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

Comparative Assessment of Efficiency for Content Based Image Retrieval Systems Using Different Wavelet Features and Pre-Classifier

Manish Chowdhury · Malay Kumar Kundu

Received: date / Accepted: date

Abstract Recently, Content Based Image Retrieval (CBIR) has emerged as an active research area having applications in various fields. There exist several states-of-the art CBIR systems that uses both spatial and transform features as input. However, as hardly any details study reported so far on the effectiveness of different transform domain features in CBIR paradigm. This motivates the current article where we have presented extensive comparative assessment of five different transform domain features considering various filter combinations. Three different feature representation schemes and three different classifiers have been used for this purpose. Extensive experiments on four widely used benchmark image databases (Oliva, Caltech101, Caltech256 and MIRFlickr25000) were conducted to determine the best combination of transform, filters, feature representation and classifier. Furthermore, we have also attempted to discover the optimal features from the best combinations using maximal information compression index (MICI). Both qualitative and quantitative evaluations show that the combination of Least Square Support Vector Machine (LSSVM) as a classifier and the statistical parametric framework based reduced feature representation in Non-Subsampled Contourlet Transform (NSCT) with "pyrexc" and "sinc" filters gives the best retrieval performances.

Keywords Content Based Image Retrieval · Fuzzy C-means Clustering Algorithm · Multi-resolution Analysis · Multi-scale Geometric Analysis · Classification

1 Introduction

Due to the rapid advancement of the information and communication technologies, thousands of digital images are being created and archived freely in the distributed databases hooked up by the internet, which becomes a part of the information to

Manish Chowdhury[1] and Malay Kumar Kundu[2]
Machine Intelligence Unit, Indian Statistical Institute
203 B.T.Road, Kolkata-108, India
Tel.: +91-33-2575-3100; Fax: +91-33-2578-3357
ail[1]st.manishc@gmail.com,[2]malay@isical.ac.in

the society [1,2]. In spite of the huge pictorial resources available on WWW, now-a-days it is possible to find the relevant images via internet using efficient Content Based Image Retrieval (CBIR) searching algorithms by different user based on their requirements [3–8]. CBIR is the task of retrieving relevant images from a large image database by measuring similarities between the query image and database images automatically based on some derived features like color, texture, shape etc., [9–12]. The performance of a CBIR system greatly depends on the effectiveness of the image representative feature vector and suitable similarity measure [13–15]. High retrieval efficiency and less computational complexity are the desired characteristics of an effective CBIR system [16–19]. But due to the enormous size of image data, searching becomes an expensive task [20,21]. Generally, to cut down searching cost in modern CBIR systems, pre-classification is often used to partition the large image database which results in reducing the search space by selecting one or few subsets of images for subsequent image similarity matching [22–24].

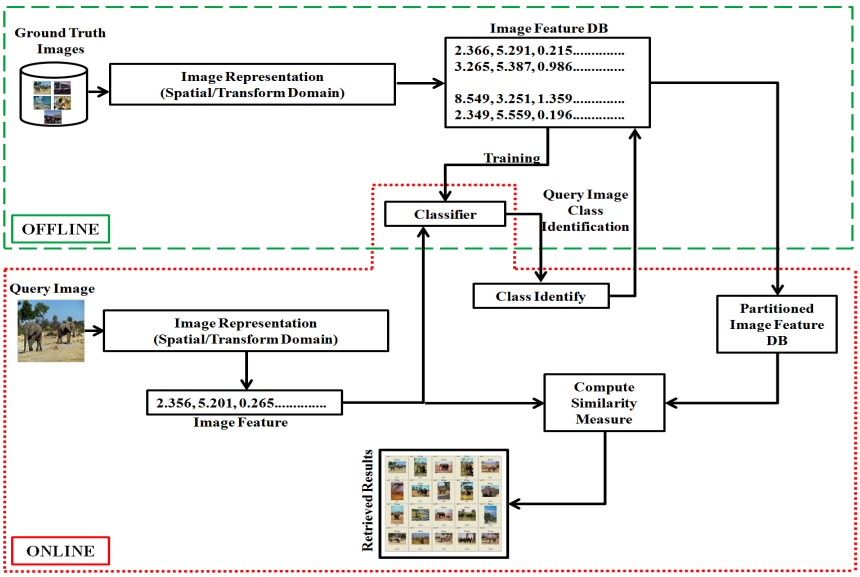


Fig. 1 General architecture of CBIR system

The block diagram of a modern pre-classification based CBIR system is shown in Fig.1. Most of these systems are based on two different operational phases: 'Offline' and 'Online'. In 'Offline' phase, the image representative feature vectors of some ground truth images are constructed either in the spatial domain or in the transform domain (sometimes a combination of the two domains). After that, the constructed image representative feature vectors are stored in a feature database with corresponding image's class information. This feature database is used to train a classifier. In the 'Online' phase, users submit a query image to the system. The image representative feature vector is calculated using the same procedure as of the 'Offline' phase. This

query image's feature vector is input to the previously trained classifier, which classifies the query image as belonging to a particular class. 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 some similarity measures. These similarities are sorted in ascending order and the corresponding images with maximum similarities are displayed to the user as the retrieved results [25–28].

Generally, the performance effectiveness of the classification and similarity measurement steps of a CBIR system greatly depends on the preceding image representative feature extraction procedure [29–32]. Image feature representation can be generally grouped into two categories: spatial and transform domain representation [33, 34]. In spatial domain approach, the image features are computed directly from image pixel intensity/color such as gradient, local standard derivation, histograms, gray-level co-occurrence matrix etc. Recently, researchers have exploited interest point detection algorithm in spatial domain which can tolerate the effect of intensity, rotation, scale and affine transform variation to some extent i.e., Scale-invariant feature transform (SIFT) [35], Speeded Up Robust Features (SURF) [36], Binary Robust Independent Elementary Features (BRIEF) [37], Oriented FAST and Rotated BRIEF (ORB) [38], etc. On the other hand, in transform domain two different types of transform (1) Unitary Transform (UT) like Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), Karhunen-Loeve Transform (KLT), Hadamard etc., and (2) Space/Spatial-Frequency Transform (SSFT) like Discrete Wavelet Transform (DWT), Gabor, etc., are used to extract the transform coefficient as image feature. These features are robust to a great extend against noise corruption and shift. In UT, spatial frequency information are generated as features from the digital image but positional information of different pixel are lost whereas in SSFT both spatial frequency and positional information are extracted and used in the transform representation [39]. One of the major advantage of the transform domain factor is the compaction of majority of image information in a fewer number of transform coefficient [40].

With the development of various transform bases, different kinds of Multi Resolution Analysis (MRA) tools have been developed and used for various applications of image processing. Wavelets and related classical multiscale transforms suffers from lack of anisotropic, shift sensitivity and limited orientation selectivity. In contrary, the recently developed various Multi-Scale Geometric Analysis (MGA) tools such as Curvelet Transform (CVT) [41], Contourlet Transform (CNT) [42], Ripplet Transform (RT) [43], Non-Subsampled Contourlet Transform (NSCT) [44] etc., have shown superior performance than Wavelet Transform (WT) and its variants in many image processing applications.

Wavelets and related classical multiscale transforms like M-band wavelet, dual-tree complex wavelet, wavelet packets etc., have been employed in CBIR paradigm extensively [45–51]. These transforms conduct decomposition over a limited direction in which the 2-D bases simply consist of all possible tensor products of 1-D basis functions [52]. Moreover, it is only well adapted to point-singularity [51], but not the curve-or-line singularities [53], because mention transforms extract limited directional information. Thus, these transforms can reveal the image features across edges, but not along the edges [51]. For 2D signal like images, there exist higher or-

der singularities i.e., curve-or-line singularities cannot be handle with conventional SSFT [51,53].

Due to these inherent shortcomings of DWT and its variants, CBIR systems based on these transforms provide sub-optimal results. To overcome from this problem, recently several advanced CBIR systems are proposed based on recently developed MRA/MGA tools: CVT, CNT, RT, NSCT [54–57]. However, there is a lack of comprehensive comparison of the effectiveness of these transforms in CBIR paradigm. In this article, we have extensively evaluate the performance efficiencies of these above mentioned recently developed advanced MRA/MGA tools (DWT, CVT, CNT, RT, NSCT) based on three different feature extraction techniques, various filters combinations and three different classifier (namely Multilayer Perceptron (MLP), Random Forest (RF) and Least Square Support Vector Machine (LSSVM)) following the general architecture of CBIR system described in Fig. 1. For similarity measurement step, in this article we have used the popular Euclidean distance (ED). Firstly, for each transform (DWT, CVT, CNT, RT, NSCT), we find out the best combination of "*Transform (Filters)* $\Rightarrow$ *Feature Representation* $\Rightarrow$ *Classifier*". After that, we have globally compared the effectiveness of the best combination for each transform. The best combinations selected from the global comparison are further being evaluated using maximal information compression index (MICI) feature selection algorithm for determining the optimal features from the finest combination. In this article, our prime objective is to find out the best combination of transform domain feature extraction technique with pre-classification followed by similarity measurement for effective CBIR application.

The rest of the paper is organized as follows: In Section 2, theoretical preliminaries of the five different MRA/MGA tools and the three different classifiers are reviewed briefly. Section 3 contains the three different features extraction techniques. Experimental results and comparisons are discussed in Section 4. Finally, the main conclusion of the paper is given in Section 5.

2 Theoretical Preliminaries

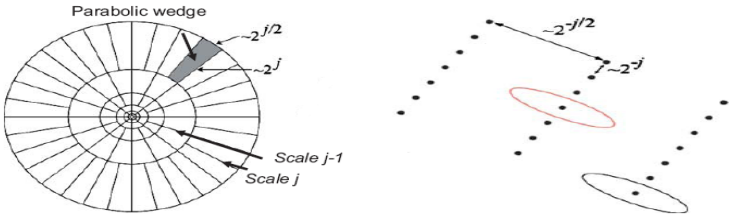
In this section, we first briefly reviewed the theoretical preliminaries of five different MRA /MGA tools. And then, we discussed about three different classifiers used in the comparative study.

2.1 Different MRA/MGA Tools

The MRA/MGA tools used in this paper are DWT, CVT, CNT, RT and NSCT. Brief descriptions of these tools are given below:

2.1.1 Discrete Wavelet Transform (DWT)

The advantage of wavelet is that it performs an MRA of a signal with localization in both time and frequency [39,40]. In addition to this, functions with discontinuities and functions with sharp spikes require fewer wavelet basis vectors in the wavelet domain than sine cosine basis vectors to achieve a comparable approximation. DWT can be implemented as a set of high-pass and low-pass filter banks. In standard dyadic wavelet decomposition, the output from the low-pass filter can subsequently be decomposed in the same way and the process continues to have finer resolution.



(a) (b)

Fig. 2 Curvelets in (a) Fourier frequency (b) Spatial domain

2.1.2 Curvelet Transform (CVT)

Traditional wavelet transform is unable to resolve $2-D$ singularities along arbitrarily shaped curves, and as a result it cannot capture curves and edges of images effectively. To overcome this problem, Candes et al., proposed the CVT with the idea of representing curve as a superposition of bases at various lengths and widths obeying the scaling law $width \approx length^2$ [41]. CVT uses a parabolic scaling law to achieve anisotropic directionality. The CVT has an almost optimal sparse representation of objects with C^2singularities, combined with other methods, excellent performance of the CVT has been shown in image processing. Fig. 2 presents the curvelet analysis method.

The CVT can be constructed by a pair of windows $W(r)$ and $V(t)$ which are defined as the radial window and angular window respectively. These are smooth, non-negative and real-valued with W as a frequency domain variable with positive real argument and is supported on polar co-ordinate r and θ in the frequency domain; V is real arguments and is supported on t. These windows will always obey the admissibility condition.

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \qquad r \in (3/4, 3/2) \tag{1}$$

$$\sum_{l=-\infty}^{\infty} V^2(t-1) = 1, \qquad t \in (-1/2, 1/2) \tag{2}$$

For each $j \leq j_0$, a polar wedge U_j is supported by W and V, applied with scale dependent window width in radial and angular direction. This polar wedge or frequency window is defined in the fourier domain by

$$U_j(r,\theta) = 2^{\frac{-3j}{4}} W(2^{-j} r) V\left(\frac{2\lfloor \frac{j}{2} \rfloor \theta}{2\pi}\right) \tag{3}$$

where $\lfloor j/2 \rfloor$ is the integer part of $j/2$ is the integer part of $j/2$. The symmetrized version of Eqn. 3 is used to obtain real-valued curvelet. A curvelet coefficient can be expressed as the inner product between an element $f \in L^2(R^2)$ and curvelet $\varphi_{j,l,k}$:

$$c(j,l,k) = < f, \varphi_{j,l,k} > = \int_{R^2} f(x) \overline{\varphi_{j,l,k}} dx = \frac{1}{2\pi^2} \int \widehat{f}(w) U_j(R_{\theta_l} w) e^{i < x_k^{(j,l)w} >} dw \tag{4}$$

where, $j = 0,1,2,.....$ is a scale parameter; $l = 0,1,2,3,.....$ is an orientation parameter; and $k = (k_1,k_2),\ k_1,k_2 \in \mathfrak{T}$ is a translation parameter. The $\varphi_j(x)$ is defined by means of its Fourier transform $\widehat{\varphi}_j(\omega) = U_j(\omega)$, where U_j is frequency window defined in the polar coordinate system in Eqn. 3. φ_j is the mother curvelet at scale 2^{-j} which are obtained by rotations and translation of φ_j. Curvelets at scale 2^{-j}, orientation θ_l and position $x_k^{(j,l)} = R_{\theta_l}^{-1}(k_1 \cdot 2^j, k_2 \cdot 2^{\frac{-j}{2}})$ can be expressed as: $\varphi_{j,l,k}(x) = \varphi(R_{\theta_l}(x - x_k^{(j,l)}))$, where $\theta_l = 2\pi \cdot 2^{\lfloor \frac{j}{2} \rfloor} \cdot l$, with $l = 0,1,...,0 \leq \theta_l < 2\pi$, R_{θ_l} is the rotation by θ_l radians and R_θ^{-1} is its inverse.

$$R_\theta = \begin{pmatrix} cos\theta & Sin\theta \\ -Sin\theta & cos\theta \end{pmatrix} \tag{5}$$

and $R_\theta^{-1} = R_\theta^T = R_{-\theta}$

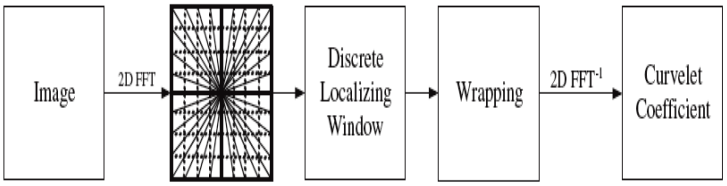


Fig. 3 The flowchart of the CVT via wrapping.

