On Content Based Image Retrieval and Its Application

A dissertation submitted for the degree of
Doctor of Philosophy (Tech.)
in
Radio Physics and Electronics

By
Manish Chowdhury

Department of Radio Physics and Electronics
University of Calcutta
2015

I would like to dedicate this thesis to my loving family ...

Declaration

I hereby declare that this dissertation is the result of my own work based on the research carried out at the Machine Intelligence Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata 700108, India. The contents of this dissertation are original except where specific reference is made to the work of others and have not been submitted in whole or in part for consideration for any other degree or qualification in any other University. Other sources are acknowledged explicitly in the text.

MANISH CHOWDHURY

Acknowledgements

It is finally done. Now, in the end, I still feel that there is much more to learn. However, I never felt alone under the able guidance of my supervisor, Professor Malay Kumar Kundu, who always helped me to acquire the necessary motivation and knowledge to achieve my goals. He provided me the freedom to pursue my own ideas and, at the same time, was rigorous in reviewing my work. I am thankful to him for granting me the opportunity to work in such a remarkable research field. He has played a fundamental role in bringing this thesis to life.

While at the Machine Intelligence Unit (MIU), Indian Statistical Institute, where this work has been carried out, I had the opportunity to meet amazing people and make new friends. I want to thank the professors of the MIU Prof. S. K. Pal, Prof. S. Bandopadhyay, Prof. S. Mitra, Prof. A. Ghosh, Prof. R. K. De, Dr. P. Maji, and Dr. B. U. Shankar for their constructive suggestions and comments. I am also thankful to Prof. C. A. Murthy for helping me in my research work by providing fruitful academic suggestions.

I would like to thank all my current and former colleagues for their helpful discussions, support and friendship specially, all my lab members S. Das, A. Bakshi, S. Chattopadhyay, B. N. Subudhi, K. R. Murthy, R. Das, T. Basu, P. P. Kundu, A. Mondal, R. Roy, B. Sharma and J. Modal and for not only their many small contributions to this work, during our informal talks, but also for keeping an amazing atmosphere inside the laboratory. I express my sincere thanks to the members of MIU, ISI for providing me all kinds of facilities whenever needed.

I would like to express my deep gratitude to my parents and all my family members (special mention needs to be made of my cousin sister Dr. Sutapa Datta and my boro-mama Mr. Sukumar Datta for their understanding, never ending love, endless patience and constant support towards me. Nothing would have been possible without their faith on me.

(MANISH CHOWDHURY)

Abstract

Searching digital information archives on the internet and elsewhere has become a significant part of our daily life. A major part of this growing body of information is digital image and video. The task of automated image retrieval is complicated by the fact that many images do not have adequate textual descriptions. Retrieval of images through the analysis of their visual content is therefore an exciting and a worthwhile research challenge. In this thesis, emphasize have been given to the different image representation technique using various recently developed Multi-Resolution Analysis (MRA) tools like curvelet transform, ripplet transform and many more. An extensive investigation has been performed to find out a proper combination of MRA feature extraction tool and the suitable classifier in order to generate compact feature vector for an efficient representation of different image class. It has been shown that a compact image feature representation plays an important role in order to obtain the high accuracy of retrieval results. However, in spite of the compact representation of the image, still there exists a semantic gap between the actual visual meaning of an image and its low level representation in terms of feature vectors. To overcome this problem, fuzzy and graph based relevance feedback mechanism have been proposed in this thesis. Finally, two image retrieval systems in real life application have been designed. In this regard, radiographic and endoscopic based image retrieval system is proposed. Extensive experiments and comparisons with state-of-the-art schemes are carried out to qualitatively as well as quantitatively evaluate the performance of the proposed schemes. In conclusion, the present thesis has addressed few existing problems in CBIR system design and developed the suitable solution for most of them.

Contents

List of Figures

List of Tables

List of Abbreviations

AP	Average Precision
AR	Average Recall
BTC	Block Truncation Coding
BCCM	Block Color Co-occurrence Matrix
BPH	Block Pattern Histogram
BoW	Bag-Of-Words
BRISK	Binary Robust Invariant Scalable Key-points
BDEE	Biased Discriminative Euclidean Embedding
BMMA	Biased Maximum Margin Analysis
CBIR	Content Based Image Retrieval
CCM	Color Co-occurrence Matrix
CLD	Color Layout Descriptor
CSD	Color Structure Descriptor
CCV	Color Coherence Vector
CVT	Curvelet Transform
CHoG	Compressed Histogram of Gradients
CSS	Curvature Scale Space
CPDH	Contour Points Distribution Histogram
CSLBP	Center Symmetric Local Binary Pattern
CSLTP	Center Symmetric Local Ternary Pattern
CBR	Case-Based Reasoning
CB-LTL	Case-Based Long Term Learning
CORI	Clinical Outcomes Research Initiative
CBD	Chessboard Distance
CBMIR	Content Based Medical Image Retrieval
DCD	Dominant Color Descriptor
DCV	Distance Coherence Vector
DWT	Discrete Wavelet Transform
DBPSP	Difference Between Pixels of Scan Pattern
DBF	Directional Filter Bank
DICOM	Digital Imaging and Communication in Medicine
EHD	Edge Histogram Descriptor
EMD	Earth Mover's Distance
ED	Euclidean Distance

FREAK	Fast Retina Key-point
FFEI	Fuzzy Feature Evaluation Index
FRF	Fuzzy Relevance Feedback
FL	Fuzzy Logic
FV	Fisher Vector
FCM	Fuzzy-C-Means
GLCM	Gray-Level Co-occurrence Matrix
GLOH	Gradient Location and Orientation Histogram
GGD	Generalized Gaussian Distribution
HOG	Histograms of Oriented Gradients
HVS	Human Visual System
HTD	Homogeneous Texture Descriptor
IGA	Interactive Genetic Algorithm
IRMA	Image Retrieval in Medical Applications
KKT	Karush-Kuhn-Tucker
KLD	Kullback-Leibler Divergence
LBP	Local Binary Pattern
LDPs	Local Derivative Patterns
LTP	Local Ternary Pattern
LTrPs	Local Tetra Patterns
LMeP	Local Mesh Pattern
LTCoP	Local Ternary Co-occurrence Pattern
LS-SVM/LSSVM	Least Square SVM
LP	Laplacian Pyramid
LM	Levenberg-Marquardt
MCM	Motif Co-occurrence Matrix
MPEG	Moving Picture Expert Group
MRF	Markov Random Field
MSD	Micro-Structure Descriptor
MICI	Maximal Information Compression Index
MSA	Multi-scale Analysis
MGA	Multi-scale Geometric Analysis
MD	Manhattan Distance
MLP	Multilayer Perceptron
MSE	Mean Squared Error

MAP	Mean Average Precision
ML	Maximum Likelihood
NSCT	Non-sub Sampled Contourlet Transform
NSD	Nested Shape Descriptor
NSP	Non-Subsampled Pyramid
NSDFB	Non-Subsampled DFB
ORB	Oriented fast and Rotated Brief
ori	Orientation
PCA	Principal Component Analysis
PSNR	Peak-Signal-to-Noise Ratio
PCNN	Pulse Coupled Neural Network
PDFB	Pyramidal DFB
PACS	Picture Archiving and Communication Systems
pyr	Pyramidal
QBIC	Query By Image Content
QFB	Quincunx Filter Banks
QP	Quadratic Programming
RT	Ripplet Transform
RF	Random Forest
ROI	Region-of-Interests
SSR	Signed Square Rooting
SCD	Scalable Color Descriptor
SPS	Similarity Positional Score
SPD	Steerable Pyramid Decomposition
SIFT	Scale-Invariant Feature Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
Semi-BMMA	Semi-supervised BMMA
TBD	Texture Browsing Descriptor
VCAD	Vector Cosine Angle Distance
VLAD	Vector of Locally Aggregated Descriptors
WT	Wavelet Transform

Chapter 1

Introduction and Scope of the Thesis

1.1 Introduction

Automatic image retrieval has become an important research problem considering its usage in handling the gigabytes of unlabeled image data that are generated and stored digitally in huge repositories, containing visual information both on the web as well as in the network computing system. These images, stored as unstructured multimedia data in different databases spread all over the world, most of which can be accessed through the internet. This growth can be attributed to the ease with which users are able to produce and archive photographs with today's consumer electronic devices. There are good reasons to believe that the rate of this growth will increase in the future, forcing search engine companies such as Google, bing and Flickr to devote vast computational resources to their image search engines. However, reliable textual information about these photographs is hardly available, making the task of building effective image retrieval systems even more challenging. Moreover, uniformity can be a major problem in manually labeled collections because a human annotator may overlook a relevant description of an image or might disagree with another person over the "correct" interpretation of that image [1, 2].

To overcome these limitations, Content Based Image Retrieval (CBIR) was proposed in the early 1990*s* [3–5]. In the recent past, Google has launched a new application, called "Google Goggles" for Google Android mobile phones[1], which is a content-based search application and allows people to search for more information about a famous landmark or work of art simply by taking a photo of that object [6, 7].

CBIR systems create machine-interpretable descriptions of an image's physical characteristics such as color, texture, shape etc. These descriptions, known as extracted features, can then be compared by a measure of similarity. The similarity between a given query image and every image in the image archive is then computed by CBIR system, and finally ranks the images in the collection according to their degree of relevance to the user's query

[1] https://play.google.com/store/apps/details?id=com.google.android.apps.unveil

[5].

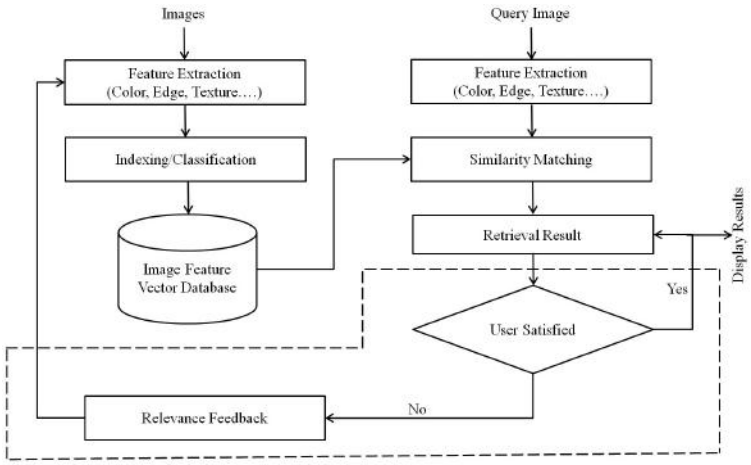


Fig. 1.1 Generic framework for content-based image retrieval

Even after designing several compact and efficient features vector for indexing each image, still there remains a wide gap between actual human perceptions model with that of feature based model used for CBIR. This is mainly due to the fact that the design of feature vector is solely based on low level visual information (like color, texture, shape etc.) without taking into account the semantic information like spatial organization, context, events etc. To overcome this shortcoming and in trying to incorporate certain amount of visual perception characteristic in the feature vector design, people have proposed the relevance feedback mechanism in CBIR design with active participation of human observers. The typical CBIR system with relevance feedback is shown in Fig.1.1, where the building blocks without the dotted line box describe the general CBIR system and the building blocks with the dotted line describe the CBIR system with relevance feedback model.

Images are being generated at an ever-increasing rate by sources such as defence and civilian satellites, military reconnaissance and surveillance flights, fingerprinting and mug-shot-capturing devices, scientific experiments, biomedical imaging, and home entertainment systems. Application areas in which CBIR has a principal activity are numerous and diverse: art galleries and museum management [8], architectural and engineering design,

law enforcement and criminal investigation [9], interior design, trademark and copyright database management [10], remote sensing and management of earth resources [11], scientific database management, weather forecasting [12], retailing [13], fabric and fashion design, Medical Imaging [14, 15] and picture archiving and communication systems. With the recent interest in multimedia systems, CBIR has attracted the attention of researchers across several disciplines [16].

1.2 A Brief Review of CBIR Systems

One of the major building blocks of a CBIR system is the feature extraction block. In the feature extraction block, an image is represented by features which allow searching for images similar to a given image. A feature is defined to capture a certain visual property of an image. Color, texture, shape, etc. are the most widely used features in CBIR applications. Several CBIR algorithms have been developed using only single features [16–18]. A single feature normally describes a specific aspect of the image content; therefore multiple visual features are necessary for the general purpose CBIR [19]. Recent research work uses multiple visual features, which can improve the retrieval performance effectively [20, 21]. Various multiple visual feature extraction schemes have been used in the literature to find the low-dimensional salient and significant features, which can effectively be used to represent the underlying image's visual characteristics [16, 22–25]. The most commonly used features in designing CBIR system are color, texture, shape, and keypoint descriptor and, brief discussion about each of which will be presented in the following paragraph. Besides that, descriptions about some commercially available CBIR system will also be presented.

Different color features in CBIR systems: The color signature of object in the foreground and also that of background makes it as one of important vision queue. Color is considered to be a robust visual feature because it is invariant to direction, size, rotation and shift. Point/pixel and block/group based are the two different color representa-

tion schemes. Pixel based color representation include color moments [26] and Color Co-occurrence Matrix (CCM) [27], where color histogram and color correlogram [28] comes under block/group based representation schemes.

In general, color spaces are classified into linear color spaces and non-linear color spaces [29, 30]. The RGB color space is an additive color space based on three primary colors: red, green and blue. Secondary colors can be generated by using the primary colors. However, the RGB space is not very efficient in dealing with real-world images, thus it is avoided in most of image retrieval algorithms because it lacks the ability of measuring the perceptual similarity. Accordingly, the HSV color space is used instead of the RGB color space because the components of hue and saturation are very close to human visual perception. The HSV model has three constituent components *i.e.*, the "hue" which refers the color, the "saturation" which refers the "vibrancy" of the color, and the "value" which refers the brightness of the color. As another color space, the YCbCr color space is divided into luminance (Y) and chrominance (Cb, Cr), while Cb and Cr denote the blue-yellow and red-green color difference, respectively. The $L*a*b$ color space is also derived from the XYZ color space to achieve perceptual uniformity. As in YCbCr, $L*a*b$ consists of one lightness dimension (L) and two chrominance dimensions ($a*, b*$) based on the color opponent process.

Color histogram is a popular color representation that has been used in many image retrieval systems. Its performance in characterizing global color content has been widely recognized. However, color histograms lacks the information about the spatial distribution of color; thus, other representations such as color correlogram and auto-correlogram have been proposed. These methods provide information about the changes of the spatial correlation of color pairs with the distance in an image, and they have shown better retrieval performance than color histograms [28]. In addition, many algorithms have been proposed in the CBIR domain based on color moments. The key idea of color moments is the use

of standard deviation and mean of distributions in each color band as a color feature. This considered as a compact characteristic and therefore it is usually used as an optimization process along with other color features. Xiaoyin *et al.* [26] has proposed a new color image retrieval method using the color moment invariant. Representative colors are computed for each image instead of being fixed in a given color space, thus this allows feature representation to be more accurate and compact. The proposed method is based on small image descriptors and adaptive to the context of the image itself by two-stage clustering approach.

CCM [31] and Motif Co-occurrence Matrix (MCM) [32] were developed by employing the spatial relation between color channels. An image can be considered as a composition of suitable "elementary structures". The elements of those pixels carry visual attributes, i.e. colors, and possess relations, i.e. distances between them. Accordingly, image contents can be characterized by an appropriate M-dimensional CCM where the attributes and relationships are represented by different matrix axes. Guoping [27] has proposed the Block Truncation Coding (BTC) as a different image coding technique which was employed in the CBIR for compressing color images. From the BTC compressed stream without decoding, two description features of image content was derived: Block Color Co-occurrence Matrix (BCCM) and Block Pattern Histogram (BPH). Both BCCM and BPH were used to compute the similarity measures of images for CBIR applications. Recently, some BCT variants [33–35] have been proposed for CBIR tasks.

The Scalable Color Descriptor (SCD), Color Layout Descriptor (CLD), Color Structure Descriptor (CSD) and Dominant Color Descriptor (DCD) [36–38] have widely been applied in image retrieval applications of Moving Picture Expert Group (MPEG)-7 color descriptors. Recently, Zeng *et al.* [39] employed the Color Coherence Vector (CCV) based on the distance histogram. They also proposed a multiscale Distance Coherence Vector (DCV) algorithm in accordance with problems where different shapes have the same descriptor and the poor performance of anti-noise of image retrieval algorithm based on the DCV. This al-

gorithm obtains relatively smoother contour curves by Gaussian function evolving contour curve, and then calculates the DCV of the original contour curve and evolved contour curves. This algorithm is invariant to the translation, rotation, and scaling transformations.

Different texture features in CBIR systems: Although texture plays an important role in image processing and computer vision task, but unfortunately, there is no precise definition of the notion texture. Petrou *et al.* define texture is the variation of data at scales smaller than the scales of interest [40]. Generally speaking, textures are complex visual patterns composed of entities, or sub-patterns that have characteristic brightness, directionality, coarseness, randomness, fineness, smoothness, granulation, slope, size, etc. Thus texture can be regarded as a similarity grouping in an image [41, 42]. Hence, texture is an important feature for describing innate surface properties of a particular object and its relationship with the surrounding regions in an image [43–46]. However, the computation complexity and retrieval accuracy are the main challenging drawbacks of texture-based CBIR systems.

Many texture-based image retrieval methods was proposed in CBIR. Some commonly used texture ques are texture descriptor based on Gabor filters, Wavelet transforms, Gray-Level Co-occurrence Matrix (GLCM) [47], Markov Random Field (MRF) [48], Edge Histogram Descriptor (EHD) [49], Steerable Pyramid Decomposition (SPD) [50], and Tamura features [51]. Gabor filters are set of wavelets and each wavelet captures energy at a specific frequency and orientation. Gabor wavelet transforms have multi-resolution and multi-orientation properties and this is optimal for measuring local spatial frequencies [52–54]. Many transform-based feature extraction techniques are used including Discrete Wavelet Transform (DWT), Curvelet Transform (CVT), Contourlet Transform (CNT), Non-sub sampled Contourlet Transform (NSCT), Ripplet Transform (RT) etc. These transform are used to extract low-level features due to its superiority in multi-resolution/geo-metric analysis and spatial frequency [55, 56].

Recently, some work in CBIR paradigm have been carried out either by using a single descriptor or by combining many algorithms to form a robust descriptor for image texture [57, 58]. Liu *et al.* [59] have developed a new image retrieval approach, namely Micro-Structure Descriptor (MSD). However, MSD model has not simulated visual attention mechanisms well. Therefore, to address this problem, recently a novel computational visual-attention model, namely salient structure model, for CBIR has been proposed in [17]. Kwitt *et al.* [60] introduced a probabilistic based texture model image retrieval. It is based on some statistical measurement done on image representation in complex wavelet domain. There are several other statistical based texture modeling is developed in [61, 62].

Different shape features in CBIR systems: Like color and textures, the shape is another fundamental unit of perception which carries semantic information of an image. Shape feature can be categorized into two groups namely contour-based and region based approach. In contour-based approach, features are computed from outer boundary while region based approach features are extracted from the entire region [63–66].

Shape analysis is a well-explored research area with many shape representation and similarity measurement techniques found in the literature [67]. Generally, shape descriptors are used with other features such as color and texture and tend to be efficient in specific applications such as for retrieval of man-made objects [68]. Shape features can be extracted using various methods such as Fourier descriptors, B-splines, Curvature Scale Space (CSS), etc [69, 70].

Generally, contour-based shape methods require a high computation time in obtaining the correspondence between contour points from two shapes using local contour information respectively. In order to solve this problem, Shu *et al.* [71] proposed a new contour-based descriptor for closed curves, namely the Contour Points Distribution Histogram (CPDH), which depicts the deformable potential at each point along a curve. In addition, developed a ground distance calculation technique, which is based on the Earth

Mover's Distance (EMD) under polar coordinates, for shape matching in order to be invariant to scale and translation. Another approach, that combines the contour-based and region-based shape methods, proposed by Chen and Xu [72], namely the rolling penetrate descriptor. This method improves the conventional methods by obtaining any desired information in a unified way rather than in a specific aspect of shape features.

In region based techniques, all the pixels within a shape region are taken into account to obtain the shape representation. Common region based methods use moment descriptors to describe shape [73–75]. These include geometric moments, Legendre moments, Zernike moments and pseudo Zernike moments. Several researchers also use the grid method to describe shape [76, 77]. Since region-based shape representations combine information across an entire object rather than exploiting information just at boundary points, they can capture interior information in a shape. Other advantages of region-based methods are, it can be employed to describe disjoint shape and robust to shape distortions. Bronstein *et al.* [78] developed an interesting approach for extracting the shape feature, namely the Shape Google. This approach was proposed in the context of non-rigid shape retrieval, and inspired by the work of Ovsjanikov *et al.* [79], who first introduced this approach.

Different key points descriptors in CBIR systems: It describes local information using key points of some image parts such as region, object of interest, edges, or corners. Recently, key point descriptors have shown their superiority for various types of applications in computer vision, as well as in CBIR paradigm [80–82]. Key point descriptors have many advantages over traditional global features since they are invariant to image scale and rotation, and provide robust matching across a wide range of different situations [83–85]. As the research attention in CBIR context has shifted to use these key points features, a comprehensive review of different key point descriptors are presented in [86].

Scale-Invariant Feature Transform (SIFT) is one of the most successful and widely applied key points descriptors in the last decade [83, 87]. The SIFT is very robust for extract-

ing distinct invariant features from images in order to carry out reliable matching between different views of objects or scenes. Many CBIR system using SIFT and its variants have been reported in the literature [88, 89].

In 2008, a new robust key point feature detector was introduced by Bay *et al.*, namely Speeded Up Robust Features (SURF) [90]. One major drawback of SIFT descriptor is it high dimensionality, and applying Principal Component Analysis (PCA) in PCA-SIFT [88] and Gradient Location and Orientation Histogram (GLOH) [89] slows down the feature computation. This descriptor is partly inspired by the SIFT descriptor so the authors have claimed that the standard SURF is faster than SIFT and more robust against different image transformations.

The bag-of-words (BoW) feature is generally based on tokenizing keypoint-based descriptor, like, SIFT; to generate a visual-word vocabulary/ codebook [91–93]. Then, the visual-word vector of an image contains the presence or absence information of each visual word in the image, for example, the number of keypoints in the corresponding cluster, that is, visual word. Since last decade, BoW has been used extensively in an image annotation [94]. However, computing the BoW representation for an image is tedious and the computation time grows linearly with the size of the codebook. Although BoW model provides good accuracy for image search, it is not scalable for large image databases, as it is intensive both in memory and computations. In order to have a more informative image representation, Peronnin *et al.* proposed a generative model based approach called Fisher Vector (FV) [95]. They estimate a parametric probability distribution over the feature space from a huge representative set of features, generally a Gaussian Mixture Model (GMM). The GMM can be interpreted as a probabilistic visual vocabulary, where each Gaussian forms a visual word or cluster. Compared to the BoW, which only records the count of local descriptors in each visual word, the FV encodes the higher order statistics, resulting in a more discriminative representation and hence better performance.

Jegou *et al.*, proposed Vector of Locally Aggregated Descriptors (VLAD) [96] as a non-probabilistic equivalent to FV. VLAD is a simplified and special case of Fisher vector, in which the residuals belonging to each of the codewords are accumulated. The concatenated residuals represent the image, using which the search is carried out via simple distance measures. A number of improvements have been proposed to make VLAD a better representation by considering vocabulary adaptation and intra-normalization [97], residual normalization and local coordinate system [98], geometry information [99] and multiple vocabularies [2]. In [100], the VLAD vectors are L2 normalized. Subsequently, a signed square rooting (SSR) normalization was introduced [96] following its use for Fisher vectors [2, 95]. To obtain the SSR normalized VLAD, each element of an un-normalized VLAD is sign square rooted and the transformed vector is L2 normalized. The main weakness of VLAD is its partial robustness to outliers. A single outlier descriptor located far from the cluster center can outweigh the combined contribution from many inlier descriptors located close to that center.

Local Binary Pattern (LBP) is widely used feature description method. It was introduced for the texture classification by Ojala *et al.* [101] and shown its utility in area of image retrieval [25, 102–106]. Various variants of the LBP also exist in the literature such as a Local Derivative Patterns (LDPs) [107], Local Ternary Pattern (LTP) [108], Center Symmetric Local Binary Pattern (CSLBP) [109], Center Symmetric Local Ternary Pattern (CSLTP) [110], etc. However, LBP, LDP, and LTP extract the information based upon the distribution of edges which are coded by only two directions (*e.g.*, positive or negative). To improve the performance of these methods, Local Tetra Patterns (LTrPs) [105] was introduced in the CBIR domain. Local Mesh Pattern (LMeP) [111] and Local Ternary Co-occurrence Pattern (LTCoP) [112] are the recent feature description methods for designing CBIR system. Recently, LDEP [113] is introduced in order to overcome the few short coming of LBP and different modified LBP methods.

In general, some of above mentioned key points descriptors are computed on the dense grid of uniformly spaced cells and their performance is improved using overlapping local contrast normalization. Dalal and Friggs [114] proposed locally normalized Histograms of Oriented Gradients (HOG) descriptors to provide better performance compared to other existing features including wavelets. Chandrasekhar *et al.* [115] proposed another alternative descriptor, namely the Compressed Histogram of Gradients (CHoG), as a low bit-rate descriptor with a 20 times reduction in the bit rate.

Another key point descriptor employed in the CBIR domain is Difference Between Pixels of Scan Pattern (DBPSP), which mainly calculates the differences among all pixels within motifs of a scan pattern. Usually it is combined with other image features and used as texture feature as describe in [116]. As an alternative to SIFT and SURF, Rublee *et al.* [117] had introduced a binary descriptor called the Oriented fast and Rotated Brief (ORB), which is based on FAST detector [118] and BRIEF descriptor [119]. In 2011, a new key point descriptor called Binary Robust Invariant Scalable Key-points (BRISK) is introduced by Leutenegger *et al.* [120]. Fast Retina Key-point (FREAK) descriptor is a very recent developed by Alahi *et al.* [121]. It is inspired by the Human Visual System (HVS), or more precisely, by the retina. Recently, a new descriptor called Nested Shape Descriptor (NSD) has been proposed in [122]. This is the first binary descriptor to outperform SIFT on the standard VGG-Affine benchmark. Furthermore, the NSD binary descriptor significantly outperforms BRISK, a state-of-the-art binary descriptor [122, 123]

Different distance measure used in CBIR systems: The similarity/dissimilarity measure between images is second major important blocks in the CBIR system after the feature extraction [124, 125]. To achieve more accurate retrieval and better performance, the CBIR system should employ an efficient similarity measure that can accurately model the similarity distance in perceptual space [49]. Finding out a suitable distance measure which can capture more accurately human perception mechanism is a challenging research issue.

Despite of fairly good success of conventional distance measure in CBIR application.

Several distance metrics, such as the Manhattan Distance (MD), the Euclidean Distance (ED) and the Vector Cosine Angle Distance (VCAD) is proposed in the literature for measuring similarity between feature vectors [126–128]. Several CBIR system was proposed using ED and MD in [129–132].

Although, Minkowski distance is most widely used metric for measuring the similarity in CBIR systems, but this measure fail to capture the true similarity/ dissimilarity measure in perceptual space [49]. The used of advance similarity/dissimilarity measure like Kullback-Leibler Divergence (KLD), EMD etc., are shown to be more efficient in capturing the perceptual distance between two images [125, 133–137].

The EMD [133] is based on the transportation problem and recently, many methods have been proposed to improve the efficiency and accuracy of EMD [138–142]. The quality of the EMD for image similarity has been demonstrated in several application scenarios: color [133], contour matching [143], texture [144], melodies [145], visual tracking [146], etc. Its high quality for image similarity search has been demonstrated in an extensive survey on distances for CBIR [133]. To benefit from the EMD's high quality retrieval in these settings, efficient storage and query processing is crucial.

CBIR systems using relevance feedback mechanism: Generally, low level features such as color, texture, shape, corner, etc., are used to represent the approximate perceptual representation of an image, using which similarity and dissimilarity of the images are computed. But this low level information is not sufficient to describe the semantic relation among different image regions, object and background and different object in the foreground as well as in the background. Moreover, overall perceptual meaning of an image depicting different events/incidence cannot be describe using such low level image representation [147–150].

Jin *et al.* [151] have proposed a feature re-weighting technique by using both the rel-

evant and the irrelevant information, to obtain more effective results. Relevance feedback is considered as a learning and classification process [152, 153], using classifiers like neural networks [154], Support Vector Machine (SVM) [155], Least Square SVM (LS-SVM/LSSVM), etc.

SVM-based relevance feedback is one of the most widely used approaches in the field of CBIR. Despite its success, it considers the positive and negative feedbacks equally which is not robust base since each sample has distinct properties. In addition, most of SVM based relevance feedback approaches do not consider unlabeled samples which may yield a weak classifier during the refinement process. Bian and Tao [156] proposed a new SVM-based relevance feedback technique using a new type of learning, referred as the Biased Discriminative Euclidean Embedding (BDEE). Zhang *et al.* [157] have also proposed another SVM-based relevance feedback scheme, namely the Biased Maximum Margin Analysis (BMMA) and Semi-supervised BMMA (Semi-BMMA). This utilizes the information of unlabeled samples and integrate the distinct properties of feedbacks. However, trained classifiers become less effective when the training samples are insufficient in number. To overcome such problems, active learning methods are also used [158, 159].

A frequent adding of images to the CBIR system may result in degrading the performance of long-term learning relevance feedback methods. Additionally, the recorded feedbacks in such methods are often sparse. To take care of these problems, Rashedi *et al.* [160] proposed a long term learning method by adopting the Case-Based Reasoning (CBR),which is called the Case-Based Long Term Learning (CB-LTL).

To consider the user preferences and subjectivity during the retrieval process, Lai and Chen [161] proposed a user-oriented mechanism based on the Interactive Genetic Algorithm (IGA) in the CBIR domain. Another relevance feedback approach based on the genetic programming was proposed by Ferreira *et al.* [162], which considered both relevant and irrelevant images that mark by the user.

Bulo *et al.* [163] developed a novel approach based on the random walker algorithm introduced in the domain of interactive image segmentation. The key idea is that the relevant and non-relevant images labelled by the user at every feedback round are treated as seed nodes for the random walker problem. The ranking score is computed for each unlabeled image as the probability that a random walker starting from that image will reach a relevant seed before encountering a non-relevant one. Recently, many graph based re-ranking algorithms in CBIR are reported in the literature [164–166].

Different commercial CBIR systems: QBIC (Query By Image Content) is the first commercial content-based image retrieval system from IBM, which uses color histograms, a moment based shape feature, and a texture descriptor [75]. Similar to QBIC, many systems have been proposed and implemented using CBIR technologies, *e.g.*, VisualSeek [167], SIMPLIcity [168], and Blobworld [131]. Such these systems have concentrated on extracting low-level features from image (*e.g.*, color, texture, and shape) in order to represent its semantics. Other systems such as WebSeek [169] and Image Rover [170] have applied the image search based on submitted query keywords, and then performed user classification by providing category browsing and search-by-example facilities. After the early success of these systems, research directions have been extended over the last years to reduce the semantic gap. Recently, a new system called Qatris iManager has been developed with acceptable performance [171]. Research efforts in the field of CBIR currently concentrate on profound and challenging problems at different disciplines such as pattern recognition, machine learning, computer vision, and artificial intelligence [172, 173].

1.3 Motivation

The accuracy of a CBIR system mainly depends on four major factors, namely (a) proper representation of an images in terms of feature vector, (b) suitable distance metrics for measuring the similarity between images (corresponding feature vectors), (c) reducing the

semantic gap that exist between image representation with low level features and actual semantic meaning of an image and (d) minimizing the searching space/time.

In view of the above facts, feature extraction is one of the major research problem in the CBIR system. Images belong to different classes in the database should be represented by feature vector, which are markedly different from each other. Therefore, the need of developing a suitable feature extraction mechanism, that capture the salient intrinsic features of an image from which development of suitable feature vector as a true representation of an image is still a challenging issue in CBIR research.

Similarity/dissimilarity measure is second important part of CBIR system development. The role of the distance metric is to measure quantitatively the degree of similarity/dissimilarity between query and image in the database in the perceptual space. However, a fundamental challenge - "how to measure perceptual similarity" - remains largely unanswered. This is due to the fact that all kinds of low level features representation of an image may not be suitable for every kind of distance measure for similarity/dissimilarity measurement in perceptual space. So choosing suitable combination of low level feature and distance measure for achieving best results is an important investigating issue.

The third important problem in CBIR research is "how to minimize the semantic gap". In order to extract the meaningful image information, the way of human perception system is used to discriminate or equate one image to other semantically.

With the ever increasing size of the database, searching time reduction is also an important research issue for future CBIR system development. One of the way to reduce the effective searching space corresponding to a query, is to pre-classify the different images in the database into number of classes. For every query image, the searching may be restricted to a small number of classes to one of which the query image may belong (due to limitation of the classifier accuracy used for pre-classification of the query image). A sincere attempt is made to address all these above mentioned issues in the present thesis.

1.4 Contribution

In the current CBIR literature, different transform has been used for image representation and based on this representation, similarity/dissimilarity measure have been used for retrieving similar kind of images from large database. The major problem in the conventional CBIR, is the semantic gap, between visual perception of a scene and feature based representation of the scene image. To bridge this gap, Relevance Feedback (RF) mechanism has come into picture. Various research groups have proposed several CBIR model for extracting the suitable feature extraction with proper RF mechanism. Following this model, three different CBIR systems has been designed in Chapter 3.

In Scheme-I of Chapter 3 [174], particular type of transform (M-band) and Fuzzy Relevance Feedback (FRF), and different similarity measure like Euclidean Distance (ED) (most commonly used in CBIR literature) along with some of the other distance measure (Manhattan Distance (MD) and Chessboard Distance (CBD)) as comparative study. But each iterations takes a considerable time which is proportional to the size of the database.

To cut down this computational cost, people uses different classifiers for approximately identify the class of the input query image into one or few number of classes (Scheme-I, Chapter 6) a very popular approach used by fuzzy set community [175]. This knowledge is used to restrict the search space for retrieval of similar image from database. Similar model has been designed in Scheme-II for Chapter 3 [176].

So naturally, a method which combines both the above stated approach will able to provide a further improvement of computational efficiency and retrieval time. Therefore, the combination of Scheme-I and II of Chapter 3 has been implemented in Scheme-III of Chapter 3 [177]. The role of the FRF mechanism, supervised pre-classification methodologies and their combination to reduce the semantic gap in a CBIR system resulting in an efficient CBIR system design. Moreover, above discuss schemes perform well for certain types of databases.

But, Scheme-I, II and III uses three different transform (M-band, Ripplet, and NSCT respectively) used for extraction of primary multi-scale image features, different types of higher level image representative in terms of secondary features (extracted from primary features), pre-classification for search space reduction can resulted in efficient CBIR design having distinct advantage for certain types of databases. So it is much important and necessary to undertake a detailed investigation to find out the best possible combination of three above mention sub-system can extract the best possible efficiency for CBIR design. This type of investigation has not been reported in any literature so far as our knowledge is concerned. This is the beginning of the research investigation taken off by us which is reported in Chapter 4.

In this thesis, there are **three novel contribution** as describe below:

- It is found in the current literature that different types of multi-scale transform are used for extracting image features whose dimension is usually reduced by standard technique like PCA, ICA etc. But it is still remains comparatively higher for achieving pre-determine retrieval accuracy. This can be further enhance using signature/secondary feature measure from primary multi-scale features distribution. Various such image signatures have been used for CBIR design. Moreover, for cut down the search space for finer retrieval, various kind of classifiers have been used for approximately ascertain the class level of the query image (classifier already train with the signature of the different level classes). So, it is most important to undertake extreme study for find out the best possible combination of multi-scale image transform, image signature and classifier to be used for search space reduction. No such study has been reported in the literature. In this respect, contribution reported in the Chapter 4 is a notable one [178].

- Only image representation technique is not sufficient for reducing the semantic gap in CBIR. Therefore, graph based relevance feedback mechanism was proposed where

nodes in the graph are representative of images and similarity are represented by corresponding edges. The proposed technique is an iterative process where at each iteration, relevant and irrelevant images are treated as seed nodes in a random walker problem. The modified edges (relation between images) are computed using Gaussian kernel. To our knowledge, Gaussian kernel has been used for the first time in computing the similarity edges. The ranking score for each unlabeled image is computed as the probability that a random walker starting from that image will reach a relevant seed before encountering a non-relevant one. k-nearest neighbor has been used for ranking the retrieved images. Gaussian kernel similarity measure has been used for the first time along with k-nearest neighbor, which shows mark improvement in retrieval accuracy over that uses only graph random walker model [163]. In this respect, it is also a novel contribution in graph based CBIR technique and which is reported in Chapter 5 of this thesis [179].

- In Scheme-I of Chapter 6 [180], is also a novel contribution due to the fact, it proposed a multi-class assignment to a query image in order to reduce the search space and enhance the retrieval accuracy by reducing the error occur due to the finite accuracy of the classifier. This is done by a new and novel weighted voting scheme computed from the retrieval image where weight are assign based on order of appearance. This is very similar to fuzzy set theoretic approach (where various degree of membership for belonging to different classes is assign to a single entry).

1.5 Benchmark Dataset

To demonstrate the effectiveness of the proposed CBIR system in this study, we have used eight different databases as follows:

- SIMPLIcity: This is the SIMPLIcity dataset [168]. It consists of 1000 images uni-

formly divided into 10 categories. The categories are: African, Ocean, Building, Bus, Dinosaurs, Elephant, Flower, Horse , Mountain, Food. It is downloaded from $http://wang.ist.psu.edu/docs/home.shtml$ website.

- Corel: It consists of 10000 images of 100 different categories image's [181]. Every category contains 100 images of size 192×128 or 128×192 in JPEG format. All images come from Corel Gallery Magic 200000. It is downloaded from $http://www.ci.gxnu.edu.cn/cbir/Dataset.aspx$ website.

- Oliva: This database consists of 2688 images classified into 8 different classes, where the number of images in each classes varies from 260 to 409 images [182]. The 8 different scene categories are: coast, mountain, forest, open country, street, inside city, tall buildings and highways. This database can be downloaded from $http://cvcl.mit.edu/database.htm$ website.

- Caltech 101: This is the Caltech 101 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 [183]. The significant variation in color, pose and lighting makes Caltech 101 database quite challenging. Most images are medium resolution, about 300×300 pixels. The significance of this database is its large inter-class variability. It is downloaded from $https://www.vision.caltech.edu/$ website.

- Caltech 256: This data set consists of images from 256 object categories and is an extension of Caltech-101 [184]. It contains from 80 to 827 images per category. The total number of images is $30,607$. It is downloaded from $http://www.vision.caltech.edu/$ website. The significance of this database is its large inter-class variability, as well as a larger intra-class variability than in Caltech-101. As example, the class "marriages" contains images of newlyweds as well as images of wedding cakes. Moreover there is no alignment amongst the object categories. In our study, we have

created a subset of the Caltech 256 dataset, obtained by randomly selecting 25 categories and 100 images per category. In addition, we created 10 copies of each image corrupted with a different type of noise: 5 images are rotated copies (with degree in $\{-10,-5,0,5,10\}$), the remaining 5 instances are respectively a smoothed copy with a circular averaging filter (radius 5 pixels), a blurred copy with camera motion blur (length 9 pixels), a copy corrupted with salt-and-pepper noise (probability 0.05), a copy corrupted by white noise (variance 0.01) and finally a copy corrupted by speckle noise (variance 0.04). In total, we have 25000 images. 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.

- MIRFlickr25000: This is a diverse collection of realistic image database known as MIRFlickr25000 [185, 186]. 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.

- Oxford5k: It consists of $5,062$ images from Flickr with Oxford landmarks and it is downloaded from $http://www.robots.ox.ac.uk/\ vgg/data/oxbuildings/$. It comes with ground-truth for 55 query images with 5 query images corresponding to a same landmark [45].

- Oxford105k: It contains Oxford5k and $100,000$ distractor images. The distractor images are downloaded from Flickr with the same resolution as the original Oxford5k images [45].

- MIRFlickr1M : This is a distractor dataset, with $1,040,801$ medium resolution (500×333) images randomly retrieved from Flickr [185, 186] and it is downloaded from $http://press.liacs.nl/mirflickr/$ website.

1.6 Scope of the Thesis

The objective of present thesis is to contribute towards developing a complete CBIR system using different transform domain features, soft computing and graph based theory.

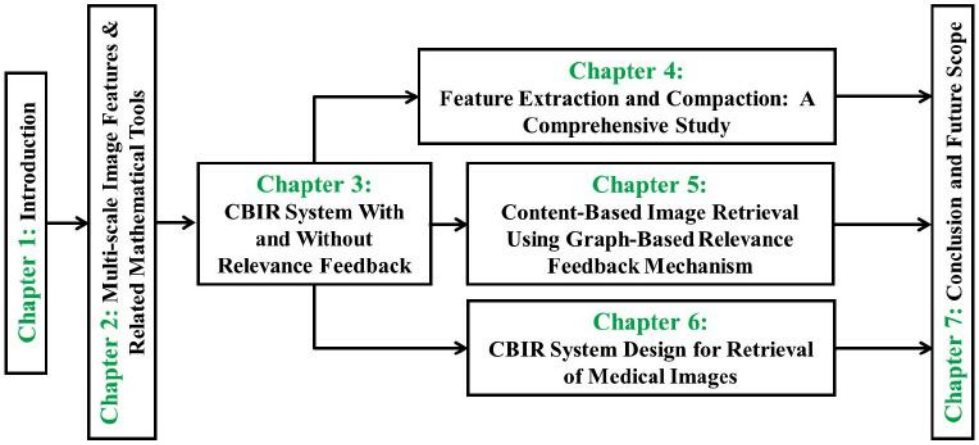


Fig. 1.2 Organization of the thesis

The present thesis comprises of six chapters, four of which describe novel contributions. Chapter-wise organization of the thesis is depicted in the flow diagram of Fig. 1.2 and is described briefly next:

- ***Chapter 2: Multi-scale Image Features and Related Mathematical Tools***

  Different low level image features representation techniques with their mathematical models are briefly discussed for the benefit of readers. Most of the mathematical models and formulation discussed in this chapter are used in subsequent chapters.

- ***Chapter 3: CBIR System With and Without Relevance Feedback*** [174, 176, 177]

  In this Chapter, we have proposed three different schemes of CBIR system. In all

these three schemes, different transform domain based features extraction mechanism have been proposed. In Scheme-I, we have proposed Fuzzy Relevance Feedback (FRF) mechanism based CBIR system. In Scheme-II, a CBIR system has been developed with pre-classification technique where the concept of partition database from the whole database has been introduced (reducing the search space). Moreover, to increase the retrieval accuracy with minimum computational complexity of CBIR system, both pre-classification (reducing the search space mechanism) and FRF mechanism have been incorporated in Scheme-III.

- ***Chapter 4: Feature Extraction and Compaction : A Comprehensive Study*** [178]

  An extensive comparative assessment of five different transforms domain features considering various filter combinations are presented. Three different feature representation schemes and three different classifiers have been used for this purpose. Extensive experiments on four widely used benchmark image databases are 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 the Maximal Information Compression Index (MICI).

- ***Chapter 5: Content-Based Image Retrieval using Graph-Based Relevance Feedback Mechanism*** [179]

  In this chapter, we have developed a CBIR system using best image representation techniques as studied in Chapter 4. Additionally, a graph theoretic relevance feedback mechanism has also been developed with minimum computational cost and satisfactory performance improvement.

