn\}$. The test statistic is computed as follows:

$$t = \overline{P} \cdot \frac{\sqrt{n}}{\sqrt{\frac{\sum_i^n (P^i - \overline{P})^2}{n-1}}} \tag{23}$$

where $\overline{P} = (1/n)\sum_1^n P^i$, $P^i = P^i_{rt} - P^i_{tr}$ and n is the number of runs (10 for this experiment) [21]. The P-value is obtained from a t-distribution table at the degree of freedom $(n-1)$, and is compared to the critical value 0.05 (i.e. 5% significant level). If the P-value is smaller than 0.05, the null hypothesis is rejected at 0.05 significance level.

Table 2, shows the average precision rates of the 2×5-fold cross validation including the advantage of MLP neural network. A hypothesis test is performed to evaluate the significance of the differences between the performances of the proposed RT based CBIR system with CBIR systems based on other transforms. The results at 5% significance level are summarized in Table 3. The null hypothesis is rejected at 0.05 significance level for the problem in hand. It means that the retrieval accuracy rates obtained using our proposed CBIR

Transform	Partition 1					Partition 2					Average
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	
Wavelet(db8)	82.24	83.34	80.98	82.09	83.17	81.31	83.06	81.24	83.39	81.41	82.22
Curvelet	82.96	83.89	81.32	82.59	83.92	81.86	83.59	81.97	83.96	81.96	82.80
Contourlet	83.36	84.49	81.55	82.91	84.21	82.01	83.89	82.31	84.24	82.13	83.11
Ripplet	83.54	84.98	82.15	83.67	84.78	82.79	84.05	83.47	84.55	82.58	**83.66**

Table 2: Average precision rates of the proposed CBIR schemes using different transforms over the 2×5-fold.

Method	T-value	P-value	Null Hypothesis (H_0)
Ripplet vs. Wavelet(db8)	12.62	5e-07	Reject
Ripplet vs. Curvelet	8.61	1.224e-05	Reject
Ripplet vs. Contourlet	5.68	3.024e-05	Reject

Table 3: The results of t-test at significance level $\alpha = 5\%$.

system based on RT are higher than those obtained using other conventional transforms and the difference is statistically significant. It is to be noted that, all P-values are smaller than 0.01, which means that the differences are actually statistically highly significant.

5 Conclusions

From our experiments, we have noticed that ripplet transform based image coding is suitable for representing low level features (color and edge) of the images. The proposed CBIR system with NN classifier using RT based features is able to improve the accuracy of the retrieval performance and to reduce the computational cost. The retrieval performance is improved further using fuzzy based RFM within 3 to 4 iteration. To overcome the problem of mis-classification in pre-classification phase, we are trying to implement fuzzy ranking membership function. The proposed mechanism could be tested for video retrieval as future scope of research.

6 Acknowledgment

The authors would like to thank the anonymous reviewers and editors for their invaluable suggestions. The authors would also like to thank Jun Xu and Dapeng Oliver Wu (Dept. of Electrical and Computer Engineering, University of Florida, USA) for helping us in the implementation of Ripplet Transform.

References

[1] D. Heesch,"A survey of browsing models for content based image retrieval", *Multimedia Tools Application*, 40(2):261-284, 2008.

[2] M. Acharyya and M. K. Kundu, "An adaptive approach to unsupervised texture segmentation using M-band wavelet tranform", *Signal Processing*, 81(7):1337-1356, 2001.

[3] J. Z. Wang, J. Li and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947-963, 2001.

[4] M. N. Do and Martin Vetterli, "The Contourlet Transform: An Efficient Directional multiresolution Image Representation", *IEEE Transactions on Image Processing*, 14(12):2091-2106, 2005.

[5] M. N. Do and Martin Vetterli, "The finite Ridgelet transform for image representation", *IEEE Transactions on Image Processing*, 12(1):16-28, 2003.

[6] E. Candes and D. Donoho, "Continuous curvelet transform: I. Resolution of the wavefront set", *Applied and Computational Harmonic Analysis*, 19(2):162-197, 2003.

[7] C. C. Yang, S. O. Prashar, J. A. Landry, J. Perret and H. S. Ramaswamy, "Recognition of weeds with image processing and their use with fuzzy logic for precision farming", *Canadian Agricultural Engineering*, 42(4):195-200, 2000.

[8] H. Pourghassem and H. Ghassemian, "Content-based medical image classification using a new hierarchical merging scheme", *Computerized Medical Imaging and Graphics*, 32(8):651-661, 2008.

[9] Y. Rui, T.S.Huang and S.Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval", *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644-655, 1998.

[10] M. K. Kundu, M. Chowdhury and M. Banerjee, "Interactive Image Retrieval with Wavelet Features", *Proc. 4th Int. Conf. on Pattern recognition and Machine Intelligence(PREMI 2011)*, Lecture Notes in Computer Science, 6744:162-172, 2011, Moscow, Russia.

[11] Z. Jin, I. King and X. Li, "Content-Based Image Retrieval by Relevance Feedback", *Proc. of VISUAL 2000*, 1929:521-529, 2000, Lyon, France.

[12] J. Xu, L. Yang and D. O. Wu, "Ripplet: A new transform for image processing", *Journal of Visual Communication and Image Representation*, 21(7):627-639, 2010.

[13] A. L. Betker, T. Szturm, and Z. Moussavi, "Application of feedforward backpropagation neural network to center of mass estimation for use in a clinical environment", *Proc. 25th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, 2714-2717, 2003.

[14] S. K. Pal and B. Chakraborty, "Intraclass and interclass ambiguities (fuzziness) in feature evaluation", *Pattern Recognition Letters*, 2(5):275-279, 1984.

[15] M. Kokare, P. K. Biswas and B. N. Chatterji, "Cosine-Modulated Wavelet Packet based Texture Features for Content based Image Retrieval", *IETE Journal of Research*, 51(6):477-483, 2005.

[16] B. Rohani and B. Nugroho, "Mahattan-Chebychev Distance Metric for MIMO Systems", *IEICE Technical Report RCS2008-138*, 49-52, 2008.

[17] J. Li, J. Z. Wang and G. Wiederhold, "IRM: Integrated Region Matching for Image Retrieval", *Proc. of the 8th ACM Int. Conf. on Multimedia*, 147-156, 2000.

[18] Y. Chen and Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1252-1267, 2002.

[19] P. S. Hiremath and J. Pujari, "Content Based Image Retrieval using Color Boosted Salient Points and Shape features of an image", *International Journal of Image Processing*, 2(1):10-17, 2008.

[20] H. N. Duc, T. L. Tien, T. D. Hong, C. B. Thu and T. N. Xuan, "Image retrieval using contourlet based interest points", *Proc. of the 10th Int. Conf. on Information Science, Signal Processing and their Applications (ISSPA 2010)*, 93-96, 2010, Kuala Lumpur.

[21] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms", *Neural Computation*, 10(7):1895-1923, 1998.