CVT obeys an anisotropy scaling relation, length $\approx 2^{\frac{-j}{2}}$ and $width = 2^{-j}$, such that $width \approx length^2$. Fast digital CVT can be implemented either by using unequispaced FFTs or using wrapping. The flowchart of the CVT implementation using wrapping is presented in Fig. 3. The wrapping mechanism is as follows: (i) apply the 2D Fast Fourier Transform (FFT) and obtain Fourier samples, (ii) form the product by multiplying the discrete localizing window $U_{j,l}$ for each scale j and angle l, (iii) wrap this product around the origin, (iv) apply the inverse 2D FFT to collect the discrete coefficients $c^D(j,l,k)$. More details of the CVT are available in [41].

2.1.3 Contourlet Transform (CNT)

CNT gives a multi-resolution, local and directional expansion of image using Pyramidal Directional Filter Bank (PDFB). The PDFB combines Laplacian Pyramid (LP) which captures the point discontinuities, with a Directional Filter Bank (DFB) which links these discontinuities into linear structures. The key idea of CNT is to find an optimal sparse representation for functions in $\mathbb{R}^2$ with curved singularities and having anisotropy scaling relation for curves [42,53].

The CNT use indices j, k, n, and l ($j \in \mathbb{Z}$, $n \in \mathbb{Z}^2$) to represent scale, direction, location, and decomposition level, respectively; m is also used to denote location. CNT are derived from nonseparable double filter banks. As for the Wavelet Filter Bank (WFB), the Contourlet Filter Bank(CFB) has an associated continuous-domain expansion in $L^2(\mathbb{R}^2)$ using the contourlet functions. LP in the CFB uses orthogonal

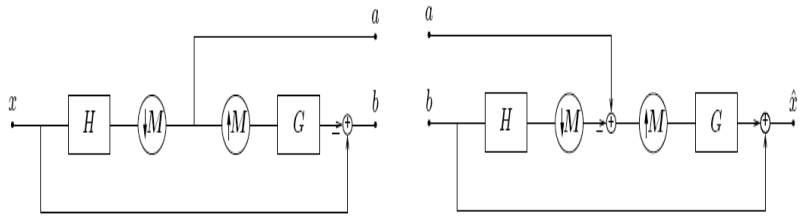


Fig. 4 Laplacian Pyramid Scheme: (a) Analysis and (b) Synthesis.

filters and downsampling by 2 in each dimension (that means M = diag(2, 2) in Fig 4).

The low pass synthesis filter G uniquely defines a unique scaling function $\phi(t) \in L^2(\mathbb{R}^2)$ satisfying the two-scale equation as follows:

$$\phi(t) = 2 \sum_{n \in \mathbb{Z}^2} g[n]\phi(2t-n) \tag{6}$$

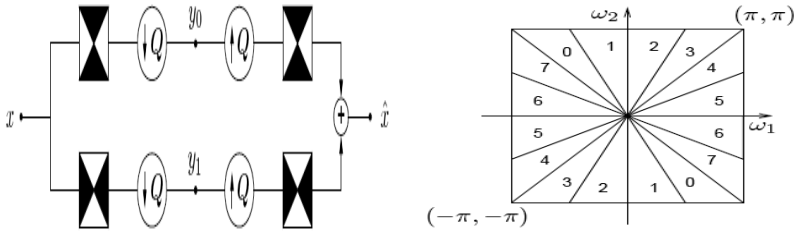
Suppose, $\phi_{j,n} = 2^{-j}\phi(\frac{t-2^j n}{2^j})$, is an orthonormal basis for approximation subspace V_j at the scale 2^j. This subspace also provides a sequence of multiresolution nested subspaces, where V_j is associated with a uniform grid of intervals $2^j \times 2^j$ that characterizes image approximation at same scale. The necessary details contain in the LP of the difference images for increasing the resolution between two consecutive approximation subspaces. Therefore, V_{j-1}, the orthogonal complement of V_j is W_j, which satisfying $V_{j-1} = V_j \oplus W_j$.

Let $F_i(z), 0 \leq i \leq 3$ be the synthesis filters for these polyphase components. These are highpass synthesis filters of the components in the over sampled filter bank LP. Then, it is associate with each of these filters to build a continuous function $\psi^i(t)$ where

$$\psi^i(t) = 2 \sum_{n \in \mathbb{Z}^2} f_i[n]\phi(2t-n) \tag{7}$$

Using $\psi^i(t)$ in Eq. 7 and assuming $\psi^{(i)}_{j,n}(t) = 2^{-j}\psi^{(i)}(\frac{t-2^j n}{2^j})$, then for all scale, $\{\psi^{(i)}_{j,n}\}$ is a tight frame for $L^2(\mathbb{R}^2)$ with frame bounds equal to 1, which realizes the optimal linear reconstruction. Since, W_j is not a shift-invariant subspace, it is being build by denoting $\mu_{j,2n+q_i}(t) = \psi^{(i)}_{j,n}(t)$, and $0 \leq i \leq 3$, where q_i are the coset representatives for downsampling by 2 in each dimension: $q_0 = (0,0)^T$, $q_1 = (1,0)^T$, $q_2 = (0,1)^T$,$q_3 = (1,1)^T$. With this notation in LP, each subspace W_j is spanned by a frame $\mu_{j,n}(t)$ that utilizes a uniform grid on $\mathbb{R}^2$ of intervals $2^{j-1} \times 2^{j-1}$.

Bamberger et al., constructed a 2D DFB that can be maximally decimated while achieving perfect reconstruction [58]. It is used in the second stage of CNT to link the edge points into linear structures, which involves modulating the input image and using Quincunx Filter Banks (QFB) with diamond-shaped filters [42]. DFB is constructed from two building modules. The first module is a two-channel quincunx filter bank with fan filters (see Fig 5(a)) that divides a 2-D spectrum into two directions i.e.,



(a) (b)

Fig. 5 (a) Two-dimensional spectrum partition using QFB with fan filters. The black regions represent the ideal frequency supports of each filter. Q is a quincunx sampling matrix. and (b) Directional filter bank: Frequency partitioning where l = 3 and there are 2^3 =8 real wedge-shaped frequency bands. Subbands $0-3$ correspond to the mostly horizontal directions, while subbands $4-7$ correspond to the mostly vertical directions

horizontal and vertical. The second module of the DFB is a shearing operator, which amounts to just reordering of image samples. Both the modules guarantee a different directional frequency partition while maintaining perfect reconstruction. Fig 5(b) shows the frequency partition of contourlets. The characteristics of quincunx DFB are as follows: (1) linear phase filter bank is use as its $1-D$ prototype filters. (2) It also has the permissible property to designed the filters having good pass-band and stop-band parameters if and only if filter bank is allowable. (3) Its high pass subbands characteristics can be considered as directional filters, with high directional selectivity and it has sub-band frequency supports having similar shapes [53].

The sampling matrics of an l-level tree structured DFB diagonal form is denoted as follows:

$$S_k^{(l)} = \begin{cases} diag(2^{l-1},2), & for\ \ 0 \le k < 2^{l-1} \\ diag(2,2^{l-1}), & for\ \ 2^{l-1} \le k < 2^l \end{cases} \tag{8}$$

DFB diagonal form is then applied to the approximation subspaces V_j

$$\rho^l_{j,k,n}(t) = \sum_{m\in\mathbb{Z}^2} d_k^{(l)}[m - S_k^{(l)} n]\phi_{j,m}(t) \tag{9}$$

where the family $\rho^l_{j,k,n}$ is an orthonormal basis of a directional subspace $V^{(l)}_{j,k}$ for each $k = 0,,2^l-1$. Furthermore, to increase the directional resolution, an extra level of decomposition by a pair of orthogonal filters $d_k^{(l)}$ (i.e., $d_k^{(l)}$ can lead to $d_{2k}^{(l+1)}$ and $d_{2k+1}^{(l+1)}$) is applied. Then by applying the directional decomposition by the family $d_k^{(l)}[m - S_k^{(l)} n]$, $0 \le k \le 2^l$ onto the detail subspace W_j as done by the CNT, we gain similar results

$$\lambda^l_{j,k,n}(t) = \sum_{m\in\mathbb{Z}^2} d_k^{(l)}[m - S_k^{(l)} n]\mu_{j,m}(t) \tag{10}$$

The family $\lambda^l_{j,k,n}$ is a tight frame of a detail directional subspace $W^{(l)}_{j,k}$ with frame bounds equal to 1 for each $k = 0,\ldots,2^l-1$ where subspaces $W^{(l)}_{j,k}$ are mutually orthogonal across scales and directions [53]. More details can be found in [42].

2.1.4 Ripplet Transform Type - I(RT)

To generalize the scaling law of the CVT and to find out which scaling law will be optimal for all types of boundaries, a method named RT was proposed by Jun Xu et al., [43]. RT is a higher dimensional generalization of the CVT. RT provides a new tight frame with sparse representation for images with discontinuities along C^d curves. 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 [43]. 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 $d = 2$.

As digital image processing needs discrete transform instead of continuous transform, here we describe the discretization of RT [43]. In the frequency domain, the corresponding frequency response of ripplet function is in the form

$$\widehat{\rho}_j(r,\omega) = \frac{1}{\sqrt{c}} a^{\frac{m+n}{2n}} W(2^{-j}\cdot r) V(\frac{1}{c}\cdot 2^{-\lfloor j\frac{m-n}{n}\rfloor}\cdot\omega - l) \tag{11}$$

where W and V are the 'radial window' and 'angular window', respectively and satisfy the following admissibility conditions:

$$\sum_{j=0}^{+\infty} |W(2^{-j}\cdot r)|^2 = 1 \tag{12}$$

$$\sum_{l=-\infty}^{+\infty} |V(\frac{1}{c}\cdot 2^{-\lfloor j(1-1/d)\rfloor}\cdot\omega - l)|^2 = 1 \tag{13}$$

given c, d and j. Here, the scale parameter a is sampled at dyadic intervals. b (position parameter) and θ (rotation parameter) 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.

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

$$H_{j,l}(r,\theta) = \{2^j \leq |r| \leq 2^{2j}, |\theta - \frac{\pi}{c}\cdot 2^{-\lfloor j(1-1/d)\rfloor}\cdot l| \leq \frac{\pi}{2}2^{-j}\} \tag{14}$$

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

$$R_{j,\overrightarrow{k},l} = \sum_{n_1=0}^{M-1}\sum_{n_2=0}^{N-1} f(n_1,n_2)\overline{\rho_{j,\overrightarrow{k},l}(n_1,n_2)} \tag{15}$$

where $R_{j,\overrightarrow{k},l}$ are the ripplet coefficients and $\overline{(.)}$ denotes the conjugate operator. The details of RT can be found in [43].

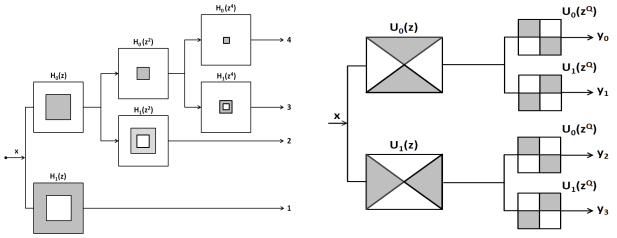


Fig. 6 (a) Non-subsampled Pyramid Filter Bank: Three-stage decomposition. (b) Four-channel NSDFB constructed with two-channel fan filter banks.

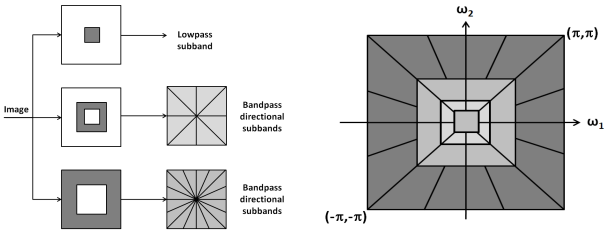


Fig. 7 Non-subsampled contourlet transform (a) NSDFB structure that implements the NSCT. (b) Idealized frequency partitioning obtained with the proposed structure.

2.1.5 Non-Subsampled Contourlet Transform (NSCT)

NSCT is a fully shift-invariant, multi-scale, and multi-direction expansion with fast implementability [44]. As opposed to the contourlet transform, which is not shift-invariant due to the presence of down-samplers and up-samplers in both the Laplacian pyramid and Directional Filter Bank (DFB) stages, NSCT achieves the shift-invariance property by using non-subsampled pyramid filter banks and non-subsampled DFB.

Non-Subsampled Pyramid Filter Bank (NSPFB) is a shift-invariant filtering structure that leads to a subband decomposition that resembles the Laplacian pyramid, which ensures the multi-scale property of the NSCT. As shown in Figure 6, it is constructed by using two-channel non-subsampled 2D filter banks, which produce a low-frequency and a high-frequency image at each NSPFB decomposition level. Filters at subsequent stages are obtained by upsampling the low-pass filters at the first stage. As a result, NSPFB can result in $\mathsf{k}+1$ sub-images, which consist of one low-frequency image and k high-frequency images whose sizes coincide with the source image, k being the number of decomposition levels. Figure 6(a) gives the NSP decomposition with $\mathsf{k}=3$ levels.

The Non-Subsampled Directional Filter Bank (NSDFB) is constructed by eliminating the downsamplers and upsamplers of the DFB and by upsampling the filters accordingly [44]. This results in a tree composed of two-channel NSDFB, described in Figure 6(b) (4 channel decomposition). At each stage of the NSPFB, the NSDFB allows a decomposition into any number of 2^l directions, l being the number of levels

in the NSDFB. This provides the NSCT with the multi-direction property and offers precise directional information. The combination between NSPFB and NSDFB is depicted in Figure 7(a). The resulting filtering structure approximates the ideal partition of the frequency plane displayed in Figure 7(b). Differently from the contourlet expansion, the NSCT has a redundancy given by $\mathsf{r} = 1 + \sum_{j=1}^{\mathsf{k}} 2^{\ell_j}$, where ℓ_j is the number of levels in the NSDFB at the jth scale. We refer to [44] for further details about NSCT.