- ***Chapter 6: CBIR system design for Retrieval of Medical Images*** [180, 187]

  Two different CBIR systems have been developed for medical applications. The systems are tested on large X-ray image and endoscopic image database.

- ***Chapter 7: Conclusion and Future Scope***

  The concluding remarks along with the scope for future scope of research are presented in Chapter 7.

We have used mathematical model/tools for designing the CBIR systems to be discussed in contributory chapters. In view of that, brief overview on different mathematical models and tools used in this thesis are discussed in Chapter 2.

Chapter 2

Multi-scale Image Features and Related Mathematical Tools

2.1 Introduction

The automatic interpretation of semantically meaningful information from low-level visual features like color, texture shape etc., is the focus of interest for most research on image retrieval. The user's requirements can be specified by querying from a database in the form of attributes like color, texture, shape, spatial relationships, a sketch, an example image or in terms of keywords. The visual contents of an image can be defined from (a) Low level properties like color, texture, shape and spatial relations [188] and (b) Semantic properties: that correspond to objects, scenes, impression, emotions and meanings associated with the combinations of low level properties [189].

Color is the most extensively used visual content for image retrieval. Commonly used color space for image retrieval include RGB, YCbCr, CIE L*a*b*, HSV, and opponent color space. The three color components of the color spaces are hue, saturation (lightness) and value (brightness). Texture is another important property of images. Texture contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. Effective representation of textures can be made from the statistical and structural properties of the brightness pattern [29]. Compared with color and texture features, shape features are usually described after images have been segmented into regions or objects. A good shape representation feature for an object should be invariant to translation, rotation and scaling [190].

The following section begins with the theoretical preliminaries of the different mathematical tools/techniques used in the thesis. A brief description of the different statistical measures used to objectively evaluate the performances of the proposed solutions is also included in this chapter.

2.2 MPEG-7 Feature Descriptors

MPEG-7 is an ISO/IEC standard developed by MPEG to facilitate effective uses of audio, visual and motion picture description to address multimedia retrieval. It is a standard for the multimedia content data that supports interpretation of the information, which can be passed onto, or accessed by a device or a computer code. MPEG-7 is broadly classified into general visual descriptors and domain specific visual descriptors. The former include color, texture, shape, and motion features, while the latter are application dependent. It can capture very closely low level visual description through number of descriptors. Both global and local descriptors are based on the multi-resolution Trace transform which constructs a set of 1-D representations of an image [191]. It is based on the MPEG-7 image signature tools [192], which form a part of the MPEG-7 standard. This set of tools was designed for robust and fast image and video retrieval [193]. MPEG-7 is not targeted at any one application in particular; rather it supports a range of applications as possible [36, 49].

Color is perhaps the one most expressive of all the visual features and has been extensively studied in the image retrieval research during the last decade. The color descriptors consist of Scalable Color Descriptor (SCD), Color Structure Descriptor (CSD), Dominant Color Descriptor (DCD) and Color Layout Descriptor (CLD). Similarly, texture is also a powerful low-level descriptor for image search and retrieval applications. MPEG-7 is considering three texture descriptors and it includes Texture Browsing Descriptor (TBD), Homogeneous Texture Descriptor (HTD) and Edge Histogram Descriptor (EHD). Shape and motion descriptor is not considered in this thesis. However in this thesis, we have used only two types of descriptor *i.e.*, CSD and EHD.

CSD represents an image by considering both color distribution (color histogram) and the local spatial structure of the color. An 8×8 structuring block is used for color structure information of the descriptor. EHD captures global spatial distribution of the edges, by dividing the image into 16 sub-images, with fixed number of blocks. Edge information is

then calculated for each block in five edge categories: Vertical, Horizontal, 45° diagonal, 135° diagonal and non directional edge. It is expressed as a 5 bin histogram for each image block. CSD and EHD consist of 256 and 80 features respectively. The details of CSD and EHD are describe in [36].

2.3 Multi-resolution Analysis Tools

In computer graphics applications, a continuous function is represented using a set of discrete samples. So, a computer with finite storage capacity, can represent an image by a finite collection of samples, and will in general, be an approximation of the original. The samples may be used to reconstruct the function that is defined over a particular domain by treating them as coefficients for a corresponding set of basis functions defined over the same domain. Thus, given the discrete samples, one can evaluate the continuous function everywhere, not just at the locations of the samples. The size of the intervals between samples dictates the resolution of the representation and the degree to which its is able to describe (resolve) fine details in the function it represents. Tiny details of frequencies higher than half the sampling rate cannot be resolved. When we zoom in to a signal/image there is no more fine detail to be seen; the signal/image usually looks blurry/blocky, depending on the nature of basis functions used. In the case of an image, for example, the pixels are usually organized in a regular rectangular grid.

Multi-resolution analysis representation often divided into two parts such as (i) Multi-scale Analysis (MSA) representation and (ii) Multi-scale Geometric Analysis (MGA) representation. MSA cover the wavelet and its variants whereas MGA cover the higher version of wavelet *i.e.*, RT, CVT, etc. Brief descriptions of MSA and MGA tools are given in the following subsection.

2.3.1 Multi-scale Analysis Tools

MSA decomposition at each level provides two images of the same size as that of the image at the previous level: the coarser scale and detail images [194]. MSA or Wavelet and its variants analysis provides an effective multi-resolution representation of a signal/image and is capable of quickly capturing the essence of a data set with only a small set of coefficients. Wavelet Transform (WT) and M-band Wavelet Transform (M-band) are discussed briefly in the following section.

2.3.1.1 Wavelet Transform

The Wavelet Transform (WT) is a signal decomposition onto a set of basis functions called wavelets. The wavelets are obtained from single prototype function by scaling and shifts. This is the standard 2 band wavelet transform [195].. The full discrete wavelet expansion of a signal $f(x) \in l_2$ (l_2 is the space of square summable functions) is given as

$$f(x) = \sum_{k \in Z} s_{J,k}\phi_{J,k}(x) + \sum_{j=1}^{J}\sum_{k \in Z} d_{j,k}\psi_{j,k}(x) \tag{2.1}$$

where ϕ and ψ are the scaling and wavelet functions, respectively and are associated with the analyzing (or synthesizing) filter h and g. $d_{j,k}$'s are the wavelet coefficients and $s_{J,k}$'s are the expansion coefficients of the coarser signal approximation of $f(x)$. It also follows from this construction that the family of sequences $\{\phi_{J,k}, \psi_{1,k}, \psi_{2,k}, ..., \psi_{J,k}\}_{k \in Z}$ constitutes an orthonormal basis. The discrete normalized basis function are defined as

$$\phi_{j,k}(x) = 2^{j/2} h_j(2^j x - k) \tag{2.2}$$

$$\psi_{j,k}(x) = 2^{j/2} g_j(2^j x - k) \tag{2.3}$$

where j and k are the scale and translation indices, respectively and the factor $2^{j/2}$ is an inner product normalization. Z corresponds to integer.

2.3.1.2 M-Band Wavelet Transform

M-band wavelet are a direct generalization of the conventional dyadic wavelets [196]. The standard dyadic wavelets are not suitable for the analysis of high frequency signals with relatively narrow bandwidth. To resolve this problem M-band wavelet were developed as a direct generalization of the 2 band wavelet of Daubechies. These M-band wavelets are able to zoom in onto narrowband high frequency components of a signal and have been found to give better energy compaction than 2 band wavelet [197, 198].

An M-band wavelet system form a tight frame for the set of square integrable functions defined over the set of real numbers $L^2(R)$. There are $M-1$ wavelets, $\psi_i(x)$, $i = 1,...,M-1$ associated with the scaling function. For any function $f(x) \in L^2(R)$, it can be shown that

$$f(x) = \sum_{i=1}^{i=M-1} \sum_{j \in Z} \sum_{k \in Z} \langle f(x), \psi_{i,j,k}(x) \rangle \psi_{i,j,k}(x) \tag{2.4}$$

where Z represents the set of integers and $\langle , \rangle$ is the inner product operator. The $\psi_{i,j,k}(x)$ functions are obtained by scaling and shifting the corresponding wavelet $\psi_i(x)$:

$$\psi_{i,j,k}(x) = M^{j/2} \psi_i(M^j x - k), i = 1,...,M-1, k \in Z, j \in Z. \tag{2.5}$$

Given a scaling function $\psi_0(x)$ in $L^2(R)$, the wavelet functions are defined as

$$\psi_i(x) = \sqrt{M} \sum_{k=0}^{k=N-1} h_i(k) \psi_0(Mx - k), i = 1,...,M-1. \tag{2.6}$$

The scaling function satisfies the recursive equation and is compactly supported in

$[0,(N-1)/(M-1)]$,

$$\psi_0(x) = \sqrt{M}\sum_{k=0}^{k=N-1} h_0(k)\psi_0(Mx-k), \tag{2.7}$$

where the sequence h_0 is the scaling filter [196, 198, 199] and satisfies the following linear and quadratic constraints:

$$\sum_{k=0}^{k=N-1} h_0(k) = \sqrt{M}, \tag{2.8}$$

$$\sum_{k=0}^{k=N-1} h_0(k)h_0(k+Ml) = \delta l. \tag{2.9}$$

The $(M-1)h_i$ vectors called the wavelet filters [196] satisfy the equation

$$\sum_{k=0}^{k=N-1} h_i(k)h_j(k+Ml) = \delta(l)\delta(i-j). \tag{2.10}$$

2.3.2 Multi-scale Geometric Analysis Tools

The success of WT is mainly due to the good performance for piecewise smooth functions in one dimension. Unfortunately, this is not the case in two dimensions. In essence, WT are good at catching point singularities, but two-dimensional piecewise smooth signals resembling images have one-dimensional singularities. That is, smooth regions are separated by edges, and while edges are discontinuous across, they are typically smooth curves. Intuitively, WT in two dimensions are obtained by a tensor-product of one dimensional wavelets and they are thus good at isolating the discontinuity across an edge, but will not see the smoothness along the edge. Therefore, the problem of WT is that it is inherently non-supportive to directionality and anisotropy. To overcome the weakness of WT in higher dimensions, in recent past, a theory called MGA for high-dimensional signals has been proposed. Using this MGA tools, different types of statistical features have been proposed for representation of images in various image processing applications. The four

different types of MGA tools *i.e.*, CVT, CNT, NSCT and RT are described in the following subsequent subsection:

2.3.2.1 Curvelet Transform

Traditional WT 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 a curve as a superposition of bases of various lengths and widths obeying the scaling law $width \approx length^2$ [200]. CVT uses a parabolic scaling law to achieve anisotropic directionality. The first-generation CVT based on multiscale ridgelets combines with a spatial bandpass filtering operation to isolate different scales. However, this transform is very complicated and redundant, including many steps, such as subband decomposition, smooth partitioning, re-normalization, and ridgelet analysis [201]. Later, a considerably simpler second-generation CVT based on frequency partition technique was proposed by the same authors [202]. The frequency domain of image is separated into disjoint wedge regions. Then, the local fourier transform is implemented on these regions.

Assume that we work throughout in two dimensions, that is, R^2, and set x as spatial variable, w as a frequency domain variable, r and θ as polar coordinates in the frequency domain [21]. Let $W(r)$ and $V(t)$ be a pair of nonnegative, real-valued, and smooth window functions, called "radial window" and "angular window", respectively. These windows will always satisfy the admissibility conditions:

$$\begin{aligned} \sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \quad r \in (\frac{3}{4}, \frac{3}{2}), \\ \sum_{t=-\infty}^{\infty} V^2(t-l) = 1, \quad t \in (\frac{-1}{2}, \frac{1}{2}). \end{aligned} \tag{2.11}$$

For each scale $j \geq j_0$, we introduce the frequency window U_j defined in the fourier

domain by

$$U_j(r,\theta) = 2^{\frac{-3j}{4}} W(2^{-j}r) V\left(\frac{2^{\lfloor \frac{j}{2} \rfloor}\theta}{2\pi}\right), \tag{2.12}$$

where $\lfloor \frac{j}{2} \rfloor$ is the integer part of $\frac{j}{2}$.

Define a "mother" curvelet as $\varphi_j(x)$, and its fourier transform $\hat{\varphi}_j(w) = U_j(w)$. Then all curvelets at scale 2^{-j} are obtained by rotations and translations of φ_j.

Introduce the equispaced sequence of rotation angles $\theta_l = 2\pi \times 2^{-\lfloor \frac{j}{2} \rfloor} \times l$, with $l = 0,1,...,0 \leq \theta_l < 2\pi$, and the sequence of translation parameters $k = (k_1,k_2) \in Z^2$. Then define curvelets at scale 2^{-j}, orientation θ_l, and position $x_k^{(j,l)} = R_{\theta_l}^{-1}(k_1 \times 2^{-j},\ k_2 \times 2^{-\frac{j}{2}})$ by

$$\varphi_{j,l,k}(x) = \varphi_j[R_{\theta_l}(x - x_k^{(j,l)})] \tag{2.13}$$

where R_θ is the rotation by θ radians.A CVT coefficient is then given by

$$c(j,k,l) = \langle f, \varphi_{j,k,l} \rangle = \int_{R^2} f(x)\overline{\varphi_{j,l,k}(x)}dx, \tag{2.14}$$

According to Plancherel's theorem, the CVT can be expressed as the integral over the frequency plane

$$\begin{aligned} c(j,l,k) &= \tfrac{1}{(2\pi)^2} \int \hat{f}(w)\overline{\hat{\varphi_{j,l,k}}(w)}dw \\ &= \tfrac{1}{(2\pi)^2} \int \hat{f}(w) U_j(R_{\theta_l} w) e^{i\langle x_k^{(j,l)}, w \rangle} dw. \end{aligned} \tag{2.15}$$

2.3.2.2 Contourlet Transform

The major drawback of WT in two dimensions is its limited ability in capturing directional information. In light of this, Do and Vetterli [203] have developed the CNT, based on an efficient 2D multiscale and Directional Filter Bank (DBF). CNT not only possess the main features of DWT, but also offer a high degree of directionality and anisotropy. It allows for different and flexible number of directions at each scale, while achieving nearly critical

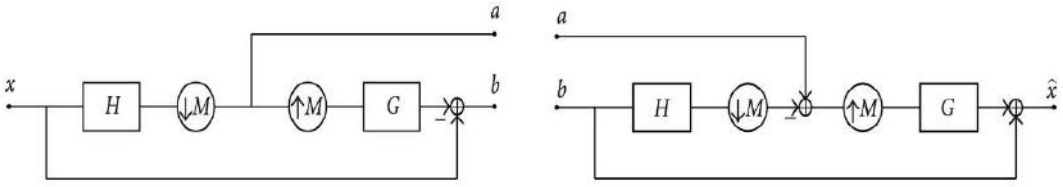


Fig. 2.1 Laplacian pyramid scheme: (a) Analysis and (b) Synthesis.

sampling. In addition, CNT uses iterated filter banks, which makes it computationally efficient ($O(N)$ operations for an N-pixels image).

CNT gives a multiresolution, local and directional expansion of image using Pyramidal DFB (PDFB). The PDFB combines Laplacian Pyramid (LP) which captures the point discontinuities, with a DFB which links these discontinuities into linear structures.

LP scheme is shown in Figure 2.1. Here, the input image x is first lowpass filtered by analysis filter H and then downsampled to produce a coarse approximation a. It is then interpolated and passed through the synthesis filter G. The resulting image is subtracted from the original image x to obtain the bandpass image b. This process can be iterated repeatedly on the coarser version of the image a. LP is a multiscale decomposition of $L^2(R^2)$ into series of increasing resolution subspaces which are orthogonal complements of each other as follows [204]:

$$L^2(R^2) = V_{j_0} \oplus (\bigoplus_{j=J_0}^{-\infty} W_j) \tag{2.16}$$

where, V_{j_0} is the approximation subspace at the scale 2^{j_0}, W_j is the detail in the finer scale 2^{j-1}. In the LP, each subspace W_j is spanned by a frame $\mu_{j,n}(t)_{n\varepsilon Z^2}$ that utilizes a uniform grid on R^2 of intervals $2^{j-1} \times 2^{j-1}$.

In 1992, Bamberger and Smith constructed a 2D DFB that can be maximally decimated while achieving perfect reconstruction[205]. 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. A l-level tree-structured DFB is equivalent to a 2^l parallel channel filter bank with equivalent filters and overall sampling matrices as shown in Figure 2.2. As shown in Figure 2.2, corresponding to the subbands indexed, the equivalent analysis and synthesis filters are denoted using H_k and G_k, $0 \leq k < 2^m$.

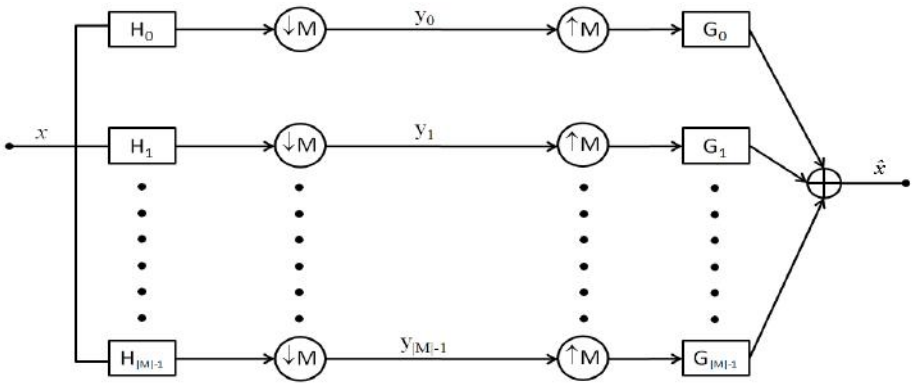


Fig. 2.2 Construction of DBF.

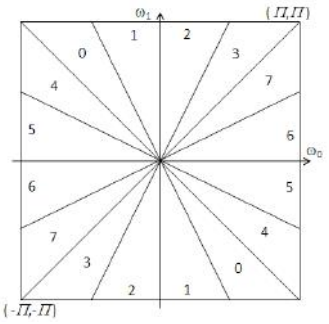


Fig. 2.3 Frequency partition of contourlet transform.

A l-level DFB generates a perfect directional basis for discrete signal in $l^2(Z^2)$ that is composed of the impulse responses of 2^l directional synthesis filters and their shift. They can be represented as follows:

$$g_k^{(l)}[n - S_k^{(l)} n]_{0 \leq k < 2^l, n \varepsilon Z^2} \tag{2.17}$$

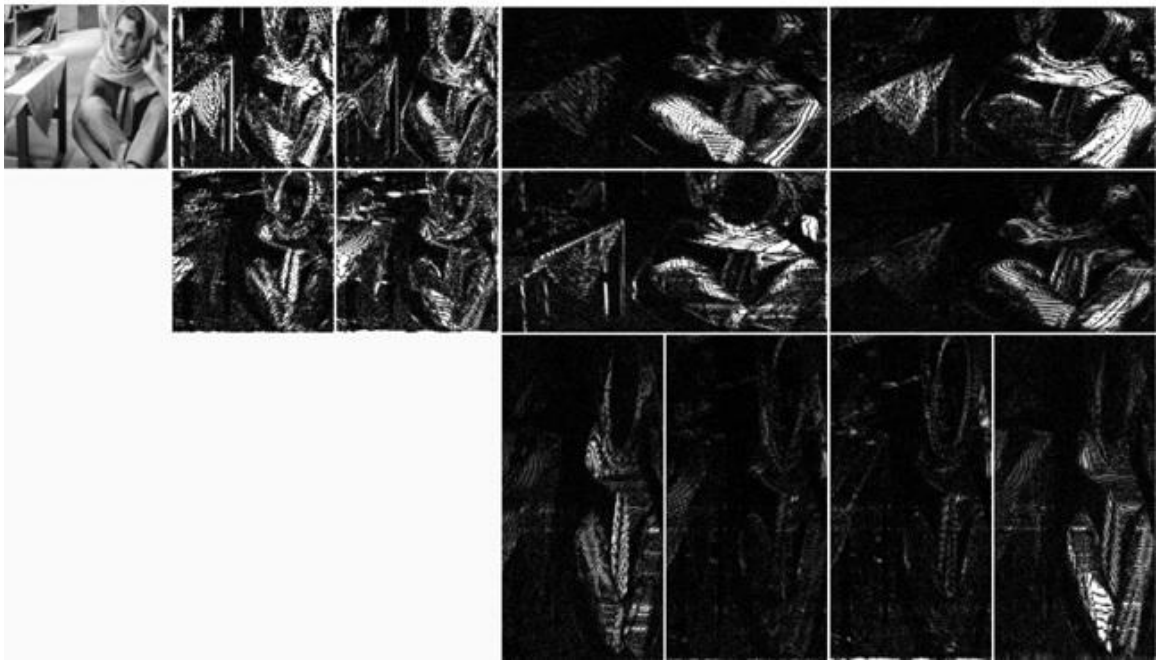


Fig. 2.4 Contourlet transform of *Barbara* image.

$$g[n] = \frac{2\Pi}{n_1}[\psi(\frac{n_1(l+1)}{N} + n_2) - \psi(\frac{n_1 l}{N} + n_2)] \tag{2.18}$$

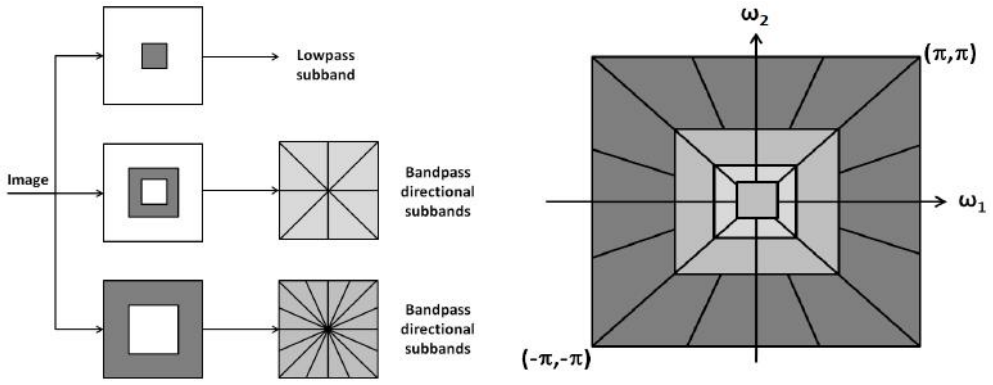
where, $N = 2^{n-2}$ and $\psi(x)$ is similar to the common *sin* function.

$$\psi(x) = \frac{1 - cos(\pi x)}{\pi x} \tag{2.19}$$

In CNT, applying a l_j-level DBF to the detail subspace W_j results in a decomposition with 2^{l_j} directional subspaces as follows:

$$W_j = \bigoplus_{k=0}^{2^{l_j-1}} W_{j,k}^{l_j} \tag{2.20}$$

A DFB is designed to capture the high frequency content like smooth contours and directional edges. Fig. 2.3, shows the frequency partition of CNT, and Fig. 2.4, presents an image along with its subbands after decomposition by CNT.



(a) (b)

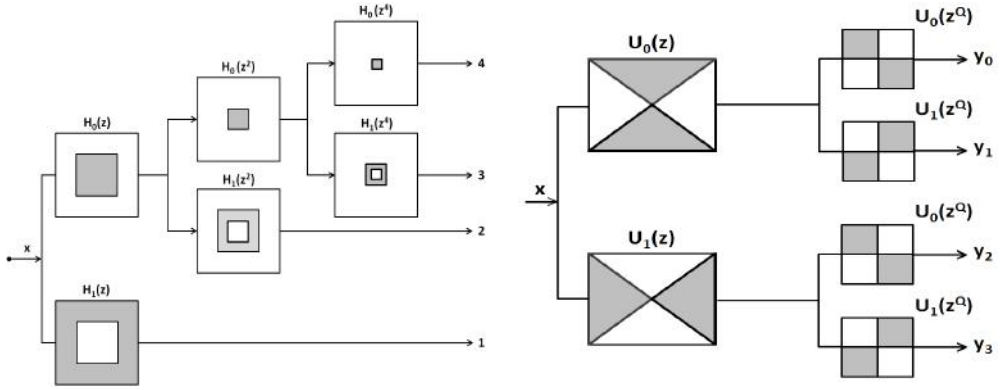
Fig. 2.5 Non-subsampled contourlet transform (a) NSFB structure that implements the NSCT. (b) Idealized frequency partitioning obtained with the proposed structure.

2.3.2.3 Nonsubsampled Contourlet Transform

In this section, we briefly describe the NSCT, which will be adopted in our system to devise a proper image representation.

NSCT is a fully shift-invariant, multi-scale, and multi-direction expansion with fast implementability [206]. As opposed to the CNT, which is not shift-invariant due to the presence of down-samplers and up-samplers in both the LP and Directional Filter Bank (DFB) stages, NSCT achieves the shift-invariance property by using Non-Subsampled Pyramid (NSP) filter banks and Non-Subsampled DFB (NSDFB).

Non-Subsampled Pyramid (NSP) filter bank is a shift-invariant filtering structure that leads to a subband decomposition that resembles the LP, which ensures the multi-scale property of the NSCT. As shown in Fig. 2.5, it is constructed by using two-channel non-subsampled 2D filter banks, which produce a low-frequency and a high-frequency image at each NSP decomposition level. Filters at subsequent stages are obtained by upsampling the low-pass filters at the first stage. As a result, NSP 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. Fig. 2.5(a) gives the NSP decomposition with $\mathsf{k}=3$ levels.



(a) (b)

Fig. 2.6 (a) Non-subsampled Pyramid Filter Bank: Three-stage decomposition. (b) Four-channel NSDFB constructed with two-channel fan filter banks.

The NSDFB is constructed by eliminating the downsamplers and upsamplers of the DFB and by upsampling the filters accordingly [206]. This results in a tree composed of two-channel NSFB, described in Fig. 2.5(b) (4 channel decomposition). At each stage of the NSP, 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 NSP and NSDFB is depicted in Fig. 2.6(a). The resulting filtering structure approximates the ideal partition of the frequency plane displayed in Fig. 2.6(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 [206] for further details about NSCT.

Among the different MGA tools NSCT has better frequency selectivity and regularity, as well as it is a flexible multi-scale, multi-directional, and shift-invariant image decomposition method [207]. NSCT coefficients are capable of capturing the fine details present in the image, which is essential in CBIR, as natural images are full of subtle image information. Moreover, it is widely accepted that the HVS bases its perception on multiple channels that are tuned to different ranges of spatial frequencies and orientations. Measurements of the receptive fields of simple cells in the primary visual cortex revealed that these

channels exhibit approximately a dyadic structure [208]. This behavior is well matched by NSCT decomposition. More details of the NSCT are available in [206].

2.3.2.4 Ripplet Transform

DWT and its variants have been used extensively for image processing applications. But the problem with DWT is that it is inherently non-supportive to directionality and anisotropy. To address these problems, Jun Xu *et al.* have proposed a new MGA tool called RT [209]. RT is a higher dimensional generalization of the curvelet transform (CVT), capable of representing images or $2D$ signals at different scales and different directions. To achieve anisotropic directionality, CVT uses a parabolic scaling law [210]. From the perspective of micro-local analysis, the anisotropic property of CVT guarantees resolving $2D$ singularities along C^2 curves. Whereas, RT provides a new tight frame with sparse representation for images with discontinuities along C^d curves [209].

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.* have generalized the scaling law of CVT, which resulted in RT. RT generalizes CVT by adding two parameters, i.e., support c and degree d. CVT is just a special case of RT with $c = 1$ and $d = 2$. The anisotropy capability of representing singularities along arbitrarily shaped curves of RT, is due to these two new parameters c and d.

As digital image processing needs discrete transform instead of continuous transform, here we describe the discretization of RT [209]. The discretization of continuous RT is based on the discretization of the parameters of ripplet functions. The scale parameter a is sampled at dyadic intervals. The position parameter b and the rotation parameter θ are sampled at equal-spaced intervals. a_j, $\overrightarrow{b}_k$ and θ_l substitute a, $\overrightarrow{b}$ and θ, respectively, and satisfy that $a_j = 2^{-j}$, $\overrightarrow{b}_k = [c \cdot 2^{-j} \cdot k_1, 2^{-j/d} \cdot k_2]^T$ and $\theta_l = \frac{2\Pi}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot l$, where

$\overrightarrow{k} = [k_1, k_2]^T$, and j, k_1, k_2, $l \in \mathbb{Z}$. $(\cdot)^T$ denotes the transpose of a vector. $d \in \mathbb{R}$, since any real number can be approximated by rational numbers, so we can represent d with $d = n/m$, $n, m \neq 0 \in \mathbb{Z}$. Usually, we prefer $n, m \in \mathbf{N}$ and n, m are both primes. In the frequency domain, the corresponding frequency response of ripplet function is in the form

$$\widehat{\rho}_j(r,\omega) = \frac{1}{\sqrt{c}} a^{\frac{m+n}{2n}} W(2^{-j} \cdot r) V(\frac{1}{c} \cdot 2^{-\lfloor j\frac{m-n}{n} \rfloor} \cdot \omega - l) \tag{2.21}$$

where, W and V are the *radial-window* and the *angular-window*, respectively. These two windows satisfy the following admissibility conditions:

$$\sum_{j=0}^{+\infty} |W(2^{-j} \cdot r)|^2 = 1 \tag{2.22}$$

$$\sum_{l=-\infty}^{+\infty} |V(\frac{1}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot \omega - l)|^2 = 1 \tag{2.23}$$

given c, d and j. These two windows partition the polar frequency domain into 'wedges'. The 'wedge' corresponding to the ripplet function in the frequency domain is

$$H_{j,l}(r,\theta) = \{2^j \leq |r| \leq 2^{2j}, |\theta - \frac{\pi}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot l| \leq \frac{\pi}{2} 2^{-j}\} \tag{2.24}$$

The discrete RT of an $M \times N$ image $X(m,n)$ is as follows:

$$R_{j,\overrightarrow{k},l} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X(m,n) \overline{\rho_{j,\overrightarrow{k},l}(m,n)} \tag{2.25}$$

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

As a generalized version of CVT, RT is not only capable of resolving 2D singularities, but it also has some useful properties:

1. It forms a new tight frame in a function space. Having good capability of localiza-

tion in both spatial and frequency domain, it provides a more efficient and effective representation for images or 2D signals.

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

Jun Xu *et al.* have showed that RT can provide a more effective representation for images with singularities along smooth curves [209]. It outperforms DCT and DWT in nonlinear approximation, when the number of retained coefficients is small. RT can achieve roughly 2 dB higher Peak-Signal-to-Noise Ratio (PSNR) on average than Joint Photographic Experts Group (JPEG), and provide better visual quality than JPEG2000 at low bit-rates, when applied to image compression. In case of image denoising application, RT performs better than CVT and DWT. RT produces high quality fused images, when applied in the medical image fusion domain [211]. All these experiments show that RT based image coding is suitable for representing texture or edges in images.

2.4 Different Tools Used for Representation, Classification and Learning

A brief discussion about Pulse Coupled Neural Network (PCNN), Fuzzy C-Means Clustering, Multilayer Perceptron (MLP), LSSVM and Random Forest (RF), which are used to different stages in developing the algorithms has been furnished. Although, a detailed discussion on the above mentioned topics are beyond the scope of this work, only the relevant portions that will be required in the rest of the thesis will be discussed.

2.4.1 Pulse Coupled Neural Network

Biological systems have always been an inspiration for developing computer vision, image and video processing algorithms. In late 1980s, during the study of cat visual cortex, Eckhorn *et al.* discovered that the midbrain in an oscillating way creates binary images that could extract different features from the visual impression [212]. Based on these binary images the actual image is created in the cat's brain. To simulate this behavior, they developed a neural network, called Eckhorn's model. Similar neural behavior was also found by Rybak *et al.* based on study of the visual cortex of guinea pig and they developed a neural network, called Rybak's model [213]. Because these models provided a simple, effective way for studying synchronous pulse dynamics in networks, they was recognized as being very potential in image processing [214–216]. Johnson *et al.* carried on a number of modifications and variations to tailor this model's performance as image processing algorithms [217, 218]. This modified neural model is called pulse-coupled neural networks (PCNN).

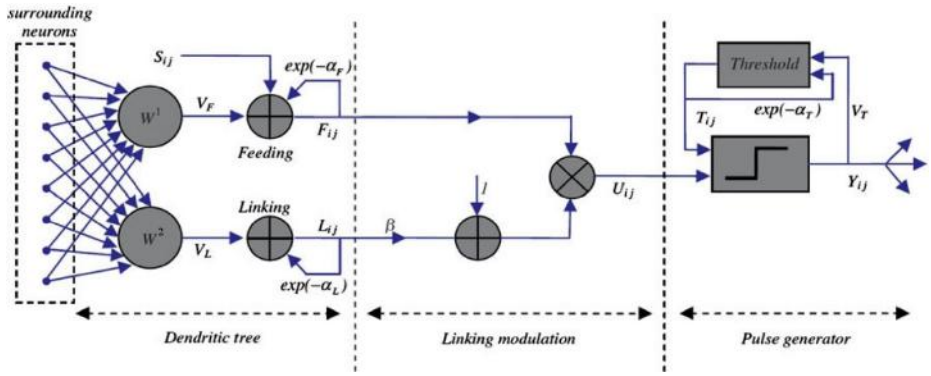


Fig. 2.7 Structure of PCNN.

The PCNN is a single layer, two-dimensional, laterally connected network of integrate-and-fire neurons, with a 1:1 correspondence between the image pixels and network neurons. This is an unsupervised neural network with self-organizing capability. The output images at different iterations typically represent some segments or edges information of the input

image. The PCNN neuron's structure is shown in Fig. 2.7. The neuron consists of an input part (dendritic tree), linking part and a pulse generator. The neuron receives the input signals from feeding and linking inputs. Feeding input is the primary input from the neuron's receptive area. The neuron receptive area consists of the neighboring pixels of corresponding pixel in the input image. Linking input is the secondary input of lateral connections with neighboring neurons. The difference between these inputs is that the feeding connections have a slower characteristic response time constant than the linking connections. The standard PCNN model is described as iteration by the following equations [214, 215]:

$$F_{i,j}[n] = e^{-\alpha_F} F_{i,j}[n-1] + V_F \sum_{k,l} w_{i,j,k,l} Y_{i,j}[n-1] + S_{i,j} \tag{2.26}$$

$$L_{i,j}[n] = e^{-\alpha_L} L_{i,j}[n-1] + V_L \sum_{k,l} m_{i,j,k,l} Y_{i,j}[n-1] \tag{2.27}$$

$$U_{i,j}[n] = F_{i,j}[n](1 + \beta L_{i,j}[n]) \tag{2.28}$$

$$Y_{i,j}[n] = \begin{cases} 1, & U_{i,j}[n] > T_{i,j}[n] \\ 0, & \text{otherwise} \end{cases} \tag{2.29}$$

$$T_{i,j}[n] = e^{-\alpha_T} T_{i,j}[n-1] + V_T Y_{i,j}[n] \tag{2.30}$$

In the Eq.(2.26) to Eq.(2.30), the indexes i and j refer to the pixel location in the image, k and l refer to the dislocation in a symmetric neighborhood around one pixel, and n denotes the current iteration (discrete time step). Here n varies from 1 to N (total number of iterations). The dendritic tree is given by Eqs.(2.26)–(2.27). The two main components F and L are called feeding and linking, respectively. $w_{i,j,k,l}$ and $m_{i,j,k,l}$ are the synaptic weight coefficients and S is the external stimulus. V_F and V_L are normalizing constants. α_F and α_L

are the time constants; generally, $\alpha_F < \alpha_L$. The linking modulation is given in Eq.(2.28), where $U_{i,j}[n]$ is the internal state of the neuron and β is the linking parameter. The pulse generator determines the firing events in the model in Eq.(2.29). $Y_{i,j}[n]$ depends on the internal state and threshold. The dynamic threshold of the neuron is Eq.(2.30), where V_T and α_T are normalized constant and time constant, respectively.

The input stimulus (the pixel intensity) is received by the feeding element and the internal activation element combines the feeding element with the linking element. The value of internal activation element is compared with a dynamic threshold that gradually decreases at iteration. The internal activation element accumulates the signals until it surpasses the dynamic threshold and then fires the output element and the dynamic threshold increases simultaneously strongly. The output of the neuron is then iteratively fed back to the element with a delay of one iteration. Based on the application and requirements, various modifications are made in the original PCNN. Moreover, it is very suitable for perfect image segmentation. For the PCNN, the neurons associated with each group of spatially connected pixels with similar intensities tend to pulse together [218]. Even when the intensity ranges of adjacent regions overlap, the PCNN can perfectly segment the regions if certain conditions are satisfied. Note that when an image is perfectly segmented, each pixel is correctly assigned to the region it belongs to. This is the basic principle of segmentation of the PCNN. Interested readers can find more details on PCNN in [215, 216].

2.4.2 Fuzzy C-Means clustering

Fuzzy C-Means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition problems. A large family of fuzzy clustering algorithms is based on minimization of the FCM objective

function, formulated as [219, 220]:

$$J_m = \Sigma_{i=1}^{c} \Sigma_{k=1}^{N} (\mu_{ik})^m \|z_k - v_i\|_A^2 \tag{2.31}$$

where μ_{ik} is the degree of membership of z_k in the cluster i, z_k is the k^{th} of d-dimensional measured data, v_i is the d-dimension center of the cluster, $\| * \|$ is any norm expressing the similarity between any measured data and the center.

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown as above, with the update of membership μ_{ik} and the cluster centers v_i by:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|z_k - v_i\|}{\|z_k - v_j\|} \right)^{2/(m-1)}}; \qquad 1 \leq i \leq c, 1 \leq k \leq N \tag{2.32}$$

and

$$v_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m z_k}{\sum_{k=1}^{N} (\mu_{ik})^m}; \qquad 1 \leq i \leq c. \tag{2.33}$$

This iteration will stop when $max_{ik} \left\{ \left| \mu_{ik}^{(j+1)} - \mu_{ik}^{(j)} \right| \right\} < \varepsilon$, where ε is a termination criterion whose value lies between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m. In fuzzy clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be in the cluster to a lesser degree than points in the center of cluster. The degree of belonging, μ_{ik}, is related inversely to the distance from z to the cluster center.

2.4.3 Multilayer Perceptron

The MLP is the standard neural network to use for supervised learning [221]. The MLP classifier used in this thesis 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 [222].

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.The number of nodes of the hidden layer in this study are computed by considering the rule of thumb. 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 network is initialized with random weights and biases, and was then trained using the Levenberg-Marquardt algorithm (LM) [221]. A momentum term is added to increases the learning rate with stability. The performance of the network is measured by Mean Squared Error (MSE). The smaller the MSE is, the better the network performs. The details of the MLP configuration used in this thesis are discussed in [154, 178, 223].

2.4.4 Least Square-Support Vector Machine

Recently, SVM have been shown to be effective for many classification problems [224]. For binary-class classifications, SVM constructs an optimal separating hyperplane between the positive and negative classes with the maximal margin. It can be formulated as a Quadratic Programming (QP) problem involving inequality constraints. The most critical drawback of SVM is its high computational complexity for high dimensional data sets. To reduce the computational demand, the LS-SVM/LSSVM is developed which attempts to minimize the least square error on the training samples while simultaneously maximizing the margin between two classes. LS-SVM avoids solving quadratic programming problem and simplifies the training procedure. While in classical SVM's many support values are zero (nonzero values correspond to support vectors), in LS-SVM, the support values are proportional to the errors. A two-norm is taken with equality instead of inequality constraints so as to ob-

tain a linear set of equations instead of a QP problem in the dual space [225, 226]. Given a training set:

$$\{(x_i, y_i)\}_{i=1}^{N} \quad and \quad y_i = \{+1, -1\} \tag{2.34}$$

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:

$$\min_{w,b,e} \mathscr{J}(w,b,e) = \frac{1}{2} w'w + \frac{1}{2} C \sum_{i=1}^{n} e_i^2 \tag{2.35}$$

subject to the equality constraints

$$y_i[w'\varphi(x_i) + b] = 1 - e_i \tag{2.36}$$

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.(2.35) is defined as follows:

$$\mathscr{L}(w, e_i, b, \alpha_i) = \min_{w,b,e} \mathscr{J}(w,b,e) - \sum_{i=1}^{n} \alpha_i \{y_i[w'\varphi(x_i) + b] - 1 + e_i\} \tag{2.37}$$

where, α_i are Lagrange multipliers. The Karush-Kuhn-Tucker (KKT) conditions for optimality: $\frac{\partial \mathscr{L}}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i y_i \varphi(x_i)$; $\frac{\partial \mathscr{L}}{\partial e_i} = 0 \rightarrow \alpha_i = C e_i$; $\frac{\partial \mathscr{L}}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0$; $\frac{\partial \mathscr{L}}{\partial \alpha_i} = 0 \rightarrow y_i[w'\varphi(x_i) + b] - 1 + e_i = 0$, is the solution to the following linear system

$$\begin{bmatrix} 0 & -Y \\ Y & \varphi\varphi' + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \overline{1} \end{bmatrix} \tag{2.38}$$

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

$$f(x) = sgn[\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b] \tag{2.39}$$

2.4.5 Random Forest

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 [227, 228].

$$\Delta E = -\sum_{i} \frac{|S_i|}{|S|} E(S_i) \tag{2.40}$$

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_{class}) of an input image if the final class distribution $p(c_i|L)$ has the maximum value [229].

$$f_{class} = \arg max\{\frac{1}{T}\sum_{t=1}^{T} P(c_i|l_t)\} \tag{2.41}$$

2.5 Mechanisms Used for Feature Selection and Evaluation

The feature selection technique is defined as the process that aims at finding adequate subsets of features for a given learning task. Feature selection is a central problem in machine learning and pattern recognition. The use of feature selection techniques may improve the accuracy of a classifier learnt from data by helping to avoid the "curse of dimensionality" issues and may speed up both the training time and improve the test (generalization) processes. There are several feature selection algorithm in the literature [230] but in this thesis we have used two feature selection algorithm: (a) PCA, (b) Maximal Information Compression Index (MICI) and (c) Fuzzy Feature Evaluation Index (FFEI). In this section, MICI and FFEI are discussed as PCA is very common and widely used in CBIR literature [231].

2.5.1 Maximal Information Compression Index

In this section, we discuss a measure called MICI to be used for feature selection, for measuring similarity between two random variables based on linear dependencies between

them.

Let $\sum$ be the covariance matrix of random variables x and y. Define MICI as $\lambda_2(x,y)$ = smallest eigenvalue of $\sum$, i.e.,

$$2\lambda_2(x,y) = var(x) + var(y) - \sqrt{(var(x)+var(y))^2 - 4var(x)var(y)(1-\rho(x,y)^2)} \quad (2.42)$$

where var() denotes the variance of a variable and cov() the covariance between two variables and $\rho(x,y)$ is the correlation coefficient of the two random variables x and y defined as

$$\rho(x,y) = \frac{cov(x,y)}{\sqrt{var(x)var(y)}} \quad (2.43)$$

The value of λ_2 is zero (theoritically) when the features are linearly dependent and increases as the amount of dependency decreases. It may be noted that the measure λ_2 is nothing but the eigenvalue for the direction normal to the principal component direction of feature pair(x,y). It is shown [232] that the maximum information compression is achieved if a multivariate data is projected along its principal component direction. The corresponding loss of information in reconstruction of the pattern is equal to the eigenvalue along the direction normal to the principal component. Hence, λ_2 is the amount of reconstruction error committed if the data is projected to a reduced dimension in the best possible way. Therefore, it is a measure of the minimum amount of information loss or the maximum amount of information compression, possible

The significance of λ_2 can also be explained geometrically in terms of linear regression. it can be easily shown [233] that the value of λ_2 is equal to the sum of the squares of the perpendicular distances of the points(x,y) to the best fit line $y = \hat{a} + \hat{b}x$, obtained by minimizing the sum of the squared perpendicular distances. The coefficients of such a best

fit line are given by $\hat{a} = \bar{x}cot\theta + \bar{y}$ and $\hat{b} = -cot\theta$, where

$$\theta = 2tan^{-1}(\frac{2cov(x,y)}{var(x)^2 - var(y)^2}) \tag{2.44}$$

The nature of errors and the best fit lines for least-square regression and principal component analysis are illustrated in [234]. The measure λ_2 possesses several desirable properties like symmetry, sensitivity to scaling and invariance to rotation. It is a property of the variable pair(x,y) reflecting the amount of error committed if maximal information compression is performed by reducing the variable pair to a single variable. Hence, it may be suitably used in redundancy reduction (see [234] for further details). The algorithmic steps for feature reduction using MICI is given below:

Algorithm:

Let D is the original number of features and $O = \{F_i, i = 1,...,D\}$ is the original feature set. Then the dissimilarity between F_i and F_j is represented by $S(F_i,F_j)$ where S can be calculated by the measure of linear dependency λ_2. Higher the value of S is, the more dissimilar the features are. Let r^{k_i} represents the dissimilarity of feature F_i and its k^{th} nearest-neighbour feature in R. Then

Step 1: Choose an initial value of $k \leq D-1$. Initialize the reduced feature subset R to the original feature set O, *i.e.*, $R \leftarrow O$.

Step 2. For each feature $F_i \in R$, compute r_i^k.

Step 3: Find feature $F_{i'}$ for which $r_{i'}^k$ is minimum. Retain this feature in R and discard k nearest features of $F_{i'}$. (Note: $F_{i'}$ denotes the feature for which removing k nearest-neighbours will cause minimum error among all the features in R. Let $\varepsilon = r_{i'}^k$.

Step 4: If $k >$ cardinality$(R) - 1$: k=cardinality$(R) - 1$.

Step 5: If k=1: **Go to Step 8.**

Step 6: **While** $r_{i'}^k > \varepsilon$ **do:**

(a) $k = k - 1$.

$r_{i'}^k = \inf_{F_i \in R} r_{i'}^k$.

(k is decremented by 1, until the "k^{th} nearest-neighbour" of at least one of the features in R is less than ε-dissimilar with the feature)

(b) **If** k=1: **Go to Step 8.**

(if no feature in R has less than ε-dissimilar "nearest-neighbour" select all the remaining features in R.)

End While

Step 7: **Go to Step 2.**

Step 8: Return feature set R as the reduced feature set.

2.5.2 Feature Evaluation Index

The feature evaluation procedure provides a deterministic approach for pinpointing those features that are indeed useful in the pattern recognition. The procedure uses mathematical, statistical and fuzzy techniques to rank the individual features. In this particular study, Fuzzy Feature Evaluation Index (FFEI) is being used extensively.

Let $\mathscr{C}_1, \mathscr{C}_2, ..., \mathscr{C}_j, ..., \mathscr{C}_m$ be the m-pattern classes in an N-dimensional (X = $\{x_1, x_2, ..., x_N\}$) feature space. Also, let n_j ($j = 1, 2, ..., m$) be the number of samples available from class C_j. A fuzzy set $\mathscr{A} = \{\mu_{\mathscr{A}}(x_i); i = 1, 2, ..., N; \mu \in [0, 1]\}$ be defined on X where $\mu_{\mathscr{A}}(x_i)$ denotes the membership of x_i to A. A measure of fuzziness for $\mathscr{A}$ in terms of entropy (H) values of the classes in order to provide a quantitative index for feature evaluation are described below.

$$H(\mathscr{A}) = (\frac{1}{n \ln 2}) \sum_i S_n(\mu_{\mathscr{A}}(x_i)) \tag{2.45}$$

where $S_n(\mu_{\mathscr{A}}(x_i))$ = -$\mu_{\mathscr{A}}(x_i) ln \mu_{\mathscr{A}}(x_i) - \{1 - \mu_{\mathscr{A}}(x_i)\} ln \{1 - \mu_{\mathscr{A}}(x_i)\}$ is the Shannon's function. Entropy [235] is dependent on the absolute values of membership (μ) and $\frac{1}{n \ln 2}$ is a normalization factor. $H(\mathscr{A})$ attains the maximum value when $\mathscr{A}(x_i)$ is most fuzzy or $\mu_{\mathscr{A}}(x_i)$ = 0.5 $\forall i$ and it attains the minimum value when $\mathscr{A}(x_i)$ = 0 or 1 $\forall i$. We have chosen

S-type membership function for modeling μ in our study.

For computing H of the class $(\mathscr{C}_|)$ along the qth feature, the parameters of the S-function can be computed as

$$b = (x_{qj})_{av} c = b + max\{|(x_{qj})_{av} - (x_{qj})_{max}|, |(x_{qj})_{av} - (x_{qj})_{min}|\} a = 2b - c \tag{2.46}$$

Here av, max, and min are used to denote the average, maximum, and the minimum value of x_{qj}, respectively. Therefore, the Feature Evaluation Index (FEI) for qth feature can thus be defined as

$$FEI_q = \frac{H_{qjk}}{H_{qj} + H_{qk}} \tag{2.47}$$

where H_{qjk} is the value of the entropy for feature q after pooling the classes $\mathscr{C}_j$ and $\mathscr{C}_k$; and H_{qj} and H_{qk} are those for the feature q computed for $\mathscr{C}_j$ or $\mathscr{C}_k$, respectively. The lower the value of FEI_q, the higher is, therefore, the quality of the qth feature in characterizing and discriminating classes $\mathscr{C}_j$ and $\mathscr{C}_k$. Interested reader may follow the further details in [236–240]

2.6 Similarity/Dissimilarity Measure

Instead of exact matching, CBIR calculates visual similarities between a query image and images in a database. Accordingly, the retrieval result is not a single image but a list of images ranked by their similarities with the query image. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent years. In this thesis, we have used five different types of distance measure *i.e.*, EMD, KLD, Chessboard Distance (CBD), MD and ED. In this section, two most important distance measure (EMD and KLD) will be discussed as ED, MD and CBD are very common and old in CBIR literature.

2.6.1 Earth Mover's Distance

EMD by its definition extends to distance between sets or distributions of elements, thereby facilitating partial matches [241]. If p_i be the centroid of each cluster of a signature $P_i = \{(p_1, w_{p_1}), ..., (p_m, w_{p_m})\}$ with m clusters and w_{p_i} the weight of each cluster, similarly q_i be the centroid of each cluster of a signature $Q = \{(q_1, w_{q_1}), ..., (q_n, w_{q_n})\}$ with n clusters and $D = [d_{ij}]$ is the ground distance matrix, where $d_{ij} = d(p_i, q_j)$. Computing EMD thus becomes finding a flow between p_i and q_j which minimizes the overall cost. A flow between P and Q is any matrix $F = (f_{i\ j})\ \varepsilon\ R^{m \times n}$. Thus, EMD is defined as the work normalized by the total flow i.e.

$$EMD(P,Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d(p_i, q_j) f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}} \tag{2.48}$$

subject to the constraints (a) $f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n$, (b) $\sum_{j=1}^{n} f_{ij} \leq w_{p_i}, 1 \leq i \leq m$, (c) $\sum_{i=1}^{m} f_{ij} \leq w_{q_j}, 1 \leq j \leq n$, and (d) $\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j})$.

2.6.2 Kullback-Leibler Divergence

Information-theoretic measures are various conceptual derivatives from the Shannon's entropy theory and treat objects as probabilistic distributions. Therefore, again, they are not applicable to features with negative values. Given a discrete probability distribution P Shannon's entropy $H(x)$ is defined as

$$H(x) = -\sum_{x} P(x) log P(x) \tag{2.49}$$

Entropy is a natural measure of complexity, since it estimates the degree of uncertainty with random variables. Intuitively it is appealing: The more uncertain we are about an outcome of an event, the more complex the phenomenon (data, image, etc.) is

Given another distribution Q, the Kullback-Leibler (K-L) divergence is defined as

$$KL(P||Q) = \sum_x P(x) log \frac{P(x)}{Q(x)} \tag{2.50}$$

KL-divergence (KLD) is also called relative entropy and it can be interpreted as the amount of extra bits that is needed to code samples from P using code from Q. If the distributions are the same, the need for extra information is zero and the divergence is zero as well. KLD is nonnegative but not symmetric and as such it can not be used directly as a measure of distance or dissimilarity between distributions. The symmetry is easy to obtain, however, just by calculating and summing the KLD from Q to P and from P to Q, hence the symmetric[1] version is simply

$$KL_{symmetric}(P,Q) = KL(P||Q) + KL(Q||P) \tag{2.51}$$

This is not a true metric but it can be used directly as measure of distance or dissimilarity between distributions

2.7 Performance Measures

In this thesis, the performance effectiveness of various proposed solutions, overcoming the problems of image retrieval have been evaluated through extensive experiments and comparisons. Both subjective (qualitative) as well as objective (quantitative) measures have been used for this purpose. In the following section, brief descriptions of the different quantitative performance measures are given, which have been used to objectively evaluate the effectiveness of the proposed solutions.

Quantitative evaluation of the proposed image classification and retrieval systems with

[1] Actually this is the original formulation that Kullback and Leibler give [242].

relevance feedback mechanism and their performance comparison with other state-of-the-art techniques have been analyzed using the following statistical measures:

Precision:

$$Precision = \frac{N_{RIR}}{N_{RIR} + N_{IRIR}} \tag{2.52}$$

Recall:

$$Recall = \frac{N_{RIR}}{T_{RID}} \tag{2.53}$$

F-Measure:

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{2.54}$$

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 in this study. Although there is no restriction of using more number of images *i.e.*,, 30 , 40 or high for do such measurement. The performance of the CBIR system is also measured by evaluating the different test-queries and calculating the Mean Average Precision (MAP). MAP is obtained by computing for each query the average of the precisions measured after each relevant image is retrieved. Average Precision (AP) and Average Recall (AR) are also used for evaluating the performance of the CBIR system in this thesis.

The tools discussed in this chapter are used for developing the different parts of the CBIR system. Using different tools, three different CBIR schemes have been developed in the next chapter. The detailed discussion of feature extraction, classification, similarity/dissimilarity measurement and relevance feedback mechanism are presented in Chapter 3

Chapter 3

CBIR System With and Without Relevance Feedback

3.1 Introduction

The basic aim of a CBIR system is to fetch a group of images similar to a query image and presented them at the output. The accuracy of the system largely depends upon the quality of visual features used to represent an image information which can capture the overall visual impression of it. To compare the similarity of images from the database with query image, different kinds of similarity measures may be used such as ED, MD, KLD, EMD etc [133].

The major problem in the conventional CBIR, is the semantic gap between visual perception of a scene and feature based representation of the scene image. To bridge this gap, user's feedback has come into picture in an interactive manner *i.e.*, train the system with the help of human assistance [147, 243–245]. Although several solutions for suitable feature extraction and RF have been proposed, but results achieved so far are not fully upto the user's expectation. Image representation by low level features is a crude mimic, the way the image is represented in the human perception mechanism. The true representation model used for human perception is not fully understand yet. So this knowledge gap (semantic gap) may be reduced to a great extent if learning process is done with the assistance from the human observer. This idea is implemented through relevance feedback mechanism in CBIR system. This model has been implemented in the Scheme-I of this Chapter with particular type of transform (M-band) and Fuzzy Relevance Feedback (FRF), and different similarity measure like Euclidean Distance (ED) (most commonly used in CBIR literature) along with some of the other distance measure (Manhattan Distance (MD) and Chessboard Distance (CBD)) as comparative study. However, each iteration takes a considerable time which is proportional to the size of the database. To cut down this cost, people use different classifiers for approximately classifying the class of the input query image. This knowledge has been used to restrict the search space for retrieval of similar images from the database. This similar kind of technique has been used in Scheme-II of this Chapter. So naturally,

a method which combines both the above stated approach will able to provide a further improvement of computational efficiency and retrieval time. Therefore, the combination of Scheme-I and II of this Chapter has been implemented in Scheme-III of this Chapter.

In this chapter, three novel CBIR Schemes ("Scheme-I", "Scheme-II" and "Scheme-III") are presented as possible solutions to the above mentioned problems in CBIR paradigm. In "Scheme-I", we have designed CBIR system using M-band wavelet transform based features with Fuzzy Relevance Feedback (FRF) based mechanism. In "Scheme-II", Gaussian distribution based RT features has been proposed with pre-classification mechanism for reducing the search space. Finally in "Scheme-III", we have designed CBIR system with both pre-classification and FRF mechanism.

3.2 Scheme-I: CBIR System Using Fuzzy Relevance Feedback [174]

Popularly used dyadic WT decomposes the signal channel into logarithmic tiling of time scale plane. Whereas, M-band wavelet transform divides the signal into a mixture of linear and logarithmic tiling of time scale plane. It gives richer parameter space having greater variety of compactly supported components. It has better resolution at high frequencies and overcomes the drawbacks of WT as discussed in [196, 197]. The M-band wavelet minimizes the computational cost by reducing the number of features for retrieval as compared to popularly used MPEG-7 visual features, like CSD and EHD [49].

Traditionally, the most commonly used similarity measures in CBIR are ED and MD. However, the main problem using ED is a scale problem whereas MD is very sensitive to the orientation of an object. In contrast, EMD measure is a variable sized description of distributions that can able to overcome this problem. In EMD, ground distance is calculated between the two signatures of same features spaces. EMD based CBIR systems have been

found to perform better than CBIR systems based on MD or ED [133].

There are several approaches on relevance feedback based CBIR scheme, reported in the literature [16]. Many researcher have used soft-computing based feature re-weighting technique in accordance to relevance and irrelevance images [151, 246]. Motivated by these ideas, we have proposed a fuzzy entropy based relevance feedback technique (human training) using EMD by considering both relevant and irrelevant images to adjust automatically the weights of the M-band wavelet features. The results are compared with MPEG-7 visual features [49] which has become almost standard norms for the evaluation of newly proposed image features for CBIR system. The details of the proposed CBIR system is discussed in the next section.

3.2.1 The Proposed Technique

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

Signature Generation: The steps of the Signature Generation part as shown in Fig.3.1 are as follows:

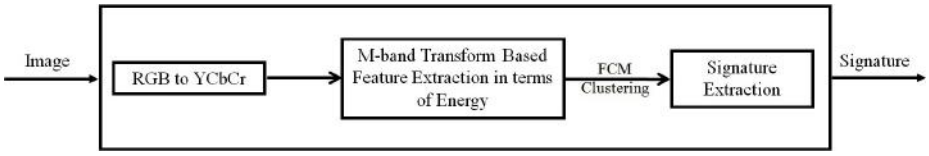


Fig. 3.1 Block Diagram of Signature Generation

1. RGB color image is converted into YCbCr color plane.

2. YCbCr image is decomposed into $M \times M$ channels without downsampling by using 1D, 16 tap 4 band orthogonal filters as a kernel for wavelet filter and then energy feature at each pixel is computed.

3. Fuzzy C-Means clustering (FCM) algorithm is used in each pixel to obtain the signature of the image.

Image Retrieval System: The salient steps of Fig.3.2 are discussed as follows:

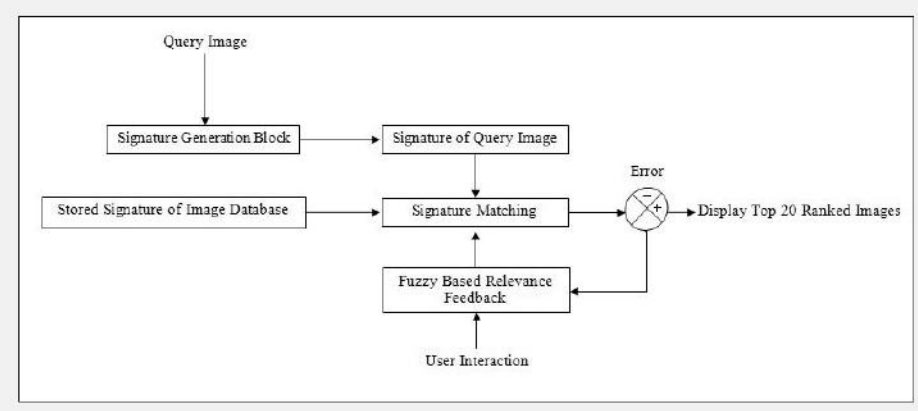


Fig. 3.2 Block Diagram of Proposed Image Retrieval System

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

Fuzzy Based Relevance Feedback (FRF): Here, in Fig.3.3, we list the main steps of the Fuzzy Based Relevance Feedback mechanism.

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

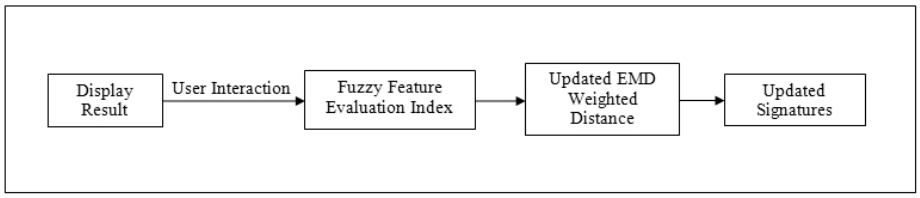


Fig. 3.3 Block Diagram of Fuzzy Based Relevance Feedback

3.2.2 M-band Wavelet Based Color and Texture Feature Extraction

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

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

Spectral Flatness Measure (SFM).

$$energy_{m_1,m_2}(i,j) = \sum_{a=1}^{N}\sum_{b=1}^{N} |Wf_{m_1,m_2}(a,b)|\, G(i-a,j-b) \tag{3.1}$$

$$1 \leq m_1 \leq M, 1 \leq m_2 \leq M$$

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

$$G(x,y) = \frac{1}{\sqrt{2\Pi\sigma}} e^{-(\frac{1}{2\sigma^2})(x^2+y^2)} \tag{3.2}$$

where σ defines the spatial support of the averaging filter.

SFM gives a measure of the global frequency content of the image. It is defined as the ratio of arithmetic mean and the geometric mean of the Fourier coefficients of the image. It has been reported in literature that the size of the neighborhood for computation of localized energy range from 11×11 to 31×31, while SFM varies from 1 to 0 [197]. We use a neighborhood size of 11×11 for SFM between 1 and 0.65, 21×21 for SFM between 0.65 and 0.35 while 31×31 for 0 to 0.35. Since images generally are formed by regions, as in from objects and their surroundings, clustering or quantizing the wavelet energy reduces the feature size retaining maximum information. Based on this basic assumption, the energy values, for each sub-band and for each plane of the color image are used as the feature for a pixel and clustered using FCM.

EMD is used as the metric for similarity matching. EMD uses a different type of signature over traditional histogram based similarity matching. Rubner *et al.* [241] has successfully shown that the EMD is an efficient metric for content based image retrieval with several advantages over other distance based similarity and dissimilarity measures. The feature vector comprises of the cluster centers location of the energy measurement

based cluster over different sub-bands and the numbers of pixels in each cluster form the image signature.

3.2.3 Fuzzy Relevance Feedback using EMD

The information obtained from the set of relevant and irrelevant images as marked by the users are used to automatically specify the weight of the component features. Weight computed are based on a measure defined as Feature Evaluation Index (FEI). The FEI which automatically estimates the importance of an individual feature can be obtained by considering a pattern classification problem as discussed in Section 2.5.2.

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

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

To compute EMD over M-band wavelet features, FCM is used to cluster the features and to obtain the signature. In real application, often there is no sharp boundary between clusters, so that fuzzy clustering is better suited for such type of data. Different clustering methods have been reported in the literature [247–249]. The feature vector comprising of the cluster centers of the energy measurement over sub-bands, with the number of pixels of the image in each cluster comprises the image signature. The number of clusters is deter-

mined, depending on the allowed computational cost. A right kind of balance is desirable between computational cost and obtained precision of the results. Increase of the number of clusters, may produce finer segmentation regions but obtained result may not always be perceptually accurate. So, in an unsupervised mode of clustering, the number of clusters is always chosen appropriately based on the expected outcome.

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

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

From the set of marked images (I_r and I_{ir}), the weight of the features computed over each Y, Cb and Cr are estimated as follows: For each cluster p_i the wavelet based sub-band features of each of Y, Cb and Cr planes are considered. The features considered to compute the feature evaluation index along each component plane *e.g.*, plane Y are $F_Y =$

$\{f_{i_{1Y}},...,f_{i_{qY}}\}$, where $i = 1,2,...,m$ (m clusters) and $q = 1,2,...,n$ ($n$sub-band features). Similarly features are considered for Cb and Cr plane. The $(FEI)_{qY}$ for the component feature q of Y plane are

$$(FEI)_{qY} = \frac{H_{qtotal}}{H_{qRel} + H_{qIrrel}}. \tag{3.3}$$

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

The overall weight factor for the Y plane is given by

$$W'_Y = \sum_{i=1}^{m}\sum_{q=1}^{n}(FEI)_{qY} \tag{3.4}$$

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

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

$$W_Y = \frac{W'_Y}{W'_Y + W'_{Cb} + W'_{Cr}} \tag{3.5}$$

Using similar formula, W_{Cb} and W_{Cr} are computed. The weights W_Y, W_{Cb} and W_{Cr} reflects user's different importance on the representative feature map for computing overall similarity between images. Multiplying with the weights, actually modify the ground distance $d(p_i,q_j)$ of Eq. 2.48 in chapter 2, but keep the distribution, *i.e.*, the number of pixels in each cluster unchanged and the total weight remains the same. So, the weighted EMD, *i.e.*, $\text{EMD}_g(P,Q)$, and $g\varepsilon(W_Y,W_{Cb},W_{Cr})$ is computed from the work flow:

$$WORK(F,P',Q') = \sum_{i=1}^{m}\sum f_{i,j}d(g(p_i),g(q_j)) \tag{3.6}$$

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

In case of irrelevant images, although the signature map may be different, the EMD distance as obtained from the work flows is nearly equal. After FRF, the ground distance (d_{ij}) is modified by the weighting factor W where $W \varepsilon [W_Y, W_{Cb}, W_{Cr}]$. The search space that varies over each plane becomes elliptic for unequal weights. As a result, irrelevant images are discarded to a large extent and more relevant images are included due to FRF mechanism because the ground distance varies in accordance to the importance of the component planes. Recently, a comprehensive and thorough overview of some new refinement techniques using the maximization of the fuzzy entropy have been proposed to enhance the performance of the learning system [250, 251].

3.2.4 Fuzzy Relevance Feedback Using Different Similarity Measures

For relative comparison of performance with different distance measure, results are also provided here using similarity measures namely ED, MD and CBD [134]. To compute these distances using the same set of M-band wavelet features, FCM is used to cluster the features. Thus, M-band wavelet features vector comprising centroids of the query signature ($\sum_{i=1}^{m} p_i$) and the stored images ($\sum_{j=1}^{m} q_j$) are considered.

(a)Euclidean Distance.

$$d_{ED} = \sum_{l=1}^{N} \parallel (p)_l - (q)_l \parallel, \tag{3.7}$$

(b)Manhattan Distance.

$$d_{MD} = \sum_{l=1}^{N} |(p)_l - (q)_l|, \tag{3.8}$$

(c)Chessboard Distance.

$$d_{CBD} = \max_{1 \leq l < N} \{|(p)_l - (q)_l|\} \tag{3.9}$$

where, N is the total number of features by considering energy measurement over sub-band of each component plane as explained in Section 3.2.3.

The results generated from the 1^{st} iteration are obtained by measuring above discussed similarity measures (ED, MD and CBD) between the M-band wavelet features of the query image and the stored images in the database. Top 20 ranked images are displayed. From the returned images of each iteration, the weight specifying the user's importance is computed in a similar fashion as explained in Section 3.2.3. And the weighted similarity measures is

(a)Euclidean Distance.

$$d_{WED} = \sum_{l=1}^{N} \parallel g(p)_l - g(q)_l \parallel, \tag{3.10}$$

(b)Manhattan Distance.

$$d_{WMD} = \sum_{l=1}^{N} |g(p)_l - g(q)_l|, \tag{3.11}$$

(c)Chessboard Distance.

$$d_{WCBD} = \max_{1 \leq l < N} \{|g(p)_l - g(q)_l|\} \tag{3.12}$$

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

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

3.2.5 Experimental Results and Discussion

To demonstrate the effectiveness of the proposed CBIR system, two standard image databases were used in the experiments such as (a) SIMPLIcity (SimDB), (b) Corel database (CorelDB) and (c) Caltech 101. In all given figures of the experimental results, the top left corner image is the query image used in the experiment. The quantitative measures used in the experiments were AP and AR and are discussed in Chapter 2 in Section 2.7.

Fig. 3.4 Retrieval results on SimDB using (Top left side image as the query image)(a) M-band wavelet features with EMD (19/20)(b) M-band wavelet features with ED (9/20) (c) M-band wavelet features with MD (12/20)(d) M-band wavelet features with CBD (8/20).

Fig.3.4(a) shows the effectiveness of the proposed scheme using M-band wavelet feature set and EMD. The performance of the proposed M-band wavelet based CBIR sys-

tem using other distances (ED, MD and CBD) is shown in Fig.3.4(b)-(d). The values (*e.g.*,(19/20),(9/20) etc.) given in the caption of the figures indicate the number of correctly retrieved images out of 20 images per frame. The graph of the Fig.3.5 shows the performance comparison of EMD against other similarity measures (ED, MD, CBD) using M-band wavelet as features in all the cases (average value obtained by considering each image of the DB as a query). It is clear from the graph of Fig.3.5, as well as from the given instances of the retrieval results of Fig.3.4(a) - (d), that M-band wavelet features with EMD performs better than M-band wavelet features with other distances, even though the CBIR using EMD method is computationally more complex than the CBIR based on other similarity measures and M-band wavelet features.

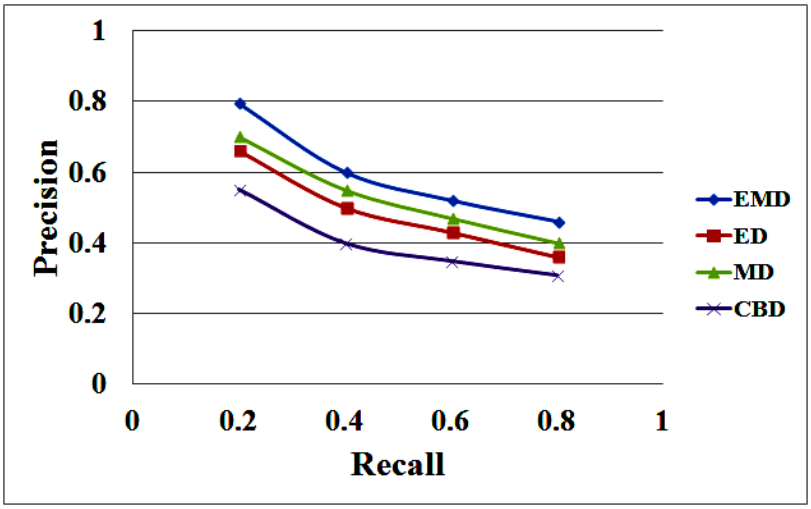
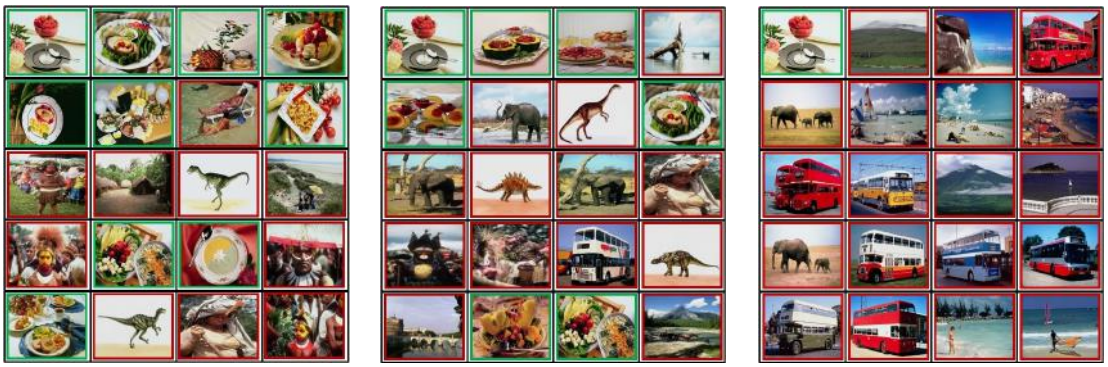


Fig. 3.5 Performance comparison graph of M-band wavelet features with EMD Versus ED, MD, CBD on SimDB.

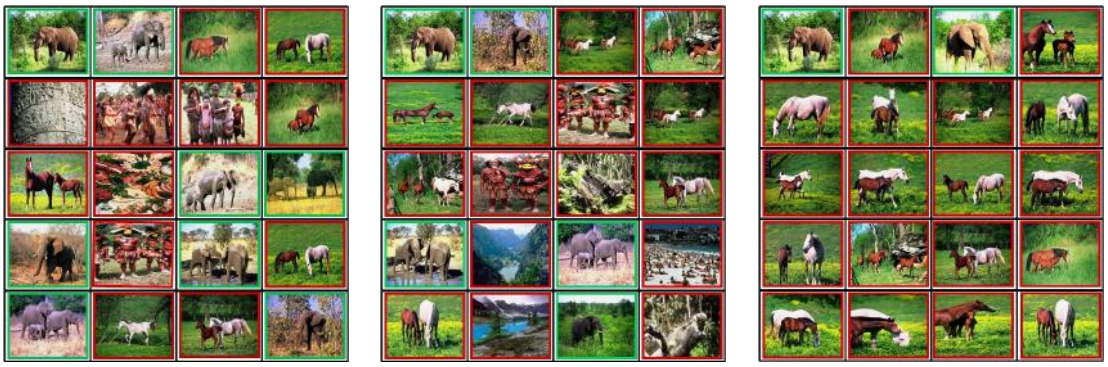
We have compared the performance effectiveness of M-band wavelet feature set with two widely used ISO MPEG-7 visual features, CSD and EHD. CSD and EHD consist of 256 and 80 features respectively. As CSD consists of 256 number of features, which is much larger than M-band (144) and EHD (80), it is expected that CSD should normally performs better than M-band and EHD.

The results of the performance comparisons are shown in Fig.3.6 and Fig.3.7, respec-



(a) (b) (c)

Fig. 3.6 Retrieval results using ED on SimDB in case of color prominent query image using (Top left side image as the query image)(a) CSD features (10/20) (b) M-band wavelet features (7/20) (c) EHD features (1/20).



(a) (b) (c)

Fig. 3.7 Retrieval results using ED on SimDB in case of edge prominent query image using (Top left side image as the query image) (a) EHD features (8/20) (b) M-band wavelet features (5/20) (c) CSD features (2/20).

tively. ED is used as the similarity measure in the comparisons, as it is the most widely used similarity measure by most of the existing CBIR system based on CSD and /or EHD. The results of the Fig.3.6 describe the performance comparisons in average cases, where CSD performs better than M-band and EHD. Similarly, the results of the performance comparisons in case of edge prominency are shown in Fig.3.7. It is clear from the Fig.3.6 that, in average case, the M-band wavelet feature set performs better than EHD but poorer than CSD, incase of color prominency. The reverse is true for EHD, where M-band wavelet feature set performs better than CSD but poorer than EHD, incase of edge prominency. M-

band wavelet feature set consists of 144 features, which is less than CSD (256) but larger than EHD (80). From the discussion above, it is obvious that even though M-band wavelet feature set is more computationally complex than EHD but less computationally complex than CSD, it has given comparable performances in both the cases.

(a) (b) (c)

Fig. 3.8 Results on SimDB using FRF mechanism (Top left side image as the query image) (a) First pass of the retrieval set Using M-band wavelet features and EMD (19/20) (b) First Iteration with weighted EMD (20/20, improved ranking)(c) Second Iteration with weighted EMD (20/20, improved ranking).

To improve the performance of the proposed CBIR system, we have used FRF mechanism. Fig.3.8 shows the improvement in the retrieval result after 1^{st} and 2^{nd} iteration using weighted EMD as the similarity measure on SimDB. It is clear from the Fig.3.8 that the performance of the proposed CBIR system has improved with every iterations of FRF mechanism, not only in the total number of correctly retrieved images, but also in image ranking. The ranking of the images have improved with every iterations of FRF mechanism as can be seen from Fig.3.8. It has been observed that normally after 2 to 3 iterations of the FRF mechanism, the retrieval results converge.

The scalability performance effectiveness of the proposed CBIR system based on M-band wavelet feature set using weighted EMD as the similarity measure and with FRF mechanism, is also evaluated using a bigger sized image database, *i.e.*, CorelDB. The results are shown in Fig.3.9. It is obvious from the results given in Fig.3.9 that the proposed

(a) (b) (c)

Fig. 3.9 Results on CorelDB using FRF mechanism (Top left side image as the query image) (a) First pass of the retrieval set Using M-band wavelet features and EMD (6/20)(b) First Iteration with weighted EMD (9/20, improved ranking)(c) Second Iteration with weighted EMD (10/20, ranking of the relevant images as compared to the query image is improved).

CBIR system works well on databases containing huge number of images. The CPU-time taken for each iteration is approximately 50*ms* for SimDB and 3*sec* for CorelDB.

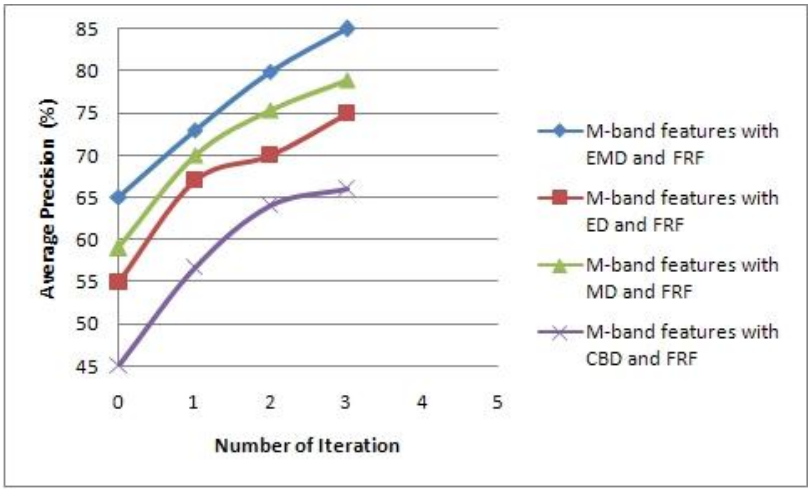


Fig. 3.10 Performance comparisons on SimDB of FRF mechanisms with M-band wavelet features with EMD similarity measures Versus ED, MD and CBD.

The graph of the Fig.3.10 is representing the average precision values computed from each frame of 20 displayed images taken over all database images. It also shows the effectiveness of the proposed FRF using M-band wavelet feature set with EMD against other similarity measures. It is obvious from the graph that the proposed CBIR system using

M-band wavelet feature set with FRF and EMD performs better than the CBIR system using M-band wavelet with FRF and other distances such as ED, MD and CBD. To further justify, we have reported our proposed CBIR system on Caltech 101 database as shown in 3.11.

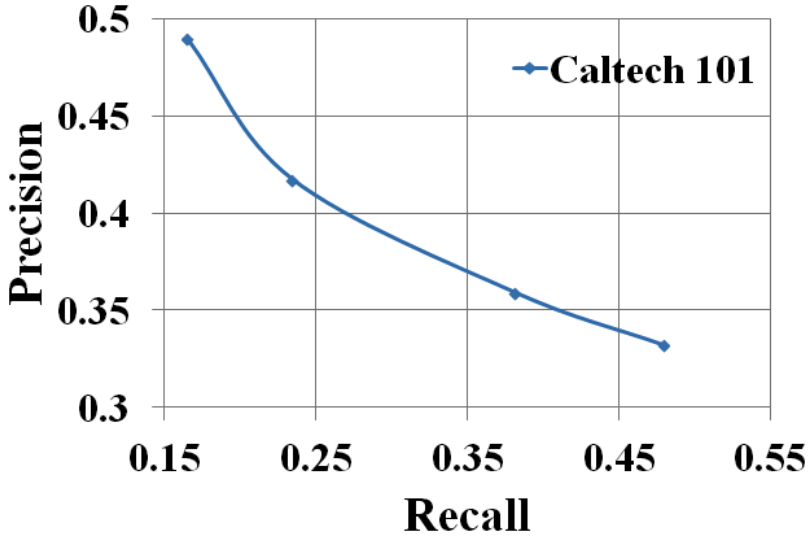


Fig. 3.11 Precision Vs Recall graph on Caltech 101 database

In this work, we have proposed a novel CBIR system based on M-band wavelet feature set using weighted EMD as the similarity measure. To improve the retrieval result we have used an interactive relevance feedback mechanism based on fuzzy feature evaluation procedure. The performance of the proposed CBIR system shows that, due to its simple structure and low computational time requirement it is well suited for real life application paradigm like the internet. But M-band wavelet suffers from lack of anisotropic, shift sensitivity and limited orientation selectivity. To overcome these limitations of M-band wavelet, a theory called MGA (like CVT, NSCT, etc.) for high-dimensional signals has been proposed in the literature [203, 206]. Moreover, the time requirement for each query searching is proportional to total number of features and images in the database. To cut down this computation cost (*i.e.*, by reducing the search space) with improved retrieval accuracy, researcher used pre-classification technique using different classes of classification, various kind of fea-

ture extraction and different simialrity/dissimilarity measure. A CBIR system using all the above mention idea is designed and discussed in "Scheme-II".

3.3 Scheme-II: CBIR System with Reduced Search Space Using Pre-Classifier [176]

High retrieval efficiency and less computational complexity, are the desired characteristics of an efficient CBIR system [16]. As it is observed from "Scheme-I", that using a FRF mechanism by incorporating FRF technique increases the retrieval accuracy but it is taking more time as its number of iterations is more. Therefore, in "Scheme-II", pre-classification based mechanism has been incorporated by introducing the concept of pre-classification technique where the concept of partition database from the whole database has been introduced.

Various feature extraction schemes have been used in the literature [172, 252]. But statistical distribution based texture feature extraction mechanism has been gaining importance from last decade [16, 135, 154]. The effectiveness of these approaches in natural color image classification is not explored well. Therefore, in "Scheme-II", we have proposed to apply the statistical framework based on GGD and KLD on RT coefficients to investigate its performance in natural color image classification problem. To improve the retrieval accuracy as well as to reduce the computational complexity, we have used LS-SVM in this scheme. The details of "Scheme-II" is discussed in the following section.

3.3.1 Ripplet Coefficient Modeled Using GGD

The images in the database prior to RT decomposition, are transformed from RGB to YCbCr color space. This ensures that the textural characterizations of the images are independent of the color characterization. RT decomposition over the intensity plane (Y) char-

acterizes the texture information, while the RT decomposition over chromaticity planes (Cb and Cr) characterizes color. Texture and color information are extracted by using RT on each color plane with 4 level (1, 2, 4, 4) decomposition. This decomposition configuration produces 11 $(= 1+2+4+4)$ subbands for each image of the DB for each color plane. As, there are 3 color planes, so altogether we get 33 $(= 3 \times 11)$ subbands for each image of the DB. The distribution of ripplet subband coefficients are then modeled with GGD which is defined as,

$$p(x;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^{\beta}} \tag{3.13}$$

where x is the ripplet subband coefficients 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 ripplet coefficients at the i^{th} subband, and $i \leq L$, the ML estimator is defined as [135].

$$L(x;\alpha,\beta) = log \prod_{i=1}^{L} p(x_i;\alpha,\beta) \tag{3.14}$$

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

$$\frac{\partial L(x;\alpha,\beta)}{\partial \alpha} = -\frac{L}{\alpha} + \sum_{i=1}^{L} \frac{\beta |x_i|^{\beta} \alpha^{-\beta}}{\alpha} = 0 \tag{3.15}$$

$$\frac{\partial L(x;\alpha,\beta)}{\partial \beta} = \frac{L}{\beta} + \frac{L\Psi(1/\beta)}{\beta^2} - \sum_{i=1}^{L} \left(\frac{|x_i|}{\alpha}\right)^{\beta} log\left(\frac{|x_i|}{\alpha}\right) = 0 \tag{3.16}$$

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.(3.15) by fixing $\beta > 0$:

$$\hat{\alpha} = \left(\frac{\beta}{L} \sum_{i=1}^{L} |x_i|^{\beta} \right)^{1/\beta} \tag{3.17}$$

Substituting this into Eq.(3.16), the shape parameter β is the solution of the following *transcendental* equation

$$1 + \frac{\Psi(1/\hat{\beta})}{\hat{\beta}} - \frac{\sum_{i=1}^{L} |x_i|^{\hat{\beta}} \log |x_i|}{\sum |x_i|^{\hat{\beta}}} + \frac{log\left(\frac{\hat{\beta}}{L} \sum_{i=1}^{L} |x_i|^{\hat{\beta}}\right)^{1/\hat{\beta}}}{\hat{\beta}} = 0 \tag{3.18}$$

which can be solved numerically using the Newton-Raphson iterative procedure. Therefore, we obtain two features from a ripplet subband (l). Considering s ($= 3l$) as the total number of subbands for an image I, we obtain 2s dimensional ripplet GGD feature vector. The final feature vector of an image I in the DB is as follows:

$$f_{vec}^{I} = [\alpha_1, \beta_1, \alpha_2, \beta_2, ..., \alpha_s, \beta_s,] \tag{3.19}$$

where α_l and β_l are the GGD parameters computed from the coefficients of the l^{th} ripplet subband and $1 \leq l \leq s$. Once the ripplet GGD feature vectors are obtained, LS-SVM classifier is used to classify images of the database. The details explanation of LS-SVM classifier is given in Section 2.4.4 of Chapter 2.

3.3.2 Similarity Measurement using KLD

After the image is classify using LS-SVM according to the ripplet GGD feature vectors, KLD is used for distance calculation between the query image and the database images.