2.2 Search Space Reduction Using Pre-classification Technique

To increase the efficiency of the retrieval system, pre-classification plays a significant role in CBIR paradigm, by reducing the search space. In this section, we briefly review three supervised classifiers, i.e., MLP, RF and LSSVM.

2.2.1 Multilayer Perceptron (MLP)

Neural network consists of several processing units, called neurons, arranged in layers that work in concert to create an end product. The MLP is the standard neural network to use for supervised learning [59]. The MLP classifier used during this study is a feed-forward neural network consisting of an input layer, hidden layer and output layer. The numbers of nodes in the input layer is equal to the dimension of the input feature vector and the number of nodes in the output layer is the number of classes to the image database. An important issues in MLP design, is choosing the number of hidden layers and also the number of nodes in these layers. Generally, one hidden layer is sufficient to solve vast majority of problems. The universal approximation theorem [60] claims that the MLP with a single hidden layer that contains finite number of hidden neurons, and with arbitrary activation function are universal approximators in $C(R^m)$. Moreover, it has been found empirically that networks with multiple hidden layers are more prone to getting caught in undesirable local minima which leads to poor performance [61]. Furthermore, MLP with single hidden layer have the property that it can approximate any function with arbitrary accuracy if a sufficient number of hidden nodes/neuron is used [62]. The determination of the number of hidden nodes in a hidden layer is a tricky problem and there is no strict guidelines in this regard. In the training phase, the number of hidden nodes in the hidden layer determines the mapping ability of the network. In one hand, using too few nodes in the hidden layers may result in high training and generalization error due to underfitting [63]. On the other hand, too many nodes in the hidden layers may result in low training error but still have high generalization error due to overfitting [63,61]. In, [64,65], authors discuss how the number of hidden units affects the bias/variance trade-off. The number of nodes of the hidden layer in this study are computed by considering the rule of thumb i.e., $\sqrt{(input\ nodes) * (output\ nodes)}$.

There are several factors that determine the optimal configuration of the network (having minimum error rate) i.e., numbers of iterations, momentum factors and different learning rates. The first layer is the input layer, second layer is the hidden layer and has a log sigmoid (log-sig) activation function, $logsig(n) = 1/(1+exp(-n))$ and the third layer, or output layer, has a linear activation function, $a = purelin(n) = n$. 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) [59]. Backpropagation is used to calculate the Jacobian JX of performance with respect to weight and bias variables X. Each variable is adjusted according to LM as shown in Eqn. (16, 17, 18), where I the identity unit matrix, E the error at the output and μ the learning parameter.

$$JJ = JX \times JX \tag{16}$$

$$Je = JX \times E \tag{17}$$

$$dX = -(JJ + \mu I)/Je \tag{18}$$

The learning function used in the proposed network is gradient descent with momentum weight/bias function as shown in Eqn. (19)

$$dW = mc * dW prev + (1 - MC) * LR * gW \tag{19}$$

where weight change dW for a given neuron was calculated from the neuron's input and error, the weight or bias (W), learning rate (LR), and momentum constant (MC) according to the gradient descent with momentum. A momentum term could be added to increases the learning rate with stability. The performance of the network is measured by Mean Squared Error (MSE), which can be quantitatively calculated. The smaller the MSE is, the better the network performs.

2.2.2 Random Forest (RF)

RF is a decision tree ensemble classifier with each tree grown using the same type of randomization. The leaf nodes of each tree are considered by estimating the posterior distribution over the image classes. Each internal node contains a test that best splits the space of data to be classified. An image is classified by sending it down every tree and aggregating the reached leaf distributions. The RF classifier has a capacity for processing huge amount of data at high training speed based upon the decision tree.

The trees are binary and constructed in a top-down manner. RF is a set of decision tree operating on a common feature space. The binary test at each node can be taken in one of two ways: (1) erratically or by a (2) greedy algorithm which picks the test that "best" separates the given training examples. In the training procedure, the evaluation measure used for selecting the "best" value is the information gain caused by partitioning the set S of examples into two subsets S_i according the given test. Here E(S) is the entropy $-\sum_{k=1}^{N} p_k log_2(p_k)$ with p_k the proportion of examples in q belonging to class k, and $|.|$ the size of the set. The process of selecting a test is repeated for each nonterminal node, using only the training examples falling in that node. Two stopping criteria's have been used in the iterative training procedure. The recursion is stopped either when the node receives too few examples, or when it reaches at the maximum depth of a tree. The maximum depth of a tree (D_{max}) and the numbers of trees (T) are the important parameters in RF [66, 67].

$$\Delta E = -\sum_{i} \frac{|S_i|}{|S|} E(S_i) \tag{20}$$

During the training stage, a leaf node has a posterior probability and the class distributions, $p(c|n)$, are estimated empirically as a histogram of the class labels, c_i, of the training examples, i, that reached node n.

When classifying the test image, it is used as input to the trained RF. The final class distribution is considered by averaging the posterior probabilities of each distribution of all trees $L = (l_1, l_2, , l_T)$ and choose c_i as the final class (f) of an input image if the final class distribution $p(c_i|L)$ has the maximum value [68].

$$f = \arg max\{\frac{1}{T}\sum_{t=1}^{T} P(c_i|l_t)\} \tag{21}$$

2.2.3 Least Squares Support Vector Machine (LSSVM)

The Support Vector Machine (SVM) is a state-of-the-art technique for pattern classification. The essential idea of SVM is to find a linear separating hyperplane which achieves the maximal margin among different classes of data. Furthermore, one can extend SVM to build nonlinear separating decision hyperplanes by using kernel trick. However, one critical drawback of SVM is its high computational complexity for high dimensional data sets. To reduce the computational demand, the least square version of SVM (LSSVM) is adopted as classifier in this paper. LSSVM avoids solving quadratic programming problem and simplifies the training procedure [69]. Considering a linearly separable binary classification problem:

$$(x_i, y_i)_{i=1}^{n} \quad and \quad y_i = \{+1, -1\} \tag{22}$$

where x_i is an n-dimensional vector and y_i is the label of this vector. LS-SVM can be formulated as the optimization problem:

$$\min_{w,b,e} \mathscr{J}(w,b,e) = \frac{1}{2}w'w + \frac{1}{2}C\sum_{i=1}^{n} e_i^2 \tag{23}$$

subject to the equality constraints

$$y_i[w'\varphi(x_i) + b] = 1 - e_i \tag{24}$$

where $C > 0$ is a regularization factor, b is a bias term, w is the weights vector, e_i is the difference between the desired output and the actual output and $\varphi(x_i)$ is a mapping function.

The lagrangian for problem of Eq.(23) is defined as follows:

$$\mathscr{L}(w, e_i, b, \alpha_i) = \min_{w,b,e} \mathscr{J}(w,b,e) - \sum_{i=1}^{n} \alpha_i\{y_i[w'\varphi(x_i) + b] - 1 + e_i\} \tag{25}$$

where α_i are Lagrange multipliers. The Karush-Kuhn-Tucker (KKT) conditions for optimality

$\frac{\partial \mathscr{L}}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i y_i \varphi(x_i)$; $\frac{\partial \mathscr{L}}{\partial e_i} = 0 \rightarrow \alpha_i = Ce_i$; $\frac{\partial \mathscr{L}}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0$; $\frac{\partial \mathscr{L}}{\partial \alpha_i} = 0 \rightarrow y_i[w'\varphi(x_i) + b] - 1 + e_i = 0$,is the solution to the following linear system

$$\begin{bmatrix} 0 & -Y \\ Y & \varphi\varphi' + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \overline{1} \end{bmatrix} \tag{26}$$

where $\varphi = [\varphi(x_1)'y_1, ..., \varphi(x_n)'y_n], Y = [y_1, ..., y_n]$, $\overline{1} = [1, ..., 1]$, and $\alpha = [\alpha_1, ..., \alpha_n]$.

For a given kernel function K(,) and a new test sample point x, the LS-SVM classifier is given by

$$f(x) = sgn[\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b] \tag{27}$$

LS-SVM was originally developed for binary classification problems. But, a number of methods have been proposed by various researchers for extension of binary classification to multi-classification problem. However, one need an appropriate method for solving this multi-class problem. It's essentially separate M mutually exclusive classes by solving many two-class problems and combining their predictions in various ways. One such technique which is commonly used is Pairwise Coupling (PWC) or "one-vs.-on" is to construct binary SVMs between all possible pairs of classes. PWC uses $M*(M-1)/2$ binary classifiers for M number of classes, each of which provides a partial decision for classifying a data point. During the testing of a feature, each of the $M*(M-1)/2$ classifiers votes for one class. The winning class is the one with the largest number of accumulated votes. Hsu et al., shows that the PWC method is more suitable for practical use [70]. The details of the LSSVM is discussed in [69].

3 Feature Extraction Mechanisms

In general, transform domain based CBIR systems use different transforms for representing the intrinsic characteristics of the images. There exist several state-of-the-art CBIR systems based on different transform domain techniques [54,47–49,56,50, 57]. Different algorithms have been developed for the extraction of features, but most commonly, first order statistical based features are used in CBIR paradigm due to less computational cost [54,71,72]. However, for better image representation, many researchers have used signature based image feature representation techniques. The advantage of using signatures is to gain an improved correlation between image representation and visual semantics [49,73]. Nevertheless, the generalized Gaussian distribution (GGD) based statistical framework is also used to model the distributions of the sub-bands, obtained by decomposing the image through various transforms and used GGD parameters (scale and shape) as feature vectors [47,57]. In this study, we have considered three different feature extraction schemes in the transform domain. Specifically, the feature extraction techniques are: (i) first order statistics (mean and standard deviation) of the sub-band coefficients, (ii) feature based on image signature computed from the sub-band coefficients and (iii) GGD parametric approach. In all the feature extraction mechanisms, an image I of size $X \times Y$ of the database is converted from RGB color space to Y-Cb-Cr color space, prior to transform based decomposition. In Y-Cb-Cr color space, the achromatic and chromatic information are separated, which is more suitable for Human Visual System (HVS). Transform decomposition over the intensity plane (Y), characterizes the texture information, while the decomposition over chromaticity planes (Cb and Cr) characterizes

color. Texture and color information are extracted by decomposing each color plane ($C_p = \{Y, Cb, Cr\}$) of the image I through a particular transform (Transform = {DWT, CVT, CNT, RT, NSCT}). In the following section, $S_j^{C_p}$ represents the j^{th} sub-band in the color plane C_p and $j = 1, 2, ..., J$, where J is the total number of sub-bands of the image plane C_p. The feature extraction mechanisms are discussed in details in the following subsections:

3.1 First Order Statistics Based Features Representation

In this feature representation technique, the feature vector of an image I is represented as follows:

$$F_1^I = [f_\mu^{S_j^Y}, f_\sigma^{S_j^Y}, f_\mu^{S_j^{Cb}}, f_\sigma^{S_j^{Cb}}, f_\mu^{S_j^{Cr}}, f_\sigma^{S_j^{Cr}}] \tag{28}$$

Here, $f_\mu^{S_j^{C_p}}$ and $f_\sigma^{S_j^{C_p}}$ represent the mean and standard deviation of the j^{th} subband of the color plane C_p respectively. The $f_\mu^{S_j^{C_p}}$ and $f_\sigma^{S_j^{C_p}}$ is computed as

$$f_\mu^{S_j^{C_p}} = \frac{1}{a \times b} \sum_{m=1}^{a} \sum_{n=1}^{N} |S_j^{C_p}(m,n)| \tag{29}$$

$$f_\sigma^{S_j^{C_p}} = \sqrt{\frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} (S_j^{C_p}(m,n) - f_\mu^{S_j^{C_p}})} \tag{30}$$

where, $S_j^{C_p}(m,n)$ represents transform coefficient at the spatial location (m,n) of the subband of size $M \times N$

3.2 Signatures Based Features Representation

In signature based image feature representation, an image I is represented as

$$F_2^I = [X_l^{S_j^Y}, w_l^{S_j^Y}, X_l^{S_j^{Cb}}, w_l^{S_j^{Cb}}, X_l^{S_j^{Cr}}, w_l^{S_j^{Cr}}] \tag{31}$$

Here, $X_l^{S_j^{C_p}}$ represents the centroid of the l^{th} cluster of the j^{th} subband of the color plane C_p. And $w_l^{S_j^Y}$ is the weight of the respective cluster. The centroids of the clusters is computed as follows:

For each sub-band $S_j^{C_p}$, a feature map $FM_j^{C_p}$ is computed by using Eq. (32) utilizing the concept of 'local energy' over a neighborhood W_{mn} of size $p \times q$, centered around a coefficient with coordinates (m,n). The size of the window W_{mn}, is determined using the Spectral Flatness Measure (SFM) as

$$FM_j^{C_p}(m,n) = \sum_{(p,q) \in W_{mn}} |S_j^{C_p}(p,q)| G(m-p, n-q) \tag{32}$$

and

$$G(m,n) = \frac{1}{\sqrt{2\Pi\sigma}} e^{-(\frac{1}{2\sigma^2})(m^2+n^2)} \tag{33}$$

where, $G(m,n)$ is a Gaussian low-pass (smoothing) filter and σ defines the spatial support of the averaging filter. Use of a Gaussian (weighting) window results in less sparse points (i.e., denser feature distributions) as compared to when uniform weighting window is used. After computing the feature maps for an image I of the database, each feature map is clustered using FCM clustering algorithm, into l different clusters. The details of this feature extraction technique are discueed in [45,49].

3.3 Gaussian Distribution Based Features Representation

The feature vector of an image I using GGD model is given by

$$F_3^I = [\alpha^{S_j^Y}, \beta^{S_j^Y}, \alpha^{S_j^{Cb}}, \beta^{S_j^{Cb}}, \alpha^{S_j^{Cr}}, \beta^{S_j^{Cr}}] \tag{34}$$

where, $\alpha^{S_j^{Cp}}$ and $\beta^{S_j^{Cp}}$ are the width and shape parameters of the GGD model. The procedure of computing these two parameters are discussed below:
Each sub-band S_j^{Cp} is modeled with GGD which is defined as,

$$p(x;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^\beta} \tag{35}$$

where x is the transformed subband coefficients (= S_j^{Cp}) and $\Gamma(.)$ is the gamma function, i.e., $\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt$, $z > 0$.