Let $p_1(x)$ and $p_2(x)$ be two continuous probability distribution functions. KLD distance

between these two probability distribution function is defined by [135]

$$D(p_1, p_2) = \int p_1(x) log\left(\frac{p_1(x)}{p_2(x)}\right) dx \tag{3.20}$$

We obtain the KLD between two GGDs substituting the GGD probability distribution function of Eq.(3.13) into Eq.(3.20), which is defined as [135]

$$\begin{aligned} D(p(,;\alpha_1,\beta_1)||p(,;\alpha_2,\beta_2)) &= \log\frac{\beta_1\alpha_2\Gamma(1/\beta_2)}{\beta_2\alpha_1\Gamma(1/\beta_1)} \\ &+ \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2}\frac{\Gamma((\beta_2+1)/\beta_1)}{\Gamma(1/\beta_1)} - \frac{1}{\beta_1} \end{aligned} \tag{3.21}$$

Ripplet coefficients of each subband are assumed to be independent. Therefore, KLD between two images is obtained by calculating the sum of all the Kullback-Leibler distances across all the ripplet subbands of these images. Let, α_j^i and β_j^i be the ripplet GGD features of subband j of image i. Using the definition of KLD between two GGDs shown in Eq. (3.21), the KLD between two images P and Q is defined by Eq.(3.22)

$$\begin{aligned} D(P,Q) &= \sum_{j=1}^{s} D(p(.;\alpha_j^{(P)},\beta_j^{(P)})||p(.;\alpha_j^{(Q)},\beta_j^{(Q)})) \\ &= \sum_{j=1}^{s} log\frac{\beta_j^{(P)}\alpha_j^{(Q)}\Gamma(1/\Gamma_j^{(Q)})}{\beta_j^{(Q)}\alpha_j^{(P)}\Gamma(1/\Gamma_j^{(P)})} + \\ &\sum_{j=1}^{s}\left(\frac{\alpha_j^{(P)}}{\alpha_j^{(Q)}}\right)^{\beta_j^{(Q)}}\frac{\Gamma((\beta_j^{(Q)}+1)/\beta_j^{(P)})}{\Gamma(1/\beta_j^{(P)})} - \sum_{j=1}^{s}\frac{1}{\beta_j^{(P)}} \end{aligned} \tag{3.22}$$

where s is the total number of ripplet subbands of an image. For a given query image, KLD is measured between the query image and each of the database images using Eq.(3.22). Database images are ranked according to ascending order of distances and are retrieved from the database.

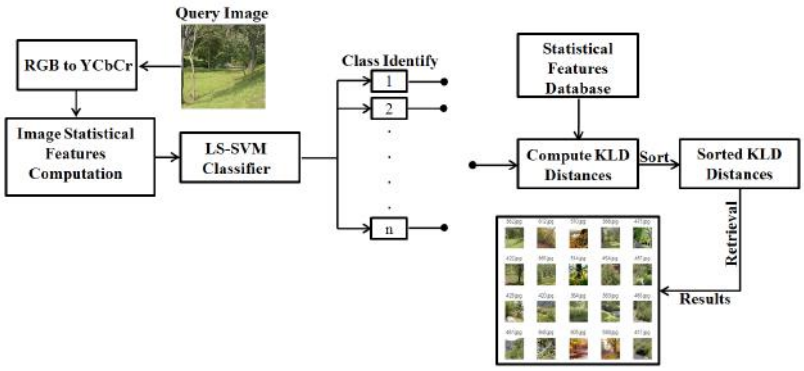


Fig. 3.12 Block Diagram of ”Scheme-II”

3.3.3 Proposed CBIR System

Here, we outlines the salient steps of the proposed CBIR system, considering the statistical features of the images in the database are already computed and stored in the ‘Statistical features Database’. The Fig. 3.12, shows the block diagram of the proposed CBIR system.

Steps:

1. Input the query image.
2. Convert the query image from RGB color plane to YCbCr color plane.
3. Compute the GGD based statistical parameters of the query image as described in Section 3.3.1.
4. Identify the class of the query image using LS-SVM, depending on the RT based statistical features.
5. As describe in Section 3.3.2, find the KLD distances between the statistical features of the query image and the statistical features of the image DB stored in “Statistical features Database”.
6. Sort the distances in ascending order and display 20 images corresponding to the first 20 sorted distances.

7. Stop.

3.3.4 Experimental Results

Extensive experiments were carried out to evaluate the performance of the proposed CBIR system, and its performance is compared with several state-of-the-art CBIR systems.

3.3.4.1 Experimental setup

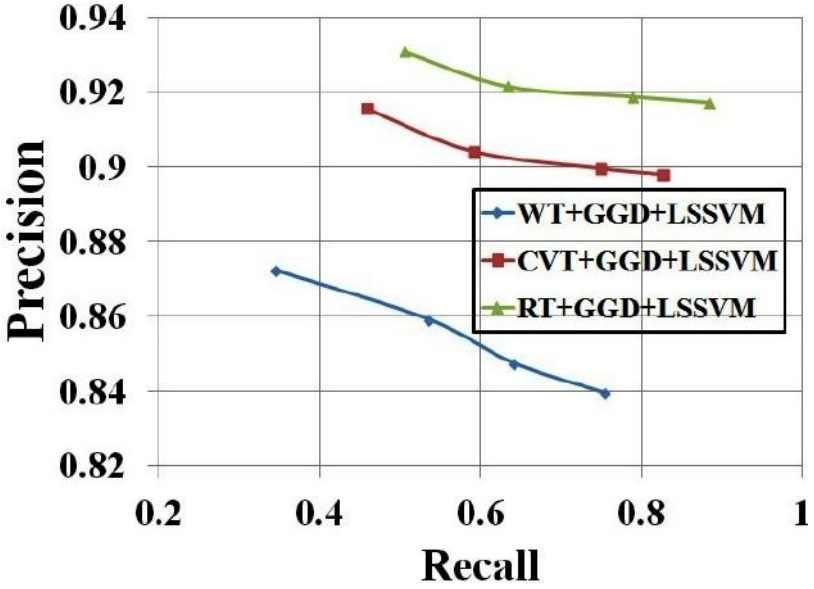
Two publicly available image databases: (1) SIMPLIcity (SimDB), (2) Oliva and (3) Caltech 101 were used in the experiments. LS-SVM classifier is trained with 70% and tested with 30% of the leveled data from the image DBs using stratified random sampling method, respectively. For training, we have used the Radial Basis Function (RBF), as the kernel. There are two tunable parameters while using the RBF kernel in LS-SVM classifier: C and γ. It is not known beforehand which values of C and γ are the best for the classification problem at hand. Hence, a $2X5$ fold Cross-Validation (CV) is conducted, where various pairs of (C, γ) are tried and the one with the lowest CV error rate is picked. We have achieved the classification accuracy of 80% for SIMPLIcity image database. With similar configuration, 75% the classification accuracy is obtained for Oliva dataset.

Quantitative evaluation of the proposed CBIR system is analyzed using two statistical measures: Mean average precision (MAP) and Mean average recall (MAR). We computed the precisions and recalls considering all the images of the used databases as the query images, and then took the average of the obtained precision and recall values over all the images as the final evaluation result.

3.3.4.2 Results and Discussion

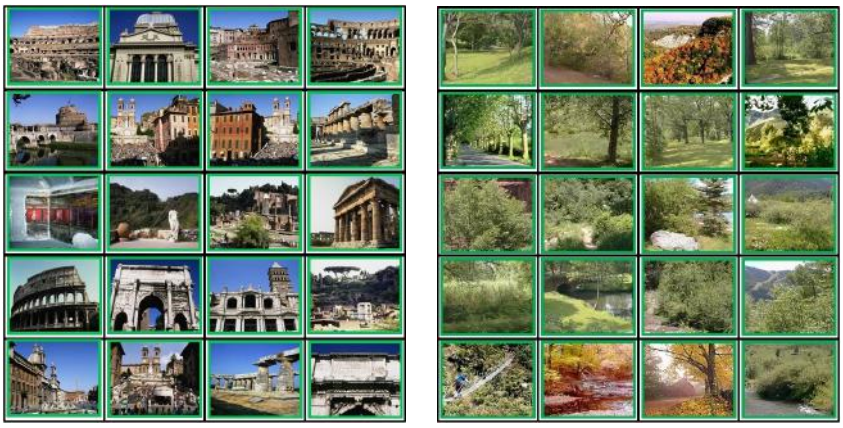
The retrieval performance of the proposed CBIR system (RT+GGD+LS-SVM) is shown in the Fig.3.13, in terms of precision and recall curve. The performance comparisons of WT

and CVT with RT are also shown in Fig.3.13 (WT+GGD+LS-SVM and CVT+GGD+LS-SVM, respectively). Here, only the images of the SIMPLIcity DB were considered for performance evaluation. It is clear from the Fig.3.13, that RT performance much better than WT and CVT in CBIR domain.



!

Fig. 3.13 Performance comparison in terms of precision and recall curves



(a) (b)

Fig. 3.14 Visual results of the proposed CBIR system for (a) SIMPLIcity and (b) Oliva database (top left image is the query image)

Fig. 3.14(a) and Fig. 3.14(b), shows two examples of the visual results obtained by the proposed CBIR system on the SIMPLIcity and Oliva dataset using query images from

"*Building*" and "*Forest*" classes, respectively. From all the given instances, we can clearly see that all the retrieved images are from the respective classes corresponding to the query images.

Table 3.1, shows the results of the performance comparisons between the proposed CBIR system with state-of-art CBIR systems namely, SIMPLIcity [168], CTCHIRS [116] and IGA-CBIR[161]. The images of the DB SIMPLIcity are used in this experiment. It can be easily seen from the Table 3.1 that our proposed CBIR system have overall higher retrieval accuracy than these existing systems. It is also to be noted that for only two categories of images ("Sea" and "Food"), the proposed CBIR system have achieved lower retrieval accuracy than the method described in [161]. The '**bold**' values of the Table 3.1 denote the highest retrieval accuracy obtained. The CBIR systems described in [116] and [161] used 71 and 86 features, whereas we have used only 66 features to represent every image of the database. It is obvious from the results given in Table 3.1 that the proposed CBIR system performs much better than the above mentioned existing CBIR schemes. The MAP and MAR of our proposed algorithm is 92.40% and 16.40% for SIMPLIcity dataset and for Oliva dataset, 85.98% and 12.45% respectively.

From our experiments, we have noticed that statistical modeling of RT based image coding is suitable for representing low level features of the images. The proposed CBIR system with LS-SVM classifier based on RT based features is able to improve the accuracy of the retrieval performance and to reduce the computational cost. To further justify, we have reported our proposed CBIR system on Caltech 101 database as shown in 3.15.

It is observed from "Scheme-II" that use of pre-classification mechanism increases the retrieval accuracy but still there is a scope for further increases in accuracy by suitable tuning mechanism with the help of FRF. As the pre-classification technique used has already achieve a more sophisticated image representation in terms of feature vector, a little human assistance/FRF (less number of iteration) can be pushup the overall efficiency of the CBIR

Class	**Wang** [168]	**CTCHIRS** [116]	**IGA-CBIR** [161]	**Proposed Method**
Africans	48	68.30	61	**89.42**
Sea	32	54	**93**	76.66
Building	35	56.15	85	**86.04**
Buses	36	88.80	71	**97.90**
Dinosaurs	95	99.25	**100**	**100**
Elephant	38	65.80	80	**93.77**
Flower	42	89.10	82	**98.08**
Horse	72	80.25	79	**100**
Mountain	35	52.15	56	**85.96**
Food	38	73.25	**99**	96.22
Average	47	72.70	80.6	**92.40**

Table 3.1 Comparisons with other existing CBIR systems in terms of average precision on SIMPLicity database

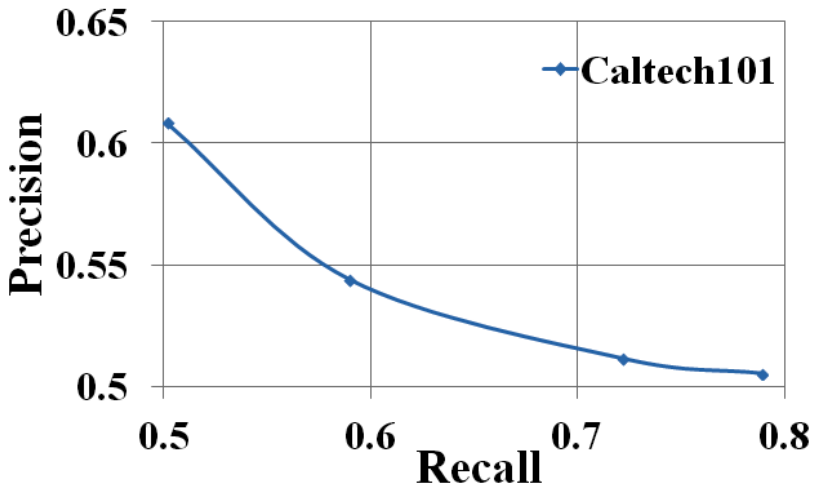


Fig. 3.15 Precision Vs Recall graph on Caltech 101 database

system at with minimum cost. The facts is demonstrated in “Scheme-III”.

3.4 Scheme-III: CBIR System with Reduced Search Space Using Pre-classifiers and Relevance Feedback [177]

M-band and RT based features extraction mechanism have been used for designing in"Scheme-I" and "Scheme-II" respectively. It is also being observed that both these features extraction mechanism perform better than dyadic WT based traditional CBIR schemes. However, both these features extraction are shift sensitive and aliasing in both space and frequency domains. To alleviate these defects of M-band and RT, a new MGA tool is proposed by combining both NSP and NSDFB, known as NSCT [206]. NSCT is also a fully decomposition (redundant) transform at ever scale. So there is no correspondence problem across the scale which is not in the case of M-band and RT.

In "Scheme-III", a novel CBIR system is proposed, where each image in the database is represented by an image signature which is computed using NSCT and FCM clustering. After feature extraction, class of the query image is identified using LS-SVM classifier. EMD is used to find the similarity between the query image's signature and other image signatures of the database. The proposed algorithm stops, if the user is satisfied with the retrieval result obtained in the first pass of the algorithm, otherwise FRF using a weighted EMD and fuzzy entropy based feature evaluation technique are used to improve the retrieval result. To evaluate the effectiveness of the proposed CBIR system, it's performance is compared against two other well-known existing CBIR systems. The details of "Scheme-III" are demonstrated in the following section.

3.4.1 The Proposed Technique

Here, we outlines the salient steps of the proposed CBIR system, considering the signatures of the images in the database are already computed and stored in the 'Features Database'. The Fig. 3.16 shows the block diagram of the proposed CBIR system. **Steps:**

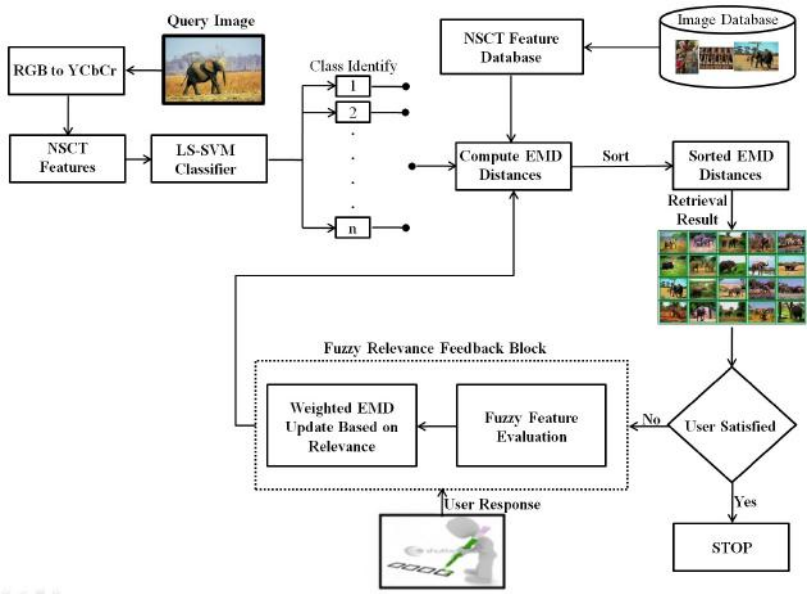


Fig. 3.16 Block Diagram of "Scheme-III".

1. Input the query image.

2. Convert the query image from RGB color plane to YCbCr color plane.

3. Compute the query image's features by the features computation method described in Section 3.4.2.

4. Identify the class of the query image depending on the features based computation as described in Section 3.4.3 .

5. Find the EMD distances between the query image's features and the features stored in 'Features Database'.

6. Sort the distances in ascending order and display 20 images corresponding to the first 20 sorted distances.

7. If the retrieved result is satisfactory then goto step 9. Otherwise, update the signatures stored in the 'Features Database', as well as the signature of the query image using the procedure described in Section 3.4.4.

8. Go to Step 5.

9. Stop.

The details demonstration of 'Scheme-III' is organized in the following section.

3.4.2 Image Signature Computation

The motivation behind using NSCT to extract the image features, is its multi-resolution, localization, multi-directionality and anisotropy properties. An image I of size $M \times N$ of the database is converted from RGB color space to Y-Cb-Cr color space, prior to NSCT decomposition. This ensures that the textural characterization of the image is independent of the color characterization. NSCT decomposition over the intensity plane characterizes the texture information, while the NSCT decomposition over chromaticity planes characterize color. NSCT is applied to Y, Cb and Cr planes of I separately. An l-level NSCT decomposition of an image plane Ip (Y, Cb, Cr), results in 2^l sub-bands, each having the same size $M \times N$ that as of the image I. Let the sub-bands of each image plane (Ip) be denoted by S_b^{Ip}, where $b = 1, 2, 3, \cdots, 2^l$.

Natural images exhibit spatial variations of the textures. So, texture based retrieval of images assume that textures region may not be homogeneous over very large areas. A localized characterization of textures thus becomes necessary. Hence, for each sub-band S_b^{Ip}, a feature map FM_b^{Ip} is computed by using Eq. (3.23) utilizing the concept of 'local energy' over a neighborhood W_{ij} of size $p \times q$, centered around a coefficient with coordinates (i, j). The size of the window W_{ij}, is determined using the SFM.

$$FM_b^{Ip}(i,j) = \sum_{(p,q)\varepsilon W_{ij}} |S_b^{Ip}(p,q)| G(i-p, j-q) \tag{3.23}$$

where, $G(i,j)$ is a Gaussian low-pass (smoothing) filter of the form

$$G(i,j) = \frac{1}{\sqrt{2\Pi\sigma}} e^{-(\frac{1}{2\sigma^2})(i^2+j^2)} \tag{3.24}$$

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 m different clusters.

In the proposed CBIR system, an image I is represented in terms of a signature

$$Sig_j = \{(C_1, w_{C_1}), (C_2, w_{C_2}), ..., (C_m, w_{C_m})\} \tag{3.25}$$

where, Sig_j is the signature of the j-th image in the database, C_i represents i-th cluster centroid ($i = 1,2,...,m$) and w_{C_i} indicates the fraction of pixels belonging to the cluster C_i. C_i constitutes the NSCT features over n feature maps of each Y, Cb and Cr plane and is represented by

$$C_i = [f^i_{1Y}, ..., f^i_{nY}, f^i_{1Cb}, ..., f^i_{nCb}, f^i_{1Cr}, ..., f^i_{nCb}] \tag{3.26}$$

In the proposed method, the number of decomposition level is set at $l = 3$, which results in $8(= 2^l)$ number of sub-bands for each image plane I_p. The number of feature maps for each image I of the database is 24 ($= 8 \times 3$; 3 color planes/image and 8 feature maps/color plane), and the number of clusters for each feature map is kept at $3(= m)$, as it gives results up to the expectation at minimum cost of computation. Increasing the number of clusters may include finer details with increase in complexity, as well as the uncertainties of characterizing the perceptual content may increase. So in the proposed system, the dimension of an image signature become $75(= 72(= 3 \times 24) + 3)$.

3.4.3 Multi-Image-class prediction using LS-SVM Classifier

The most critical drawback of SVM is its high computational complexity for high dimensional data sets. To reduce the computational demand, LS-SVM is adopted as classifier in Scheme-III. LS-SVM avoids solving quadratic programming problem and simplifies the training procedure [253].

LS-SVM was originally develop 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.-one" 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 than the other methods discussed in [254]. Hence, we use the one against one multi-class image classification method based on the LS-SVM tool by combining all pairwise comparison of binary classifiers [253].

3.4.4 Relevance Feedback Mechanism

The proposed algorithm stops at first pass if the user is satisfied with the retrieval result of the first pass. Otherwise, the information obtained from the set of relevant and irrelevant images as marked by the users are used to automatically specify the weights of the component features. Weights computations are based on a FEI. The FEI which automatically estimates the importance of an individual feature can be obtained by considering a pattern

classification problem as discussed in Section 3.2.3 and Section 2.5.2.

In conventional CBIR approaches an image I is usually represented by a set of features, $F = \{f_q\}_{q=1}^{N}$, where f_q is the q^{th} feature component in the N dimensional feature space. In the proposed system, the number of classes are two of which one class constitute the relevant images $I_r = \{I_{r1}, I_{r2}, ..., I_{rm}\}$ and the other consists of the irrelevant images I_{ir} = $\{I_{ir1}, I_{ir2}, ..., I_{irm}\}$. H_{qj} is computed from $I_r^{(q)}$= $\{I_{r1}^{(q)}, I_{r2}^{(q)}, I_{r3}^{(q)}, ..., I_{rk}^{(q)}\}$. Similarly, H_{qk} is computed from the set of images where, $I_{ir}^{(q)}$ =$\{I_{ir1}^{(q)}, I_{ir2}^{(q)}, I_{ir3}^{(q)}, ..., I_{irk}^{(q)}\}$. H_{qkj} is computed combining both the sets. Images are ranked according to the similarity measures. The user marks the relevant and irrelevant set from 20 returned images, for automatic evaluation of the FEI [235].

The results retrieved from the 1^{st} pass are obtained by measuring EMD between the signature of the query image and the stored images in the database. A better retrieval is then obtained if the user is not satisfied with the result, by using a weighted distance from user relevance feedback mechanism at successive iterations. In this cases, weights of each component features of different planes is determined from the feature evaluation mechanism. From the set of marked and unmarked images, the weights of the features computed over each Y, Cb and Cr are estimated and updated at each iteration. The details FRF mechanism are discussed in Section 3.2.3.

3.4.5 Experiment Results and Discussion

We evaluated the effectiveness of the proposed CBIR algorithm on SimDB (as described in Section 3.2.5) and Caltech 101 database. Semantic categorization of this dataset is determined by the creators which reflect the human perception of image similarity. In this proposed approach, we have used 3 level NSCT decomposition with “9-7" pyramidical and “pkva" directional filters. The experiments were carried out on a Dell Precision T7400 PC with 4GB RAM and was implemented using MATLAB R2008a. To make a fair com-

parison, the retrieval results obtained using the proposed CBIR system, are compared with some of the state of the art CBIR systems, such as Wang [168] and Lai [161].

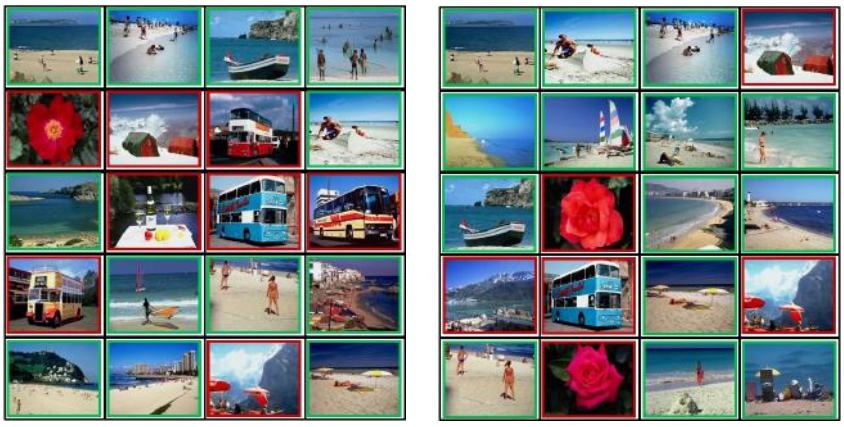
LS-SVM classifier is trained with 70% and tested with 30% of the leveled data using stratified random sampling method from the image database. For the training, we use the RBF as the kernel. The two tunable parameters of LS-SVM are C and γ. In our experiment, C = 100 and $\gamma = 0.02$ with accuracy 75% is selected for SIMPLIcity image database.

In image retrieval systems, user consist of a query formulation part. Commonly used query formations are: category browsing, query by concept, query by sketch, and query by example. Our system follows the query by example procedure. In query by example procedure, the user formulates a query by providing an example image to the system. The system initially constructs an image representative feature vector based on various high level and low level features of the image. This image representative feature vector is used in the subsequent steps of the system to classify the image. After the class identification of the query image, similar images are retrieved from the image database. If the user is satisfied with the retrieved results without using RFM then the retrieval process stops at that point. If, however the user is not satisfied, he/she can select top query based relevant images as positive feedbacks and the remaining as negative examples for updating the weight parameters and revising the features. Using these feedbacks, FEIs and weights of the features are recomputed as described in Section 3.4.4. Then, all the images are re-sorted based on the weighted similarity metric. If the user is still not content with the result, s/he can repeat the process.

Fig. 3.17 and Fig. 3.18, shows the retrieval results from SIMPLicity database without using the LS-SVM classifier, where the results are based on the combination of "*NSCT + EMD + FRF*". In the example of Fig. 3.17, an image from the "Beach" class was used as the query image, similarly an image from "Elephant" class was input to the system "*NSCT + EMD + FRF*" as the query image in the example of Fig. 3.18. It is clear from the results

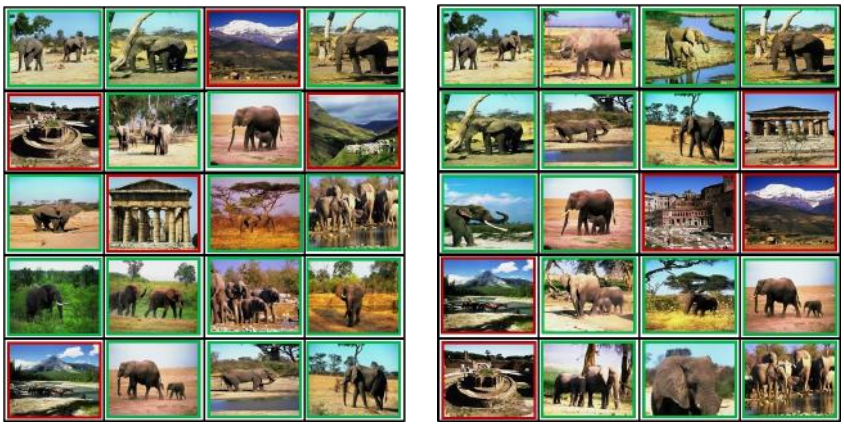
given in Fig. 3.17 and Fig. 3.18 that the performance of the CBIR system based on the combination "*NSCT + EMD + FRF*" was improved with every iteration of the used FRF mechanism. But the overall system performance of the combination "*NSCT + EMD + FRF*" is not satisfactory as is evident from the high number of misclassified images (6/20 in Fig. 3.17 and 5/20 in Fig. 3.18 after 2nd iteration of FRF).

Therefore, to improve the performance of the retrieval system we used the "*NSCT + LSSVM + EMD + FRF*" combination incorporating the advantages of the LS-SVM classifier. Fig. 3.19 and Fig. 3.20, show the examples of the retrieval results obtained by our proposed system based on the combination "*NSCT + LSSVM + EMD + FRF*", using the same query images (query images from "Beach" and "Elephant" classes). It is evident from the results given in Fig. 3.19 and Fig. 3.20 that our proposed system based on the combination "*NSCT + LSSVM + EMD + FRF*" performs better than the combination "*NSCT + EMD + FRF*". From the given results shown in Fig. 3.19 and Fig. 3.20, it is clear that in these cases not only all the retrieved images are from the same class as of the query image, but also with every iteration of the FRF mechanism the ranking of the retrieval images gets improved.



(a) (b)

Fig. 3.17 Results on SIMPLIcity using NSCT+EMD+FRF (Top left side image is the query image) (a) First pass of the retrieval set Using NSCT features and EMD (12/20) (b) Second Iteration with weighted EMD (14/20, improved ranking).



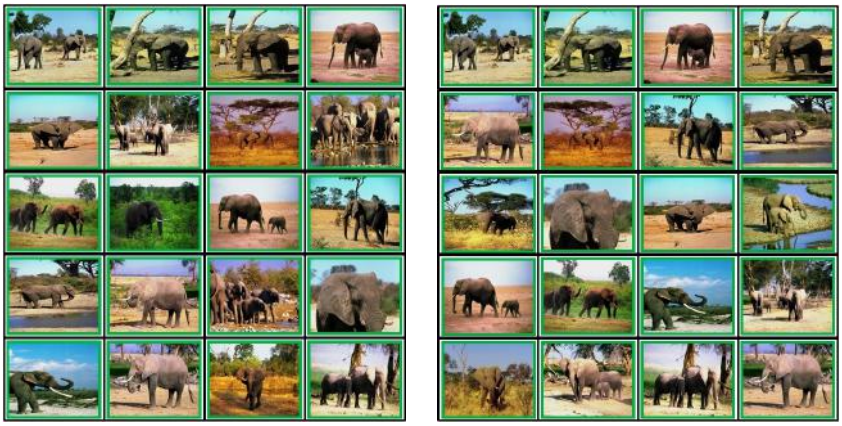
(a) (b)

Fig. 3.18 Results on SIMPLIcity using NSCT+EMD+FRF (Top left side image is the query image) (a) First pass of the retrieval set Using NSCT features and EMD (15/20) (b) First Iteration with weighted EMD (15/20, improved ranking).

(a) (b)

Fig. 3.19 Results on SIMPLIcity using NSCT+LSSVM+EMD+FRF (Top left side image is the query image) (a) First pass of the retrieval set Using NSCT features and EMD (20/20) (b) Second Iteration with weighted EMD (20/20, improved ranking).

To make a fair comparison with other existing CBIR systems, we computed two statistical measures: average precision, and average recall as the evaluation metrics. We compared our method with two other existing methods such as [168] and [161]. Only the retrieval results obtained from low level visual features (without using FRF) were used to evaluate the superiority of our approach than the two aforementioned methods. Table. 3.2 demonstrates that our method outperformed for all the classes in comparison with the method proposed



(a) (b)

Fig. 3.20 Results on SIMPLIcity using NSCT+LSSVM+EMD+FRF (Top left side image is the query image) (a) First pass of the retrieval set Using NSCT features and EMD (20/20) (b) First Iteration with weighted EMD (20/20, improved ranking).

by Wang *et al.* [168] and for some classes the method proposed by Lai *et al.* [161] gives better result. But in overall, our method yields the best average precision value. The "Bold" values in the Table. 3.2 indicate the highest retrieval performance in terms of average precision.

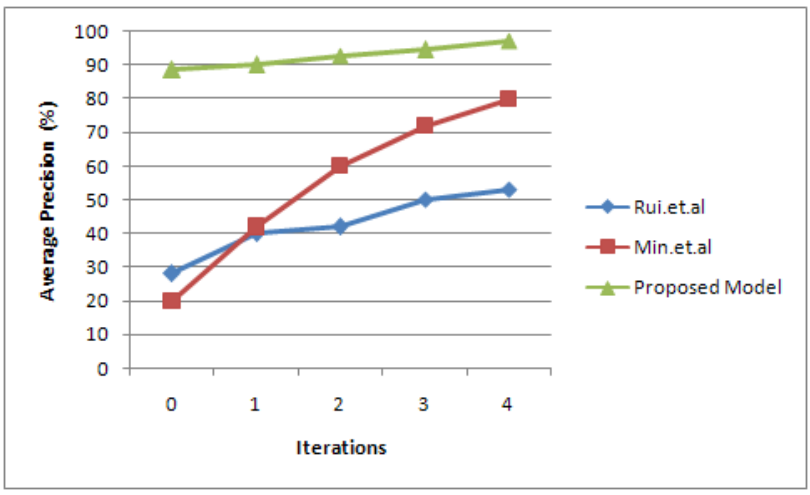


Fig. 3.21 Comparison with others Relevance feedback Mechanisms

We also compared our FRF model with two existing methods of relevance feedback mechanism. In [255], Rui *et al.* employed the weight assignment technique and obtained 52% average precision in three to four iterations from SIMPLIcity database. The other

Class	**Wang** [168]	**IGA-CBIR** [161]	**Proposed Method**
Africans	48	**99**	84.83
Sea	32	**85**	71.78
Building	35	**93**	83.95
Buses	36	80	**88.41**
Dinosaurs	95	71	**100**
Elephant	38	**100**	96.42
Flower	42	82	**99.74**
Horse	72	79	**98.11**
Mountain	35	56	**69.45**
Food	38	61	**92.79**
Average	47	80.6	**88.55**

Table 3.2 Comparisons with other existing CBIR systems without Relevance feedback in terms of average precision

method proposed by Min *et al.*, [38] used fuzzy based SVM in their relevance feedback model and achieved up to 80% average precision in three to four iterations whereas, our FRF model have achieved nearly 98% average precision in the same number of iterations as shown in Fig. 3.21. Therefore, in " Scheme-III", high retrieval accuracy has been achieved with minimum iteration.

In this work *i.e.*, "Scheme-III", we have presented an efficient interactive CBIR system. NSCT and FCM is used to extract a compact image signature based on the geometrical information. To enhance the retrieval result accuracy, pre-classification using LS-SVM classifier is used to classify the query image and retrieved the similar types of images from that class only. In addition, FRF mechanism with fuzzy entropy based feature evaluation technique is used with EMD to improve the retrieval results further. Experimental results of the "Scheme-III" have shown the significant improvement in retrieval performance. To further justify, we have reported "Scheme-III" on Caltech 101 database as shown in 3.22.

It is observed that the various combination of different transform used for extraction of primary multi-scale image features, different types of higher level image representa-

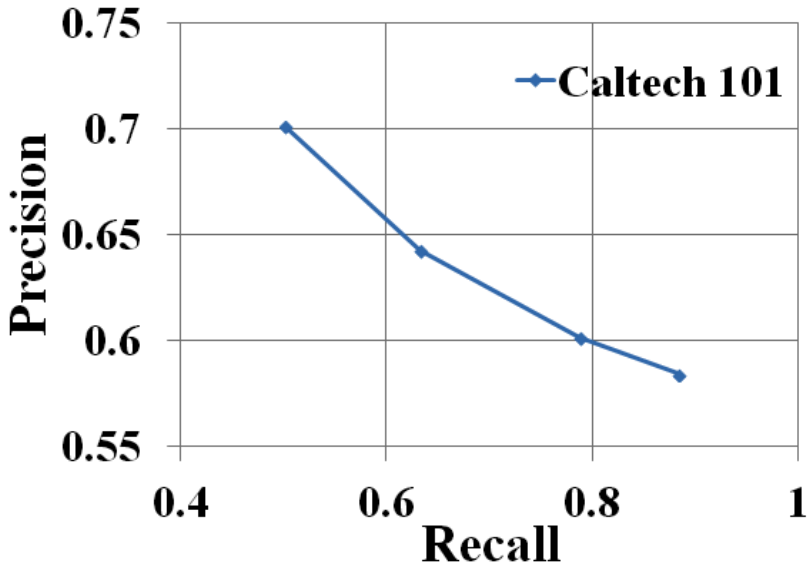


Fig. 3.22 Precision Vs Recall graph on Caltech 101 database

tive in terms of secondary features (extracted from primary features), different types of pre-classification for search space reduction can resulted in efficient CBIR design having distinct advantage for certain types of databases. So it is much important and necessary to undertake a detailed investigation to find out the best possible combination of three above mention sub-system can extract the best possible efficiency for CBIR design. This type of investigation has not been reported in any literature so far as our knowledge is concerned. This is the genesis of the research investigation taken off by us which is reported in Chapter 4.

∗∗**Major portion of this Chapter taken from the following publication.**

- M. K. Kundu, M. Chowdhury and M. Banerjee, "Interactive Image Retrieval using M-band wavelet, Earth Mover's Distance and Fuzzy Relevance Feedback", International Journal of Machine Learning and Cybernetics (Springer), Vol. 3, No. 4, pp.285–296, 2012 (**IF:1.110**).

- M. Chowdhury, S. Das and M. K. Kundu, "A ripplet Transform Based Statistical Framework for Natural Color Image Retrieval Interactive", Proc.17th Int. Conf.

Image Analysis and Processing (ICIAP 2013), Lecture Notes in Computer Science series (LNCS 8156), Vol-I, pp.492–502, Castel dell'Ove, Naples, Italy, September 9-13, 2013.

- M. Chowdhury, S. Das and M. K. Kundu, "Compact Image Signature Generation: An Application in Image Retrieval", Proc. 5th Int. Conf. on Computer Science and Inf. Tech. (CSIT 2013), IEEE Press, pp.1–7, Applied Science University, Amman, Jordan, March 27–28, 2013.

Chapter 4

Feature Extraction and Compaction: A Comprehensive Study

4.1 Introduction

People have been using various multi-scale transforms for extraction of image features for CBIR application. Although dyadic WT and its different variants are very popular as feature extractor tools, but DWT is not invariant to rotation, translation and shift. Additionally, there also exists a correspondence problem across the scale which restricts its uses in image analysis to a great extent. For 2D signal like images, there exist higher order singularities *i.e.*, curve-or-line singularities which cannot be handled with conventional dyadic WT [256, 257]. To overcome these problems (*i.e.*, lack of anisotropic, shift sensitivity and limited orientation selectivity), recently several advanced CBIR systems are proposed based on recently developed MRA/MGA tools: CVT, CNT, RT, NSCT [154, 258–260]. However, there is a lack of comprehensive studies of the above mentioned transforms in CBIR paradigm.

In this chapter, we have extensively studied about which transform is best possible as a feature extractor for a particular class of the image. Moreover, it is also been studied how the different parameters setting of a particular transform can affect the overall efficiency of the system. The next obvious step will be to derive an best possible secondary feature representation. In order to obtain higher efficiency in terms of computations cost and retrieval accuracy, in this regards, we have used three different approaches, namely, (1) First Order Statistics Based Features Representation, (2) Signatures Based Features Representation and (3) Gaussian Distribution Based Features Representation.

The combination of different transform and secondary features representation are number of data to find out the best possible configuration. Based on the secondary features, the entire data in the image database may be pre-classified into a number of classes which enhance both speed and accuracy of the CBIR system. In this regard, a number of classification may be used, namely, MLP, RF, etc. In this chapter, we have extensively studied the combination of classifier with a different kind of secondary feature representation based on

different transform in order to design a CBIR system and test them on a different benchmark database. This is done, in order to find out the best possible feature representation schemes with a particular combination of parameters and pre-classification which can generate the best possible results for particular classes of images in the different benchmark database. The detailed study is demonstrated in the following sections.

4.2 Feature Extraction Mechanisms [178]

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 [135, 154, 174, 259–262]. 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 [154, 263, 264]. 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 [174, 265]. Nevertheless, the 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 [135, 260]. 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 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 = {M-band, 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:

4.2.1 First Order Statistics Based Features Representation

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

$$F_1^I = [f_\mu^{S_j^Y}, f_\sigma^{S_j^Y}, f_\mu^{S_j^{Cb}}, f_\sigma^{S_j^{Cb}}, f_\mu^{S_j^{Cr}}, f_\sigma^{S_j^{Cr}}] \tag{4.1}$$

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

$$f_\mu^{S_j^{C_p}} = \frac{1}{a \times b} \sum_{m=1}^{a} \sum_{n=1}^{N} |S_j^{C_p}(m,n)| \tag{4.2}$$

$$f_\sigma^{S_j^{C_p}} = \sqrt{\frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} (S_j^{C_p}(m,n) - f_\mu^{S_j^{C_p}})} \tag{4.3}$$

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

4.2.2 Signatures Based Features Representation

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

$$F_2^I = [X_l^{S_j^Y}, w_l^{S_j^Y}, X_l^{S_j^{Cb}}, w_l^{S_j^{Cb}}, X_l^{S_j^{Cr}}, w_l^{S_j^{Cr}}] \tag{4.4}$$

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 is computed by using Eq. (3.23). 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 discussed in Section 3.2.2 and Section 3.4.2 of previous chapter.

4.2.3 Gaussian Distribution Based Features Representation

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

$$F_3^I = [\alpha^{S_j^Y}, \beta^{S_j^Y}, \alpha^{S_j^{Cb}}, \beta^{S_j^{Cb}}, \alpha^{S_j^{Cr}}, \beta^{S_j^{Cr}}] \tag{4.5}$$

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

$$p(x;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^\beta} \tag{4.6}$$

where x is the transformed subband coefficients (= $S_j^{C_p}$) 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 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. The details of feature extraction mechanism are discussed in Section 3.3.1 of previous chapter.

4.3 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.

4.3.1 Experimental Setup

We have studied the performance effectiveness of different combinations of transform domains, filters, feature representation and classifiers on four different benchmark image databases: (a) Oliva (ImgDb1) (b) Caltech 101 (ImgDb2) (c) Caltech 256 (ImgDb3) and (d) MIRFlickr25000 (ImgDB4).

For M-band, four different types of filter families are considered in our experiments: Meyer, article, Modulated Lapped Transform (MLT) and Shannon. For CVT, CNT, RT and NSCT different pyramidal (pyr) and orientation (ori) filter combinations are considered. The number of decomposition level of M-band, considered in the study is one and $M = 4$, while that of the advanced 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 work 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, three 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 [178]. Furthermore, we have considered different parameter configurations in MLP: numbers of iterations (*iter*), momentum factors (μ), and different learning rates (η) for different values and selected the best-performing configuration by using 2 $\times$ 5-fold cross validation (CV) technique.

Table 4.1 Parameters of Classifiers

Image Database	Classifier	Parameters	M-band			CVT			CNT			RT			NSCT		
			F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
ImgDb1	MLP	iter	1600	2200	1700	1100	1800	1200	2000	3000	2000	2000	3200	1800	2200	3800	2300
		μ	0.62	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	0.72	0.74	0.69	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	210	330	200	200	280	210	220	300	225	225	340	225	230	350	200
		D_{max}	25	30	25	20	30	25	20	30	20	25	30	25	25	30	25
	LSSVM	C	120	130	110	80	100	100	150	200	150	180	200	100	100	200	140
		γ	0.0215	0.0256	0.0357	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	2200	2600	2400	2000	2500	2000	2200	2900	2200	2300	3500	2100	2400	4000	2500
		μ	0.625	0.732	0.665	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	0.724	0.764	0.752	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	230	350	230	200	320	200	230	360	230	240	360	230	240	370	240
		D_{max}	30	40	30	20	35	20	25	30	20	20	35	20	20	30	25
	LSSVM	C	210	240	230	240	210	200	230	280	210	240	270	240	220	290	260
		γ	0.263	0.252	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	3200	4000	3100	2900	3400	3000	3200	4300	3300	3500	4500	3200	3300	4400	3200
		μ	0.645	0.872	0.825	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	0.787	0.755	0.821	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	510	590	550	410	500	560	600	570	500	540	600	500	530	600	500
		D_{max}	50	55	55	35	50	50	50	50	50	50	50	50	45	50	40
	LSSVM	C	330	370	340	320	360	350	400	370	330	350	360	320	320	400	300
		γ	0.335	0.521	0.332	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	3300	4000	3500	3200	3400	3000	3500	4500	3200	3400	4700	3200	3200	5000	3400
		μ	0.677	0.833	0.818	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	0.782	0.755	0.829	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	480	550	490	390	520	580	590	600	540	520	600	530	520	600	500
		D_{max}	50	60	50	45	50	50	40	45	40	40	40	45	50	50	45
	LSSVM	C	330	370	340	310	350	3750	420	320	310	360	370	310	330	420	310
		γ	0.327	0.587	0.415	0.339	0.338	0.389	0.445	0.539	0.423	0.476	0.528	0.389	0.3814	0.520	0.327

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 best possible configuration, having a better classification performance in terms of accuracy and computational time. In LSSVM, we have used the 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 4.1.

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 chapter. In this study, we have used different transform *e.g.*, M-band, CVT, CNT, RT and NSCT; Filters *e.g.*, Meyer, article, MLT, Shannon, and (pyr, ori); Feature Representation in terms of 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 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 EMD [133] 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 [140, 174]. The computation cost of similarity/dissimilarity measure for EMD is $O(N^3 logN)$ [141] against ED is $O(N)$ [266]. 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.