Here, the scale parameter α models the width of the probability distribution function (PDF) peak (standard deviation), while the shape parameter β is inversely proportional to the decreasing rate of the peak. These two parameters need to be estimated for feature vector creation. As Maximum Likelihood (ML) estimator is best suited for estimating heavy-tailed distribution like GGD for both small and large samples, we have used ML estimator in our proposed scheme.

For the sample set $x = (x_1, x_2, x_3, ..., x_k)$, x_i is the transform subband coefficients at the i^{th} subband, and $i \leq L$, the ML estimator is defined as [47].

$$L(x;\alpha,\beta) = log \prod_{i=1}^{L} p(x_i;\alpha,\beta) \tag{36}$$

GGD parameters are defined with the following equations, which have a unique root in probability

$$\frac{\partial L(x;\alpha,\beta)}{\partial \alpha} = -\frac{L}{\alpha} + \sum_{i=1}^{L} \frac{\beta |x_i|^\beta \alpha^{-\beta}}{\alpha} = 0 \tag{37}$$

$$\frac{\partial L(x;\alpha,\beta)}{\partial \beta} = \frac{L}{\beta} + \frac{L\Psi(1/\beta)}{\beta^2} - \sum_{i=1}^{L} \left(\frac{|x_i|}{\alpha}\right)^\beta log\left(\frac{|x_i|}{\alpha}\right) = 0 \tag{38}$$

where $\Psi(.)$ is the digamma function, i.e.,$\Psi(z) = \Gamma'(z)/\Gamma(z)$. α has a unique, real, positive solution and can be obtained from Eq.(37) by fixing $\beta > 0$:

$$\hat{\alpha} = \left(\frac{\beta}{L}\sum_{i=1}^{L} |x_i|^\beta\right)^{1/\beta} \tag{39}$$

Substituting this into (38), the shape parameter β is the solution of the following *transcendental* equation

$$1+\frac{\Psi(1/\hat{\beta})}{\hat{\beta}}-\frac{\sum_{i=1}^{L}|x_i|^{\hat{\beta}}\log|x_i|}{\sum|x_i|^{\hat{\beta}}}+\frac{log\left(\frac{\hat{\beta}}{L}\sum_{i=1}^{L}|x_i|^{\hat{\beta}}\right)^{1/\hat{\beta}}}{\hat{\beta}}=0 \tag{40}$$

which can be solved numerically using the Newton-Raphson iterative procedure.

4 Experimental Results and Comparisons

Extensive experiments were carried out to determine the best combination of transform domain, filters, feature representation and classifier. In this section, we first describe the experimental setup followed by results and comparisons.

4.1 Experimental Setup

We have studied the performance effectiveness of different combinations of transform domains, filters, feature representation and classifiers on three different benchmark image databases: Details of the datasets are given below:

- **ImgDb1:-** This is the Oliva database consists of 2688 images classified into 8 different classes, where the number of images in each classes varies from 260 to 409 images [74]. The 8 different scene categories are: coast, mountain, forest, open country, street, inside city, tall buildings and highways.
- **ImgDb2:-** This is the Caltech101 diverse objects database containing $9,196$ images which are classified into 101 categories, where the number of images in each category varies from 31 to 800 images [75]. The significant variation in color, pose and lighting makes Caltech101 database quite challenging.
- **ImgDb3:-** This is a subset of the Caltech256 dataset [76], obtained by randomly selecting 25 categories and 100 images per category is focused on single object only. In addition, we created 10 copies of each image corrupted with random rotation, blurring, salt-and-pepper and white noise. In total, we have 25000 images in Caltech256 dataset. This database contains 25 distinct object categories: binocular, bag, Ak47, guitar, jetplane, video-projector, telephone-box, waterfall, giraffe, hibiscus, teddy-bear, football-helmet, computer-monitor, horse, elephant, ostrich, dog, train, eagle, car, ship, zebra, dolphin, school-buses and sea-beaches.
- **ImgDb4:-** This is a diverse collection of realistic image database known as MIR-Flickr25000 [77]. This database contains $25,000$ color images and it is downloaded from $http://press.liacs.nl/mirflickr/$ website. This database contains 25 distinct image categories: animal, dog, bird, sky, clouds, snow, night, sea, river, lake, beach, building, bridge, street, city, graffiti, street-art, transport, trees, flower, landscape, people, girl, boy and sunset.

For DWT, four different types of filter families are considered in our experiments: Haar (dbN, N = 1), Daubechies (dbN, N = 2, 3, 6, 8, 10), Coiflets (coifN, N = 1,2,3,4,5) and Biorthogonal (biorN.N, N.N = 1.3, 2.2, 3.5, 4.4, 6.8). For CVT, CNT, RT and NSCT different pyramidal (pyr) and orientation (ori) filter combinations are considered. The number of decomposition level of DWT, considered in the study is three, while that of the advanced MRA/MGA tools (CVT, CNT, RT, and NSCT) can be set to the integer exponent of 2. For example $\{1,2,2\}$ means that the images

are decomposed into three levels, and the number of orientation from coarse to fine resolution are 1, 2, 2, respectively. It is important to point out that, all experiments are carried out in this paper with same set of decomposition level. This is because; large decomposition levels are sensitive to noise. Moreover, large decomposition levels consume more time and have higher memory requirements. When the number of decomposition levels is too small, the spatial information of an image cannot be captured well. The different filter combinations and used decomposition levels for four different MRA/MGA tools are shown below:

- CVT = {pyr: (5/3, 9/7); ori: (9/7, 5/3, pkva)}; decomposition = {1,1,2};
- CNT = {pyr: (5/3, 9/7); ori: (9/7, 5/3, pkva)}; decomposition = {1,2,2};
- RT = {pyr: (5/3, 9/7); ori: (9/7, 5/3, pkva)}; decomposition = {1,2,2};
- NSCT = {pyr: (9/7, pyrexc); ori: (sinc, pkva)}; decomposition = {1,2,2};

We represent the three feature representation schemes as F1, F2 and F3 where F1 represent the first order statistics based features, F2 = signature based features and F3 = GGD based features; Feature Representation = {F1, F2, F3}.

We have tested our network with two or more number of hidden layers. But it is observed that by increasing the number of hidden layers computational complexity of the MLP network increases many fold without proportional improvement in the performance. Hence, a single layer is selected in the present experiment and the number of nodes in the hidden layer is selected as discussed in Section 2.2.1. Furthermore, we have considered different parameter configurations in MLP: numbers of iterations ($iter$), momentum factors (μ), different learning rates (η) for different values and selected the best-performing configuration by 10-fold cross validation (CV) technique. Similarly, CV is also performed to obtain the initial parameters in RF: (i) number of trees (T) and (ii) maximum depth of the tree (D_{max}), to obtain the optimal configuration, having a better classification performance in terms of accuracy and computational time. In LSSVM, we have used the Radial Basis Function (RBF), $K(x_i,x_j) = exp(-\gamma \parallel x_i - x_j \parallel^2), \gamma > 0$, as the kernel for training. There are two tunable parameters while using the RBF kernel in LSSVM classifier: C and γ. The kernel parameter γ controls the shape of the kernel and regularization parameter C controls the tradeoffs between margin maximization and error minimization. It is not known beforehand which values of C and γ is the best for the classification problem at hand. Hence, a CV is conducted, where various pairs of (C, γ) are tried and the one with the lowest CV error rate is picked. After finding the best values for the parameters C and γ, these values are used to train the LSSVM model, and the test set is used to measure the error rate of the classification system. In all the three classifier CV is done on 60% of the train dataset and remnant 40% is used as test dataset for evaluating the performance of the classifiers used in our study. The selected parameters of the classifiers are shown in Table 1.

For detail assessment of system performance, we have selected different combinations of transforms, filters, feature representations and classifiers which is represented as "*Transform (Filters)* $\Rightarrow$ *Feature Representation* $\Rightarrow$ *Classifiers* in this article. In this study, we have used different transform e.g., DWT, CVT, CNT, RT and NSCT; Filters e.g., dbN, coifN, biorN.N, and (pyr, ori); Feature Representation in terms of

Table 1 Parameters of Classifiers

Image Database	Classifier	Parameters	DWT			CVT			CNT			RT			NSCT		
			F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
ImgDb1	MLP	iter	1500	2000	1600	1100	1800	1200	2000	3000	2000	2000	3200	1800	2200	3800	2300
		μ	0.55	0.53	0.56	0.54	0.51	0.54	0.65	0.95	0.75	0.66	0.85	0.672	0.714	0.769	0.821
		η	0.7	0.72	0.67	0.66	0.71	0.72	0.72	0.725	0.731	0.689	0.72	0.734	0.687	0.674	0.741
	RF	T	200	320	200	200	280	210	220	300	225	225	340	225	230	350	200
		D_{max}	20	30	20	20	30	25	20	30	20	25	30	25	25	30	25
	LSSVM	C	100	100	100	80	100	100	150	200	150	180	200	100	100	200	140
		γ	0.02	0.025	0.035	0.045	0.036	0.038	0.045	0.251	0.161	0.175	0.214	0.021	0.0241	0.219	0.177
ImgDb2	MLP	iter	2100	2400	2200	2000	2500	2000	2200	2900	2200	2300	3500	2100	2400	4000	2500
		μ	0.65	0.712	0.675	0.613	0.701	0.661	0.667	0.789	0.642	0.665	0.775	0.641	0.69	0.81	0.85
		η	0.714	0.742	0.75	0.732	0.82	0.744	0.74	0.831	0.713	0.747	0.825	0.731	0.779	0.85	0.88
	RF	T	220	340	220	200	320	200	230	360	230	240	360	230	240	370	240
		D_{max}	25	35	25	20	35	20	25	30	20	20	35	20	20	30	25
	LSSVM	C	200	220	220	240	210	200	230	280	210	240	270	240	220	290	260
		γ	0.252	0.242	0.231	0.32	0.301	0.281	0.34	0.42	0.324	0.351	0.294	0.210	0.246	0.46	0.45
ImgDb3	MLP	iter	3000	3800	3100	2900	3400	3000	3200	4300	3300	3500	4500	3200	3300	4400	3200
		μ	0.675	0.812	0.715	0.693	0.781	0.751	0.721	0.844	0.711	0.689	0.921	0.722	0.725	0.91	0.882
		η	0.781	0.729	0.812	0.751	0.807	0.866	0.843	0.889	0.752	0.783	0.924	0.697	0.727	0.944	0.913
	RF	T	480	520	450	410	500	560	600	570	500	540	600	500	530	600	500
		D_{max}	40	45	45	35	50	50	50	50	50	50	50	50	45	50	40
	LSSVM	C	300	340	310	320	360	350	400	370	330	350	360	320	320	400	300
		γ	0.325	0.491	0.321	0.337	0.345	0.394	0.441	0.522	0.40	0.483	0.510	0.379	0.384	0.521	0.328
ImgDb3	MLP	iter	3200	4000	3300	3200	3400	3000	3500	4500	3200	3400	4700	3200	3200	5000	3400
		μ	0.673	0.823	0.718	0.691	0.784	0.763	0.724	0.841	0.713	0.684	0.919	0.721	0.726	0.921	0.880
		η	0.772	0.723	0.823	0.758	0.812	0.852	0.841	0.891	0.754	0.778	0.91	0.701	0.721	0.945	0.924
	RF	T	470	530	460	390	520	580	590	600	540	520	600	530	520	600	500
		D_{max}	40	50	40	45	50	50	40	45	40	40	40	45	50	50	45
	LSSVM	C	320	350	330	310	350	3750	420	320	310	360	370	310	330	420	310
		γ	0.321	0.492	0.325	0.339	0.338	0.389	0.445	0.539	0.423	0.476	0.528	0.389	0.3814	0.520	0.327

F1, F2 and F3; and the Classifiers used are MLP, RF and LSSVM. They are used in different combination that are discussed in the experimental section.

There is no general rule for selecting appropriate similarity measure for CBIR paradigm. In all the experiments, we have used Euclidean distance (ED) as the similarity measure instead of other distance measures as it is less computationally complex. In the current work, the primary objective is to select the best combination of wavelet based image filters and classifiers that may provide better results for CBIR system using any simplest distance measure. In the present experiments, we have studied the large varieties of wavelet based image filters for feature extraction and varieties of classifiers for initial classification of different class of images in large image databases. The main intention of such classification is to use a fraction of the total database for searching images similar to the input query image (class) which will drastically reduce the total searching time. It is reported in the literature that Earth Mover's Distance (EMD) [78] improved the performance of CBIR because it capture human perception mechanism more accurately than any other similarity/dissimilarity measure but its computation cost is much higher than ED [49,79]. The computation cost of similarity/dissimilarity measure for EMD is $O(N^3 logN)$ [80] against ED is $O(N)$ [81]. So, we restrict our attention on ED similarity/dissimilarity measure in the current study. From the above discussion, it can be inferred that if ED is giving

satisfactory results with our selected best combination then the use of other distance measures will increase the retrieval results furthermore with additional computational cost. But depending upon the application, user may prefer speed over accuracy or vice versa.

4.2 Evaluation Criteria

To find out the best possible combination "*Transform (Filters) ⇒ Feature Representation ⇒ Classifiers*", three statistical measures are used in this paper: Average Precision (AP), Average Recall (AR) and F-Measure (F). We have computed the precisions and recalls considering all the images of the used databases as the query images, and then take the average of the obtained precision and recall values over all the images as the final evaluation result. The three statistical measures are defined as follows:

$$Precision\ (P) = \frac{N_{RIR}}{N_{RIR} + N_{IRIR}} \tag{41}$$

and

$$Recall\ (R) = \frac{N_{RIR}}{T_{RID}} \tag{42}$$

$$F = 2 * (\frac{P * R}{P + R}) \tag{43}$$

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. During the experiments, top 20 retrieved images were used to compute the precision and recall values.

4.3 Results and Discussion

In this section, firstly we have evaluated the performances of the different combinations "*Transform (Filters) ⇒ Feature Representation ⇒ Classifiers*",individually. After finding the best combination for each MRA/MGA tool, we have compared these best combinations to find out the overall winner in terms of F-Measure and the best combination are further analyzed by AP Vs AR graph.