4.3.2 Evaluation Criteria

To find out the best possible combination "*Transform (Filters)* $\Rightarrow$ *Feature Representation* $\Rightarrow$ *Classifiers*", three statistical measures are used in this work: AP, AR and F-Measure. 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.

4.3.3 Results and Discussion

In this section, firstly we have evaluated the performances of the different combinations "*Transform (Filters)* $\Rightarrow$ *Feature Representation* $\Rightarrow$ *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.

4.3.3.1 Retrieval with M-band Wavelet Transform

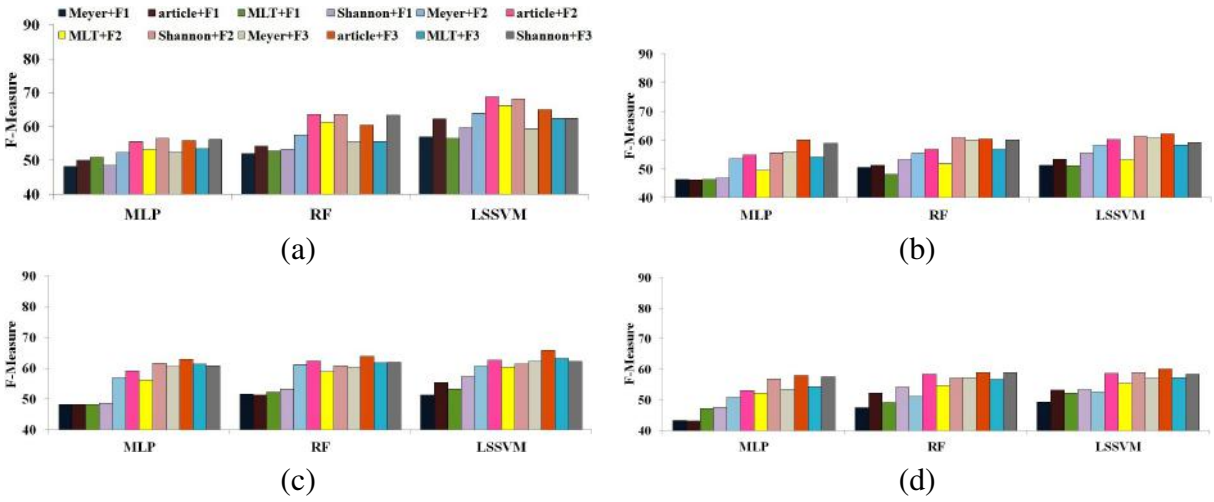


Fig. 4.1 F-Measure based performance graphs of M-band on (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d) ImgDb4

Fig. 4.1 represents the retrieval results obtained using different combinations of "*Transform (Filters) ⇒ Feature Representation ⇒ Classifiers*" for M-band wavelet transform. The graph of Fig. 4.1(a) shows that the combination of "*M-band(article) ⇒ F2 ⇒ RF*" is giving satisfactory performance on ImgDb1 database. The combination of "*M-band(article) ⇒ F3 ⇒ LSSVM*" is giving acceptable results on ImgDb2 and ImgDb4 image databases as shown in Fig. 4.1(b) and (d) respectively. But from Fig. 4.1(c), it has been observed that the combination"*M-band(article) ⇒ F3 ⇒ RF*" is showing the best performance as compared to other combination for ImgDb3 image database. Therefore, "*article*" filter show good performance with F2 and F3 in all four image databases. Due to the symmetry, orthogonality, regularity, coding gain, best possible time-frequency localization properties and the appropriate dealing with image signals with smooth changing patterns, "*article*"

filter performs better than the other member of the M-band wavelet filter family. The property of "symmetry" provides phase linearity and also useful in avoiding boundary artifacts in images. 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 4.2 for global comparison.

4.3.3.2 Retrieval with Curvelet Transform

The performance graphs of CVT using various combinations of "*Transform (Filters)* $\Rightarrow$ *Feature Representation* $\Rightarrow$ *Classifiers*" are shown in Fig. 4.2. From the Fig. 4.2(a) and Fig. 4.2(c), it can be seen that, the combination of " *CVT (9/7, pkva)* $\Rightarrow$ *F3* $\Rightarrow$ *MLP* " performs best for ImgDb1 and ImgDb3 image databases respectively. But for the ImgDb2 database, the combination of "*CVT (9/7, pkva)* $\Rightarrow$ *F3* $\Rightarrow$ *RF*" is giving the best performance as shown in Fig. 4.2(b). The combination of "*CVT (9/7, pkva)* $\Rightarrow$ *F3* $\Rightarrow$ *LSSVM*" is showing the good performance on ImgDb4 as shown in Fig. 4.2(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 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.

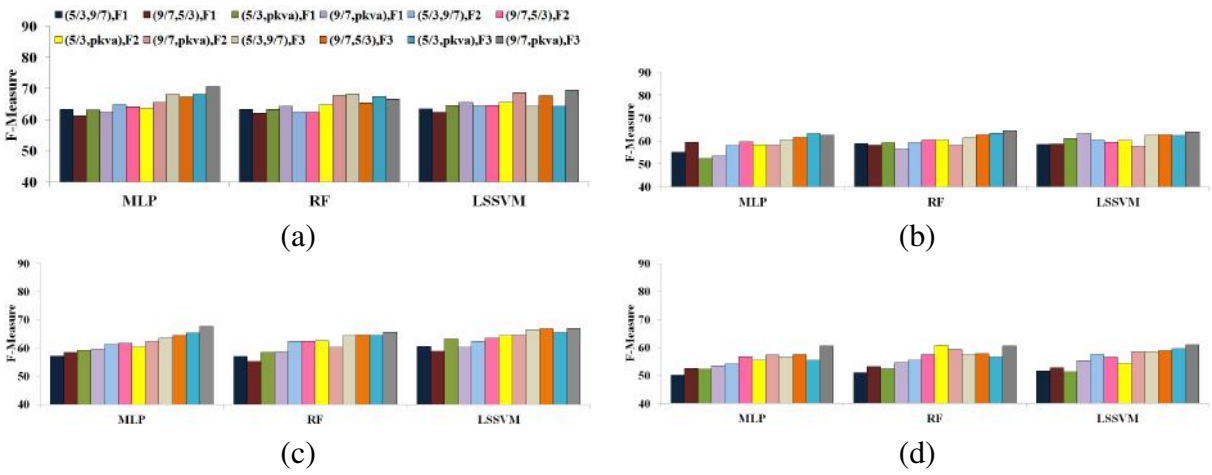


Fig. 4.2 F-Measure based performance graphs of M-band on (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d) ImgDb4

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 4.2 for global comparison.

4.3.3.3 Retrieval with Contourlet Transform

The implementation of the CNT is based on pyramid filtering (pyr) and orientation filtering (ori) as discussed in Section 2.3.2.2. We have compared different combinations of pyramid and orientation filters with three different types of classifiers. The graph of Fig. 4.3(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. 4.3(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. 4.3(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.

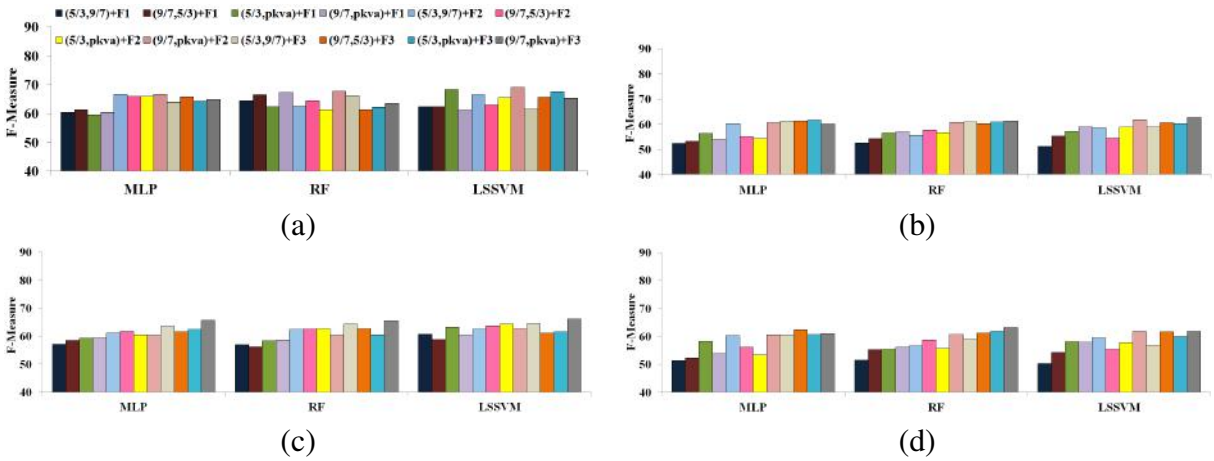


Fig. 4.3 F-Measure based performance graphs of M-band on (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d) ImgDb4

Comparing Fig. 4.2 and Fig. 4.3, 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 [200]. 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 4.2 for global comparison.

4.3.3.4 Retrieval with Ripplet Transform

The retrieval results using RT are shown in the graph of Fig. 4.4(a)-(d) in terms of F-Measure values. From the Fig. 4.4(a) and Fig. 4.4(c), it can be seen that the combination of " *RT (9/7, pkva)⇒ F3 ⇒ LSSVM* " is giving the best results as compared with other com-

bination in ImgDb1 and ImgDb3 image database. We can see from the graph in Fig. 4.4(b) and Fig. 4.4(d) is that the combination of " *RT (9/7, pkva)* $\Rightarrow$ *F3* $\Rightarrow$ *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.

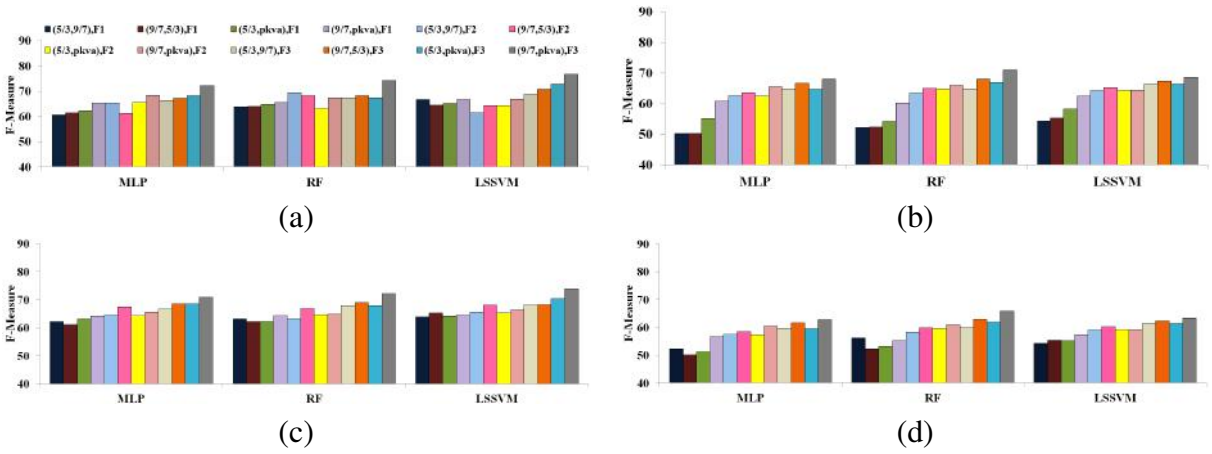


Fig. 4.4 F-Measure based performance graphs of M-band on (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d) ImgDb4

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 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. 4.2, Fig. 4.3 and Fig. 4.4. The best results are listed in Table 4.2 for global comparison.

4.3.3.5 Retrieval with Nonsubsampled Contourlet Transform

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 work. For each filter grouping, the setting of decomposition levels is same as CNT.

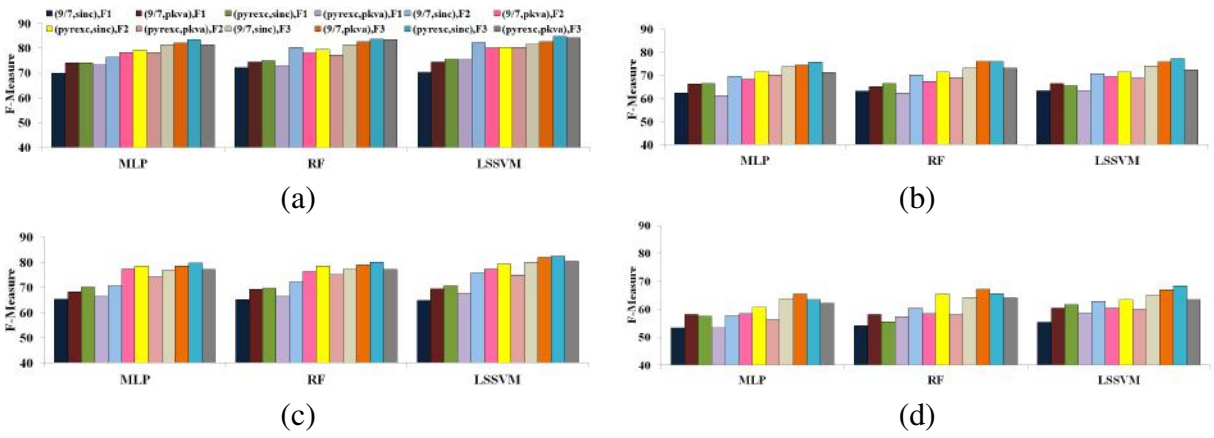


Fig. 4.5 F-Measure based performance graphs of M-band on (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d) ImgDb4

Fig 4.5(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 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. 4.5, 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. 4.5(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. 4.3 and Fig. 4.5, 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 4.2 for global comparison.

4.3.3.6 Global Comparison

Table 4.2 lists the overall best results of each type of transforms. From Table 4.2, it has been observed that NSCT is performing better than RT, CVT, CNT and M-band. This is mainly due to the multi-directional, shift-invariant and flexible multi-scale image decomposition property of NSCT. Table 4.2 also shows that, RT generally performs better than M-band, CVT and CNT. This is because RT has higher energy concentration ability as well as anisotropy property, which guarantee to capture singularities along various curves.

Table 4.2 The best results of each multi-resolution transform

Image Database	Transform	Filters		Feature	Classifier	F-Measure
ImgDb1	M-band	article	-	F2	LSSVM	68.78
	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	M-band	article	-	F3	LSSVM	62.18
	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	M-band	article	-	F3	RF	65.91
	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	M-band	article	-	F3	LSSVM	60.15
	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**

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

Feature representation plays a vital role in CBIR paradigm. 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 M-band is 16, therefore in three planes; we have total 48 sub-bands. Considering two statistical parameters, in total an image is represented by 96 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 M-band, the number of feature maps for each image of the database is 48 ($= 16 \times 3$; 3 color planes per image and 16 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 147 $(= 144(= 3 \times 48) + 3)$. 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 4.2 with the size of the features vector as discussed above, 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 [267, 268]. 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 4.2, 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 best possible 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 [269]. 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 [270].

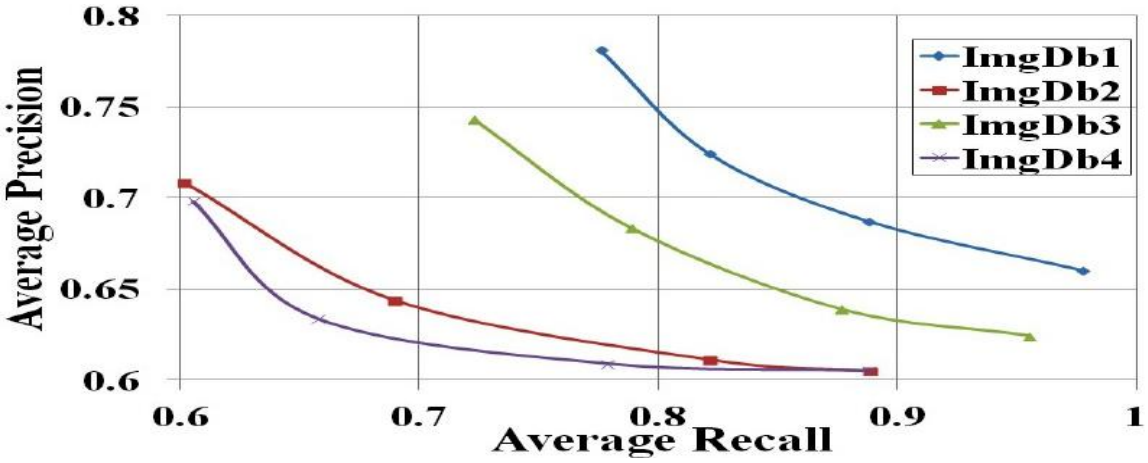


Fig. 4.6 Average precision Vs Average Recall graph of the combination "NSCT (pyrexc, pkva)⇒ F3 ⇒ LSSVM'

Fig. 4.6, 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 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 work. 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

Table 4.3 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

and object based image retrieval. Therefore, the combination " *NSCT (pyrexc, 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 chapter adopts MICI [234] based feature evaluation mechanism to choose the best possible features from the best combination. Table 4.3 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 best possible 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) [271], HOG [272] and SIFT [83] as shown in Table 4.4. Here, we have used LSSVM as classifier. It is seen from the Table 4.4 that the classification

Table 4.4 Comparisons with other existing CBIR systems in terms of Classification Accuracy (%)

Image Database	SPM	HOG	SIFT	NSCT(pyrexc,sinc)⇒F3
Feature Dimension	4200	128	128	66
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 4.5 Approximate feature extraction time of each image in each multi-resolution transform

Transform	Time(in seconds)		
	F1	F2	F3
M-band	0.98	2.00	1.95
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

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 from Table 4.4 that the feature dimension used for SPM, HOG and SIFT are very high as compared to that of "*NSCT (pyrexc, sinc) ⇒ F3*". 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 computed 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 (Mband, CVT, CNT, RT and NSCT) are shown in Table 4.5. From this

Table 4.6 Approximate processing time in best combination on different image databases

Image Database	Mechanism	Processing Time (in seconds)
ImgDb1	M-band(article)⇒F2⇒RF	4.68
	CVT(9/7,pkva)⇒F3⇒MLP	4.71
	CNT(9/7,pkva)⇒F2⇒LSSVM	5.86
	RT(9/7,pkva)⇒F3⇒LSSVM	5.39
	NSCT(pyrexc, sinc)⇒F3⇒LSSVM	5.83
ImgDb2	M-band(article)⇒F3⇒LSSVM	6.30
	CVT(9/7,pkva)⇒F3⇒RF	6.75
	CNT(9/7,pkva)⇒F3⇒LSSVM	7.81
	RT(9/7,pkva)⇒F3⇒RF	8.62
	NSCT(pyrexc, sinc)⇒F3⇒LSSVM	8.89
ImgDb3	M-band(article)⇒F3⇒LSSVM	9.85
	CVT(9/7,pkva)⇒F3⇒MLP	10.73
	CNT(9/7,pkva)⇒F3⇒LSSVM	11.64
	RT(9/7,pkva)⇒F3⇒LSSVM	11.95
	NSCT(pyrexc, sinc)⇒F3⇒LSSVM	12.40
ImgDb4	M-band(article)⇒F3⇒LSSVM	9.14
	CVT(9/7,pkva)⇒F3⇒LSSVM	10.80
	CNT(9/7,pkva)⇒F3⇒RF	11.31
	RT(9/7,pkva)⇒F3⇒RF	11.67
	NSCT(pyrexc, sinc)⇒F3⇒LSSVM	12.42

Table 4.7 Average processing time using best possible 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

table, it can be inferred that M-band is taken minimum time and NSCT is taking the maximum time for feature extraction among other different transforms discussed in this chapter. Table 4.6 shows the processing time of best combination. From the Table 4.6, it can be seen that the combination "*NSCT(pyrexc, sinc)⇒F3⇒LSSVM*" performs satisfactorily on time complexity 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 4.7 shows the average CPU processing time using best possible features (= 33) and ED similarity measure.

Fig. 4.7 to Fig. 4.10, shows the examples of the visual results obtained by the selected

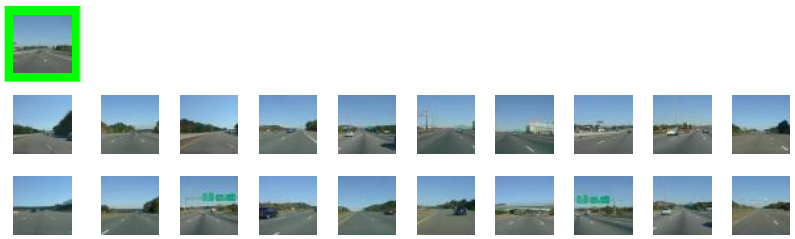


Fig. 4.7 Visual results of the CBIR system using the best combination of "*NSCT (pyrexc, sinc)* $\Rightarrow$ *F3* $\Rightarrow$ *LSSVM*" and ED distance measure for ImgDb1 database (image marked by the greenbox is the query image)

Fig. 4.8 Visual results of the CBIR system using the best combination of "*NSCT (pyrexc, sinc)* $\Rightarrow$ *F3* $\Rightarrow$ *LSSVM*" and ED distance measure for ImgDb2 database (image marked by the greenbox is the query image)

Fig. 4.9 Visual results of the CBIR system using the best combination of "*NSCT (pyrexc, sinc)* $\Rightarrow$ *F3* $\Rightarrow$ *LSSVM*" and ED distance measure for ImgDb3 database (image marked by the greenbox is the query image)

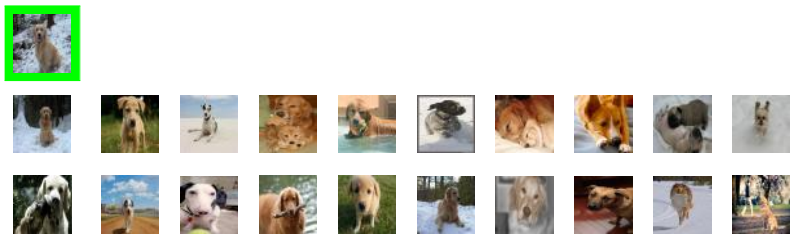


Fig. 4.10 Visual results of the CBIR system using the best combination of "*NSCT (pyrexc, sinc)* $\Rightarrow$ *F3* $\Rightarrow$ *LSSVM*" and ED distance measure for ImgDb4 database (image marked by the greenbox is the query image)

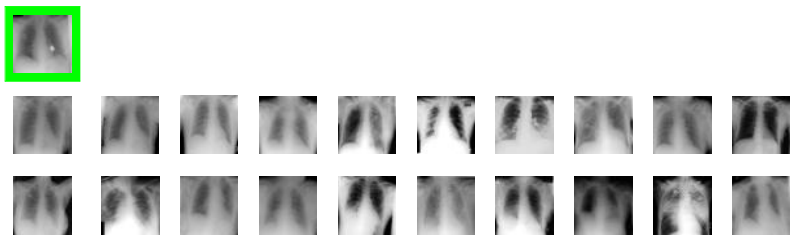


Fig. 4.11 Visual results of the CBIR system using the best combination of "*NSCT (pyrexc, sinc)*$\Rightarrow$ *F3* $\Rightarrow$ *LSSVM*" and ED distance measure for medical database (image marked by the green box is the query image)

best combination *i.e.*, " *NSCT (pyrexc , sinc)* $\Rightarrow$ *F3* $\Rightarrow$ *LSSVM*", with ED similarity measure and best possible 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. 4.11 also shows the visual results on IRMA medical database [273] using the combination " *NSCT (pyrexc, sinc) $\Rightarrow$ F3 $\Rightarrow$ LSSVM*" with best possible features (33) and ED similarity measure. Fig. 4.12 shows the visual results results using the combination " *NSCT (pyrexc, sinc) $\Rightarrow$ F3 $\Rightarrow$ LSSVM*" with ED similarity measure and FRF mechanism (as discussed in Chapter 3) on ImgDb3. Fig. 4.12(a)-(c) shows that how ranking of the retrieved images are improved with user interactions.

In this chapter, we have presented an extensive comparative study of the effectiveness

(a) (b) (c)

Fig. 4.12 Retrieval results using best combination of "*NSCT (pyrexc, sinc)⇒ F3 ⇒ LSSVM*" and FRF mechanism on ImgDb3 (Top left side image as the query image and images marked in red and green indicates the ranking of relevant images) (a) First Pass (20/20) (b) 1st iterations (20/20, improved ranking) (c) 2nd Iterations (20/20, improved ranking).

Database	YCbCr	CIElab
ImgDb1	85.674	87.152
ImgDb2	78.653	79.351
ImgDb3	82.431	83.119
ImgDb4	70.284	71.027

Table 4.8 Best mechanism (*NSCT (pyrexc, sinc) ⇒ F3 ⇒ LSSVM*) in different color space *w.r.t* F-Measure

of five multi-resolution transforms using different filters and three different classifiers in CBIR application. For each multi-resolution transform, the best possible 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 best possible 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 to discover the best possible 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 M-bandWT. Contrary to this, M-bandWT takes minimum time as compared with other transforms. To the best in our knowledge, this is the first attempt, where different transforms have been compared in the same CBIR platform. To further justify the best transform (NSCT), we have compared two different color space (YCbCr and CIELab) in four different color space as shown in Table 4.8. Here, we have extracted F3 features in CIElab color space and compared with earlier F3 (in YCbCr color space), considering the same feature vector dimension. From the table, it is found that using CIElab color space substantially increases the accuracy of the CBIR system as compared with YCbCr.

Therefore, after extensive experimental investigation of five different transform in this chapter, it is observed that although M-band WT as a feature extractor takes minimum computation time, but the accuracy of a CBIR system based on M-band WT is comparatively lower than that of the system based on other MGA/MRA tools like CNT, NSCT, etc. It is also observed that the system designed using NSCT is having high accuracy with high computational cost. Due to the complex nature of its bases function/kernel, its complexity is higher. In view of the above observation, we will going to describe a new relevance feedback based CBIR system with a graph theoretic approach using NSCT and with its best possible combination of parameters. This work has been published in [178].

$**$**Major portion of this Chapter taken from the following publication.**

- M.Chowdhury and M. K. Kundu, "Comparative assessment of efficiency for content based image retrieval systems using different wavelet features and pre-classifier", Multimed. Tools Appl.(Springer), Vol. 72, No. 3, pp. 1–36, 2014(**IF:1.331**).

Chapter 5

Content-Based Image Retrieval using Graph-Based Relevance Feedback Mechanism

5.1 Introduction

In chapter 3, we have proposed an auto-updation of feature important components based on Fuzzy Relevance Feedback (FRF) mechanism. In the weight updating approach, the weights of the different components of the multidimensional feature vectors components are altered based on a score guided by the factor relevant/irrelevant as decided by the observer/user. The respective weight associated with the feature component is increased or decreased depending upon its presence in relevant and irrelevant images. Let us take an example for irrelevant images: One feature component may be close to the query whereas other feature component may be far away from the query feature. But the overall magnitude of the similarity vector distance may be close to the relevant ones. After the characteristic features of the query and the stored images, the relative weight associated with the feature component is getting updated. The feature components having larger relative weights are expected to dominate over the feature space such that the rank of the irrelevant ones is pulled down. Beside the relative weight, individual feature component values also greatly influence the outcome of similarity computation. For example: if one or few feature components have values far greater than that of other feature components, then in the computation of similarity distance they will dominate the overall outcome. In this process, some of the important contributions of the lower value feature component will be overshadowed, resulting in an erroneous ranking of the images. To overcome this problem in CBIR systems, a good re-ranking algorithm with satisfactory time complexity is needed to be developed [163, 274, 275].

In this chapter, we propose a new CBIR system which is based on graph theoretic model. In the proposed system, we have used a higher order representation of features (secondary features) which is computed from the distribution of primary multiscale transform component features extracted using NSCT on an image. This will results in a lower dimensional features vectors (much much less than multiscale primary NSCT based features).

The features dimension is further reduced by applying Maximal Information Compression Index (MICI) on the feature vector [234]. We then employ a transductive, graph-based ranking approach that exploits the relevance feedback information. It relies on a sparse graph representation in which the database images and the query image are nodes and edge weights are expressed in terms of the Gaussian kernel distance between the image's feature representation. Following the method discussed in [163], ranking is initially carried out by a simple k-nearest-neighbor approach, while the subsequent rankings, which account for the user's feedback, are given according to the probability that a random walk starting from a node in the graph will reach a relevant image before hitting a non-relevant one. Extensive experiments using the proposed NSCT-based features and this graph theoretic relevance feedback mechanism on four different databases show the superiority of our proposed method over different recent approaches.

5.2 Research Contribution [179]

The major contribution of this chapter is as follows:

The proposed work based on an idea that creates a graph where nodes are representative of images and similarity are represented by corresponding edges. The proposed technique is an iterative process where at each iteration, relevant and irrelevant images are treated as seed nodes in a random walker problem. The modified edges (relation between images) are computed using Gaussian kernel. To our knowledge, gaussian kernel has been used for the first time in computing the similarity edges. The ranking score for each unlabeled image is computed as the probability that a random walker starting from that image will reach a relevant seed before encountering a non-relevant one. The details are explain in the following subsection.

5.2.1 NSCT based Image Representation

An efficient representation of images usually leads to improvements in terms of storage, computational complexity and performance of image processing algorithms. One way to obtain such representations is by means of image transformation methods. Conventional transformation methods such as Fourier transform and wavelet transform suffer from discontinuities in images like the one introduced by edges. To overcome this problem, we adopt the NSCT method described in the previous section, which exhibits a larger degree of shift-invariance and anisotropy as well as better directional decomposition and frequency localization.

Before extracting the NSCT-based features, we map the RGB images to the CIE Lab color space in order to deal with a color metric that better matches the human perception [276]. This is important as human will play a key role during the retrieval phase. Moreover, by decoupling the brightness intensity and the chromatic channels we ensure an independence between textural and color characterizations. Indeed, the NSCT decomposition over the intensity channel (L) captures the texture information, while the same decomposition over the chromatic coordinates (a and b) characterizes color information. Texture and color information are extracted by using NSCT on the Lab channels with a 4 level (1, 2, 4, 4) decomposition. This decomposition configuration provides for each image in the database 11 $(= 1+2+4+4)$ subbands per channel (in total 33 subbands per image). We consider a 4-level NSCT decomposition because the Peak Signal to Noise Ratio (PSNR), which measures the quality of the reconstructed image, does not improve significantly when the level of decomposition is greater than 4. Consequently, the designed level provides a good compromise between having a low loss of information and a compact feature-representation. Note that this empirical evidence is not new, as it was already mentioned in [277]. Each subband S^j, $0 \leq j \leq 33$ is then summarized in terms of its mean

(f^j_{mean}), standard deviation (f^j_{std}) and energy (f^j_{energy}) as follows:

$$f^j_{mean} = \frac{1}{\mathsf{hw}} \sum_{m=1}^{\mathsf{h}} \sum_{n=1}^{\mathsf{w}} \mathtt{S}^j_{mn}\,, \tag{5.1}$$

$$f^j_{std} = \sqrt{\frac{1}{\mathsf{hw}} \sum_{m=1}^{\mathsf{h}} \sum_{n=1}^{\mathsf{w}} (\mathtt{S}^j_{mn} - f^j_{mean})^2}\,, \tag{5.2}$$

$$f^j_{energy} = \frac{1}{\mathsf{hw}} \sum_{m=1}^{\mathsf{h}} \sum_{n=1}^{\mathsf{w}} (\mathtt{S}^j_{mn})^2\,, \tag{5.3}$$

where, $\mathtt{S}^j_{mn}$ represents the NSCT coefficient at the spatial location (m,n) of the subband $\mathtt{S}^j$, begin a $\mathsf{h} \times \mathsf{w}$-dimensional matrix. Finally, each image I in the database is abstracted in terms of a feature vector comprising the aforementioned statistics for each subband. This yields the following 99-dimensional vector:

$$f^I = \left(f^1_{\text{mean}}, ..., f^{33}_{\text{mean}}, f^1_{\text{std}}, ..., f^{33}_{\text{std}}, f^1_{\text{energy}}, ..., f^{33}_{\text{energy}}\right)\,. \tag{5.4}$$

5.2.2 Features selection

The success of the image recognition task depends on the feature representation adopted. Large feature vectors are prone to contain helpful information, but on the other side an excess of poorly informative features can interfere with the positive outcome of the learning algorithm. Moreover, from a computational perspective, large feature representations increase the recognition complexity. For this reason, a good practice consists in trying to reduce the feature space in a way to ideally preserve only the most informative features (*feature selection*).

In our work, we retain the relevant NSCT features by adopting a feature selection algorithm based on the MICI [234]. This unsupervised method does not rely on searching

techniques, thus avoiding over-consumption of calculation time, but it removes redundancies among the features by evaluating their mutual dissimilarity in terms of MICI. In details, given two random variables x and y, let Σ be their covariance matrix with entries Σ_{xx}, Σ_{yy} and $\Sigma_{xy} = \Sigma_{yx}$. The value of MICI for x and y is defined as the smallest eigenvalue $\lambda_2(x,y)$ of the covariance matrix Σ, which is given by

$$\lambda_2(x,y) = \frac{1}{2}\left[\Sigma_{xx} + \Sigma_{yy} - \sqrt{(\Sigma_{xx} - \Sigma_{yy})^2 + 4\Sigma_{xy}^2}\right] . \tag{5.5}$$

The value of $\lambda_2(x,y)$ is zero when the features represented by x and y are linearly dependent (hence, similar) and it increases as the degree of dependency decreases. The algorithm clusters the features into homogeneous groups by exploiting this MICI-based feature dissimilarity and retains a representative feature for each cluster. The set of representatives forms the reduced set of features and its size is controlled by a user-defined parameter (see [234] for further details).

Hereafter, we denote by $\overline{f}^I$ the reduced NSCT feature vector associated to image I, obtained after the application of the aforementioned feature selection procedure.

5.2.3 Graph based image re-ranking

Human perception of image similarity is subjective, semantic- and task-dependent. Hence, a CBIR system cannot deliver satisfactory results in general by relying merely on features extracted in an unsupervised way, but the user has to be incorporated in the retrieval loop as a source of feedback. By doing so, the system can iteratively adapt the query results to meet user-specific requirements.

The solution that we adopt has been proposed in [179] and consists in a random-walk-based re-ranking algorithm that relies on a graph abstraction of the images dataset. Let $\mathscr{V} = \{0,\ldots,\mathsf{n}\}$ and let $\mathscr{I} = \{I_u\}_{u\in\mathscr{V}}$ be a set of $\mathsf{n}+1$ images, I_0 being the query image. We

represent our image database in terms of a *(edge-weighted) graph* $G = (\mathscr{V}, \mathscr{E}, \mathtt{W})$, where $\mathscr{V}$ is the set *vertices* corresponding to image indices and $\mathscr{E} \subseteq \mathscr{V} \times \mathscr{V}$ is a set of *edges*, each one being augmented with a nonnegative weight. The *weights* are provided by the $\mathtt{n} \times \mathtt{n}$ matrix $\mathtt{W} = (\mathtt{W}_{ij})$, where each entry $\mathtt{W}_{ij}$ holds the weight of edge $(i, j) \in \mathscr{E}$ or zero if $(i, j) \notin \mathscr{E}$. In our specific case, we use a Gaussian kernel function to determine similarities among pairs of images abstracted in terms of the reduced NSCT features, *i.e.*,

$$\mathtt{W}_{ij} = \exp\left(\frac{\left\|\overline{f}^{I_i} - \overline{f}^{I_j}\right\|^2}{\sigma^2}\right), \tag{5.6}$$

where the variance σ^2 determines the similarity scale level. The *Laplacian matrix* $\mathtt{L}$ of graph G is defined as $\mathtt{L} = \mathtt{D} - \mathtt{W}$, where $\mathtt{D}$ is a diagonal matrix having diagonal entries $\mathtt{D}_{ii} = \sum_{j \in \mathscr{V}} \mathtt{W}_{ij}$.

The user, who issues the query, can determine whether any image in the database is relevant or not. We denote by $\varphi_i \in \{0, 1\}$, $i \in \mathscr{V}$, the feedback that the user would give to the ith image, namely $\varphi_i = 1$ if the image is relevant and $\varphi_i = 0$ if it is non-relevant. Clearly, we always assume the query image itself to be relevant, *i.e.*, $\varphi_0 = 1$. At each feedback iteration $t > 0$, the user is given the opportunity to indicate which images among the provided ones are relevant for the query. By doing so, the system gathers more and more information about the relevancy of the images in the database. We denote by $\mathscr{L}^{(t)} \subseteq \mathscr{V}$ the indices of images that have been labelled as relevant/non-relevant by the user up to the tth feedback round. Initially, only the query image is labelled, *i.e.*, $\mathscr{L}^{(0)} = \{0\}$, and the set of labeled images can never shrink as the number of feedback rounds increases, *i.e.*, $\mathscr{L}^{(t-1)} \subseteq \mathscr{L}^{(t)}$ for all $t > 0$.

Consider now a generic feedback round $t > 0$. The idea behind the re-ranking algorithm is to determine for each image I_j, $j \in \mathscr{V}$, the probability π_j that a random walk on graph G starting in j will reach a relevant image before hitting a non-relevant one. The ranking can

then be obtained by sorting the images in decreasing order of π_j. If we assume the transition probability between two vertices $(i,j) \in \mathscr{E}$ of the graph G to be proportional to the edge weight $\mathtt{W}_{ij}$, the (column) vector of probabilities $\pi = (\pi_0, \ldots, \pi_\mathsf{n})^\top$ can be determined as the solution of the following *convex*, quadratic optimization problem [163]:

$$\begin{aligned} \text{minimize} \quad & \pi^\top \mathtt{L}\, \pi \\ \text{s.t.} \quad & \pi_i = \varphi_i \quad \forall i \in \mathscr{L}^{(t)} \\ & 0 \leq \pi_i \leq 1 \quad \forall i \in \mathscr{V}\,. \end{aligned} \tag{5.7}$$

Without loss of generality, assume the labelled images to have the smallest indices in $\mathscr{V}$, *i.e.*, $\mathscr{L} = \mathscr{L}^{(t)} = \{0, \ldots, \mathsf{m}\}$, in a way to allow the following decompositions for $\mathtt{L}$ and π into parts involving labelled (L) and unlabelled (U) images:

$$\mathtt{L} = \begin{bmatrix} \mathtt{L}_{LL} & \mathtt{L}_{LU} \\ \mathtt{L}_{UL} & \mathtt{L}_{UU} \end{bmatrix} \qquad \pi = \begin{pmatrix} \pi_L \\ \pi_U\,. \end{pmatrix}$$

Given this reordering of the indices, the solution π^* of Eq.(5.7) can be easily found by solving the following system of linear equations:

$$\mathtt{L}_{UU}\pi_U = -\mathtt{L}_{UL}\pi_L\,. \tag{5.8}$$

The system is nonsingular if the graph is connected or if every connected component contains a labeled image [278]. By applying some sparsifying strategy on graph G, the system can be efficiently solved also in case of large-scale image datasets. Once π_U and thus π is computed, we can sort the images in decreasing order of π_i to get the most relevant images ranked first.

Interestingly, the minimization problem in Eq.(5.7) is a combinatorial formulation of the *Dirichlet problem* [279], *i.e.*, the problem of finding an harmonic function $\pi(\cdot)$ that

satisfies the Laplace equation $\nabla^2\pi = 0$ subject to some boundary conditions $\pi(x) = \varphi(x)$ for all $x \in \Omega$ and some fixed function φ. The harmonic function that satisfies the boundary conditions minimizes the *Dirichlet integral* $D(\pi) = \frac{1}{2}\int_\Omega |\nabla u|^2 d\Omega$, for the Euler-Lagrange equation for $D(\cdot)$ gives the Laplace equation. Translated into a graph setting, we have that the function $\pi : \mathscr{V} \to \mathbb{R}$ assigns values to vertices of the graph (so π can be regarded as a vector π) and the boundary conditions are fixed values of $\pi_i = \varphi_i$ on a subset of vertices $i \in \Omega$. In our case Ω is the set of labelled images $\mathscr{L}^{(t)}$ and φ_i is the feedback for the ith image. Under this discretized setting, the Dirichlet integral is given by $\pi^\top \mathtt{L}\pi$, if we take the Laplacian matrix $\mathtt{L}$ as the combinatorial Laplacian operator. Accordingly, its minimization under the boundary constraints, *i.e.*, the solution to Eq.(5.7), gives the solution to the (combinatorial) Dirichlet problem.[1] More details about the connection to the combinatorial Dirichlet problem and other connections to discrete potential theory are discussed in [279].

5.3 Summary of the algorithm

In this section we summarize the algorithm underlying our CBIR system with relevance feedback. We start by preprocessing the database of images as follows. We convert each image I in the database into a NSCT feature f^I as detailed in Section 2.3.2.3. By applying the feature selection criterion described in Section 5.2.2, we determine a reduced set of NSCT features, which allows us to lower the dimensionality of the data. As a result, each image I will be associated to a reduced feature vector $\overline{f}^I$. Next, we pre-construct a graph $G_0 = (\mathscr{V} \setminus \{0\}, \mathscr{E}, \mathtt{W})$ having images as vertices and edge-weights representing the similarity of two images computed in terms of the distance between their reduced NSCT feature vectors. The graph construction procedure follows the description in Section 5.2.3 with the

[1] Note that the conditions $0 \leq \pi_i \leq 1$ in (5.7) are actually superfluous, for they are always satisfied as long as $0 \leq \varphi_i \leq 1$.

only exception that we do not consider the query image yet (indeed we excluded 0 from the vertex set).

When the user performs a query by presenting an image I_0, we convert it into a reduced NSCT feature vector and we connect it to graph G_0, obtaining graph $G = (\mathscr{V}, \mathscr{E}, \mathtt{W})$ defined in Section 5.2.3. The initial query result consist of s images being the most similar to the query image. Here, s is a user-defined parameter that determines the size of the query result set. The user is then given the opportunity of refining the query result by providing a feedback, which consists in marking the relevant images. The feedback is used by the system to setup the optimization problem in Eq.(5.7). Once the solution π is computed by simply solving the linear system of equalities in Eq.(5.8), the images are re-ranked in decreasing order of π_i and the s best ranked images are returned to the user. This process is reiterated until the user is satisfied with the query result.

5.4 Experimental Results

We conduct extensive experiments to evaluate the performance of the proposed CBIR system with relevance feedback. We compare the performance of our system with several state-of-the-art CBIR systems. The details description of the evaluating process are describe as follows.

5.4.1 Experimental setup

We evaluate the proposed CBIR system on images taken from three publicly available image datasets. They are organized into semantic categories in a way to reflect the human perception of image similarity. Databases are (a) SIMPLIcity (ImgDb1) (b) Oliva (ImgDb2) (c) Caltech 256 (ImgDb3) and (d) MIRFlickr25000 (ImgDB4).

Our system uses a 4-level NSCT decomposition as described in Section 5.2.1 and Sec-

tion 2.3.2.3 using the CIELab color space. Our choice for the Lab color space has been motivated by the analysis conducted in Subsection 5.4.2. The feature selection phase has been tuned according to the procedure presented in Subsection 5.4.3. The performance on each database is evaluated by considering each image as the query image and by measuring the average classification *P* and *R* as discussed in Section 2.7 of Chapter 2.