4.3.1 Retrieval with DWT

Fig. 8 represents the retrieval results obtained using different combinations of "*Transform (Filters) ⇒ Feature Representation ⇒ Classifiers*" for DWT. The graph of Fig. 8(a) shows that the combination "*DWT (bior6.8) ⇒ F2 ⇒ LSSVM*" is giving satisfactory performance on ImgDb1 database. The combination of " *DWT (bior6.8) ⇒ F3 ⇒ LSSVM* " is giving acceptable results on ImgDb2 and ImgDb4 image databases as shown in Fig. 8(b) and Fig. 8(d) respectively. But from Fig. 8(c), it has been observed that the combination "*DWT (bior6.8) ⇒ F3 ⇒ RF* " is showing the best performance as compared to other combination for ImgDb3 image database.

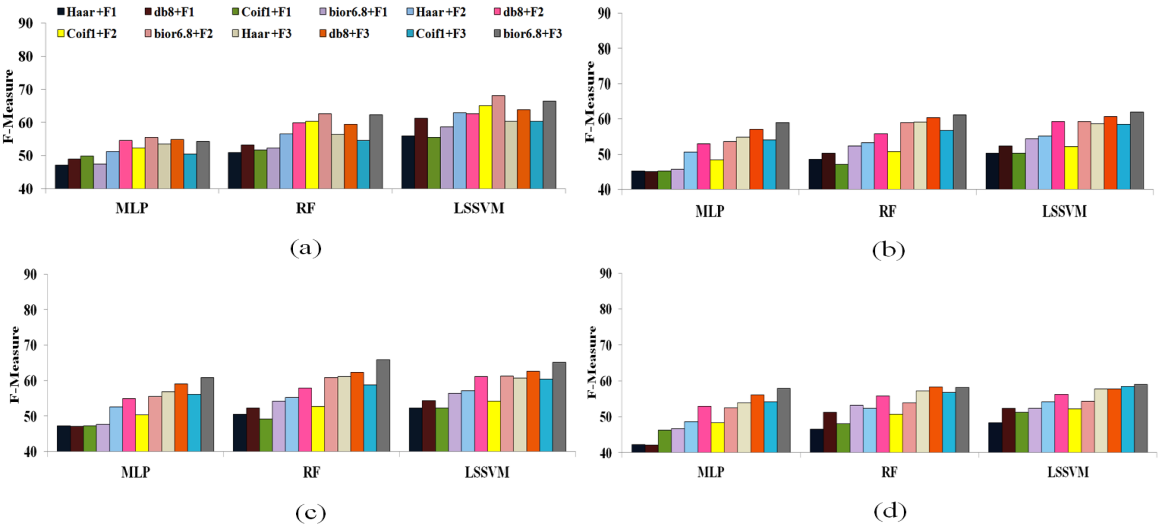


Fig. 8 F-Measure based performance graphs of DWT on (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d) ImgDb4

Therefore, "*bior6.8*" filters show good performance with F2 and F3 in all four image databases. Due to the symmetry, optimal time-frequency localization properties and the appropriate dealing with image signals with smooth changing patterns, "*bior6.8*" filter performs better than the other member of the DWT filter family. The property of "symmetry" provides linear phase and also useful in avoiding boundary artifacts in images. Moreover, "bior6.8" is the short filter that helps to minimize the computational complexity. From the graph, it can also be observed that F2 is giving good performance on scene (ImgDb1) database whereas F3 is giving satisfactory performance on both object (ImgDb2 and ImgDb3) and realistic image (ImgDb4) databases. The best results are listed in Table 2 for global comparison.

4.3.2 Retrieval with CVT

The performance graphs of CVT using various combinations of "*Transform (Filters) ⇒ Feature Representation ⇒ Classifiers*" are shown in Fig. 9. From the Fig. 9(a) and Fig. 9(c), it can be seen that, the combination of " *CVT (9/7, pkva) ⇒ F3 ⇒ MLP* " performs best for ImgDb1 and ImgDb3 image databases respectively. But for the ImgDb2 database, the combination of "*CVT (9/7, pkva) ⇒ F3 ⇒ RF*" is giving the best performance as shown in Fig. 9(b). The combination of "*CVT (9/7, pkva) ⇒ F3 ⇒ LSSVM*" is showing the good performance on ImgDb4 as shown in Fig. 9(d). Due to strong orientation-sensitive property, CVT is suitable for detecting curved singularities in images. Therefore, CVT is very useful in representing the edges of images. CVT has multi-resolution, band pass, and directional property, which possesses the three characteristics of good image representation as compared with DWT with less number of feature vectors.

The filter combination "*9/7*" and "*pkva*" filters are giving satisfactory results on both scene and object based image databases. The acronym "*9/7*" refers to the length

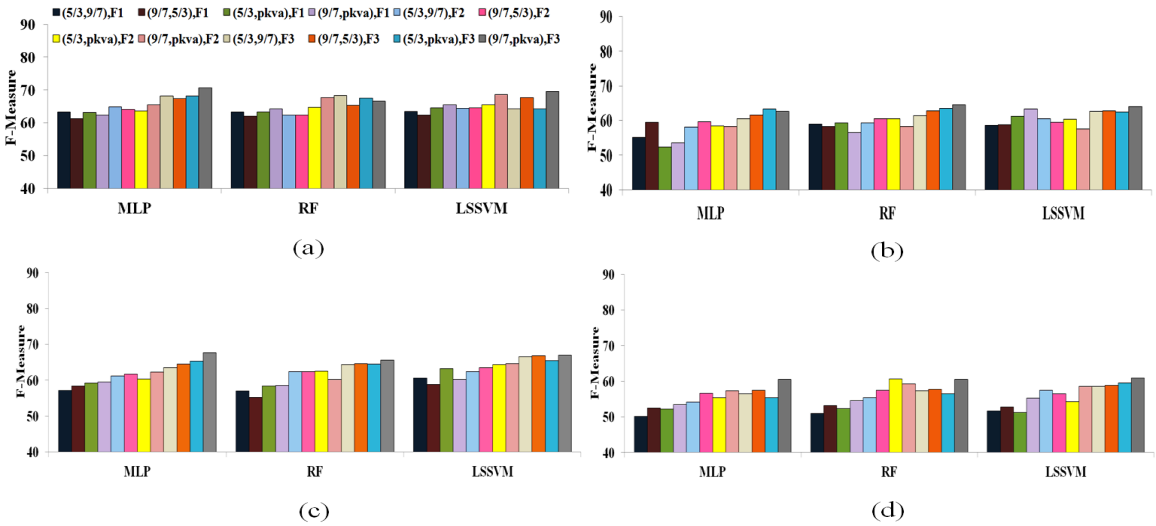


Fig. 9 F-Measure based performance graphs of CVT on (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d) ImgDb4

of the dual scaling filter (denote the number of filter taps) having four vanishing moments. The filter with fewer vanishing moments gives less smoothing and remove less details, but the filter with more vanishing moments produce distortion. It is known that the ratios of the "*9/7*" scaling functions are always closer to one, which indicates that they are closer to orthogonality with less computational complexity, compactly supported and easy implementation. On the other side, "*pkva*" filter is used to capture the high frequency content of the images like smooth contours and directional edges. It is of quincunx/fan filter type with best PSNR performance. Thus, the combination of "*9/7*" and "*pkva*" filter is best for preserving the more subtle image information as compared with other combination of the filters. Along with this filter combination, F3 is giving satisfactory results on both scene and object image database. These results are listed in Table 2 for global comparison.

4.3.3 Retrieval with CNT

The implementation of the CNT is based on pyramid filtering (pyr) and orientation filtering (ori) as discussed in Section 2.1.3. We have compared different combinations of pyramid and orientation filters with three different types of classifiers. The graph of Fig. 10(a) shows that the combination "*CNT (9/7, pkva)* $\Rightarrow$ *F2* $\Rightarrow$ *LSSVM*" is giving satisfactory results on ImgDb1 database, as compared with other combinations in the same database. It can be seen from the Fig. 10(b)-(c) that the combination "*CNT (9/7, pkva)* $\Rightarrow$ *F3* $\Rightarrow$ *LSSVM*" is giving the best results for ImgDb2 and ImgDb3 database. However, the combination "*CNT (9/7, pkva)* $\Rightarrow$ *F3* $\Rightarrow$ *RF*" is showing the acceptable results on ImgDb4 image database as depicted in Fig. 10(d). From the graph, it has been observed that "*9/7*" and "*pkva*" filter are giving satisfactory performance on scene, object and realistic scenario based image database. F2 and F3 are giving best performances on scene and object image databases particularly. However, as ImgDb4

is consist of realistic use scenario (i.e., images contain a variety of objects, people, lighting conditions and poses) where F3 is also giving satisfactory performance.

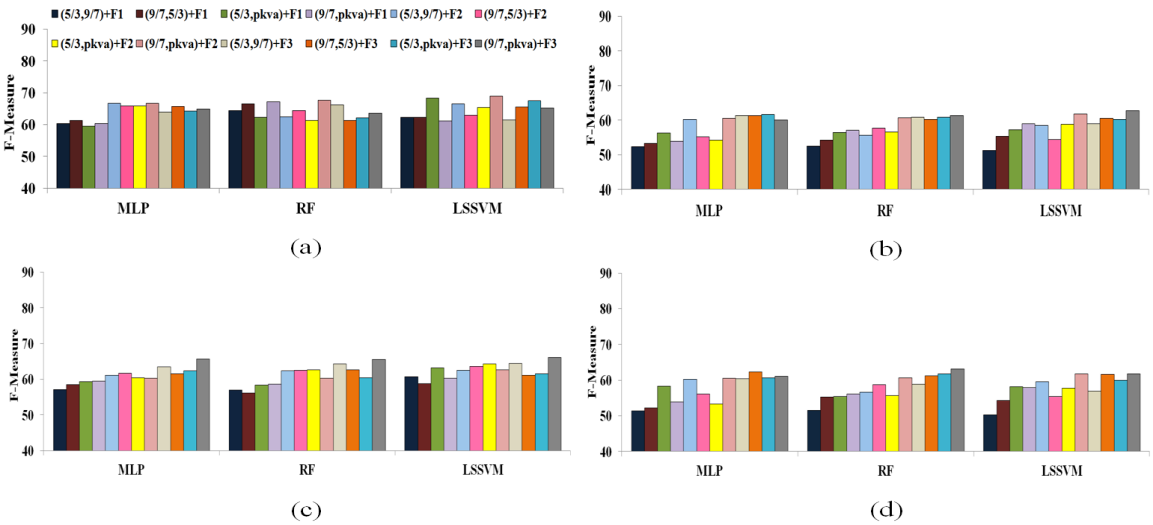


Fig. 10 F-Measure based performance graphs of CNT on (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d) ImgDb4

Comparing Fig. 9 and Fig. 10, it can be inferred that retrieval accuracy using CNT is lower than CVT. This is due to less directional features of CNT which leads to artifacts in images [41]. The key difference between CNT and CVT is that the CVT is initially developed in the continuous domain and then discretized for sampled data, whereas CNT starts with a discrete-domain construction and then studies its convergence to an expansion in the continuous domain. Due to the discrete domain construction, CNT is computationally less expensive than CVT. These results are listed in Table 2 for global comparison.

4.3.4 Retrieval with RT

The retrieval results using RT are shown in the graph of Fig. 11(a)-(d) in terms of F-Measure values. From the Fig. 11(a) and Fig. 11(c), it can be seen that the combination of " *RT (9/7, pkva)⇒ F3 ⇒ LSSVM* " is giving the best results as compared with other combination in ImgDb1 and ImgDb3 image database. We can see from the graph in Fig. 11(b) and Fig. 11(d) is that the combination of " *RT (9/7, pkva) ⇒ F3 ⇒ RF*" performs fairly in ImgDb2 and ImgDb4 image database. F3 based on RT features performs quite satisfactory on scene, object and realistic scenario based image databases with same filter combination.

The RT is a generalized version of CVT with addition of two parameters i.e., support and degree. The parameter "support" controls the number of directions in the high-pass bands and the parameter "degree" controls how the number of directions changes across bands. Therefore, these parameters help RT to achieve the anisotropy capability that guarantees to capture singularities along the arbitrary-shaped curves

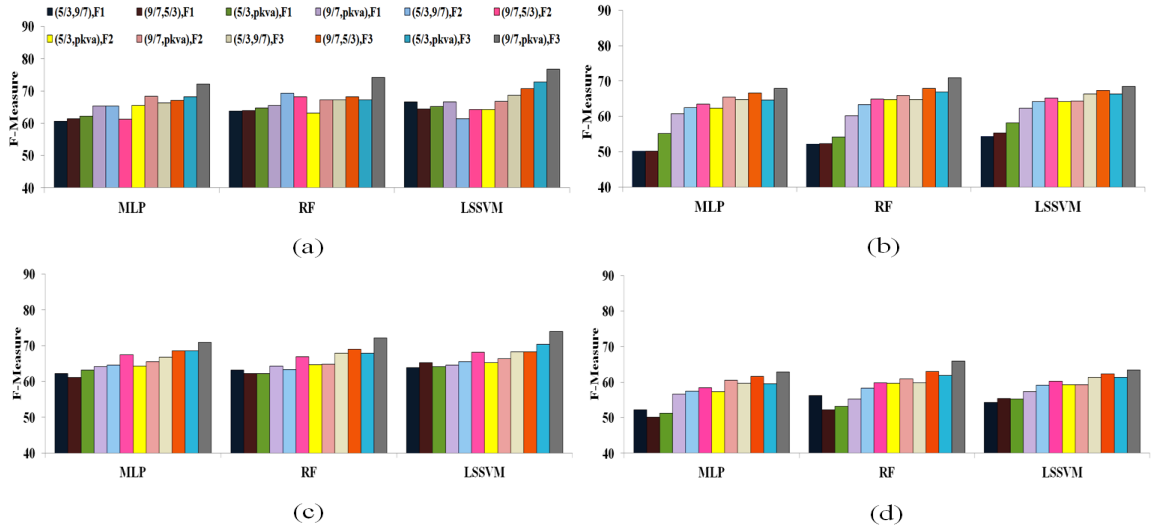


Fig. 11 F-Measure based performance graphs of RT on (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d) ImgDb4

effectively with different scales and different directions. Hence, edges are represented more efficiently in RT. Furthermore, RT is well localized in both spatial and frequency domains with higher energy concentration ability as compared with CVT. Since the RT successively approximates images from coarse to fine resolutions, it provides hierarchical representation of images. Moreover, RT provides a new tight frame with sparse representation for images with discontinuities along C^d curves as compared to C^2 curves in CVT and CNT (for resolving the 2D singularities). Due to all these characteristics and added advantage of "*9/7*" and "*pkva*" filters, RT performs quite satisfactorily in scene, object and realistic scenario based image database as compared with CVT and CNT as observed form the Fig. 9, Fig. 10 and Fig. 11. The best results are listed in Table 2 for global comparison.

4.3.5 Retrieval with NSCT

The implementation of the NSCT is based on NSPFB and NSDFB. Different from CNT, these filters are upsampled in each scale. Two categories of pyramid filters ("*9/7*" and "*pyrexc*") and two categories of directional/orientation filters ("*sinc*" and "*pkva*") are compared. We investigated all four groups of two filters in this paper. For each filter grouping, the setting of decomposition levels is same as CNT.

Fig 12(a)-(d) shows that LSSVM classifier performs better as compared to other classifiers (RF and MLP) in respect of F-measure for the combination of " *NSCT (pyrexc, sinc)* $\Rightarrow$ *F3* " in all the four Image databases (ImgDb1, ImgDb2, ImgDb3 and ImgDb4). For NSPF, we can see that "*pyrexc*" and for NSDFB, "*sinc*" perform the best for all four image databases (ImgDb1, ImgDb2, ImgDb3 and ImgDb4). NSPF is set as the "*pyrexc*" filter which is derived from 1D using the maximally flat mapping function with two vanishing moments but exchanging two high pass filter. Similarly, NSDFB is set as "*sinc*" filter which removes all frequency components above a given cutoff frequency, without affecting lower frequencies. Therefore, the smooth region

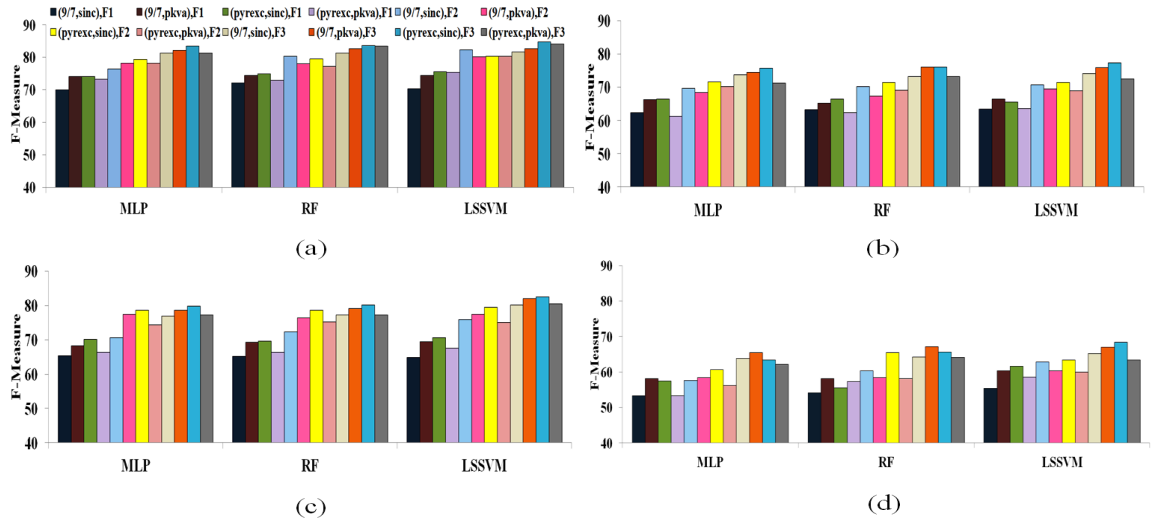


Fig. 12 F-Measure based performance graphs of NSCT on (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d) ImgDb4

are efficiently represented by the small size low pass image by using the "*sinc*" filter while the smooth edges are efficiently represented by "*pyrexc*" filter. Thus , by choosing the "*pyrexc*" and "*sinc*" filter combination with linear phase approximation, image is more efficiently represented in NSCT by means of good frequency localization and better image subband decomposition.

From the graph of Fig. 12, it has also been observed that "*pyrexc*" and "*sinc*" filters with F3 are giving promising results as compared with the others filters and features (F1 and F2) combination on both benchmark scene (ImgDb1) and object (ImgDb2 and ImgDb3) image databases. Moreover, it has also been seen that F3 is giving satisfactory performance on realistic scenario based image database (ImgDb4). Therefore, from the graph of Fig. 12(a)-(d) it can be inferred that the combination "*NSCT (pyrexc, pkva)* $\Rightarrow$ *F3* $\Rightarrow$ *LSSVM*" is giving the superior results for all four image databases.

Comparing Fig. 10 and Fig. 12, it is seen that the retrieval accuracy is higher in NSCT as compared to CNT because NSCT features has fully shift-invariant, multiscale, and multidirection information of the images. It has better frequency selectivity thereby achieving better subband decomposition. LSSVM classifier is giving the best results on all three image database. These results are listed in Table 2 for global comparison.

4.3.6 Global Comparison

Table 2 lists the overall best results of each type of transforms. From Table 2, it has been observed that NSCT is performing better than RT, CVT, CNT and DWT. This is mainly due to the multi-directional, shift-invariant and flexible multi-scale image decomposition property of NSCT. Table 2 also shows that, RT generally performs better than DWT, CVT and CNT. This is because RT has higher energy concentration ability as well as anisotropy property, which guarantee to capture singularities along

Table 2 The best results of each multi-resolution transform

Image Database	Transform	Filters		Feature	Classifier	F-Measure
ImgDb1	DWT	bior6.8	-	F2	LSSVM	68.14
	CVT	9/7	pkva	F3	MLP	70.65
	CNT	9/7	pkva	F2	LSSVM	68.95
	RT	9/7	pkva	F3	LSSVM	76.81
	NSCT	pyrexc	sinc	F3	LSSVM	**84.74**
ImgDb2	DWT	bior6.8	-	F3	LSSVM	61.97
	CVT	9/7	pkva	F3	RF	64.56
	CNT	9/7	pkva	F3	LSSVM	62.75
	RT	9/7	pkva	F3	RF	70.87
	NSCT	pyrexc	sinc	F3	LSSVM	**77.29**
ImgDb3	DWT	bior6.8	-	F3	RF	65.85
	CVT	9/7	pkva	F3	MLP	67.82
	CNT	9/7	pkva	F3	LSSVM	66.14
	RT	9/7	pkva	F3	LSSVM	73.92
	NSCT	pyrexc	sinc	F3	LSSVM	**82.47**
ImgDb4	DWT	bior6.8	-	F3	LSSVM	58.91
	CVT	9/7	pkva	F3	LSSVM	60.97
	CNT	9/7	pkva	F3	RF	63.14
	RT	9/7	pkva	F3	RF	65.84
	NSCT	pyrexc	sinc	F3	LSSVM	**68.42**

various curves. Therefore, on average NSCT is the best transform as compared with other transforms (DWT, CVT, CNT and RT) for representing the low level features of images.

Table 3 Dimension of the various features in different transforms

Transform \ Feature	F1	F2	F3
DWT	60	93	60
CVT	54	84	54
CNT	66	102	66
RT	66	102	66
NSCT	66	102	66

Feature representation plays a vital role in CBIR paradigm. The dimension of different features (F1, F2 and F3) using different transforms (DWT, CVT, CNT, RT and NSCT) are shown in Table 3. The dimension of F1 and F3 feature set are same because, in these two feature representation techniques, we have computed two statistical parameters for F1 (mean and standard deviation) and F3 (scale and shape) feature set, respectively. The total number of sub-bands in DWT is 10, therefore in three planes; we have total 30 sub-bands. Considering two statistical parameters, in total an image is represented by 60 feature vectors. Similarly, F1 and F3 feature set are computed for CVT, CNT, RT and NSCT with 9, 11, 11 and 11 sub-bands respectively. But for F2 feature computation in DWT, the number of feature maps for each image of the database is 30 ($= 10 \times 3$; 3 color planes per image and 10 feature map per color plane), and the number of clusters for each feature map is kept at 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. Therefore, a right

kind of balance is desirable between computational cost and precision of the results. Increasing the number of clusters may include finer segmentation details. As a result, the uncertainties of characterizing the perceptual content may increase. So, in an unsupervised mode of clustering, the number of clusters should always be chosen appropriately based on the expected outcome. Therefore, the dimension of F2 feature becomes 93 $(= 90(= 3 \times 30) + 3)$. Similarly, F2 feature set is computed for other transforms (CVT, CNT, RT and NSCT) with different number of sub-bands as discussed above.

Comparing Table 2 and Table 3, F3 feature set is giving satisfactory performance on almost all the transform domains with less number of feature vectors. GGD based F3 feature vectors model the texture, edge, color and geometrical invariant information (in all direction) as a probability inference problem, which captures the statistical pattern of the images with minimum computational time. Use of GGD based image representative feature vectors is motivated by psychological research on HVS [82, 83]. The dimension of F1 and F3 features set is small as compared to F2 feature set. F1 feature set is the first order statistical based features used for representing the images. F1 feature set cannot capture the edge information accurately as described by the HVS. F2 feature set is dependent on the number of the cluster in the FCM, whereas F3 feature set does not depend on any parameters. One of the major drawbacks of F2 feature set that makes it computationally expensive is the large dimensionality as compared to F1 and F3 feature set. Notwithstanding, the FCM works well on the majority of noise-free images, it is very responsive to noise and other imaging artifacts, since it does not consider any information about spatial perspective. To compensate this drawback of FCM, a preprocessing image smoothing step has been incorporated by considering the SFM in F2 features computation. But, this process makes the F2 feature set computation more complex. Therefore, F3 feature is chosen as a best feature vector as compared to other discussed methods.

Table 2, also shows the performance comparison of three different types of classifier i.e., MLP, RF and LSSVM. LSSVM classifier is performing quite satisfactorily as compared with RF and MLP classifiers, whereas retrieval results using RF classifier show better results than MLP classifier. This is because the optimal architecture of MLP classifier is not easy to obtain; moreover, it is inclined to become stuck in local minima during the training stage. In addition, the MLP classifier method shows its inferiority on high dimensional data. RF classifier performed better than MLP classifier, but required long computation time due to the generation of a large number of bootstrapped trees for decision making. Thus, RF classifier has faster classification speed, but is significantly slow in training and the complexity of training can grow exponentially with the number of classes [84]. On the other hand, LSSVM classifier uses least square loss function to obtain a set of linear equations in dual space so that learning rate is faster and the complexity of calculation in convex programming (in SVM) is relaxed [85].

Fig. 13, shows the AP Vs AR graph considering frames of different sizes (5,10,15 and 20) obtained from four image databases, respectively. We can clearly see from the graph that the retrieval accuracy (in terms of AP and AR) is higher in ImgDb1 and ImgDb3 as compared to ImgDb2 and ImgDb4 image databases. The reason for lower AP and AR values in ImgDb2 and ImgDb4 image databases is because of the large

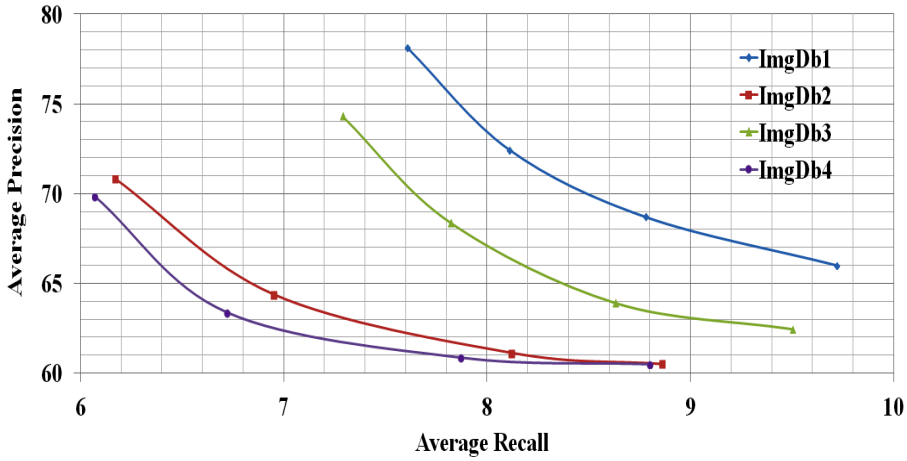


Fig. 13 Average precision Vs Average Recall graph of the combination "NSCT (pyrexc, pkva)⇒ F3 ⇒ LSSVM'

intraclass variation, for example in ImgDb2: images of the class brain is almost similar to the class bonsai; class aeroplane is very much similar to the class helicopter, and many more. It has also been observed that ImgDb2 database consists of images on which object information are textureless (e.g., beaver and cougar), object that camouflage well in their environment (e.g., crocodile) or thin objects (e.g. ant). Similarly in ImgDb4, some of the images that belongs to different animal classes have close similarity with that of some images in dog classes (overlapping classes). Moreover in different animal classes not only real life animal pictures are there but also it contain hand drawn images as well as different parts of the animal body as individual images. So, this sort of aberration in the picture collection contradicts the very claim of the real life image collection. Thus, ImgDb2 and ImgDb4 image databases are more challenging than ImgDb1 and ImgDb3 database.