5.4.2 Effect of color space transformation

In this experiment, we evaluate the quality of the NSCT-based low-level features computed using different color spaces, namely RGB, YCbCr and CIELab. Fig. 5.1(a)-(d) shows the retrieval performance obtained with the different color spaces on the datasets mentioned before in terms of average precision/recall at varying scope sizes, *i.e.*, the number of images returned by the query, namely 20, 40, 60 and 100. The results show that the CIELab color space outperforms the other ones on all the databases. This is probably due to its perceptual uniformity property, while RGB and YCbCr suffer from color closeness which is not supported by the HVS [280]. The advantage of adopting the CIELab color space will be evident also in the quantitative experiments of CBIR with relevance feedback that we conduct in Subsection 5.4.5.

5.4.3 Feature Evaluation

In this section, we tune for each dataset the number of features that should be selected by the algorithm presented in Subsection 5.4.3. The idea is to select the number of features that allows for the best category discrimination. To this end, we train a multi-class LS-SVM classifier using a one-versus-one strategy with a 10-fold cross-validation on the parametrization. A share of 60% labeled samples from the datasets forms the training set, while the remaining 40% is reserved for the test set. We evaluate the performance of the classifier using different shares of selected features (20%, 40%, 60% and 100%) in terms

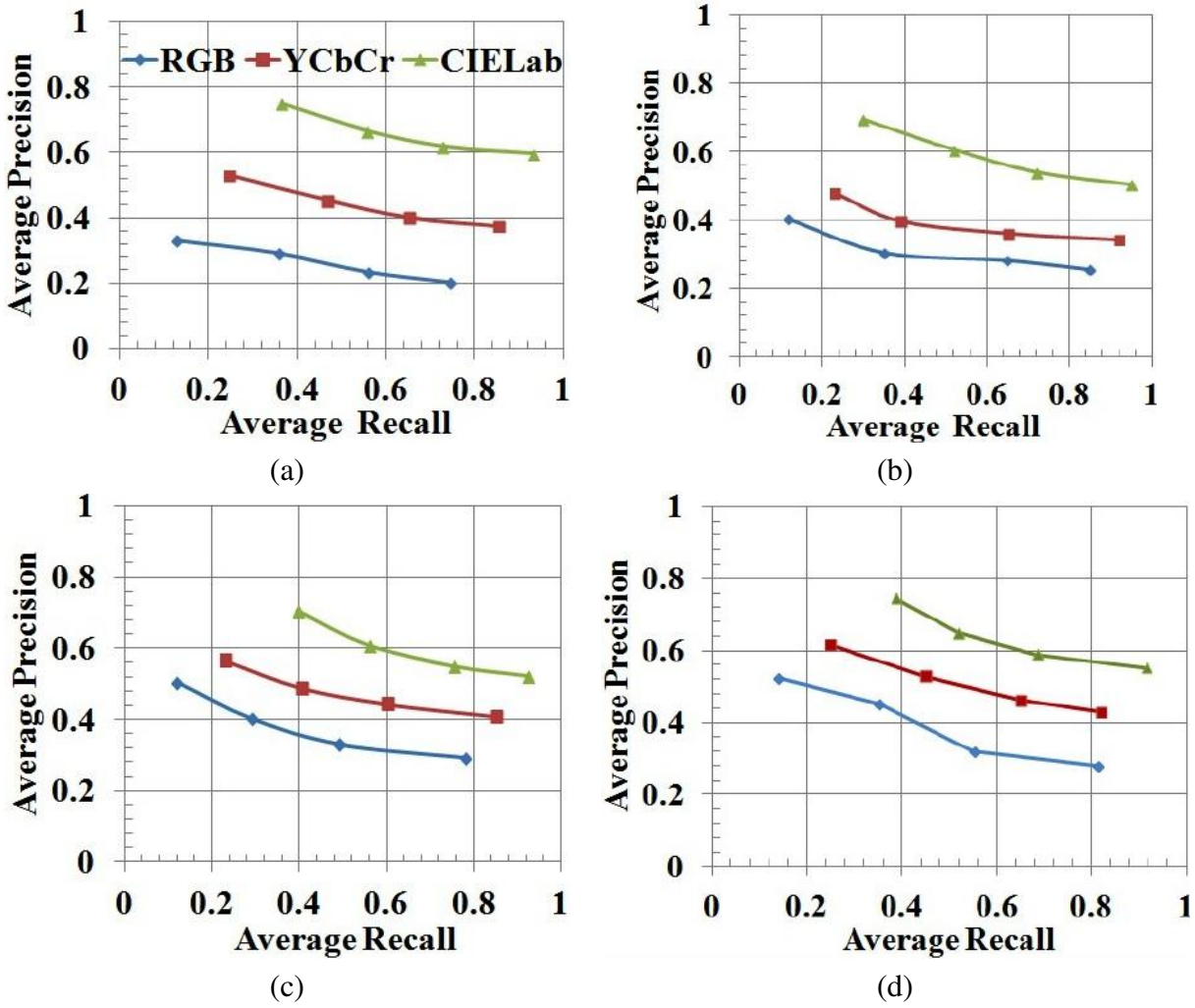


Fig. 5.1 Average Precision Vs Average Recall Graph for Different Color Model (RGB, YCbCr and CIELab) (a) ImgDb1 (b) ImgDb2 (c) ImgDb3 (d)ImgDb4

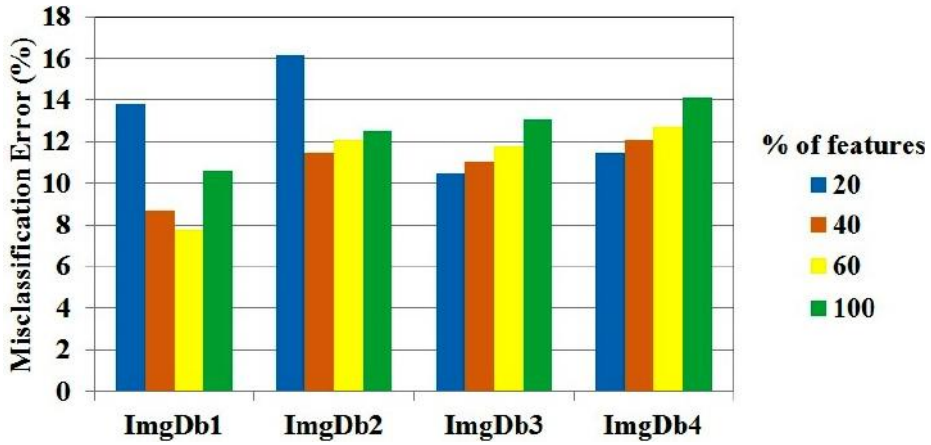


Fig. 5.2 Results obtained on three datasets with a LS-SVM trained considering different share of features.

of misclassification error on each dataset (ImgDb1-4). According to the results reported in Fig. 5.2, we can clearly see that the feature selection step has an impact on the performance. The best results are achieved with a share of 60%, 40% and 20% features on ImgDb1, ImgDb2, ImgDb3 and ImgDb4, respectively. Hence, the feature selection algorithm will be tuned for the subsequent experiments in a way to return the optimal shares of features that we found for each dataset.

5.4.4 Qualitative Results

In Figures 5.3-5.6 we provide a sample query from each dataset before and after some feedback steps. Fig. 5.3(a) shows the initial retrieval results obtained on ImgDb1 dataset. Here, the top-left image is the query image. As it turns out, $17/20$ images are from the same class "flower" and have an appearance similar to the query flower. The remaining ones are misclassified images from the class "food", which mostly arise due to the prominent role of the color feature in both the flower and food classes. Nevertheless, also the shape component is partially causing the misclassification, for some erroneous images do have a flower like shape (see,*e.g.*, bottom-right image in (a)). After 1 feedback iteration only 1

error remains, while we achieve 100% accuracy within the 8th feedback iteration.

(a) (b) (c)

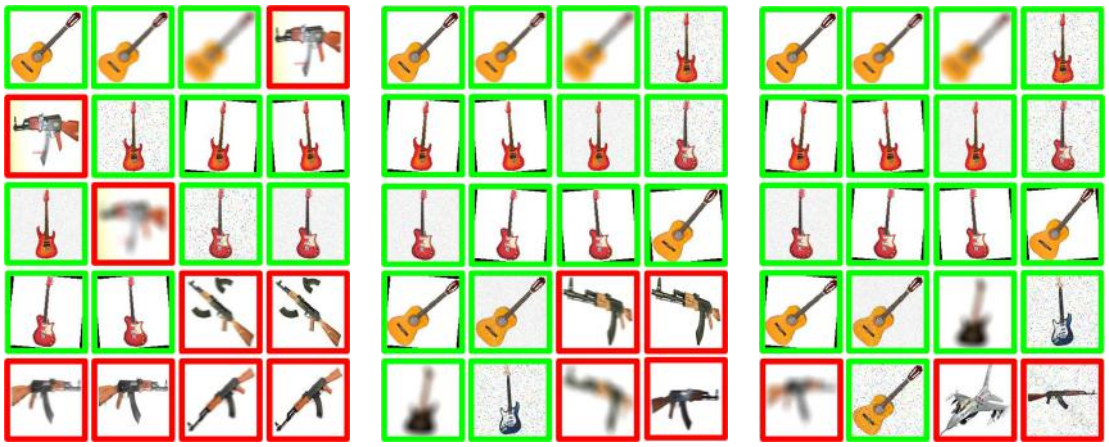
Fig. 5.3 Retrieval results on ImgDb1 with the query image in the top-left corner. (a) Initial result (85% accuracy) (b) 1st iteration (95%) (c) 8th iteration (100%)

Fig. 5.4(a) shows the initial retrieval results obtained on ImgDb2 dataset. Here, $15/20$ images are similar to the query image (top-left), while 5 images are from different classes. Also in this case, the color cue plays an important role (*e.g.*, color of road and sky), after the first human intervention (see Fig.5.4(b)), the results are improved by reducing the errors to just 3. We achieve 100% accuracy within 4 iterations as shown in Fig. 5.4(c).

(a) (b) (c)

Fig. 5.4 Retrieval results on ImgDb2 with the query image in the top-left corner. (a) Initial result (75% accuracy) (b) 1st iteration (85%) (c) 4th iteration (100%)

Fig.5.5(a)-(c) shows the retrieval results obtained on ImgDb3 using a guitar query image. The proposed algorithm retrieves both images that represent similar guitars as well as images of the same guitar that have been corrupted (*e.g.*, blurred and rotated), thus demonstrating the robustness of the method to different types of noise. In this case, the distinctive feature is shape and not just color. We experience a high misclassification rate in the initial retrieval result (45%), which decreases to 20% after 1 iteration and to 15% after 4 feedback rounds.



(a) (b) (c)

Fig. 5.5 Retrieval results on ImgDb3 with the query image in the top-left corner. (a) Initial result (55% accuracy) (b) 1st iteration (80%) (c) 4th iteration (85%)

Fig. 5.6(a)-(c) shows the retrieval results obtained on ImgDb4 using pink flower as a query image. Here, images retrieved resembles the pink flower like structure but some images are irrelevant. Here, also color plays a important role in initial retrieval results like Fig. 5.4. But with every iterations results have been improved from 65% to 85% within 5^{th} iterations. Finally, Figures 5.7, 5.8 and 5.9 show the results of a sample query on ImgDb1 obtained by our approach, by the Ego Similarity Measure (ESM) [281] and by the Semi-Supervised Learning (SSL) approach of Yang *et al.* [282]. Both ESM and SSL are recent approaches for CBIR with relevance-feedback: SSL is a state-of-the-art, graph-based ranking mechanism that is closer in spirit to our method thus being a direct competitor, whereas

(a) (b) (c)

Fig. 5.6 Retrieval results on ImgDb4 with the query image in the top-left corner. (a) Initial result (65% accuracy) (b) 1st iteration (75%) (c) 5th iteration (85%)

ESM is among the most recent approaches based on a weight-updating scheme. As we can see from the plots, our approach achieves the best accuracy (90%) already after the first iteration, followed by SSL with 85% and ESM with 80%. Also in terms of ranking quality, our approach outperforms the competitors, for the first error appears at the 17th position compared to the 4th and 9th position of ESM and SSL, respectively. After 4 feedback iterations, our approach achieves the highest accuracy and the best ranking quality compared to ESM and SSL, which in turn obtain an accuracy of 85% and 90%. It is interesting to noice that most of the errors are in favor of the "African people" and "building" class, probably due to the presence of color combinations or edge distributions, that resemble the ones of the typical bus images.

5.4.5 Quantitative results

We conduct in this section some quantitative experiments to show the effectiveness of our approach.

In the first place we do not involve the user in the retrieval loop and, therefore, we simply measure the quality of the initial retrieved set of images. To the sake of comparison, we

(a) (b) (c)

Fig. 5.7 Retrieval results on ImgDb1 obtained by our approach with the query image in the top-left corner. (a) Initial result (80% accuracy) (b) 1st iteration (90%) (c) 4th iteration (100%)

(a) (b) (c)

Fig. 5.8 Retrieval results on ImgDb1 obtained by ESM with the query image in the top-left corner. (a) Initial result (80% accuracy) (b) 1st iteration (80%) (c) 4th iteration (85%)

(a) (b) (c)

Fig. 5.9 Retrieval results on ImgDb1 obtained by SSL with the query image in the top-left corner. (a) Initial result (80% accuracy) (b) 1st iteration (85%) (c) 4th iteration (90%)

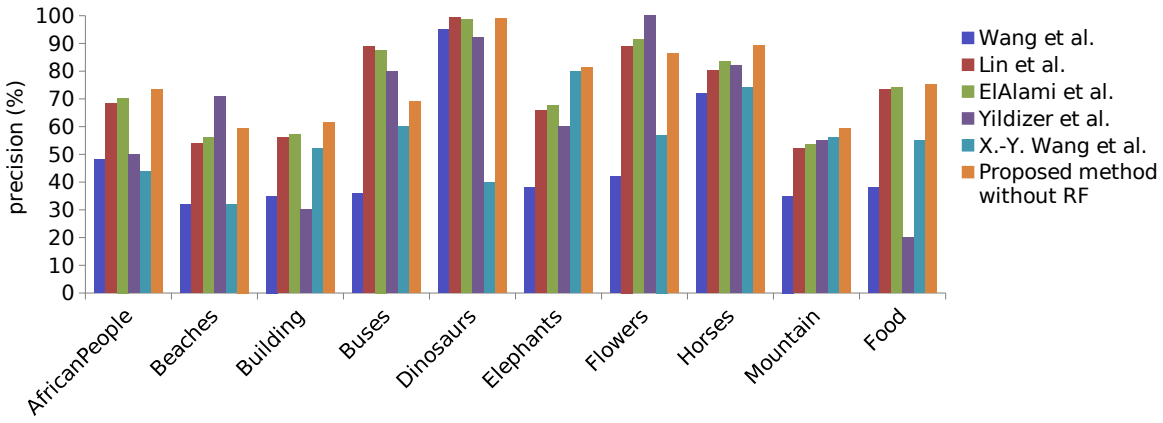


Fig. 5.10 Comparisons with other existing CBIR systems in terms of average precision on ImgDb1

evaluate our approach on ImgDb1 against the following approaches: [168], [116], [283], [284] and [285]. Note that we limit the analysis under this setting to ImgDb1 because no score was reported for the other two datasets in the competitors' works. We report the average precision obtained on 10 random queries per class (in total 80 queries) with a scope size of 20 images. As we can see in Fig.5.10, our approach gives an overall satisfactory performance compared to the other state-of-the-art approaches. We outperform the competitors in $6/10$ classes and obtained comparable results on the "Dinosaurs" class. As for the remaining classes, we never fall below 50% and we exhibit on average a good performance.

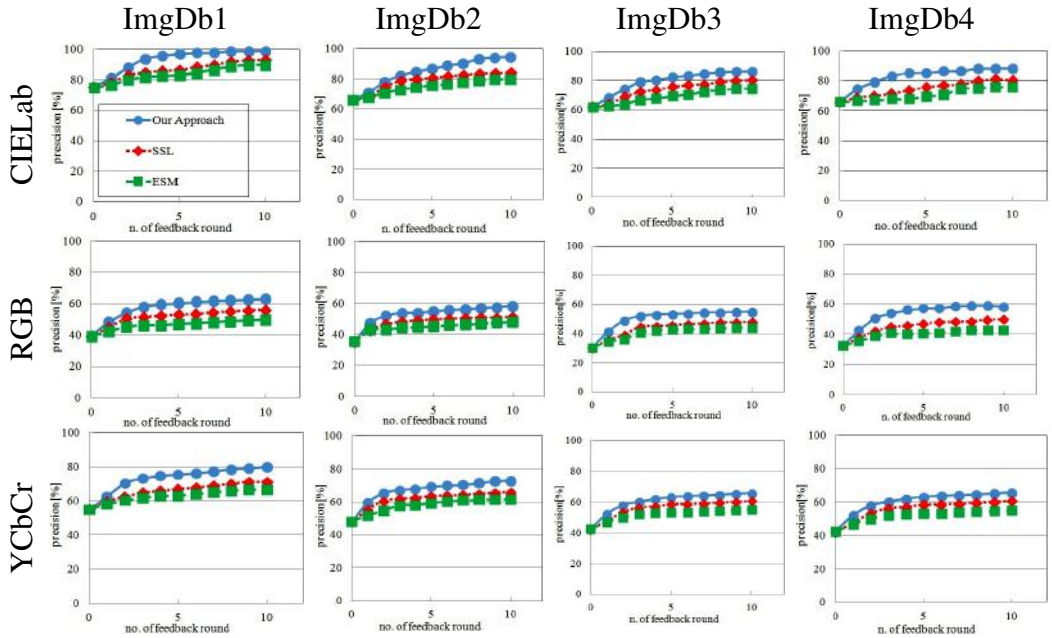


Fig. 5.11 Plots of the average precision of different CBIR with RF approaches using CIELab/RGB/YCbCr color spaces and scope size 20 on datasets ImgDb1, ImgDb2, ImgDb3 and ImgDb4.

We conduct a second experiment on all datasets involving also the relevance feedback. We compared our approach against the ESM and SSL relevance feedback methods mentioned in the previous section, considering a scope size of 20 images. In Figures 5.11, we report the results obtained by our approach, ESM and SSL in terms of average precision with respect to 10 random queries per image category. Moreover, we compare the

performance of the three approaches by varying also the starting color spaces, namely CIELab, RGB and YCbCr. All approaches start from the same baseline in terms of the initial query result, which is obtained from the NSCT-based representation proposed in this work. This gives indeed a good starting point when we consider the CIELab color space with an average precision above 60%, whereas the starting performance is lower with RGB and YCbCr. This gap in performance between the color spaces is consistent with the preliminary experiment that we conducted in Subsection 5.4.2. Our relevance feedback mechanism outperforms the competitors on all the experiments conducted. In general, the results on small datasets involving less categories like ImgDb1 are better in absolute terms that the one obtained on larger ones like, *e.g.*, ImgDb3. Nevertheless, the improvements deriving from the user's feedback in our method are around 25% for all datasets when we consider the CIELab color spaces. The performances are get worse when we move from CIELab to YCbCr and RGB. As for the competitors, we have that SSL outperforms ESM, thus highlighting the superiority of graph-based ranking mechanisms over the simpler weight-updating schemes.

5.4.6 Running Time

The experiments have been conducted with Matlab 2012b on a Windows 7 machine equipped with an Intel $i7-2600$ CPU 3.40 GHz and 4Gb of RAM. In this section we report the query time, which is an important component since the relevance feedback algorithm will be run several times between user interactions.

In Fig. 5.12, we report the average execution time in seconds of each approach on the different datasets with the NSCT-based features as a function of each relevance feedback round. Our approach is slightly slower than ESM, but faster than SSL. It is worth mentioning, however, that our current implementation employs a rather dense graph representation for the image datasets. In [163], it is indeed shown that sparsifying the graph allows to

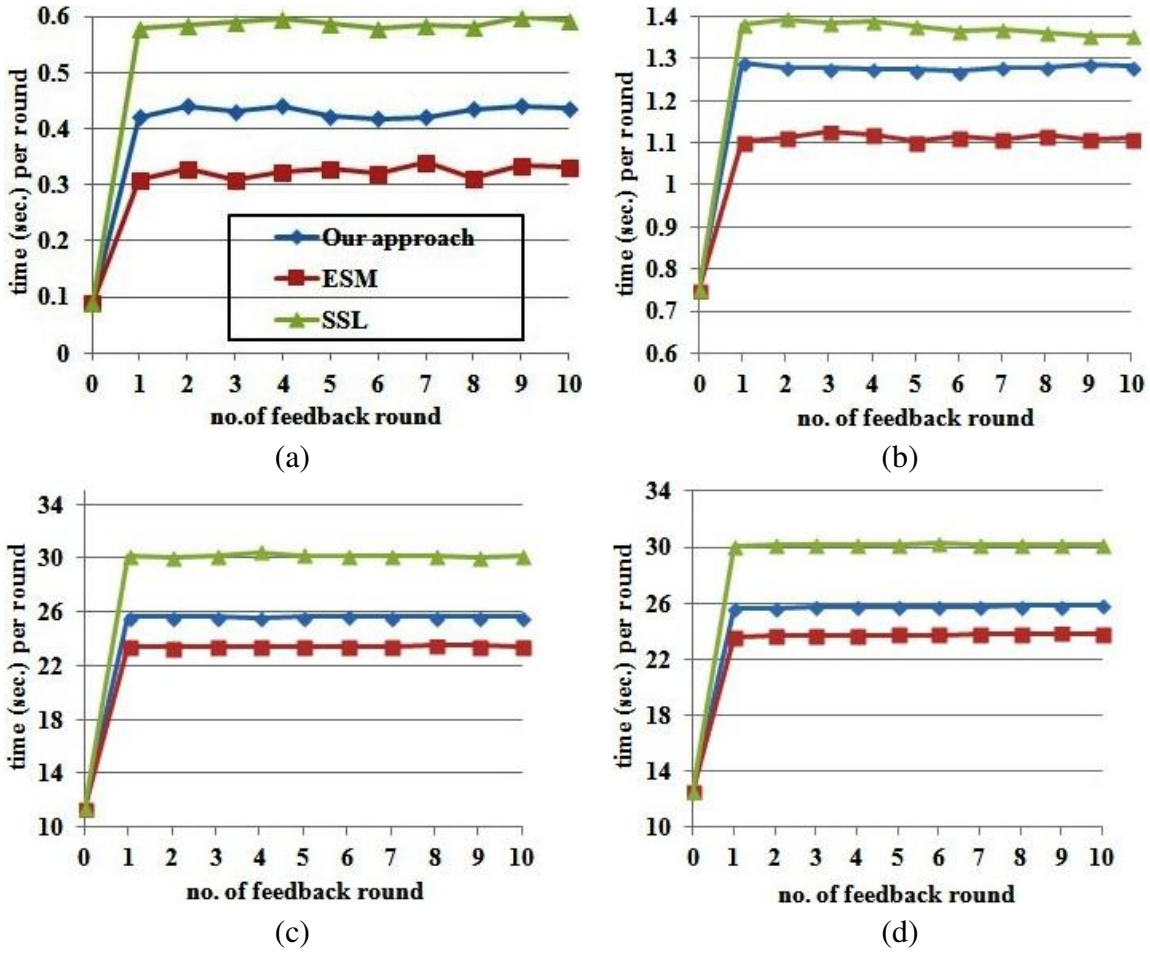


Fig. 5.12 Plots of the average running time in seconds at the different feedback rounds on dataset ImgDb1, ImgDb2, ImgDb3, and ImgDb4, using CIELab color space.

have a performance boost both in terms of execution time and accuracy.

In this work, we presented an efficient interactive CBIR system with relevance feedback, which is based on NSCT features that allow to capture geometrical information about the images. To enhance the retrieval accuracy, we performed feature selection using an unsupervised algorithm. The new reduced feature abstractions are used to retrieve similar images from the dataset, given a query image. In addition, we employed a random-walk-based method to incorporate a relevance feedback mechanism and thus better fit the retrieval results to the user's needs.

Experimental evaluations of the proposed framework have shown that our approach is effective and improves the retrieval performance of CBIR systems significantly compared to other state-of-the-art methods. In more details, the adoption of a feature representation, which derives from a flexible multi-scale, multi-directional and shift-invariant image decomposition method like NSCT, endows our system with the ability of capturing similarities within images in either shape or color cues also in the presence of changes in scale, rotation and translation. The experiment on CBIR without user intervention that we have reported in Fig.5.10 showed how this feature abstraction choice allows to achieve good performances on classes of images like,*e.g.*, horses and elephants, where both color and shape cues play an important role. At the same time, it also showed encouraging discriminant capabilities on classes of images that exhibit a larger variability in either cues like, *e.g.*, buildings and food. Part of the efficacy of the method is also due to the feature selection step, which in our case was based on MICI. As one can read from the experiment we conducted in Fig.5.2 on a supervised classification task, this step contributes by reducing the dimensionality of the feature representation in a way to keep the positive signal and remove redundancies that might negatively affect the subsequent retrieval process in terms of quality of the result and execution time. Finally, a graph-based ranking system has been proposed to include the user in the retrieval loop in order to reduce the semantic gap.

The quantitative experiments on CBIR with relevance feedback that we have conducted on three different datasets, and reported in Fig.5.11, have shown that the proposed ranking mechanism effectively exploits the user's feedback to improve the quality of the retrieved set of images on all settings we considered.

We have discussed in detailed about a CBIR system using different approaches in the current and the previous chapter. The performance of those proposed systems was tested with different benchmarking databases. It will be desirable to show how the proposed systems perform on real life dataset. In this regard, we are going to describe two different CBMIR systems for X-ray and endoscopic images respectively as practical implementation. The details of this work has been published in [179].

∗∗**Major portion of this Chapter taken from the following publication.**

- M. K. Kundu, M. Chowdhury and S. Rota Bul, "A graph-based relevance feedback mechanism in content-based image retrieval", Knowledge Based System (Elsevier), Vol. 73, pp 254–264, 2015 (**IF: 3.325**).

Chapter 6

CBIR System Design for Retrieval of Medical Images

6.1 Introduction

Over the last few decades, extensive attempts have been made by the researchers to develop Content Based Medical Image Retrieval (CBMIR) systems for efficient management and retrieval of medical image data. In general, digital medical images are stored in centralized networked servers [286–288]. In large hospitals and diagnostic centers, several terabytes of medical images are generated and stored every year for diagnosis of diseases, research and education etc., [14, 289]. In today's health-care framework, these images are archived in Picture Archiving and Communication Systems (PACS) [290, 291]. For precise diagnosis and treatment planning, often medical personnel have to browse through similar content images in these archives. This procedure necessitates the development of novel and intelligent techniques for searching images accurately and efficiently from a very large medical image database [292–294].

Generally, in a large PACS image database, images can be accessed using alphanumeric description and textual meta information, provided in the header of each image data. Similarly, the Digital Imaging and Communication in Medicine (DICOM) header information is also used for textual search [291, 292, 295]. For each DICOM image, header text information is generated manually, which comprises different parameters, modality and information regarding diagnoses. So it is obvious that the accuracy of DICOM system will not be very good due to the involvement of manual annotations which suffers from human fatigue [295].

This may create a large amount of error due to the individual human perception. So, final perceptual description will have a semantic gap between actual and perceived information and textual description. The system itself can measure the similarity between the images by extracting visual information using image vector generated from perceptual image information [291].

In this chapter, attempts have been made to design robust and efficient CBMIR systems

on both X-ray and color endoscopic image databases respectively. In "Scheme-I", we have discussed a radiographic image retrieval system whereas "Scheme-II" is dedicated to endoscopy color image retrieval. In this chapter, we have used NSCT based feature extraction mechanism in both Scheme-I and II along with this, PCNN based segmentation technique has also been used for extracting the shape features in "Scheme-I". Moreover, we have used relevance feedback mechanism for "Scheme-I" and search space reduction mechanism for "Scheme-II".

6.2 Notable Research Contribution

The major contribution of this chapter is as follows:

- From the extensive study of Chapter 4, it was found that NSCT is the best among multiscale transform used for extracting image features which is having good degree of translational, shift and rotational invariance. Accordingly, we have applied NSCT as a basic transform for extracting image features in both Scheme-I and II of this Chapter. In the Scheme-I of Chapter 6, we have used a model for CBIR system which is very similar to that of Scheme-III of Chapter 3 with following exception.

  (a) We have used here weighted voting scheme for approximate pre-classification of the query image into number of possible classes in order to reduce the error of the conventional classifier (classification accuracy is much much less than ideal 100%) such as LSSVM, K-NN etc.

  (b)In Fuzzy Relevance Feedback (FRF) model, index of fuzziness in place of fuzzy entropy is used as estimation for FRF.

- Scheme-II of this Chapter has been the extension of Scheme-II of Chapter 3. In Scheme-II of Chapter 3, ripplet transform has been applied on whole image and GGD based secondary features has been computed. On the contrary, in the Scheme-II of

this Chapter, ripplet transform has been applied on whole image and the transform coefficient is further divided into number of blocks of size $w \times w$. The GGD for each block is computed separately which are used for secondary features representation of endoscopic images. This partitioning of transform block could able to capture some positional information based on the block position along with global information from GGD. This new combination gives improved results compared to our earlier model proposed in Scheme-II of Chapter 3.

The details are explain in the subsequent section.

6.3 Scheme I: Interactive Radiographic Image Retrieval System [180]

Several CBMIR system have been reported in the literature that uses the different X-ray databases such as spine radiographs [296], MedGIFT [297], Image Retrieval in Medical Applications (IRMA) [290], etc. In medical CBIR system having multiple classes in the database, there exist a problem of correct classification or categorization of individual candidate precisely if there exist a problem that if similarity between a candidate within a class is comparable to that of some candidates of other classes [175, 291–293]. In "Scheme-I", we have used X-ray images of IRMA database around 15000 images which are categorizes into 193 classes. It is generally found in the X-ray images like IRMA that there exit a large variability within a particular class which is of the same order as that of across the classes. This makes the task of finding the similarity based on similarity metric between the query images and the image of the same classes in the database bit difficult. This often results in wrong selection of image against a particular query.

Many CBMIR systems have been proposed [291, 298–301] that use MLP [302], SVM [301], RF [303], etc., for coarse classification of category so that overall cost of computation could

be reduced without sacrificing accuracy. But due to the use of relatively expensive classifiers, overall cost of computation remains fairly high. Moreover, there is no classifier exist whose accuracy is absolute, *i.e.*, 100%(ideally). So this accuracy gap of any classifier which may cause a greater degree of error in subsequent retrieval process could be reduced using suitable uncertainty handling mechanism. One such scheme used in the proposed work is known as Similarity Positional Score (SPS), which is described in "Scheme-I".

In "Scheme-I", we have proposed a two-stage method for image retrieval and two different kinds of image features are used for image representation in the two different stages. Similar procedure is also followed for the query image. In the proposed method, each of the images in the database is represented in two different representation forms (1) Using compact PCNN based global shape (PCNN-shape) feature set and (2) compact NSCT based local texture (NSCT-texture) feature set. The compaction of feature set is done class wise in the IRMA image database using MICI. In the first stage of retrieval, PCNN-shape features are used for coarsely classifying the input image to one or more existing classes in the IRMA database depending upon its visual proximity as measured by similarity positional score (inversely proportional to distance metric). Based on this score, top probable classes similar to the class of the query image are identified and also restricting the final search only within the apparent classes which comprises a very small portion of the whole IRMA image database (partitioned database). The Basic aim of this stage is to reduce the misclassification error of any classifier which classifies the data into a particular class instead of probable classes. For the finer retrieval (*i.e.*, in the second stage or in final search), compact NSCT-texture features have been used for retrieving similar images from the partitioned database, thus reducing the overall search space which results in a faster and more accurate retrieval. In addition, we have developed a novel *fuzzy index* based relevance feedback mechanism (*fuzzy-index*-RFM) to handle the uncertainty and ambiguity arising out of user feedback which is highly individualistic in nature. Extensive experiments show

that the proposed system is efficient and effective for radiographic image retrieval. The proposed mechanism of "Scheme-I" is describe in the followings section.

6.3.1 Feature Extraction for Image Representation

The performance and efficiency of CBMIR system is largely dependent on how effectively the image is represented in the database in terms of feature vector. There are certain images which can be easily distinguishable by the global feature characteristics such as shape, geometric property, orientation etc., whereas there are some other images which can be distinguishable by local feature characteristics such as color, texture, positional distribution etc. But judicious combination of both kind of features can help in achieving high accuracy at lower computation cost [304]. Feature dimension is also an important issue for achieving greater computational efficiency. Lower the dimension of feature vector use, higher will be computational efficiency [14, 293]. Furthermore, total computational cost depends upon number of comparison to be made for matching which is proportional to total number of images in the database consider for this purpose. Keep in mind above mention problem, a two stage algorithm is used in the proposed technique.

In this work, two different feature representation schemes namely PCNN-shape and NSCT-texture are used in two stages respectively. For both, the input images are gray level X-ray images taken from the IRMA image database. In the proposed scheme, the first stage is coarse classification where a PCNN based segmentation scheme is used for extracting the shape features. Due to its adaptive thresholding nature, PCNN serves as an efficient tool for edge sharpening as well as for ambient noise removal in the mega voltage X-ray images [305]. But before doing segmentation, there is a need of pre-processing the image due to intensity variations, low contrast and high rate of noise in X-ray images [302]. Thus, in pre-processing, we should enhance the quality of X-ray images by a combination of methods for reducing noise and improving the contrast. The median filter for elimina-

tion of noise in images is a very suitable method. It is also a very well known denoising and smoothing operator [306]. Hence, at first, median filter is used for noise reduction. After noise reduction, adaptive histogram equalization is used for contrast enhancement, considering the exponential distribution of histogram. But it is important to note that using adaptive histogram equalization provides improvement of the image contrast with no destructive effects on the areas with high contrast, but it increases the noise in the image. As a result, after adaptive histogram equalization, a median filter is used again on the X-ray images to further reduce the noise [302].

After pre-processing, segmentation is carried out using PCNN [307] and 10 global shape features are computed in each iteration. These features are perimeter, area, euler-number, orientation, eccentricity, major-axis-length, minor-axis-length, compactness and two zernike moments (*i.e.*, amplitude and phase). Therefore, considering 10 iterations, each image is represented with 100 PCNN-shape features:

$$FM^{I}_{PCNN-shape} = \left(f^{1}_{\text{PCNN-shape}}, \cdots, f^{100}_{\text{PCNN-shape}} \right). \tag{6.1}$$

In the second stage of the algorithm for retrieval of similar images, NSCT-texture features are used for finer retrieval. Based on this search, a group of 20 closest similar images are retrieved. In the literature, it is reported that using several different transforms like the WT, CVT, CNT etc., the transform domain feature are used for retrieval of medical images [14, 252, 293, 308]. But those techniques lack shift-invariance and anisotropy properties. To overcome these problems, we adopt NSCT, which is a fully shift-invariant, multiscale, and multidirection expansion with fast implementability and anisotropy as well as has better directional decomposition and frequency localization capability [206]. NSCT is constructed by combining the NSPFB and NSDFB, and the resulting filtering structure approximates the ideal partition of the frequency plane. NSCT coefficients are capable of capturing the fine details present in the image, which is very much essential in im-

age retrieval, as radiographic images are full of subtle image information. Moreover, it has been reported that HVS has characteristics like multiscale, rotational invariance and anisotropy [208]. This behavior is well matched by NSCT decomposition. Based on the above facts, NSCT is used in the proposed CBMIR system as the feature representative scheme for finer retrieval. Detailed information about PCNN and NSCT is available in [307] and [206] respectively.

NCST is used for image representation following the procedure describe below:

Algorithm 1 Algorithmic Steps to Compute NSCT based features

1: Images (I) of size $N \times N$ are pre-processed using enhancement mechanism
2: After pre-processing, images are decomposed into l number of high and low frequency sub-bands of size $n \times n$.
3: Each sub-band is sub-divided into N number of blocks of size $w \times w$, where $N = n^2/w^2$.
4: Standard deviation of each block is used to represent the coefficients of that block.
5: The Low Frequency Sub-band (LFS) that contains the DC information is ignored due to its relatively small contribution in terms of standard deviation compare to that of AC sub-bands. So the effective number of subbands consider is $2^l - 1$

In the present case, total dimension of NSCT-texture feature vectors is $2^l - 1 \times \frac{n^2}{w^2} = 448$, where $w = 16$, $n = 128$ and $l = 3$:

$$FM^I_{NSCT-texture} = \left(f^1_{\text{NSCT-texture}}, \cdots, f^{448}_{\text{NSCT-texture}}\right). \tag{6.2}$$

After feature computation, in "Scheme-I", we retain the relevant PCNN-shape and NSCT-texture based features by adopting a feature selection algorithm based on the MICI (see Section 2.5.1 of Chapter 2 for further details). Finally, we denote by $FMR^I_{PCNN-shape}$ and $FMR^I_{NSCT-texture}$ the reduced PCNN-shape and NSCT-texture feature vector associated to image I, obtained after the application of the aforementioned MICI algorithm.

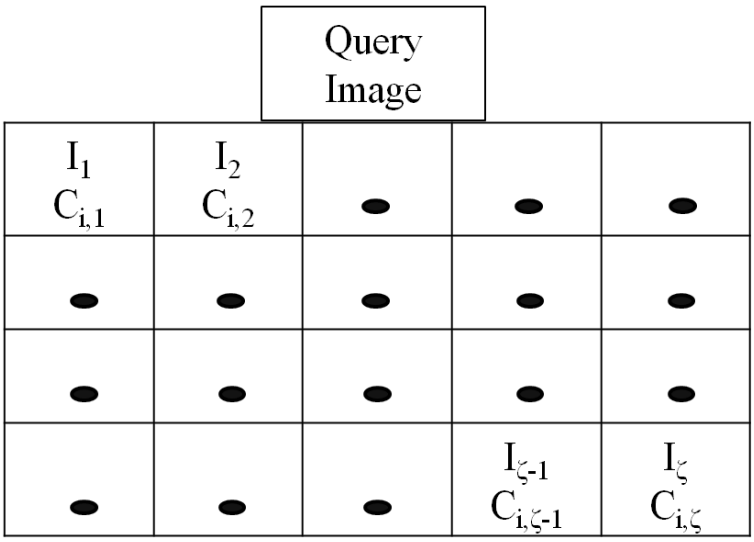


Fig. 6.1 Representation of Retrieved Images with $\zeta = 20$ with class level information of the Retrieved Images in ζ.

6.3.2 Computation of Possible Class Label(s) of Query Image

The basic aim of this operation is to identify a few approximate classes to which query image may belong. This facilitates the searching over a small number of images belong to those classes for ultimate similarity estimate. In the present work, all the images in the database are leveled and a prior assigned to the pre-number of classes. In the first phase, a query image is presented to the system. Based on the PCNN-shape features, most similar images are retrieved according to Euclidean distance measure with respect to query image. It is obvious that retrieved images on the top of the scope size (ζ) are most similar to the query images. Therefore, more score is given to the top retrieved image and least score is given to the last image in ζ. Out of total ζ number of retrieved images, in the present work, we have considered only first 20 (here $\zeta = 20$, but can be selected as per convenience) images for the purpose of assigning a few number of possible class level (depends upon sample from different classes in the first 20 retrieved images) to the query image. This possible class information about the query image will be used to restrict the search space for the finer retrieval spanning over a small portion of the database comprising

of few number of classes.

Considering the scope size ($\zeta =$) 20, Fig.6.1 shows the retrieved image representation where I_1 is most similar and I_ζ is least similar to the query image. Possible class levels $(C_{i,j})$ of the retrieved images are also shown where $C_{i,j}$ represents that the positional image I_j belongs to the class i; where, $1 \leqslant i \leqslant$ Maximum number of classes and $1 \leqslant j \leqslant \zeta$. Using the class level and positional information of the retrieved images in ζ, similarity positional scores are computed. $C_{i,j}$ depends on from which class of the partition database this particular image belongs, same number of the class may occur at the different position of ζ and also the image belong to different class may occur once or in few number in the ζ.

Similarity positional score (s_i) for each class level (i) is computed based on the position of the retrieved images and ζ. At the beginning, we assume that the importance of all different classes as in the database with respect to the retrieved images are same and initial s_i is zero. Suppose the retrieved images $I_{j_1}, I_{j_2}, ..., I_{j_m}$ of I belong to the same class i i.e the corresponding cells $C_{i,j_1}, C_{i,j_2}, ..., C_{i,j_m}$ of $\mathbf{C}$ contain the same class level i, where j_m represent the position of the retrieved images in ζ and m is the number of retrieved images in particular class. Then the s_i is computed as:

$$s_i = \frac{j_1 + j_2 + ... + j_m}{\zeta}; 1 \leqslant j_1, ..., j_m \leqslant \zeta \tag{6.3}$$

Let the number of distinct classes present in $\mathbf{C}$ be $i_1, i_2, ..., i_n$; where $1 \leqslant i_1, ..., i_n \leqslant$ Maximum number of classes. Then calculate $s_{i_1}, s_{i_2}, ..., s_{i_n}$ and represented as $A = [s_{i_1}, s_{i_2}, ..., s_{i_n}]$. Now sort the set A in decreasing order based on the value of $s_{i_1}, s_{i_2}, ..., s_{i_n}$ and obtain the set B as $B = [s_{i'_1}, s_{i'_2}, ..., s_{i'_n}]$ where $s_{i'_1} \geq s_{i'_2} \geq ... \geq s_{i'_n}$ and $\{i'_1, i'_2, ..., i'_n\}$ is a permutation of $\{i_1, i_2, ..., i_n\}$.

For fixed ζ (Here, $\zeta = 20$) threshold can be calculated in the following ways

$$\delta = 1 + \frac{\zeta - 1}{\zeta} + \frac{\zeta - 2}{\zeta} + ... + \frac{\zeta - \zeta'}{\zeta}; \text{where } \zeta' = \frac{\zeta}{2} + 1 \tag{6.4}$$

Now choose the first p classes $i'_1, i'_2, ..., i'_p$ from the set $\{i'_1, i'_2, ..., i'_n\}$ where $p \leq n$ and $s_{i'_1} \geq s_{i'_2} \geq ... \geq s_{i'_p}$ such that

$$s_{i'_1} + s_{i'_2} + ... + s_{i'_p} \geq \delta \tag{6.5}$$

6.3.3 Fuzzy Index Based Relevance Feedback Mechanism

Human perception of image similarity is subjective, semantic and task dependent. Hence, CBMIR system cannot deliver satisfactory result in general by merely relying on low-level features extracted in an unsupervised way. The user's response has to be incorporated in the retrieval loop as source of feedback. By doing so, the system can iteratively adopt to the query results to meet user-specific requirements. In formal CBMIR system, an image I is usually represented by an N dimensional feature vector: $I = \{f_q | \ q = 1, 2, 3, ..., N, \ f_q \in \mathbb{R}\}$, and represents the q^{th} feature component. The commonly used decision function for measuring similarity between query image (I_{qr}) and the database image I, is represented as

$$D(I, I_{qr}) = \sum_{q=1}^{N} w_q \| f_q(I) - f_q(I_{qr}) \| \tag{6.6}$$

where $\| f_q(I) - f_q(I_{qr}) \|$ is the distance between q^{th} feature component and w_q is the weight assigned to it. The weight should be adjusted such that, the feature have small variation over the relevant images and large variation over the irrelevant images. Rahman *et al.*, proposed a new relevance feedback mechanism where the query parameters is updated dynamically and adjust the proposed matching functions [309]. A weight-updating mechanism have been developed by using the feedback history information [310]. In [301], weight are updated at each iterations by considering both the precision and the rank order

information of the relevant images in the individual result lists based on the user information. As, human perception and responses are subjective as well as ambiguous in nature. Therefore, this subjectivity and ambiguity has to be properly handled for designing an efficient *fuzzy-index*-RFM. Fuzzy set theoretic [235] models try to mimic human reasoning and the capability of handling uncertainty arising from ill-defined, incomplete, defective and imprecise input. Fuzzy logic based RFM is proved to be very effective in CBMIR based applications [311]. In light of this, we have developed a novel relevance feedback mechanism based on index of fuzziness (*fuzzy-index*-RFM).