ImgDb1 database consist of scene images. Scene images are of two types: man-made and natural scene images. Man-made scenes are characterized by horizontal and vertical structure, where more anisotropy and more directional based features play an important role (e.g., tall buildings). Similarly, natural scenes are rich in both color and texture information (e.g., forest). Therefore, scene images should be represented by a set of features that describe the global spatial structure of the scene, as spatial information is very useful for scene based image retrieval. ImgDb2 and ImgDb3 are object based database consist of images having single object. From the literature of object based image retrieval, it is clear that color, texture, shape and geometrical invariant information are mainly used for object based recognition/retrieval and classification. Shape information is outside the scope of this paper. ImgDb4 is a complex image database consist of a variety of image classes like scene, man-made object (building, bridge, etc.), natural object (bird, dog, etc.), hand drawn images and many others which makes the task of image retrieval harder in ImgDb4 than from any other standard benchmark image database. As NSCT based F3 features (with the combination "*pyrexc*" and "*sinc*" filters) incorporates color, texture, anisotropy, more directional and geometrical shift-invariant information which are important for both scene and object based image retrieval. Therefore, the combination " *NSCT (pyrexc,*

Table 4 Performance with reduced features set

Image Database	Steps	NSCT(pyrexc,sinc)⇒F3⇒LSSVM	
		No. of features (F3)	F-Measure
ImgDb1	W/O feature Evaluation	66	84.74
	With feature Evaluation	33	84.67
	With feature Evaluation	17	70.14
ImgDb2	W/O feature Evaluation	66	77.29
	With feature Evaluation	33	77.00
	With feature Evaluation	17	60.73
ImgDb3	W/O feature Evaluation	66	82.47
	With feature Evaluation	33	82.43
	With feature Evaluation	17	65.27
ImgDb4	W/O feature Evaluation	66	68.42
	With feature Evaluation	33	68.14
	With feature Evaluation	17	42.31

pkva) $\Rightarrow$ *F3* $\Rightarrow$ *LSSVM* " performs better than the other combinations in ImgDb1, ImgDb2, ImgDb3 and ImgDb4 image databases.

It is generally believed that a better image recognition can be achieved with more feature descriptors used, but this is not always true. Not all features are helpful for image recognition. Ill features are actually interfering signals and cause a drop in the recognition rate, especially if the effect of the ill features exceeds that of the effective ones. The goal of feature selection is to select the best features from the original ones. It can not only achieve high recognition rate but can also simplify the calculation of image retrieval. This paper adopts Maximal Information Compression Index (MICI) based feature evaluation mechanism to choose the optimal features from the best combination [86]. Table 4 summarizes the performance of " *NSCT (pyrexc, sinc)* $\Rightarrow$ *F3* " based reduced features on the three image databases using MICI. Therefore, from the above detailed discussion," *NSCT (pyrexc , sinc)* $\Rightarrow$ *F3* $\Rightarrow$ *LSSVM*" combination are selected for achieving the best retrieval results in large color image database with 33 optimal features.

To further justify our proposed model, we have compared "*NSCT (pyrexc, sinc)* $\Rightarrow$ *F3*" combination with three different types of spatial features based method such as Spatial Pyramid Matching (SPM) [87], Histograms of Oriented Gradient (HOG) [88] and SIFT [35] as shown in Table 5. Here, we have used LSSVM as classifier. It is seen from the Table 5 that the classification accuracy for SPM is slightly better than our proposed system by a factor of 0.128% in ImgDb1 and for ImgDb2, SIFT perform marginal better than our proposed system by a factor of 0.095%. However, for ImgDb3 and ImgDb4 the proposed technique outperform the SPM, HOG and SIFT by a considerable amount so far as classification accuracy is concerned. If we consider the average classification accuracy of the proposed technique for all the databases together, then it is observed that our selected combination is superior to that of SPM, HOG and SIFT based method by a factor of atleast 1% or more. It is also to be noted that the feature dimension used for SPM, HOG and SIFT are 4200, 128 and 128 respectively, which are very high as compared to that of "*NSCT (pyrexc, sinc)* $\Rightarrow$ *F3*"(66 features dimension). So, it is obvious that the computational cost of the proposed technique will be much lower to that of SPM, HOG and SIFT.

Different algorithms have been implemented using MATLAB R2014a on a Dell Precision T7400 workstation. The total processing time of the query images is com-

Table 5 Comparisons with other existing CBIR systems in terms of Classification Accuracy (%)

Image Database	SPM	HOG	SIFT	NSCT(pyrexc,sinc)⇒F3
ImgDb1	**85.802**	75.652	82.153	85.674
ImgDb2	74.048	73.414	**78.748**	78.653
ImgDb3	72.194	78.596	81.534	**82.431**
ImgDb4	58.615	63.025	70.239	**70.284**
Average	72.665	72.672	78.169	**79.261**

Table 6 Approximate feature extraction time of each image in each multi-resolution transform

Transform	Time(in seconds)		
	F1	F2	F3
DWT	0.85	1.88	1.82
CVT	1.47	2.27	2.21
CNT	1.25	2.07	2.00
RT	1.94	2.78	2.71
NSCT	2.83	3.32	3.25

Table 7 Approximate processing time and the memory usage in best combination on different image databases

Image Database	Mechanism	Processing Time (in seconds)	Memory Usage (in MB)
ImgDb1	DWT(bior6.8)⇒F2⇒LSSVM	4.66	1.80
	CVT(9/7,pkva)⇒F3⇒MLP	4.71	1.03
	CNT(9/7,pkva)⇒F2⇒LSSVM	5.86	1.94
	RT(9/7,pkva)⇒F3⇒LSSVM	5.39	1.26
	NSCT(pyrexc, sinc)⇒F3⇒LSSVM	5.83	1.28
ImgDb2	DWT(bior6.8)⇒F3⇒LSSVM	6.39	3.60
	CVT(9/7,pkva)⇒F3⇒RF	6.75	3.38
	CNT(9/7,pkva)⇒F3⇒LSSVM	7.81	4.16
	RT(9/7,pkva)⇒F3⇒RF	8.62	4.15
	NSCT(pyrexc, sinc)⇒F3⇒LSSVM	8.89	4.18
ImgDb3	DWT(bior6.8)⇒F3⇒RF	10.38	10.3
	CVT(9/7,pkva)⇒F3⇒MLP	10.73	9.60
	CNT(9/7,pkva)⇒F3⇒LSSVM	11.64	11.32
	RT(9/7,pkva)⇒F3⇒LSSVM	11.95	11.31
	NSCT(pyrexc, sinc)⇒F3⇒LSSVM	12.40	11.35
ImgDb4	DWT(bior6.8)⇒F3⇒LSSVM	10.66	10.4
	CVT(9/7,pkva)⇒F3⇒LSSVM	10.80	9.64
	CNT(9/7,pkva)⇒F3⇒RF	11.31	11.34
	RT(9/7,pkva)⇒F3⇒RF	11.67	11.33
	NSCT(pyrexc, sinc)⇒F3⇒LSSVM	12.42	11.36

puted by considering the feature extraction, classification and retrieval process. Time taken for extraction of three different types of features (F1, F2 and F3) for each image using five different transform (DWT, CVT, CNT, RT and NSCT) are shown in Table 6. Table 7 shows the processing time and memory usage of best combination. From the Table 7, it can be seen that the combination "*NSCT(pyrexc, sinc)⇒F3⇒LSSVM*" performs satisfactorily on time complexity as well as memory usage with a quite high F-Measure as compared to other selected combinations. The average retrieval time of selected combination "*NSCT(pyrexc, sinc)⇒F3⇒LSSVM*" is further being reduced by using the MICI feature selection algorithm. Due to the addition of MICI feature selection step the processing time could have been increased. But because of the reduced feature vector, retrieval and classification time decreases enormously and hence the overall processing time reduces. Table 8 shows the average CPU processing time using optimal features (= 33) and ED similarity measure.

Table 8 Average processing time using optimal features and ED similarity measure

Image Database	Methods	Time(sec)
ImgDb1	NSCT(pyrexc,sinc)⇒F3⇒LSSVM	4.21
ImgDb2	NSCT(pyrexc,sinc)⇒F3⇒LSSVM	6.67
ImgDb3	NSCT(pyrexc,sinc)⇒F3⇒LSSVM	10.28
ImgDb4	NSCT(pyrexc,sinc)⇒F3⇒LSSVM	10.30

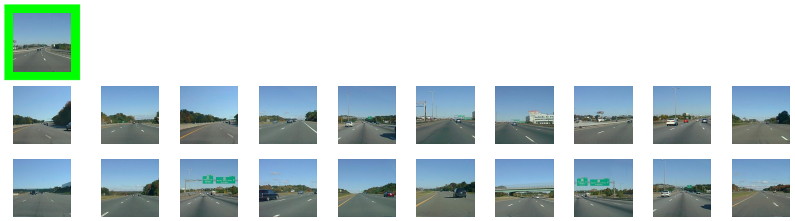


Fig. 14 Visual results of the CBIR system using the best combination of "*NSCT (pyrexc, sinc)* ⇒ *F3* ⇒ *LSSVM*" and ED distance measure for ImgDb1 database (image marked by the greenbox is the query image)

Fig. 15 Visual results of the CBIR system using the best combination of "*NSCT (pyrexc, sinc)* ⇒ *F3* ⇒ *LSSVM*" and ED distance measure for ImgDb2 database (image marked by the greenbox is the query image)

Fig. 14 to Fig. 17, shows the examples of the visual results obtained by the selected best combination i.e.," *NSCT (pyrexc , sinc)* ⇒ *F3* ⇒ *LSSVM*", with ED similarity measure and optimal features (33) on the ImgDb1, ImgDb2, ImgDb3 and ImgDb4 database using query images from "*Road*","*Piano*","*binocular*" and "*Dog*" classes, respectively. From all the four given instances, we can clearly see that all the retrieved images are from the respective classes corresponding to the query images. Fig. 18 also shows the visual results on IRMA medical database [89] using the combination " *NSCT (pyrexc, sinc)* ⇒ *F3* ⇒ *LSSVM* " with optimal features (33) and ED similarity measure.

5 Conclusion

In this paper, we have presented an extensive comparative study of the effectiveness of five multi-resolution transforms using different filters and three different classifiers in CBIR application. For each multi-resolution transform, the optimal settings are presented in four image benchmark database i.e., scene database (Oliva), object database (Caltech101 and Caltech256) and realistic database (MIRFlickr25000), respectively. Then the optimal settings are compared against each other globally to find out the best combination. The experimental results indicate the appropriate transform with appropriate filter with the best classifier. Furthermore, we have also attempted

Fig. 16 Visual results of the CBIR system using the best combination of "*NSCT (pyrexc, sinc)* ⇒ *F3* ⇒ *LSSVM*" and ED distance measure for ImgDb3 database (image marked by the greenbox is the query image)

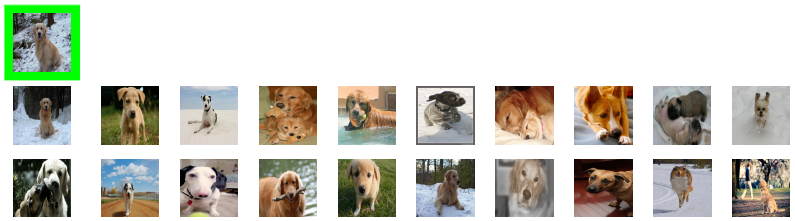


Fig. 17 Visual results of the CBIR system using the best combination of "*NSCT (pyrexc, sinc)* ⇒ *F3* ⇒ *LSSVM*" and ED distance measure for ImgDb4 database (image marked by the greenbox is the query image)

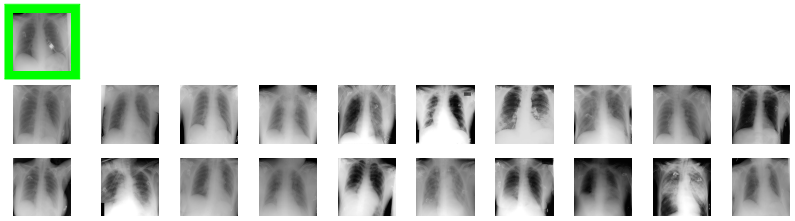


Fig. 18 Visual results of the CBIR system using the best combination of "*NSCT (pyrexc, sinc)*⇒ *F3* ⇒ *LSSVM*" and ED distance measure for medical database (image marked by the greenbox is the query image)

to discover the optimal features from the best combinations using unsupervised feature selection algorithm. The extensive experimental results indicate that the NSCT based GGD statistical features with LSSVM classifier combination performs usually the best, followed by the RT, CVT, CNT and the DWT. To the best in our knowledge, this is the first attempt, where different transforms have been compared in the same CBIR platform.

Acknowledgements The authors would like to thank all the anonymous reviewers and the associate editor for their valuable comments. The work is mainly funded by Machine Intelligence Unit, Indian Statistical Institute, Kolkata-108 (Internal Academic Project) for providing facilities to carry out this work. Malay K. Kundu acknowledges the Indian National Academy of Engineering (INAE) for their support through INAE Distinguished Professor fellowship.

References

1. T. Kajiyama and S. Satoh, "Construction of image retrieval systems focused on user knowledge interaction," in Proc. of the 18th ACM Int. Conf. on Multimedia, (Firenze, Italy), pp. 1673–1676, 2010.
2. H. Zhang, Z.-J. Zha, Y. Yang, S. Yan, Y. Gao, and T.-S. Chua, "Attribute-augmented semantic hierarchy: towards bridging semantic gap and intention gap in image retrieval," in Proc. of the 21st ACM Int. Conf. on Multimedia, (Barcelona, Spain), pp. 33–42, 2013.
3. S.-F. Chang, G. Auffret, J. Foote, C.-S. Li, B. Shahraray, T. F. Syeda-Mahmood, and H. Zhang, "Multimedia access and retrieval: the state of the art and future directions," in Proc. of the 7th ACM Int. Conf. on Multimedia, (Orlando, FL, USA), pp. 443–445, 1999.
4. R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Comput. Surv., vol. 40, no. 2, pp. 1–60, 2008.
5. D. Heesch, "A survey of browsing models for content based image retrieval," Multimedia Tools Appl., vol. 40, no. 2, pp. 261–284, 2008.
6. H. Kosch and P. Maier, "Content-based image retrieval systems - reviewing and benchmarking," J. Digital Inf. Management, vol. 8, no. 1, pp. 54–64, 2010.
7. P. Sandhaus and S. Boll, "Semantic analysis and retrieval in personal and social photo collections," Multimedia Tools Appl., vol. 51, no. 1, pp. 5–33, 2011.
8. M. Wang and X.-S. Hua, "Active learning in multimedia annotation and retrieval: A Survey," ACM Trans. on Intelligent Syst. and Tech., vol. 2, no. 2, pp. 10:1–10:21, 2011.
9. I. Elsayad, J. Martinet, T. Urruty, and C. Djeraba, "Toward a higher-level visual representation for content-based image retrieval," Multimedia Tools Appl., vol. 60, no. 2, pp. 455–482, 2012.
10. M. Grgic, S. Grgic, and M. Ghanbari, "A new approach for retrieval of natural images," J. of Electrical Engg., vol. 52, no. 5-6, pp. 117–124, 2001.
11. C. Shahabi and M. Safar, "An experimental study of alternative shape-based image retrieval techniques," Multimedia Tools Appl., vol. 32, no. 1, pp. 29–48, 2007.
12. X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, "Content-based image retrieval by integrating color and texture features," Multimedia Tools Appl., vol. 68, no. 3, pp. 545–569, 2014.
13. C. Beecks, S. Kirchhoff, and T. Seidl, "On stability of signature-based similarity measures for content-based image retrieval," Multimed Tools Appl., vol. 71, no. 1, pp. 349–362, 2014.
14. W. Jiang, G. Er, Q. Dai, and J. Gu, "Similarity-based online feature selection in content-based image retrieval.," IEEE Trans. Image Process., vol. 15, no. 3, pp. 702–712, 2006.
15. H. Mai and M. Kim, "Utilizing similarity relationships among existing data for high accuracy processing of content-based image retrieval," Multimedia Tools Appl., vol. 72, no. 1, pp. 331–360, 2014.
16. S. Atnafu, R. Chbeir, and L. Brunie, "Efficient content-based and metadata retrieval in image database," J. Universal Computer Science, vol. 8, no. 6, pp. 613–622, 2002.
17. C. Beecks, T. Skopal, K. Schöffmann, and S. Thomas, "Towards large-scale multimedia exploration," in Proc. of the 5th Int. Workshop on Ranking in Databases, (Seattle, WA, USA), pp. 31–33, 2011.
18. Y. Gong, C. H. Chuan, and G. Xiaoyi, "Image indexing and retrieval based on color histograms," Multimedia Tools Appl., vol. 2, no. 2, pp. 133–156, 1996.
19. Y. Tao and W. I. Grosky, "Image indexing and retrieval using object-based point feature maps," J. of Visual Languages and Computing, vol. 11, no. 3, pp. 323–343, 2000.
20. G. Ciocca, C. Cusano, S. Santini, and R. Schettini, "Halfway through the semantic gap: Prosemantic features for image retrieval," Inf. Sci., vol. 181, no. 22, pp. 4943–4958, 2011.
21. R. Yan, B. Huet, and R. Sukthankar, "Large-scale multimedia retrieval and mining," IEEE MultiMedia, vol. 18, no. 1, pp. 11–13, 2011.
22. Y. Huang, J. Zhang, H. Huang, and D. Wang, "Medical image retrieval based on unclean image bags," Multimedia Tools Appl., pp. 1–23, 2013.
23. L. Kovacs, A. Utasi, and T. Sziranyi, "VISRET A Content Based Annotation, Retrieval and Visualization Toolchain," in Proc. of the 11 Int. Conf. on Advanced Concepts for Intelligent Vision Systems, vol. 5807, (Bordeaux, France), pp. 265–276, Springer Berlin Heidelberg, 2009.
24. O. Marques, L. M. Mayron, G. B. Borba, and H. R. Gamba, "An attention-driven model for grouping similar images with image retrieval applications," EURASIP J. Appl. Signal Process., vol. 2007, no. 1, pp. 116–116, 2007.
25. J. J. Ashley, R. Barber, M. D. Flickner, J. L. Hafner, D. Lee, C. W. Niblack, and D. Petkovic, "Automatic and semiautomatic methods for image annotation and retrieval in query by image content(QBIC)," in Proc. of SPIE in Storage and Retrieval for Image and Video Databases III, vol. 2420, (San Diego/La Jolla, CA, USA), pp. 24–35, 1995.

26. N. P. Kozievitch, J. Almeida, R. d. S. Torres, N. A. Leite, M. A. Gonalves, U. Murthy, and E. A. Fox, "Towards a formal theory for complex objects and content-based image retrieval," J. of Inf. and Data Management, vol. 2, no. 3, pp. 321–336, 2011.
27. O. Marques and B. Furht, "MUSE: A content-based image search and retrieval system using relevance feedback," Multimedia Tools Appl., vol. 17, no. 1, pp. 21–50, 2002.
28. W. Xiong, B. Qiu, Q. Tian, C. Xu, S. H. Ong, K. W. C. Foong, and J.-P. Chevallet, "MultiPRE: a novel framework with multiple parallel retrieval engines for content-based image retrieval," in Proc. of the 13th ACM Int. Conf. on Multimedia, (Singapore), pp. 1023–1032, 2005.
29. F. Fleites, S.-C. Chen, and K. Chatterjee, "A semantic index structure for multimedia retrieval," Int. J. Semantic Computing, vol. 6, no. 2, pp. 155–178, 2012.
30. H. Li, X. Wang, J. Tang, and C. Zhao, "Combining global and local matching of multiple features for precise item image retrieval," Multimedia Syst., vol. 19, no. 1, pp. 37–49, 2013.
31. F.-h. Tang and H. H.-S. Ip, "Image fusion enhancement of deformable human structures using a two-stage warping-deformable strategy: A content-based image retrieval consideration," Inf. Sys. Frontiers, vol. 11, no. 4, pp. 381–389, 2009.
32. K.-H. Yap, K. Wu, and C. Zhu, "Knowledge propagation in collaborative tagging for image retrieval," Sig. Proc. Syst., vol. 59, no. 2, pp. 163–175, 2010.
33. C. A. Hernández-Gracidas, L. E. Sucar, and M. Montes-Y-Gómez, "Improving image retrieval by using spatial relations," Multimed Tools Appl., vol. 62, no. 2, pp. 479–505, 2013.
34. M. S. Kankanhalli, B. M. Mehtre, and H. Y. Huang, "Color and spatial feature for content-based image retrieval," Pattern Recogn. Lett., vol. 20, no. 1, pp. 109–118, 1999.
35. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.
36. H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in Proc. of 9th European Conference on Computer Vision, vol. 3951, (Austria), pp. 404–417, 2006.
37. C. M., V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," in Proc. of 11th European Conference on Computer Vision, vol. 6314, (Greece), pp. 778–792, 2010.
38. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in Proc. of Int. Conf. on Computer Vision, (Barcelona), pp. 2564–2571, 2011.
39. S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press, 3rd ed., 2008.
40. M. Vetterli and J. Kovačevic, Wavelets and Subband Coding. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.
41. J. Ma and G. Plonka, "The curvelet transform," IEEE Signal Process. Mag., vol. 27, no. 2, pp. 118–133, 2010.
42. M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," IEEE Trans. Image Process., vol. 14, no. 12, pp. 2091–2106, 2005.
43. J. Xu, L. Yang, and D. Wu, "Ripplet: A new transform for image processing," J. Vis. Comun. Image Represent., vol. 21, no. 7, pp. 627–639, 2010.
44. A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," IEEE Trans. Image Process., vol. 15, no. 10, pp. 3089–3101, 2006.
45. M. Acharyya and M. K. Kundu, "An adaptive approach to unsupervised texture segmentation using M-band wavelet tranform," Signal Processing, vol. 81, no. 7, pp. 1337–1356, 2001.
46. M. Banerjee, M. K. Kundu, and P. Maji, "Content based image retrieval using visually significant point features," Fuzzy Sets and Syst., vol. 160, no. 23, pp. 3323–3341, 2009.
47. M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and kullbackleibler distance," IEEE Trans. Image Process., vol. 11, no. 2, pp. 146–158, 2002.
48. M. M. Eltoukhy, I. Faye, and B. B. Samir, "A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram," Comput. Biol. Med., vol. 40, no. 4, pp. 384–391, 2010.
49. M. K. Kundu, M. Chowdhury, and M. Banerjee, "Interactive image retrieval using M-band wavelet, earth movers distance and fuzzy relevance feedback," Int. J. of Machine Learning and Cybernetics, vol. 3, no. 4, pp. 1–12, 2011.
50. G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Fast wavelet-based image characterization for highly adaptive image retrieval," IEEE Trans. Image Process., vol. 21, no. 4, pp. 1613–1623, 2012.
51. J. Zhang, Y. Wang, Z. Zhang, and C. Xia, "Comparison of wavelet, gabor and curvelet transform for face recognition," Optica Applicata, vol. 41, no. 1, pp. 183–193, 2011.
52. S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," Inf. Fusion, vol. 12, no. 2, pp. 74–84, 2011.

53. H. Shan, J. Ma, and H. Yang, "Comparisons of wavelets, contourlets and curvelets in seismic denoising," J. of Applied Geophysics, vol. 69, no. 2, pp. 103–115, 2009.
54. M. Chowdhury, S. Das, and M. K. Kundu, "Novel CBIR system based on ripplet transform using interactive neuro-fuzzy technique," Electronic Letters on Computer Vision and Image Analysis, vol. 11, no. 1, pp. 1–13, 2012.
55. J. Lan, Y. Guan, Z. Tang, and J. Zhang, "Texture image retrieval based on nonsubsampled contourlet transform and matrix f-norm," Applied Mathematical Sciences, vol. 7, no. 53, pp. 2613–2619, 2013.
56. A. Mosleh and F. Zargari, "A new content based image retrieval method using contourlet transform," J. of Computer and Robotics, vol. 2, no. 1, pp. 45–51, 2010.
57. I. J. Sumana, G. Lu, and D. Zhang, "Comparison of curvelet and wavelet texture features for content based image retrieval," Proc. of Int. Conf. on Multimedia and Expo, pp. 290–295, 2012.
58. R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," IEEE Trans. Signal Proc., vol. 40, no. 4, pp. 882–893, 1992.
59. 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," in Proc. of the 25th Annual Int. Conf. of the IEEE Engg. in Medicine and Biology Society, vol. 3, pp. 2714–2717, 2003.
60. B. C. Csáji, "Approximation with artificial neural networks," MSc thesis, Eötvös Loránd University, Hungary, pp. 1–45, 2001.
61. S. Haykin, Neural Networks and Learning Machines. Upper Saddle River, NJ, USA: Prentice Hall PTR, 3rd ed., 2009.
62. B. Liu and Y. Jiang, "A multitarget training method for artificial neural network with application to computer-aided diagnosis," Med. Phys., vol. 40, no. 1, pp. 1–9, 2013.
63. A. Kasapis, "MLPs and pose, expression classification," Proc. of UNiS Report, pp. 1–87, 2003.
64. S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," Neural Computation, vol. 4, no. 1, pp. 1–58, 1992.
65. D. Tzikas and A. Likas, "An incremental bayesian approach for training multilayer perceptrons," in Proc. of the 20th Int. Conf. on Artificial Neural Networks, vol. 6352, (Thessaloniki, Greece), pp. 87–96, 2010.
66. A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in Proc. of Int. Conf. on Computer Vision, pp. 1–8, 2007.
67. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
68. B. C. Ko, J. Lee, and J. Y. Nam, "Automatic medical image annotation and keyword-based image retrieval using relevance feedback.," J. Digit. Imaging, vol. 25, no. 4, pp. 454–465, 2012.
69. J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," Neural Process. Lett., vol. 9, no. 3, pp. 293–300, 1999.
70. C. Hsu and C. J. Lin, "A comparison of methods for multi-class support vector machines," IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 425–425, 2002.
71. I. Sumana, M. Islam, D. Zhang, and G. Lu, "Content based image retrieval using curvelet transform," in Proc. of the 10th IEEE Int. Workshop on Multimedia Signal Processing, (Cairns, Queensland, Australia), pp. 11–16, 2008.
72. R. Vieux, J. Benois-Pineau, and J.-P. Domenger, "Content based image retrieval using bag-of-regions: an efficient approach," in Proc. of Int. Conf. on Multimedia Modeling, (Klagenfurt, Austria), pp. 1–11, 2012.
73. M. Lamard, G. Quellec, L. Bekri, B. Cochener, C. Roux, and G. Cazuguel, "Content based image retrieval based on wavelet transform coefficients distribution," in Proc. of 29th Annual Int. Conf. of IEEE Engg. in Medicine and Biology Society, (Lyon, France), pp. 4532 – 4535, 2007.
74. A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," Int. J. of Comp. Vision, vol. 42, no. 3, pp. 145–175, 2001.
75. L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," Comput. Vis. Image Und., vol. 106, no. 1, pp. 59–70, 2007.
76. G. Griffin, A. Holub, and P. Perona, "Caltech 256 object category dataset," Technical Report UCB/CSD-04-1366, California Institute of Technology, 2007.
77. M. J. Huiskes and M. S. Lew, "The MIR Flickr retrieval evaluation," in Proc. of Int. Conf. on Multimedia Inf. Retrieval, (New York, NY, USA), ACM, 2008.
78. Y. Rubner and C. Tomasi, Perceptual Metrics for Image Database Navigation. Norwell, MA, USA: Kluwer Academic Publishers, 2001.
79. O. Pele and M. Werman, "Fast and robust earth mover's distances," in Proc. of Int. Conf. on Computer Vision, pp. 460–467, Sept 2009.

80. S. Shirdhonkar and D. Jacobs, "Approximate earth movers distance in linear time," in Proc. of Computer Vision and Pattern Recogn, pp. 1–8, June 2008.
81. D. T. Vollmer, T. Soule and M. Manic, "A distance measure comparison to improve crowding in multimodal optimization problems," in Proc. of 3rd Int. Symp. on Resilient Control Systems, pp. 31–36, Aug 2010.
82. G. Kanizsa, "Organization in Vision: Essays on Gestalt Perception," Praeger Publishers Inc., 1972.
83. P. J. Kellman and T. F. Shipley, "A theory of visual interpolation in object perception," Cognitive Psychology, vol. 23, no. 2, pp. 141–221, 1991.
84. S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in Proc. of Computer Vision and Pattern Recogn, (Anchorage, Alaska, USA), pp. 1–8, 2008.
85. P. Ou and H. Wang, "Prediction of stock market index movement by ten data mining techniques," Modern Applied Science, vol. 3, no. 12, pp. 28–42, 2009.
86. P. Mitra, C. A. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 3, pp. 301–312, 2002.
87. S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in Proc. of Computer Vision and Pattern Recogn, vol. 2, (New York, NY, USA), pp. 2169–2178, 2006.
88. T. Kobayashi, A. Hidaka, and T. Kurita, "Selection of histograms of oriented gradients features for pedestrian detection," in Neural Information Processing, pp. 598–607, Berlin, Heidelberg: Springer-Verlag, 2008.
89. H. Müller and T. M. Deserno, "Content-based medical image retrieval," in Biomedical Image Processing, Biological and Medical Physics, Biomedical Engineering, pp. 471–494, Springer Berlin Heidelberg, 2011.