Assuming, k similar images $I_s = \{I_1, I_2, \ldots, I_k\}$ are returned to the user by the proposed system. From these returned images, the user marks the relevant and irrelevant images. Let, $I_c = \{I_r, I_{ir}, I_s\}$, where I_r = relevant images, I_{ir} = irrelevant images, I_s = system's returned images: $I_r \cup I_{ir} = I_s$ and $I_r \cap I_{ir} = \varnothing$. The information from I_r, I_{ir} and I_s are combined to compute the relative importance of the individual features by fuzzy Feature Evaluation Index (FEI).

A fuzziness measure of a fuzzy set express the average amount of ambiguity in making a decision whether an element belongs to the set or not [235]. One such measure, the index of fuzziness denotes the degree of ambiguity or fuzziness present in a set by measuring the distance between the membership values of the fuzzy elements of the fuzzy set A and its nearest ordinary set $\tilde{A}$. The index of fuzziness is defined as

$$\begin{aligned} \nu_p(A) &= (\frac{2}{n^{\frac{1}{p}}}) d^p(A, \tilde{A}) \\ &= (\frac{2}{n^{\frac{1}{p}}}) [\sum_{i=1}^{n} \{min(\mu_A(x_i), 1 - \mu_A(x_i))\}^p]^{1/p} \end{aligned} \tag{6.7}$$

where $d(A, \tilde{A})$ denotes the distance between the fuzzy set A and its nearest ordinary set $\tilde{A}$ and n is the number of elements in the fuzzy set. An ordinary set $\tilde{A}$ nearest to the fuzzy set

A is defined as

$$\begin{aligned} \mu_{\tilde{A}}(x_i) &= 0 \text{ if } \mu_A(x_i) \leq 0.5 \\ &= 1 \text{ if } \mu_A(x_i) > 0.5 \end{aligned} \tag{6.8}$$

The value of p depends on the type of distance function used. For example, $p = 1$ for generalized Hamming distances whereas $p = 2$ for Euclidean distance. The corresponding indices of fuzziness are called the "linear index of fuzziness" $\nu_{L1}(A)$ and "quadratic index of fuzziness" $\nu_Q(A)$, respectively. In this proposed CBMIR system we have used the "quadratic index of fuzziness" for finding out the relative importance w_q of the individual feature f_q. The $\mu_A(x_i)$ is computed using a standard S-type membership function [235]:

In our case, A represents the q^{th} column of $I_c = \{I_r, I_{ir}, I_s\}$ with x_i denoting the corresponding feature value and $\mu(x_i)$ as the corresponding fuzzy membership value.

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

$$FEI_q = \frac{\nu_Q(f_q^{I_s})}{\nu_Q(f_q^{I_r}) + \nu_Q(f_q^{I_{ir}})} \tag{6.9}$$

Lower value of FEI_q, indicates better quality of importance of the qth feature in recognizing and discriminating different classes. The precision of retrieval can be improved with these values. The weight w_q is computed as a function of the evaluated (FEI_q) as shown below:

$$w_q = FEI_q \tag{6.10}$$

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

6.3.4 Summary of Proposed CBMIR System

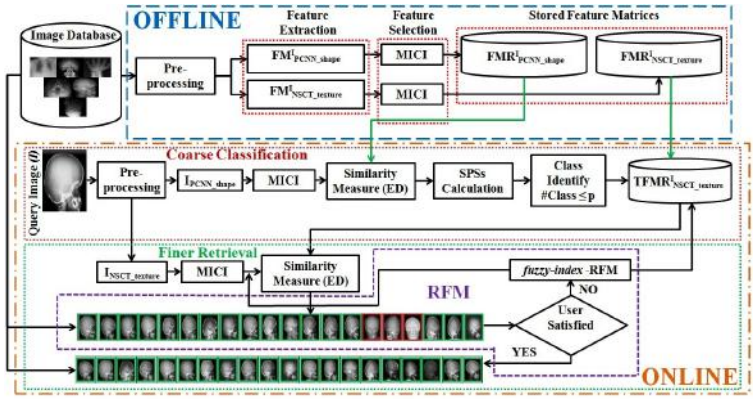


Fig. 6.2 Overview of the proposed CBMIR system.

The proposed CBMIR system consists of two main phases as shown in the block diagram of Fig. 6.2: an "OFFLINE" phase and an "ONLINE" phase. In the "OFFLINE" phase two different features ($FMI^{I}_{PCNN-shape}$ and $FMI^{I}_{NSCT-texture}$) are extracted from each image in the stored image database after pre-processing. The dimension of the extracted feature vectors are reduced by MICI feature selection scheme. The reduced feature vectors are then stored in two different feature database $FMR^{I}_{PCNN-shape}$ and $FMI^{I}_{NSCT-texture}$ along with the class levels of the images. The "ONLINE" phase of the proposed system consists of two steps described below.

Coarse Classification:

1. Input the query image I.
2. Compute the Euclidean distances between compact PCNN-shape feature vector of query image ($I_{PCNN-shape}$) and each reduced feature vector of $FMR^{I}_{PCNN-shape}$.
3. Sort the distances in ascending order and display the top k images from the database.
4. The class level information of the top k images corresponding to the first k sorted distances are retrieved.

5. Similarity positional score is computed for the top k images and sorted in ascending order.

6. Choose the first p classes which satisfy Eq. (6.5).

Finer Retrieval:

1. Construct a temporary feature database $TFMR^{I}_{NSCT-texture}$ from the $FMR^{I}_{NSCT-texture}$ database consisting only the feature vectors corresponding to the images belonging to the top p identified image classes.

2. Compute the Euclidean distance between compact NSCT based query feature vector $I_{NSCT-texture}$ and each reduced feature vector of $TFMR^{I}_{NSCT-texture}$.

3. Sort the distances in ascending order and display the top k images according to the sorted distances from the selected image classes database.

4. If the user is satisfied with the displayed result, then stop.

5. Otherwise, take the user's feedback response by marking the relevant images from the irrelevant images in the displayed images.

6. Compute the fuzzy feature evaluation index based upon the user's response (described in section 6.3.3).

7. Go to step 2, taking into consideration the ranked fuzzy features to recompute the Euclidean distance.

6.3.5 Experimental Results and Discussion

Extensive experiments were carried out to evaluate the performance of the proposed CBMIR system. We conducted several different experiments to evaluate the effectiveness of the proposed system both quantitatively and qualitatively.

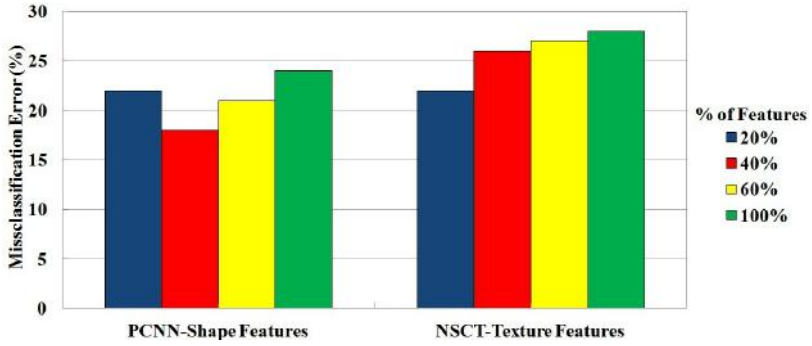


Fig. 6.3 Results obtained on IRMA datasets with a LS-SVM trained considering different share of PCNN-Shape Features and NSCT-Texture Features.

6.3.5.1 Experimental Setup

The experiments were carried out on a Dell Precision T7400 PC with 4GB RAM and were implemented using MATLAB R2014b. All the experiments were performed on the diverse medical image collection available from the IRMA dataset [290, 295]. The X-ray images are noisy with irregular brightness and contrast, and sometimes contain dominant visual artifacts such as artificial limbs and X-ray frame borders. This database contains $14,410$ images divided into 193 distinct categories, where the number of images in each category varies largely (from 1 to 2314 images). In the experiments, we only took the categories which contain atleast 50 images. This resulted in a subset of the database containing 10902 images of 57 categories. The images in this database are organized into semantic categories as could be selected by a human observer. The decomposition parameter of NSCT was set to $levels = [1,2,4,4]$, and we used "*pyrexc*" and "*sinc*" as the pyramid filter and orientation filter, respectively. The performance on the database was evaluated by considering each image as the query image and by measuring the classification accuracy as well as AP and AR. During the experiments, top 20 retrieved images were used to compute the precision and recall values.

6.3.5.2 Effectiveness: Feature Representation and Selection

The first set of experiments we conducted were to evaluate the effectiveness of the feature representation technique used in the proposed CBMIR system. For these experiments, we only considered the results obtained through the proposed CBMIR system without *fuzzy-index*-RFM.

At first, we tune for the IRMA dataset, the number of PCNN-shape and NSCT-texture features that should be extracted by the algorithm presented in Section 6.3.1 and furthermore top features are selected by Maximal Information Compression Index (MICI) algorithm. The idea is to select the number of features that allows for the best category discrimination in both shape and texture. To this end, we train a multi-class LS-SVM classifier using a one-versus-one strategy with a 10-fold cross-validation on the parametrization. A share of 60% labeled samples from the datasets forms the training set, while the remaining 40% is reserved for the test set. We evaluate the performance of the classifier using different shares of selected features (20%, 40%, 60% and 100%) in terms of misclassification error on the IRMA dataset. According to the results reported in Fig. 6.3, we can clearly see that the feature selection step has an impact on the performance. The best results are achieved with a share of 40% PCNN-shape and 20% NSCT-texture feature on IRMA dataset. Hence, the feature selection algorithm will be tuned for the subsequent experiments in a way to return the optimal shares of PCNN-shape and NSCT-texture features that we found for IRMA dataset.

Moreover, we have also evaluated the performance of NSCT with other different transforms such as WT, CVT and CNT. The performance comparisons of WT, CVT, CNT and NSCT are shown in Fig. 6.4 (WT-texture, CVT-texture, CNT-texture and NSCT-texture, respectively). Here, only the subset of the IRMA image database were considered for performance evaluation. It is clear from the Fig. 6.4, that NSCT-texture performs much better than WT-texture, CVT-texture and CNT-texture in CBMIR domain.

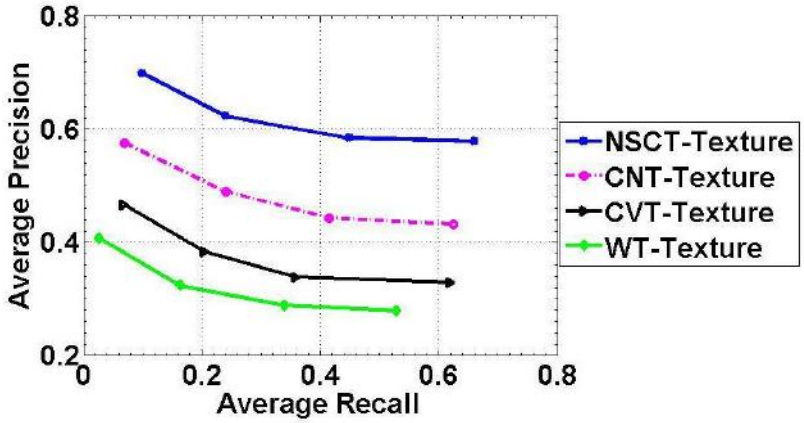


Fig. 6.4 Performance comparison in terms of AV and AR curves.

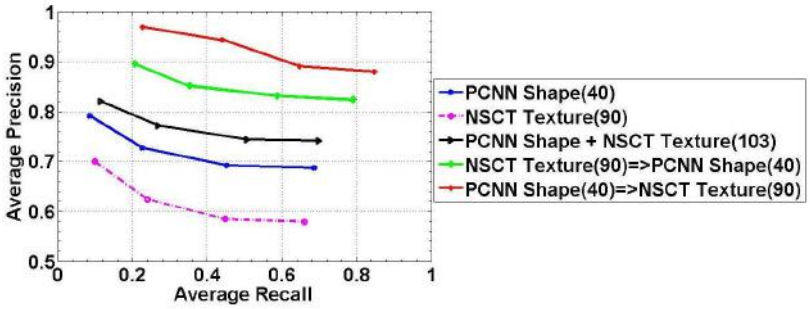


Fig. 6.5 Average Precision vs. Average Recall.

The graph of Fig. 6.5 shows the performance of compact PCNN-shape and NSCT-texture feature vectors with various combinations, considering various numbers of images per frame (5, 10, 15 and 20 images/frame). The graph shows the performance of the CBMIR system using compact PCNN-shape and NSCT-texture features, combination of compact PCNN-shape + NSCT-texture features, compact NSCT-texture followed by compact PCNN-shape and compact PCNN-shape followed by compact NSCT-texture, respectively. Here, "⇒" means "followed by", PCNN-shape ⇒ NSCT-texture means PCNN-shape followed by NSCT-texture. Compact NSCT performs worse than compact PCNN-shape features whereas the combination of compact PCNN-shape + NSCT-texture performs better than only PCNN-shape or NSCT-texture based feature representation schemes. Therefore, the proposed feature representation scheme (PCNN-shape followed by NSCT-texture) pro-

Different Methods	With All Feature Vectors		With Compact Feature Vectors	
	AP(%)	Time(sec)	AP(%)	Time(sec)
PCNN-shape	84.69	1.112	83.86	0.571
NSCT-texture	75.12	2.915	74.45	1.146
PCNN-shape ⇒ NSCT-texture	93.12	3.142	92.93	1.289

Table 6.1 Comparisons with different methods in terms of time and AP

vides the best retrieval results. In other words, we can say that the feature vectors used in the proposed system are reasonably smaller than the existing techniques. The advantage of using a low dimensional feature vector is that it can increase the retrieval speed when large image databases are searched with simple similarity measures by compromising only a very little retrieval accuracy. Table 6.1 shows the average retrieval time using all and optimal features with Euclidean distance measure. From Table 6.1, it can be seen that the "PCNN-shape⇒ NSCT-texture" performs well in terms of classification accuracy while having a satisfactory on time complexity as compared to the other selected combinations.

The reason behind using using two-stage CBMIR system is that PCNN-shape considers the shape characteristics of the image, whereas, the block based NSCT-texture feature representation scheme considers the local texture information. For most of the radiographic images, there exists prominent distinguishable shape characteristics. But, the NSCT based subtle local texture information often fails to distinguish between two visually dissimilar images. Thus, empirically, the dimensions of PCNN-shape and NSCT-texture feature vectors are reduced from 100 and 448 to 40 and 90 respectively. Therefore, in short, using the compact set of PCNN-shape features, an initial matching of the input query image is done with that of similar images in the database in order to find out the coarse classes to which the query image may belong. This process will select one or a few classes (a small portion of the whole database) for finer matching in the second stage. And in the second stage, most similar images are retrieved using the compact NSCT-texture features from the partitioned database.

Moreover, we have also compared our proposed technique of classification *i.e.*, SPS

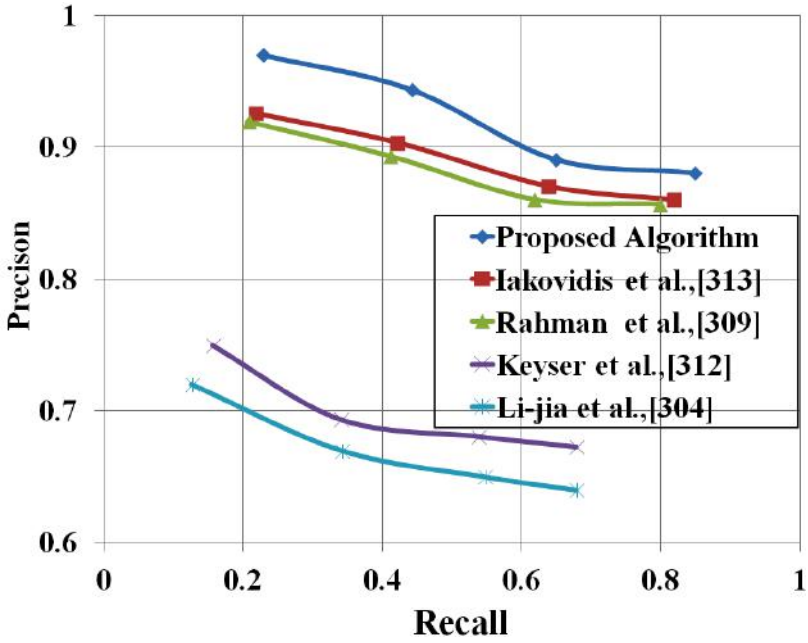


Fig. 6.6 Comparisons with other existing CBMIR systems in terms of AP Vs AR on the subset of IRMA database.

with k-nearest neighbor ($k-NN$) classification. Choice of k is very critical in $K-NN$ classifier. A small value of k means that noise will have a higher influence on the result and a large value make it computationally expensive (*i.e.*, that points that are near might have similar densities or classes). Performance of a nearest neighbor classifier depends on the distance function and the value of the neighborhood parameter k. Euclidean metric is the most popular choice for the distance function. Therefore, considering $k = 57$, euclidean metric as a distance function and the combination "PCNN-shape $\Rightarrow$ NSCT-texture" as a features, we have achieved the accuracy of 56% as compared with 92.93% with SPS.

To further justify our proposed two-stage model, we have compared our proposed two stage CBMIR system without involving the user in the retrieval loop and, therefore, we simply measure the quality of the initial retrieved set of images. For the sake of comparison, we evaluate our approach on the subset of the IRMA dataset against the following approaches: [304, 309, 312, 313]. We report the AP vs AR graph obtained on 10 random queries per class with a scope size of 20 images. As we can see in Fig.6.6, our approach

shows satisfactory overall performance compared to the other state-of-the-art approaches.

6.3.5.3 Effectiveness: Fuzzy Index Based RFM

To test the retrieval effectiveness of the *fuzzy-index*-RFM, the top 20 retrieved images were used to compute the average precision values. The graph of Fig. 6.7, shows the effectiveness of the proposed system in terms of average precision in different phases and number of iterations. We can clearly see the incremental improvement in the retrieval accuracy achieved by the proposed system over the different stages of the system. The proposed CBMIR system achieves near about 78% average precision after the coarse retrieval (using the compact PCNN-shape features). Based on the retrieval results provided by the coarse retrieval steps, the system works in a reduced search space (consisting of only the images of the top p classes returned by coarse retrieval steps). The AP increases to approximately 93% in the finer retrieval stage (using compact NSCT-texture features). If the user is not satisfied with the retrieval result, based on the feedback response of the user, the proposed *fuzzy-index*-RFM tries to improve the result further. From the graph of Fig. 6.7, we can see that the proposed *fuzzy-index*-RFM achieves the highest retrieval accuracy in only three or four iterations.

Fig. 6.8 and Fig. 6.9 shows the visual results (considering $\zeta = 20$) achieved by the different steps of the proposed CBMIR system for a particular query image (top image marked by black box). Row 2 shows the coarse classification results, when retrieval is done on whole database. Row 3 shows the results using limited classes, which has been selected using similarity position score. Based on the user feedback, final results are shown in the last row of the respective query image. We can see that all the retrieved images after the 3^{rd} iteration of the *fuzzy-index*-RFM are almost from the same class as that of the query image with improved ranking. But in Fig. 6.9, after using the proposed *fuzzy-index*-RFM, only one irrelevant image has been present in the top 20 retrieved images as shown in row

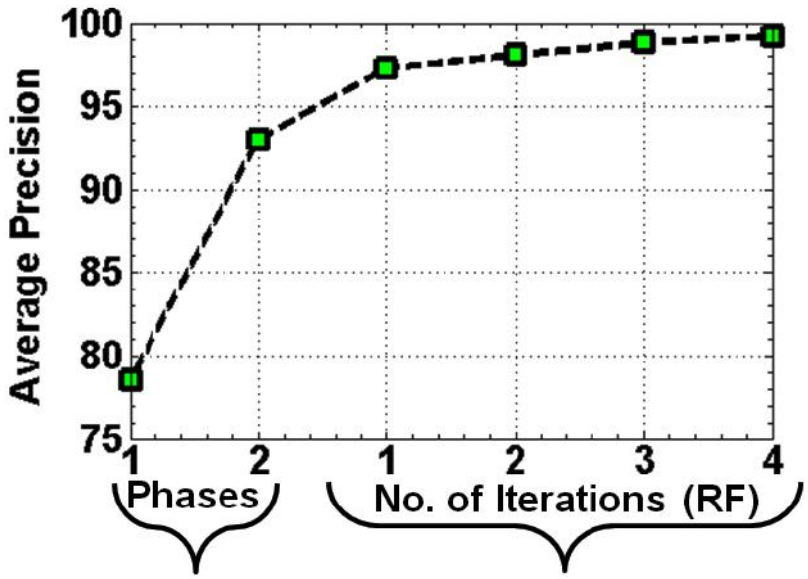


Fig. 6.7 Effectiveness of the proposed FRF mechanism.

4. This is because pathology information has not being considered in our proposed system. Hence, from both qualitative and quantitative results, we can say that our system achieves AP of 92.9325% without using *fuzzy-index*-RFM and it increases to 98.375% by using the proposed *fuzzy-index*-RFM on the subset of 10902 IRMA images from 57 classes.

In this scheme, we have proposed an efficient CBMIR system for radiographic images based on a novel approach of two-stage retrieval system. Main importance is given on constructing effective feature vector representing the image's content and search-space reduction, without compromising the retrieval performance. Furthermore, to handle the uncertainty and ambiguity of user response a novel *fuzzy-index*-RFM is incorporated in the proposed system. Experimental evaluations of the proposed framework have shown that our approach is effective and improves the retrieval performance of CBMIR systems significantly compared to other state-of-the-art methods.

To replace the human assistance mechanism, we have proposed a new reducing search space mechanism for retrieval of color endoscopic image retrieval in "Scheme-II".

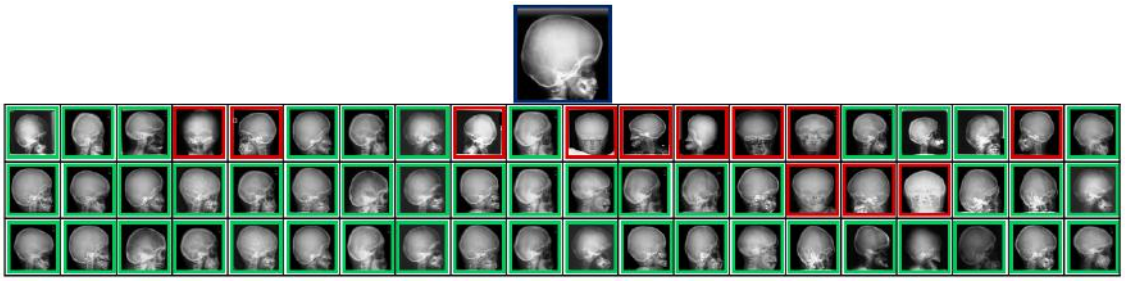


Fig. 6.8 Visual performance of the proposed *fuzzy-index*-RFM (black box → query image, green box → relevant image and red box → irrelevant image): Top image is the query image. Row 2 shows the Coarse Classification Results (11/20). Row 3 shows the Results when Selected Search from Limited Number of Classes(17/20, Ranking Improved) using similarity positional score. Row 4 shows the Refinement of Second Result with *fuzzy-index*-RFM after 3^{rd} Iteration (20/20, Ranking Improved).

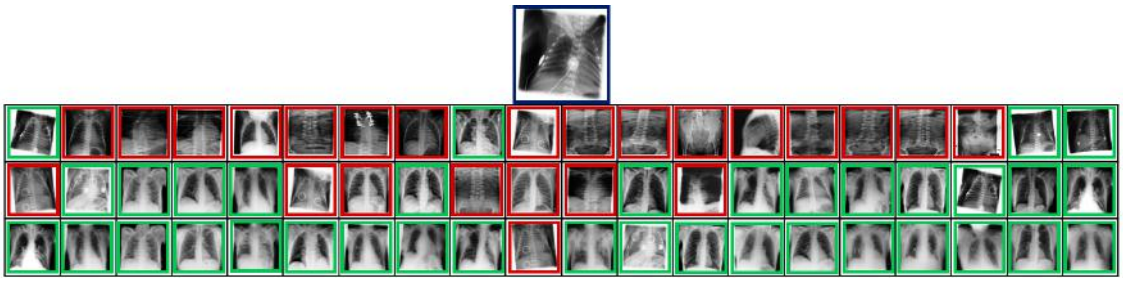


Fig. 6.9 Visual performance of the proposed *fuzzy-index*-RFM (black box → query image, green box → relevant image and red box → irrelevant image): Top image is the query image. Row 2 shows the Coarse Classification Results (4/20). Row 3 shows the Results when Selected Search from Limited Number of Classes(13/20, Ranking Improved) using similarity positional score. Row 4 shows the Refinement of Second Result with *fuzzy-index*-RFM after 3^{rd} Iteration (19/20, Ranking Improved).

6.4 Scheme II: Endoscopic Image Retrieval Using Multiscale Image Features [187]

The Clinical Outcomes Research Initiative (CORI), is a research proposal that acquires information that will improve the quality of clinical practice in gastroenterology. The data collected under this proposal has been fully annotated by a gastroenterologist. The image archive in the CORI can be used as a resource for prospective research on topics such as colon polyp surveillance or to develop research hypotheses in endoscopic images related

research [314].

Even though statistical modeling has been used extensively in CBIR domain as describe in Chapter 3 of Section 3.3, but the effectiveness of these approaches is not explored well in the field of natural color medical image retrieval. Moreover, in Chapter 3 of Section 3.3, GGD based statistical (secondary) features has been derived from RT coefficients (primary features). But from Chapter 4, it has been found that GGD features extracted from NSCT with pyrexc and sinc filter combination is best. Therefore, in Scheme-II, we propose a statistical model based on GGD and KLD over NSCT subbands coefficients to investigate its performance in color endoscopy image retrieval problem. But, in addition to color information, our approach extends the feature extraction process by using a window, generating more local features. After feature extraction, subset of optimal features is selected using PCA algorithm. Moreover, the computational complexity is further being reduced using LSSVM based classifier. Preliminary studies on a endoscopic color image database consisting of 276 images of 6 different types of endoscopic classes show promising retrieval result. The details of "Scheme-II" is describe as follows:

6.4.1 NSCT based localized GGD features

The images in the database prior to NSCT decomposition, are transformed from RGB to CIELab color space. This ensures that the textural characterizations of the images are independent of the color characterization. NSCT decomposition over the intensity plane (L) characterizes the texture information, while the NSCT decomposition over chromaticity planes (a and b) characterizes color. Texture and color information are extracted by using NSCT on each color plane with l level of decomposition (here, l = 3) which results in 10 high pass and 1 low pass sub-band images. Therefore, this configuration produces sb $(= 11)$ subbands for each color plane. As, there are 3 color planes, so altogether we get S $(= 3 \times sb)$ subbands for each image of the database. Each subband contains the

transform coefficients of size $n \times n$, where n is the size of the NSCT subband images. Now, for the computation of the local feature vectors, each such subbands are sub-divided into N number of blocks using window size $w \times w$, where $N = \frac{n^2}{w^2}$. The distribution of the coefficients of each N blocks of each subband coefficients are then modeled with GGD. The details of GGD based image representation mechanism is describe in Section 3.3.1 of Chapter 3. Therefore, each block of each NSCT subband (*i.e.*, *sb*) is represented by two GGD parameters. Considering S number of subbands of each image I and N number of blocks for each subband, we obtain $2 \times S \times \frac{n^2}{w^2}$ dimension of feature vectors, where $n = 256$ and $w = 64$.

Finally, each image in the database is represented by 1056 dimension of localized feature vectors. However, this localized feature vectors actually reduced the discriminatory capacity of the classification accuracy because of the redundancy among different feature vectors. Therefore, we need to apply some dimension-reduction technique to reduce the the feature vector dimension and improve the retrieval accuracy with minimum computational complexity. In this study, we have used the well-known PCA for this purpose. PCA reduces the dimensionality of the search to a basis set of prototypes that best describes the images [292, 309]. This process results in an overall reduction of feature dimension by around 82% (from 1056 to 195). Therefore, each image is represented with 195 compact localized statistical feature vectors.

6.4.2 Classification using LSSVM

Once the NSCT based GGD feature vectors are obtained, LSSVM classifier is used to classify images of the database. The details of LSSVM is describe in Section 2.4.4 of Chapter 2.

6.4.3 KLD based Similarity Measure

After the image is classified using LSSVM according to the NSCT based GGD feature vectors, KLD is used for distance calculation between the query image and the database images. Database images are ranked according to ascending order of distances and are retrieved from the database. The details of implementing is describe in Section 3.3.2 of Chapter 3.

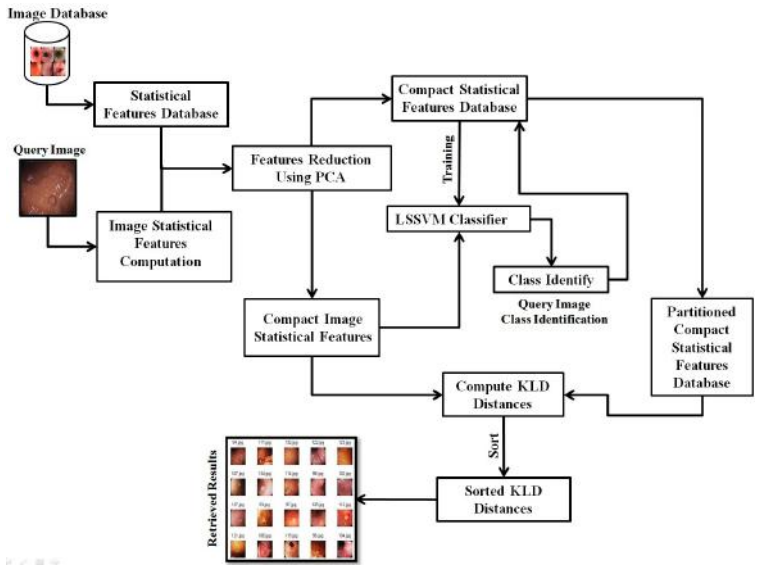


Fig. 6.10 Block Diagram of the Proposed CBMIR System -"Scheme-II"

6.4.4 The Summary of CBMIR System

Here, we outlines the salient steps of the proposed endoscopic based CBMIR system. The Fig. 6.10, shows the block diagram of the proposed CBMIR system.

Steps:

1. Input the query image.
2. Compute the GGD based compact image statistical features of the query image as described in Section 6.4.1.

3. Identify the class of the query image using LSSVM, depending on the GGD based compact statistical features as described in Section 6.4.2.

4. As describe in Section 6.4.3, after the pre-classification of the query image, find the KLD distances between the compact statistical features of the query image and the compact local statistical features of the image database stored in "Partitioned Compact Statistical Features Database".

5. Sort the distances in ascending order and display 20 images corresponding to the first 20 sorted distances.

6. Stop.

6.4.5 Experimental Results

Extensive experiments were carried out to evaluate the performance of the proposed CB-MIR system. We conducted several different experiments to evaluate the effectiveness of the proposed system both quantitatively and qualitatively.

6.4.5.1 Experimental setup

In our experiments, we have used endoscopy image dataset consist of 276 images which grouped into 6 categories (46 images per category). The datasets are heterogeneous as they have different sizes along with 4 different types of abnormality (diseases)and 2 normal types as shown in Fig. 6.11.

The number of decomposition level of NSCT, considered in this study is 3. The details about NSCT decomposition are discussed in [178, 206]. Moreover, in this experiment, we have considered the local features of the endoscopic images. Local features is important for image retrieval because the similarity between two images perceived by human observers

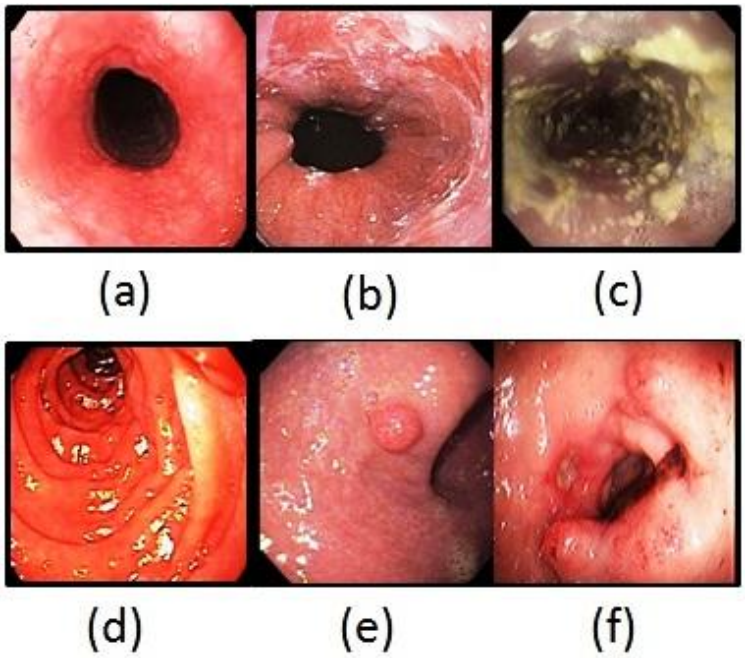


Fig. 6.11 Examples of categories from endoscopy image datasets (a) Normal Esophagus (b) Barrett's Esophagus (c) Esophageal Inflammation (d) Normal Upper Gastrointestinal (e) Upper Gastrointestinal Polyps and (f) Upper Gastrointestinal Ulcers

strongly depends on the layout of the image as well as the spatial relations among depicted objects [315].

In LSSVM, we have used the RBF as the kernel for training. Considering the CV technique, the best value of C and γ are computed. The details of these parameters are discussed in previous chapters. 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 our study, CV is done on 70% of the train dataset and remnant 30% is used as test dataset for evaluating the performance of the classifiers used in our study. We have achieved the classification accuracy of 71.45% for endoscopic image database.

Quantitative evaluation of the proposed CBMIR system is analyzed using two statistical measures: AP, and AR. 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 as discussed in

Section 2.7 of Chapter 2.

6.4.5.2 Feature Evaluation

The graph of the Fig. 6.12 shows the effect of PCA based NSCT+GGD (*i.e.*, GGD based localized statistical features on NSCT subbands) feature representation schemes verses classification accuracy. From the graph, it can be inferred that with nearly 18% of NSCT+GGD based features have satisfactory classification accuracy as compared to 100% of their respective features. Therefore, we have selected 195 distinguished NSCT+GGD feature vectors for our proposed CBMIR system on endoscopic images.

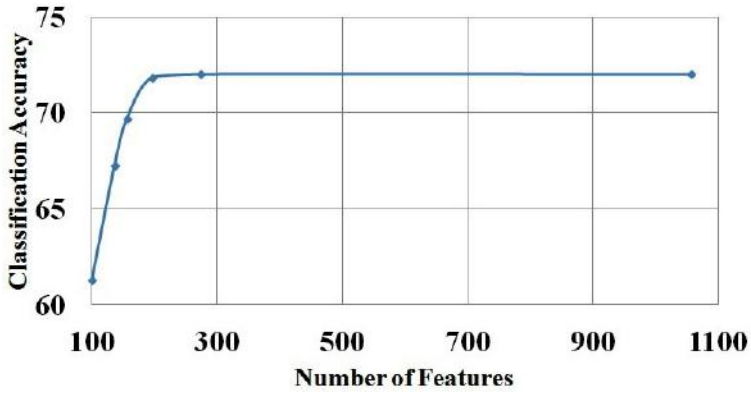


Fig. 6.12 Classification Accuracy Vs Number of Features

6.4.6 Results and Discussion

The retrieval performance of the proposed CBMIR system (NSCT+GGD+LSSVM) is shown in the Fig.6.13, in terms of precision and recall curve. The performance comparisons of WT, CVT and CT with NSCT are also shown in Fig.6.13 (WT+GGD+LSSVM, CVT+GGD+LSSVM and CT+GGD+LSSVM, respectively). Here, endoscopy image database were considered for performance evaluation. It is clear from the Fig.6.13, that NSCT performance much better than WT, CVT and CT in proposed endoscopic CBMIR system.

Fig. 6.14, Fig. 6.15 and Fig. 6.16, shows three examples of the visual results obtained by

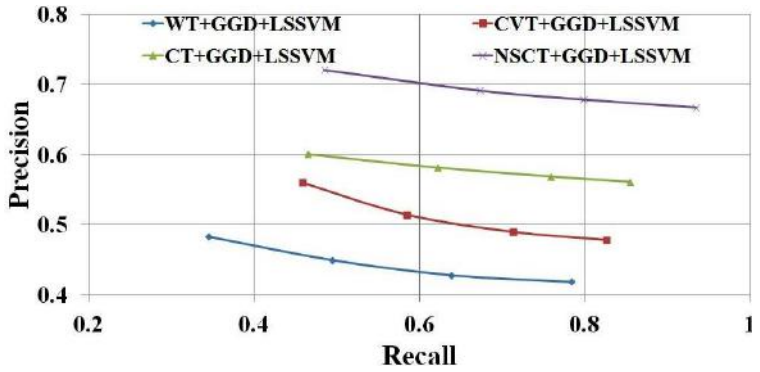


Fig. 6.13 Performance comparison in terms of precision and recall curves

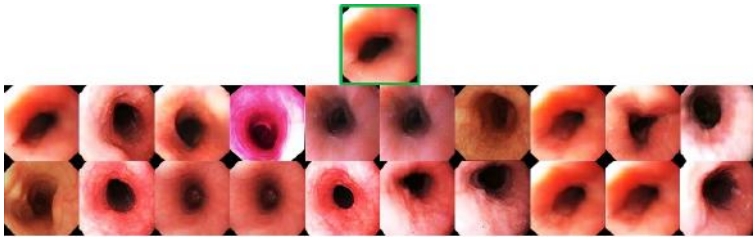


Fig. 6.14 Visual results of the proposed CBIR system for Normal Esophagus (top image is the query image)

the proposed CBMIR system on endoscopy image database using query images from "*Normal Esophagus*", "*Upper Gastrointestinal Polpys*" and "*Barrett's Esophagus*" classes, respectively. Fig. 6.15 and Fig. 6.16 shows the diseased endoscopic images where as Fig. 6.14 shows the esophagus or gastrointestinal tract of healthy persons. In "*Barrett's Esophagus*", a pink esophageal lining (mucosa) appear that extends a short distance (usually less than 2.5 inches) up the esophagus from the gastroesophageal junction and finding intestinal type cells (goblet cells) on biopsy of the lining whereas in "*Upper Gastrointestinal Polpys*" is a mass of tissue protruding abnormally from the mucous membrane (protective lining of epithelial cells for secretion and absorption) of the esophagus. From the given instances, it can be inferred that the local color, texture and geometrical invariant information can distinguished the normal esophagus or gastrointestinal tract endoscopic images from the abnormal ones. Our proposed NSCT+GGD features can represent those information efficiently and in turn increases the efficiency of the proposed CBMIR system. This fact is

Fig. 6.15 Visual results of the proposed CBIR system for Upper Gastrointestinal Polpys (top image is the query image)

corroborated by all the instances of the results obtained from our system where we can clearly see that all the retrieved images are from the respective classes corresponding to the query images.

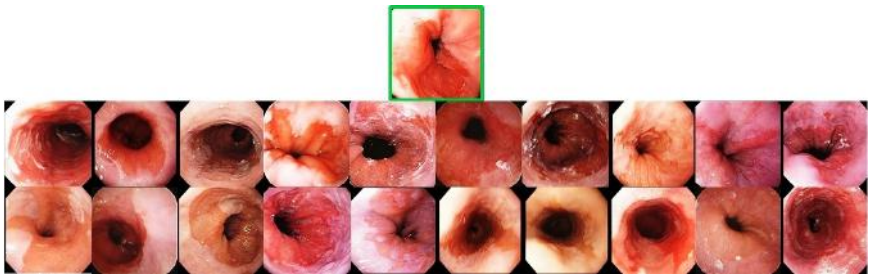


Fig. 6.16 Visual results of the proposed CBIR system for Barrett's Esophagus (top image is the query image)

The proposed algorithm has been implemented using MATLAB R2014a on a Dell Precision T7400 workstation. The total processing time of the query images is computed by considering the feature extraction, classification and retrieval process. Average time taken for extraction of GGD features of each image is 19.770591 seconds where as processing time is 0.862763 seconds, considering 1056 features. However, the processing time of the proposed algorithm is further being reduced by using the PCA algorithm.

From our experiments, we have noticed that statistical modeling of NSCT based image coding is suitable for representing localized low level features of the endoscopic images. The proposed CBMIR system with LSSVM classifier based on compact GGD based localized statistical features on NSCT coefficients is able to achieve the satisfactory results with

minimum computational cost. The details of Scheme-I and Scheme-II are found in [180] and [187] respectively.

In the last chapter, we have summarized overall contribution of the present thesis, its novelty and future scope of investigation.

∗∗**Major portion of this Chapter taken from the following publication.**

- M. K. Kundu, M. Chowdhury and S. Das, "Interactive Radiographic Image Retrieval System", Computer Methods and Programs in Biomedicine (Major Revision Submitted), 2016.

- M.Chowdhury and M. K. Kundu, "Endoscopic Image Retrieval System Using Multiscale Image Features", Proc. of the 2nd Int. Conf. on Perception and Machine Intelligence, pp. 64–70, 2015.

Chapter 7

Conclusion and Future Scope

7.1 Conclusion

This thesis presents a number of new and novel results. Current trends in CBIR design is to use different multiscale transform as feature extraction tools. The extracted transform feature are used as an input to different kinds of classifiers for compact generations of signature of the particular class of the image. Different combination of MRA tools and classifiers has been used for various kinds of images as reported in the literature. But no extensive study has been done so far on which MRA tool to be associated with a particular type of classifier in order to get the best result for CBIR design for a particular class of the image. In the present thesis, extensive study has been made on this issue and it was found that the combination of NSCT with parameters 'pyrexc' and 'sinc' along with LSSVM classifier can produce best results for most of the natural image classes.

We have also shown that one of the major problems of conventional CBIR is the semantic gap, that exist between low level image representation and actual visual response. To bridge this gap, relevance feedback mechanism is used. In this present thesis, we have proposed relevance feedback model using fuzzy set theoretic approach, which used both transform feature vector, as well as pre-classified feature vector (using some standard classifier like LSSVM) as input. It is found that the number of the iterations requires to achieve the same level of accuracy is much lower in the case of pre classified feature vector which makes the CBIR system very efficient and accurate.

In the fuzzy based relevance feedback mechanism, the relative weight associated with feature component is getting updated, but if values of some component of the feature vector are relatively small, then they will have insignificant influence in the final result of the similarity computation. As a result, some of the important contributions of the lower valued feature components may be overshadowed by the higher valued features, resulting in an erroneous ranking of the images. To overcome this problem, we have introduced a graph theoretic based re-ranking relevance feedback mechanism in this thesis. This method

gives quite satisfactory retrieval accuracy and can outperforms the weight updation based relevance feedback mechanism.

The thesis also presented two interesting applications on medical CBIR. These CBMIR systems have been designed for radiographic X-ray image database and for color endoscopic image database respectively. In reality, as there exists no classifier, whose accuracy is absolute (100%) So the accuracy of any classifier based CBIR system is always restricted by the accuracy limit of the classifier used. To reduce this accuracy gap, the similarity positional score has been proposed in radiographic X-ray based image retrieval system. This approach has helped the CBMIR system to achieve fairly high accuracy. We have also presented, CBMIR system for endoscopic image retrieval. In this system, NSCT (extract color and texture features) with GGD are used for compact, feature vector generation, and a preclassifier LSSVM is using auto learning. It is found from the initial results that accuracy achieved using these features is quite promising with small size (300) database. Experimentation with larger endoscopic database extensive performance evaluation is the future scope.

7.2 Future work

The following research works can be undertaken based on the results obtained in this thesis as future research task.

1. NSCT based features with the LSSVM classifier combination can also be used for accurate interframe dissimilarity measurement. This can be used for achieving results with higher accuracy in various video analytic problem like abrupt and gradual shot detection, fading, zooming etc.

2. In Content Based Video Retrieval system (CBVR), relevance feedback mechanism can be used to bridge the gap between semantic notions used for search relevance and

the low level representation of video frame content. Based on the result obtained in this thesis, it can infer that a graph theoretic based re-ranking mechanism in relevance feedback can be used with suitable modification for achieving higher accuracy in CBVR system design.

3. The design of CBMIR system for endoscopic image may be improved further with relevance feedback mechanism based on graph theoretic based re-ranking and using PCNN based shape features beside color and texture. The robustness of such system to be tested large size endoscopic image database ($2000 - 5000$), which was not available during our thesis work.

References

[1] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. of ICCV - Volume 2*, ICCV '03, (Washington, DC, USA), pp. 1470–, IEEE Computer Society, 2003.

[2] H. Jégou and O. Chum, "Negative evidences and co-occurrences in image retrieval: the benefit of PCA and whitening," in *Proc of ECCV*, (Firenze, Italy), Oct. 2012.

[3] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *J. Vis. Commun. Image Represent.*, vol. 10, no. 1, pp. 39–62, 1999.

[4] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, 2000.

[5] O. Marques and B. Furht, *Content-based image and video retrieval*. Kluwer Academic Publishers, 2002.

[6] A. Diaz, "Through the google goggles: Sociopolitical bias in search engine design," in *Web Search* (A. Spink and M. Zimmer, eds.), vol. 14 of *Information Science and Knowledge Management*, pp. 11–34, Springer Berlin Heidelberg, 2008.

[7] X. Yang, X. Qian, and T. Mei, "Learning salient visual word for scalable mobile image retrieval," *Pattern Recogn.*, vol. 48, no. 10, pp. 3093–3101, 2015.

[8] M. F. A. Fauzi, *Content-based image retrieval of museum images*. PhD thesis, University of Southampton, 2004.

[9] J. Lee, R. Jin, A. Jain, and W. Tong, "Image retrieval in forensics: Tattoo image database application," *IEEE Multimedia Mag.*, vol. 19, no. 1, pp. 40–49, 2012.

[10] C.-H. Wei, Y. Li, W.-Y. Chau, and C.-T. Li, "Trademark image retrieval using synthetic features for describing global shape and interior structure," *Pattern Recogn.*, vol. 42, no. 3, pp. 386–394, 2009.

[11] J. A. Piedra-Fernandez, G. Ortega, J. Z. Wang, and M. Canton-Garbin, "Fuzzy content-based image retrieval for oceanic remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5422–5431, 2014.

[12] S. Srinivasulu and P. Sakthivel, "Extracting spatial semantics in association rules for weather forecasting image," in *Trendz in Inf. Sci. & Computing*, pp. 54–57, Dec 2010.

[13] J. Li, "The application of cbir-based system for the product in electronic retailing," in *Proc. of Int. Conf. on Computer-Aided Industrial Design Conceptual Design*, vol. 2, pp. 1327–1330, 2010.

[14] K. H. Hwang, H. Lee, and D. Choi, "Medical image retrieval: Past and Present," *Healthc. Inform. Res.*, vol. 18, no. 1, pp. 3–9, 2012.

[15] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: Hashing-based image retrieval," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 496–506, 2015.

[16] 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. 5:1–5:60, 2008.

[17] G.-H. Liu, J.-Y. Yang, and Z. Li, "Content-based image retrieval using computational visual attention model," *Pattern Recogn.*, vol. 48, no. 8, pp. 2554–2566, 2015.

[18] R. C. Veltkamp and M. Tanase, "Content-based image retrieval systems: A survey," *Technical Report UU-CS-2000-34*, 2001.

[19] S. Pandey, P. Khanna, and H. Yokota, "An effective use of adaptive combination of visual features to retrieve image semantics from a hierarchical image database," *J. Vis. Commun. Image Represent.*, vol. 30, pp. 136–152, 2015.

[20] L. Xie, J. Wang, B. Zhang, and Q. Tian, "Fine-grained image search," *IEEE Trans. Multimedia*, vol. 17, no. 5, pp. 636–647, 2015.

[21] J. Zhang, D. Li, Y. Zhao, Z. Chen, and Y. Yuan, "Representation of image content based on RoI-BoW," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 37–49, 2015.

[22] L. Zhang, L. Wang, and W. Lin, "Generalized biased discriminant analysis for content-based image retrieval," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 1, pp. 282–290, 2012.

[23] J. Yu, D. Liu, D. Tao, and H. S. Seah, "On combining multiple features for cartoon character retrieval and clip synthesis," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 5, pp. 1413–1427, 2012.

[24] G. Boato, D.-T. Dang-Nguyen, O. Muratov, N. Alajlan, and F. G. De Natale, "Exploiting visual saliency for increasing diversity of image retrieval results," *Multimed. Tools Appl.*, pp. 1–22, 2015.

[25] M. Verma, B. Raman, and S. Murala, "Local extrema co-occurrence pattern for color and texture image retrieval," *Neurocomputing*, vol. 165, pp. 255–269, 2015.

[26] X. Duanmu, "Image retrieval using color moment invariant," in *Proc. of 7th Int. Conf. on Inf. Tech.: New Generations*, pp. 200–203, 2010.

[27] G. Qiu, "Color image indexing using btc," *IEEE Trans. Image Process.*, vol. 12, no. 1, pp. 93–101, 2003.

[28] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. of CVPR*, pp. 762–768, 1997.

[29] X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, "Content-based image retrieval by integrating color and texture features," *Multimed. Tools Appl.*, vol. 68, no. 3, pp. 545–569, 2014.

[30] R.-H. Miao, J.-L. Tang, and X.-Q. Chen, "Classification of farmland images based on color features," *J. Vis. Commun. Image Represent.*, vol. 29, pp. 138–146, 2015.

[31] V. Kovalev and M. Petrou, "Multidimensional co-occurrence matrices for object recognition and matching," *Graph Model Im. Proc.*, vol. 58, no. 3, pp. 187–197, 1996.

[32] N. Jhanwar, S. Chaudhuri, G. Seetharaman, and B. Zavidovique, "Content based image retrieval using motif cooccurrence matrix," *Image Vision Comput.*, vol. 22, no. 14, pp. 1211–1220, 2004.

[33] J. Mathews, M. S. Nair, and L. Jo, "A novel color image coding technique using improved btc with k-means quad clustering," in *Advances in Signal Processing and Intelligent Recognition Systems*, pp. 347–357, Springer, 2014.

[34] J.-M. Guo, H. Prasetyo, and J.-H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 466–481, 2015.

[35] J.-M. Guo and H. Prasetyo, "Content-based image retrieval using features extracted from halftoning-based block truncation coding," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1010–1024, 2015.

[36] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: multimedia content description interface*, vol. 1. John Wiley & Sons, 2002.

[37] H. Shao, Y. Wu, W. Cui, and J. Zhang, "Image retrieval based on mpeg-7 dominant color descriptor," in *'Proc. of 9th Int. Conf. for Young Computer Scientists*, pp. 753–757, 2008.

[38] R. Min and H. Cheng, "Effective image retrieval using dominant color descriptor and fuzzy support vector machine," *Pattern Recogn.*, vol. 42, no. 1, pp. 147–157, 2009.

[39] Z. Jiexian, L. Xiupeng, and F. Yu, "Multiscale distance coherence vector algorithm for content-based image retrieval," *Sci. World J.*, vol. 2014, pp. 1–13, 2014.

[40] M. Petrou and P. García Sevilla, *Image processing: dealing with texture*. Chichester, UK: John Wiley & Sons, 2006.

[41] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. Orlando, FL, USA: Academic Press, Inc., 2nd ed., 1982.

[42] M. D. Levine and M. D. Levine, *Vision in man and machine*. McGraw-Hill New York, 1985.

[43] N. Shrivastava and V. Tyagi, "Content based image retrieval based on relative locations of multiple regions of interest using selective regions matching," *Inform. Sci.*, vol. 259, pp. 212–224, 2014.

[44] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," in *Proc. of CVPR*, 2008.

[45] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. of CVPR*, 2007.

[46] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval.," in *Proc. of ICCV*, pp. 1–8, 2007.

[47] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[48] P. Lu, X. Peng, X. Zhu, and R. Li, "Finding more relevance: Propagating similarity on markov random field for object retrieval," *Signal Process.-Image*, vol. 32, pp. 54–68, 2015.

[49] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, 2001.

[50] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. of ICIP*, pp. 444–447, 1995.

[51] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst., Man, Cybern.*, vol. 8, no. 6, pp. 460–473, 1978.

[52] Y. Meyer, "Wavelets algorithms and applications," *SIAM Rev.*, vol. 36, no. 3, pp. 526–528, 1994.

[53] G. Papakostas, D. Koulouriotis, and V. Tourassis, *Feature extraction based on wavelet moments and moment invariants in machine vision systems*. INTECH Open Access Publisher, 2012.

[54] L. Chen, G. Lu, and D. Zhang, "Effects of different gabor filter parameters on image retrieval by texture," in *Proc. of MMM*, pp. 273–278, 2004.

[55] M. Vetterli and J. Kovačevic, *Wavelets and Subband Coding*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.

[56] Y. Dong, D. Tao, X. Li, J. Ma, and J. Pu, "Texture classification and retrieval using shearlets and linear regression," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 358–369, 2015.

[57] S. A. Chatzichristofis and Y. S. Boutalis, "FCTH: Fuzzy color and texture histogram - A low level feature for accurate image retrieval," in *Proc. of 9th Int. Workshop on Image Analysis for Multimedia Interactive Services*, pp. 191–196, 2008.

[58] X. Wang and Z. Wang, "A novel method for image retrieval based on structure element's descriptor," *J. Vis. Commun. Image Represent.*, vol. 24, no. 1, pp. 63–74, 2013.

[59] G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, "Image retrieval based on micro-structure descriptor," *Pattern Recogn.*, vol. 44, no. 9, pp. 2123–2133, 2011.

[60] R. Kwitt and A. Uhl, "Lightweight probabilistic texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 241–253, 2010.

[61] N.-E. Lasmar and Y. Berthoumieu, "Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2246–2261, 2014.

[62] X.-Y. Wang, Z.-F. Chen, and J.-J. Yun, "An effective method for color image retrieval based on texture," *Comput. Stand. Inter.*, vol. 34, no. 1, pp. 31–35, 2012.

[63] X.-Y. Wang, Y.-J. Yu, and H.-Y. Yang, "An effective image retrieval scheme using color, texture and shape features," *Comput. Stand. Inter.*, vol. 33, no. 1, pp. 59–68, 2011.

[64] G. B. de Souza and A. N. Marana, "HTS and HTSn: New shape descriptors based on hough transform statistics," *Comput. Vis. Image Und.*, vol. 127, pp. 43–56, 2014.

[65] W. Chen, Q. Li, and K. Dahal, "Roi image retrieval based on multiple features of mean shift and expectation–maximisation," *Digit. Signal Process.*, vol. 40, pp. 117–130, 2015.

[66] S. W. Teng, R. M. N. Sadat, and G. Lu, "Effective and efficient contour-based corner detectors," *Pattern Recogn.*, vol. 48, no. 7, pp. 2185–2197, 2015.

[67] M. Yang, K. Kpalma, and J. Ronsin, "A survey of shape feature extraction techniques," *Pattern Recogn.*, pp. 43–90, 2008.

[68] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Trans. Multim. Comput.*, vol. 2, no. 1, pp. 1–19, 2006.

[69] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recogn.*, vol. 40, no. 1, pp. 262–282, 2007.

[70] J. Vogel and B. Schiele, "Performance evaluation and optimization for content-based image retrieval," *Pattern Recogn.*, vol. 39, no. 5, pp. 897–909, 2006.

[71] X. Shu and X.-J. Wu, "A novel contour descriptor for 2d shape matching and its application to image retrieval," *Image Vision Comput.*, vol. 29, no. 4, pp. 286–294, 2011.

[72] Y. W. Chen and C. L. Xu, "Rolling penetrate descriptor for shape-based image retrieval and object recognition," *Pattern Recogn. Lett.*, vol. 30, no. 9, pp. 799–804, 2009.

[73] C. H. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 4, pp. 496–513, 1988.

[74] S. X. Liao and M. Pawlak, "On image analysis by moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 3, pp. 254–266, 1996.

[75] C. W. Niblack, R. Barber, W. Equitz, M. D. Flickner, E. H. Glasman, and *et al*, "QBIC project: querying images by content, using color, texture, and shape," in *Proc. of SPIE: Storage and Retrieval for Image and Video Databases*, vol. 1908, pp. 173–187, 1993.

[76] M. Safar, C. Shahabi, and X. Sun, "Image retrieval by shape: a comparative study," in *Proc. of ICME*, vol. 1, pp. 141–144, 2000.

[77] K. Chakrabarti, M. Ortega-Binderberger, K. Porkaew, and S. Mehrotra, "Similar shape retrieval in MARS," in *Proc. of ICME*, vol. 2, pp. 709–712, 2000.

[78] A. M. Bronstein, M. M. Bronstein, L. J. Guibas, and M. Ovsjanikov, "Shape Google: Geometric words and expressions for invariant shape retrieval," *ACM Trans. Graphic*, vol. 30, no. 1, pp. 1:1–1:20, 2011.

[79] M. Ovsjanikov, A. M. Bronstein, M. M. Bronstein, and L. J. Guibas, "Shape Google: a computer vision approach to isometry invariant shape retrieval," in *Proc. of ICCV*, pp. 320–327, 2009.

[80] T.-E. Kim and M. H. Kim, "Improving the search accuracy of the vlad through weighted aggregation of local descriptors," *J. Vis. Commun. Image Represent.*, vol. 31, pp. 237–252, 2015.

[81] G. Serra, C. Grana, M. Manfredi, and R. Cucchiara, "Gold: Gaussians of local descriptors for image representation," *Comput. Vis. Image Und.*, vol. 134, pp. 22–32, 2015.

[82] M. Jiang, S. Zhang, H. Li, and D. N. Metaxas, "Computer-aided diagnosis of mammographic masses using scalable image retrieval," *IEEE Trans. Biomed. Eng*, vol. 62, no. 2, pp. 783–792, 2015.

[83] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[84] S. Candemir, E. Borovikov, K. Santosh, S. Antani, and G. Thoma, "RSILC: Rotation and scale invariant, line based color aware descriptor," *Image Vision Comput.*, vol. 42, pp. 1–12, 2015.

[85] T. Takahashi and T. Kurita, "Mixture of subspaces image representation and compact coding for large-scale image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 7, pp. 1469–1479, 2015.

[86] D. Mukherjee, Q. M. J. Wu, and G. Wang, "A comparative experimental study of image feature detectors and descriptors," *Mach. Vis. Appl.*, vol. 26, no. 4, pp. 443–466, 2015.

[87] J. Yao, B. Yang, and Q. Zhu, "Near-duplicate image retrieval based on contextual descriptor," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1404–1408, 2015.

[88] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. of CVPR*, vol. 2, pp. 506–513, 2004.

[89] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 27, no. 10, pp. 1615–1630, 2005.

[90] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Und.*, vol. 110, no. 3, pp. 346–359, 2008.

[91] N. Ali *et al.*, "A novel image retrieval based on visual words integration of SIFT and SURF," *PlosOne*, vol. 11, no. 6, pp. 1–20, 2016.

[92] C.-F. Tsai, "Bag-of-words representation in image annotation: A review," *ISRN Artificial Intelligence*, vol. 2012, no. 376804, pp. 1–19.

[93] J. Mukherjee, J. Mukhopadhyay, and P. Mitra, "A survey on image retrieval performance of different bag of visual words indexing techniques," in *Students Technology Symposium (TechSym), 2014 IEEE*, pp. 99–104, Feb 2014.

[94] D. Local Feature Detectors and I. R. A. Survey, "Yusuke uchida," *arXiv:1607.08368*, 2016.

[95] F. Perronnin and C. R. Dance, "Fisher kernels on visual vocabularies for image categorization.," in *Proc. of CVPR*, 2007.

[96] H. Jegou *et al.*, "Aggregating local image descriptors into compact codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, pp. 1704–1716, Sept. 2012.

[97] R. Arandjelovic and A. Zisserman, "All about VLAD," in *Proc. of CVPR*, pp. 1578–1585, 2013.

[98] J. Delhumeau, P.-H. Gosselin, H. Jegou, and P. Perez, "Revisiting the VLAD image representation," in *ACM Multimedia*, Oct 2013.

[99] G. V. for Large Scale Image Search, "Z. wang and w. di and a. bhardwaj and v. jagadeesh and r. piramuthu," *arXiv:1403.3829*, 2014.

[100] H. Jegou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *Proc. of CVPR*, pp. 3304–3311, 2010.

[101] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 24, no. 7, pp. 971–987, 2002.

[102] L. Sørensen, S. B. Shaker, and M. De Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 559–569, 2010.

[103] D. Unay, A. Ekin, and R. S. Jasinschi, "Local structure-based region-of-interest retrieval in brain MR images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 4, pp. 897–903, 2010.

[104] V. Takala, T. Ahonen, and M. Pietikäinen, "Block-based methods for image retrieval using local binary patterns," in *Proc. of the 14th Scand. Conf. on Image Analysis*, pp. 882–891, 2005.

[105] S. Murala, R. Maheshwari, and R. Balasubramanian, "Local tetra patterns: a new feature descriptor for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2874–2886, 2012.

[106] M. Subrahmanyam, R. Maheshwari, and R. Balasubramanian, "Local maximum edge binary patterns: a new descriptor for image retrieval and object tracking," *Signal Processing*, vol. 92, no. 6, pp. 1467–1479, 2012.

[107] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 533–544, 2010.

[108] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, 2010.

[109] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recogn.*, vol. 42, no. 3, pp. 425–436, 2009.

[110] R. Gupta, H. Patil, and A. Mittal, "Robust order-based methods for feature description," in *Proc. of CVPR*, pp. 334–341, 2010.

[111] S. Murala and Q. Wu, "Local mesh patterns versus local binary patterns: biomedical image indexing and retrieval," *IEEE J. Biomed. Health Inform*, vol. 18, no. 3, pp. 929–938, 2014.

[112] S. Murala and Q. J. Wu, "Local ternary co-occurrence patterns: a new feature descriptor for mri and ct image retrieval," *Neurocomputing*, vol. 119, pp. 399–412, 2013.

[113] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local diagonal extrema pattern: a new and efficient feature descriptor for CT image retrieval," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1215–1219, 2015.

[114] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. of CVPR*, vol. 1, pp. 886–893, 2005.

[115] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and et al., "CHOG: Compressed histogram of gradients a low bit-rate feature descriptor," in *Proc. of CVPR*, pp. 2504–2511, 2009.

[116] C.-H. Lin, R.-T. Chen, and Y.-K. Chan, "A smart content-based image retrieval system based on color and texture feature," *Image Vision Comput.*, vol. 27, no. 6, pp. 658–665, 2009.

[117] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. of ICCV*, pp. 2564–2571, 2011.

[118] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 32, no. 1, pp. 105–119, 2010.

[119] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," *Proc. of ECCV:Part IV*, pp. 778–792, 2010.

[120] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. of ICCV*, pp. 2548–2555, 2011.

[121] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *Proc. of CVPR*, pp. 510–517, 2012.

[122] N. Khan, B. McCane, and S. Mills, "Better than SIFT?," *Mach. Vis. Appl.*, vol. 26, no. 6, pp. 819–836, 2015.

[123] J. Byrne and J. Shi, "Nested shape descriptors," in *Proc. of ICCV*, pp. 1201–1208, 2013.

[124] S. Santini and R. Jain, "Similarity measures," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 21, no. 9, pp. 871–883, 1999.

[125] C. Beecks, M. S. Uysal, and T. Seidl, "A comparative study of similarity measures for content-based multimedia retrieval," in *Proc. of ICME*, pp. 1552–1557, 2010.

[126] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, "Efficient color histogram indexing for quadratic form distance functions," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 17, no. 7, pp. 729–736, 1995.

[127] A. Vailaya, A. Jain, and H. J. Zhang, "On image classification: City images vs. landscapes," *Pattern Recogn.*, vol. 31, no. 12, pp. 1921–1935, 1998.

[128] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. of Comp. Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[129] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc .of ICIP*, vol. 2, pp. 815–818, 1997.

[130] W.-Y. Ma and B. S. Manjunath, "NETRA: A toolbox for navigating large image databases," *Multimed. Syst.*, vol. 7, no. 3, pp. 184–198, 1999.

[131] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 24, no. 8, pp. 1026–1038, 2002.

[132] H. Voorhees and T. Poggio, "Computing texture boundaries from images," *Nature*, vol. 333, no. 6171, pp. 364–367, 1988.

[133] Y. Rubner and C. Tomasi, *Perceptual Metrics for Image Database Navigation*. Norwell, MA, USA: Kluwer Academic Publishers, 2001.

[134] H. Liu, D. Song, S. Rüger, R. Hu, and V. Uren, "Comparing dissimilarity measures for content-based image retrieval," in *Proc. of the 4th Asia Inf. Retrieval Conf. on Inf. Retrieval Tech.*, pp. 44–50, Springer, 2008.

[135] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and kullback leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, 2002.

[136] J. R. Hershey and P. A. Olsen, "Approximating the kullback leibler divergence between gaussian mixture models," in *Proc. of ICASSP*, vol. 4, pp. IV–317–IV–320, 2007.

[137] C. Beecks, A. M. Ivanescu, S. Kirchhoff, and T. Seidl, "Modeling image similarity by gaussian mixture models and the signature quadratic form distance," in *Proc. of ICCV*, pp. 1754–1761, 2011.

[138] H. Ling and K. Okada, "An efficient earth mover's distance algorithm for robust histogram comparison," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 29, no. 5, pp. 840–853, 2007.

[139] O. Pele and M. Werman, "A linear time histogram metric for improved sift matching," in *Proc. of ECCV*, pp. 495–508, 2008.

[140] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proc. of ICCV*, pp. 460–467, 2009.

[141] S. Shirdhonkar and D. W. Jacobs, "Approximate earth mover's distance in linear time," in *Proc. of CVPR*, pp. 1–8, 2008.

[142] F. Wang and L. J. Guibas, "Supervised earth mover's distance learning and its computer vision applications," in *Proc. of ECCV*, pp. 442–455, 2012.

[143] K. Grauman and T. Darrell, "Fast contour matching using approximate earth mover's distance," in *Proc. of CVPR*, vol. 1, pp. 220–227, 2004.

[144] S. Lazebnik, C. Schmid, and J. Ponce, "Sparse texture representations using affine-invariant neighborhoods," in *Proc. of CVPR*, vol. 2, pp. 319–324, 2003.

[145] R. Typke, R. C. Veltkamp, and F. Wiering, "Searching notated polyphonic music using transportation distances," in *Proc. of the 12th Annual ACM Int Conf on Multimedia*, pp. 128–135, 2004.

[146] V. Karavasilis, C. Nikou, and A. Likas, "Visual tracking using the earth mover's distance between gaussian mixtures and kalman filtering," *Image Vision Comput.*, vol. 29, no. 5, pp. 295–305, 2011.

[147] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. Cir. and Sys. for Video Technol.*, vol. 8, no. 5, pp. 644–655, 1998.

[148] G. T. Papadopoulos, K. C. Apostolakis, and P. Daras, "Gaze-based relevance feedback for realizing region-based image retrieval," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 440–454, 2014.

[149] A. Kovashka, D. Parikh, and K. Grauman, "Whittlesearch: Interactive image search with relative attribute feedback," *Int. J. of Comp. Vision*, pp. 1–26, 2015.

[150] A. Lakshmi, M. Nema, and S. Rakshit, "Long term relevance feedback: A probabilistic axis re-weighting update scheme," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 852–856, 2015.

[151] Z. Jin, I. King, and X. Li, "Content-based image retrieval by relevance feedback," in *Proc. of the 4th Int. Conf. on Visual Inf. Systems*, pp. 521–529, 2000.

[152] M. Behnam and H. Pourghassem, "Optimal query-based relevance feedback in medical image retrieval using score fusion-based classification," *J. Digit. Imaging*, vol. 28, no. 2, pp. 160–178, 2015.

[153] Z. Shi, Q. He, and Z. Shi, "Bayes-based relevance feedback method for CBIR," in *Innovations in Hybrid Intelligent Systems* (E. Corchado, J. Corchado, and A. Abraham, eds.), vol. 44 of *Advances in Soft Computing*, pp. 264–271, Springer Berlin Heidelberg, 2007.

[154] M. Chowdhury, S. Das, and M. K. Kundu, "Novel CBIR system based on ripplet transform using interactive neuro-fuzzy technique," *Elect. Letters on Comp. Vision and Image Anal.*, vol. 11, no. 1, pp. 1–13, 2012.

[155] X.-Y. Wang, H.-Y. Yang, Y.-W. Li, W.-Y. Li, and J.-W. Chen, "A new svm-based active feedback scheme for image retrieval," *Eng. Appl. Artif. Intel.*, vol. 37, pp. 43–53, 2015.

[156] W. Bian and D. Tao, "Biased discriminant euclidean embedding for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 545–554, 2010.

[157] L. Zhang, L. Wang, and W. Lin, "Semisupervised biased maximum margin analysis for interactive image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2294–2308, 2012.

[158] B. Demir and L. Bruzzone, "A novel active learning method in relevance feedback for content-based remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2323–2334, 2015.

[159] X. He, O. King, W.-Y. Ma, M. Li, and H.-J. Zhang, "Learning a semantic space from user's relevance feedback for image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 1, pp. 39–48, 2003.

[160] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "Long term learning in image retrieval systems using case based reasoning," *Eng. Appl. Artif. Intel.*, vol. 35, pp. 26–37, 2014.

[161] C.-C. Lai and Y.-C. Chen, "A user-oriented image retrieval system based on interactive genetic algorithm," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3318–3325, 2011.

[162] C. D. Ferreira, J. Santos, R. d. S. Torres, M. A. Gonçalves, R. C. Rezende, and et al., "Relevance feedback based on genetic programming for image retrieval," *Pattern Recogn. Lett.*, vol. 32, no. 1, pp. 27–37, 2011.

[163] S. R. Bulo, M. Rabbi, and M. Pelillo, "Content-based image retrieval with relevance feedback using random walks," *Pattern Recogn.*, vol. 44, no. 9, pp. 2109–2122, 2011.

[164] B. Xu, J. Bu, C. Chen, C. Wang, D. Cai, and et al., "EMR: a scalable graph-based ranking model for content-based image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 1, pp. 102–114, 2015.

[165] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas, "Query specific rank fusion for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 37, no. 4, pp. 803–815, 2015.

[166] J.-H. Su, W.-J. Huang, P. S. Yu, and V. S. Tseng, "Efficient relevance feedback for content-based image retrieval by mining user navigation patterns," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 3, pp. 360–372, 2011.

[167] J. R. Smith and S.-F. Chang, "VisualSEEk: a fully automated content-based image query system," in *Proc. of the 4th ACM Int. Conf. on Multimedia*, pp. 87–98, 1997.

[168] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 23, no. 9, pp. 947–963, 2001.

[169] J. R. Smith and S.-F. Chang, "Visually searching the web for content," *IEEE Multimedia Mag.*, no. 3, pp. 12–20, 1997.

[170] S. Sclaroff, M. La Cascia, S. Sethi, and L. Taycher, "Unifying textual and visual cues for content-based image retrieval on the world wide web," *Comput. Vis. Image Und.*, vol. 75, no. 1, pp. 86–98, 1999.

[171] M. Barrena, A. Caro, M. Durán, P. Rodríguez, J. P. Arias-Nicolás, and et al., "Qatris imanager: a general purpose cbir system," *Mach. Vision Appl.*, vol. 26, no. 4, pp. 423–442, 2015.

[172] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: an experimental comparison," *Inform. Retri.*, vol. 11, no. 2, pp. 77–107, 2008.

[173] J. E. Beaudoin, "Content-based image retrieval methods and professional image users," *J. Assoc. Inf. Syst.*, 2015.

[174] M. K. Kundu, M. Chowdhury, and M. Banerjee, "Interactive image retrieval using M-band wavelet, earth mover's distance and fuzzy relevance feedback," *Int. J. Mach. Learn. and Cyber.*, vol. 3, no. 4, pp. 1–12, 2011.

[175] D. P. Mandal, C. A. Murthy, and S. K. Pal, "Formulation of a multivalued recognition system," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 22, no. 4, pp. 607–620, 1992.

[176] M. Chowdhury, S. Das, and M. K. Kundu, "A ripplet transform based statistical framework for natural color image retrieval," in *Image Analysis and Processing, ICIAP 2013* (A. Petrosino, ed.), vol. 8156 of *Lecture Notes in Computer Science*, pp. 492–502, Springer Berlin Heidelberg, 2013.

[177] M. Chowdhury, S. Das, and M. K. Kundu, "Compact image signature generation: An application in image retrieval," in *Proc. of Int. Conf. on Computer Science and Inf. Tech.*, pp. 1–7, March 2013.

[178] M. Chowdhury and M. K. Kundu, "Comparative assessment of efficiency for content based image retrieval systems using different wavelet features and pre-classifier," *Multimed. Tools Appl.*, vol. 72, no. 3, pp. 1–36, 2014.

[179] M. K. Kundu, M. Chowdhury, and S. Rota Bulò, "A graph-based relevance feedback mechanism in content-based image retrieval," *Knowl-Based Syst.*, vol. 73, pp. 254–264, 2015.

[180] M. K. Kundu, M. Chowdhury, and S. Das, "Interactive radiographic image retrieval system," *Computer Methods and Programs in Biomedicine (Major Revision Submitted)*, 2016.

[181] G.-H. Liu, J.-Y. Yang, and Z. Li, "Content-based image retrieval using computational visual attention model," *Pattern Recogn.*, vol. 48, no. 8, pp. 2554–2566, 2015.

[182] 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.

[183] 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.

[184] G. Griffin, A. Holub, and P. Perona, "Caltech 256 object category dataset," *Technical Report UCB/CSD-04-1366, California Institute of Technology*, 2007.

[185] M. J. Huiskes and M. S. Lew, "The MIR Flickr retrieval evaluation," in *Proc. of MIR*, pp. 39–43, 2008.

[186] M. D. H. Jégou and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *Proc. of ECCV*, 2008.

[187] M. Chowdhury and M. K. Kundu, "Endoscopic image retrieval system using multiscale image features," in *Proc. of the 2nd Int. Conf. on Perception and Machine Intelligence*, pp. 64–70, 2015.

[188] H. Li, X. Wang, J. Tang, and C. Zhao, "Combining global and local matching of multiple features for precise item image retrieval," *Multimed. Syst.*, vol. 19, no. 1, pp. 37–49, 2013.

[189] F. F. Fleites, S. C. Chen, and K. Chatterjee, "A semantic index structure for multimedia retrieval," *Int. J. Semantic Comput.*, vol. 6, no. 2, pp. 155–178, 2012.

[190] C. Shahabi and M. Safar, "An experimental study of alternative shape-based image retrieval techniques," *Multimed. Tools Appl.*, vol. 32, no. 1, pp. 29–48, 2007.

[191] V. Chandrasekhar and Others, "Comparison of local feature descriptors for mobile visual search," in *Proc. IEEE Int. Conf. Image Process.*, pp. 3885–3888, 2010.

[192] M. Bober and P. Brasnett, "MPEG-7 Visual Signature Tools," in *Proc. of the 2009 IEEE Int. Conf. on Multimedia and Expo*, ICME'09, (Piscataway, NJ, USA), pp. 1540–1543, IEEE Press, 2009.

[193] P. Kakar and N. Sudha, "Exposing postprocessed copypaste forgeries through transform invariant features," *Trans. Info. For. Sec.*, vol. 7, pp. 1018–1028, June 2012.

[194] Y. Chibani and A. Houacine, "Multiscale versus multiresolution analysis for multisensor image fusion," *Proc. Eusipco 98*, 1998.

[195] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 3rd ed., 2008.

[196] C. S. Burrus, A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transform: A Primer*. Prentice Hall International Editions, 1998.

[197] M. Acharyya and M. K. Kundu, "An adaptive approach to unsupervised texture segmentation using M-band wavelet transform," *Signal Process.*, vol. 81, no. 7, pp. 1337–1356, 2001.

[198] Z. Wang, D. Li, Q. Li, and V. Tao, "M-band wavelet transform for the generation of high-resolution multispectral images," *Can. J. Remote Sens.*, vol. 34, no. 1, pp. 33–45, 2008.

[199] A. Broumandnia, J. Shanbehzadeh, and M. R. Varnoosfaderani, "Persian/arabic handwritten word recognition using m-band packet wavelet transform," *Image Vision Comput.*, vol. 26, no. 6, pp. 829–842, 2008.

[200] J. Ma and G. Plonka, "The curvelet transform," *IEEE Signal Process. Mag.*, vol. 27, no. 2, pp. 118–133, 2010.

[201] E. J. Candes and D. L. Donoho, "Continuous curvelet transform i: Resolution of the wavefront set," *Applied and Computational Harmonic Analysis*, vol. 19, no. 2, pp. 162–197, 2005.

[202] E. J. Candes and D. L. Donoho, "Continuous curvelet transform i: Discretization and frames," *Applied and Computational Harmonic Analysis*, vol. 19, no. 2, pp. 198–222, 2005.

[203] 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.

[204] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, 1983.

[205] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: theory and design," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 882–893, 1992.

[206] A. L. D. 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.

[207] S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," *Inform. Fusion*, vol. 12, no. 2, pp. 74–84, 2011.

[208] A. M. Garcia-Perez, "The perceived image: Efficient modelling of visual inhomogeneity," *Spat. Vis.*, vol. 6, no. 2, pp. 89–99, 1992.

[209] J. Xu, L. Yang, and D. Wu, "Ripplet: A new transform for image processing," *J. Vis. Commun. Image R.*, vol. 21, no. 7, pp. 627–639, 2010.

[210] J. L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, 2002.

[211] S. Das, M. Chowdhury, and M. K. Kundu, "Medical image fusion based on ripplet transform type-I," *Prog Electromagn Res B*, vol. 30, pp. 355–370, 2011.

[212] R. Eckhorn, H. J. Reitboec, M. Arndt, and P. Dicke, "Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex," *Neural Comput.*, vol. 2, no. 3, pp. 293–307, 1990.

[213] I. A. Rybak, N. A. Shevtsova, L. N. Podladchikova, and A. V. Golovan, "A visual cortex domain model and its use for visual information processing," *Neural Networks*, vol. 4, pp. 3–13, 1991.

[214] J. L. Johnson and M. L. Padgett, "PCNN models and applications," *IEEE Transactions on Neural Network*, vol. 10, pp. 480–498, May 1999.

[215] Z. Wang, Y. Ma, F. Cheng, and L. Yang, "Review of pulse-coupled neural networks," *Image Vision Comput.*, vol. 28, no. 1, pp. 5–13, 2010.

[216] M. M. Subashini and S. K. Sahoo, "Pulse coupled neural networks and its applications," *Expert Syst. Appl.*, vol. 41, pp. 3965–3974, 2014.

[217] J. L. Johnson and D. Ritter, "Observation of periodic waves in a pulse-coupled neural network," *Optical Letter*, vol. 18, no. 15, pp. 1253–1258, 1993.

[218] H. S. Ranganath, G. Kuntimad, and J. L. Johnson, "Pulse coupled neural networks for image processing," in *Proc. of Southeast Conf on Visualize the Future*, pp. 37–43, 1995.

[219] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. of Cyb.*, vol. 3, no. 3, pp. 32–57, 1973.

[220] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 1981.

[221] 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 Eng. in Medicine and Biology Society*, vol. 3, pp. 2714–2717, 2003.

[222] S. Haykin, *Neural Networks and Learning Machines*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 3rd ed., 2009.

[223] M. Chowdhury, S. Das, and M. K. Kundu, "Interactive content based image retrieval using ripplet transform and fuzzy relevance feedback," in *Perception and Machine Intelligence* (M. K. Kundu, S. Mitra, D. Mazumdar, and S. K. Pal, eds.), vol. 7143 of *Lecture Notes in Computer Science*, pp. 243–251, Springer Berlin Heidelberg, 2012.

[224] V. Vapnik, *The nature of statistical learning theory*. New-York: Springer-Verlag, 1995.

[225] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.

[226] T. V. Gestel, J. Suykens, B. Baesens, S. Viaene, J. Vanthienen, and et al., "Benchmarking least squares support vector machine classifiers," *Mach. Learn.*, vol. 54, no. 1, pp. 5–32, 2004.

[227] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *Proc. of ICCV*, pp. 1–8, 2007.

[228] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[229] 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.

[230] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, 2014.

[231] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Elsevier, 4th ed., 2008.

[232] P. A. Devijver and J. Kittler, *Pattern recognition: A statistical approach*, vol. 761. Prentice-Hall London, 1982.

[233] C. R. Rao, *Linear statistical interference and its application*. John Wiley & Sons New York, 1973.

[234] P. Mitra, C. 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.

[235] S. K. Pal and D. K. Dutta-Majumder, *Fuzzy Mathematical Approach to Pattern Recognition*. New York, NY, USA: Halsted Press, 1986.

[236] R. K. De, N. R. Pal, and S. K. Pal, "Feature analysis: Neural network and fuzzy set theoretic approaches," *Pattern Recognition*, vol. 30, no. 10, pp. 1579–1590, 1997.

[237] N. R. Pal and J. C. Bezdek, "Measuring fuzzy uncertainty," *IEEE Trans. Fuzzy Systems*, vol. 2, no. 2, pp. 107–118, 1994.

[238] S. K. Pal and B. Chakraborty, "Fuzzy set theoretic measure for automatic feature evaluation," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 16, no. 5, pp. 754–760, 1986.

[239] S. K. Pal, J. Basak, and R. K. De, "Fuzzy feature evaluation index and connectionist realization," *Information Sciences*, vol. 105, no. 1, pp. 173–188, 1998.

[240] J. Basak, R. K. De, and S. K. Pal, "Fuzzy feature evaluation index and connectionist realization II: Theoretical analysis," *Information Sciences*, vol. 111, no. 1, pp. 1–17, 1998.

[241] Y. Rubner, C. Tomasi, and L. J. Guibas, "The Earth Mover's distance as a metric for image retrieval," *Int. J. Comput. Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[242] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, pp. 79–86, 1951.

[243] L. Wang, K. L. Chan, and P. Xue, "A criterion for optimizing kernel parameters in KBDA for image retrieval," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 35, no. 3, pp. 556–562, 2005.

[244] M. Banerjee and M. K. Kundu, "Handling of impreciseness in gray level corner detection using fuzzy set theoretic approach," *Applied Soft Computing*, vol. 8, no. 4, pp. 1680–1691, 2008.

[245] M. Banerjee and M. K. Kundu, "Edge based features for content based image retrieval," *Pattern Recognition*, vol. 36, no. 11, pp. 2649–2661, 2003.

[246] 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.

[247] J. Liang and W. Song, "Clustering based on steiner points," *Int. J. Mach. Learn. and Cyber.*, vol. 3, no. 2, pp. 141–148, 2012.

[248] A. J. Graaff and A. P. Engelbrecht, "Clustering data in stationary environments with a local network neighborhood artificial immune system," *Int. J. Mach. Learn. and Cyber.*, vol. 3, no. 1, pp. 1–26, 2012.

[249] G. Guo, S. Chen, and L. Chen, "Soft subspace clustering with an improved feature weight self-adjustment mechanism," *Int. J. Mach. Learn. and Cyber.*, vol. 3, no. 1, pp. 39–49, 2012.

[250] X.-Z. Wang and C.-R. Dong, "Improving generalization of fuzzy IF–THEN rules by maximizing fuzzy entropy," *IEEE Trans. on Fuzzy Syst.*, vol. 17, no. 3, pp. 556–567, 2009.

[251] W. Jun, W. Shitong, and F.-l. Chung, "Positive and negative fuzzy rule system, extreme learning machine and image classification," *Int. J. Mach. Learn. and Cyber.*, vol. 2, no. 4, pp. 261–271, 2011.

[252] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Wavelet optimization for content-based image retrieval in medical databases," *Med. Image Anal.*, vol. 14, no. 2, pp. 227–241, 2010.

[253] J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002.

[254] 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.

[255] Y. Rui and T. Huang, "Optimizing learning in image retrieval," in *Proc. of CVPR*, vol. 1, pp. 236–243, 2000.

[256] J. Zhang, Y. Wang, Z. Zhang, and C. Xia, "Comparison of wavelet, gabor and curvelet transform for face recognition," *Opt. Appl.*, vol. 41, no. 1, pp. 183–193, 2011.

[257] H. Shan, J. Ma, and H. Yang, "Comparisons of wavelets, contourlets and curvelets in seismic denoising," *J. Appl. Geophys.*, vol. 69, no. 2, pp. 103–115, 2009.

[258] J. Lan, Y. Guan, Z. Tang, and J. Zhang, "Texture image retrieval based on nonsubsampled contourlet transform and matrix f-norm," *Appl. Math. Sci.*, vol. 7, no. 53, pp. 2613–2619, 2013.

[259] A. Mosleh and F. Zargari, "A new content based image retrieval method using contourlet transform," *J. Computer and Robotics*, vol. 2, no. 1, pp. 45–51, 2010.

[260] I. J. Sumana, G. Lu, and D. Zhang, "Comparison of curvelet and wavelet texture features for content based image retrieval," in *Proc. of ICME*, pp. 290–295, 2012.

[261] 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.

[262] 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.

[263] I. J. Sumana, M. M. Islam, D. Zhang, and G. Lu, "Content based image retrieval using curvelet transform," in *Proc. of 10th Int. Workshop on MMSP*, pp. 11–16, 2008.

[264] R. Vieux, J. Benois-Pineau, and J.-P. Domenger, "Content based image retrieval using bag-of-regions: an efficient approach," in *Proc. of MMM*, pp. 1–11, 2012.

[265] 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 Eng. in Medicine and Biology Society, pp. 4532–4535, 2007.

[266] D. T. Vollmer, T. Soule, and M. Manic, "A distance measure comparison to improve crowding in multi-modal optimization problems," in *Proc. of ISRCS*, pp. 31–36, 2010.

[267] G. Kanizsa, "Organization in Vision: Essays on Gestalt Perception," *Praeger Publishers Inc.*, 1972.

[268] P. J. Kellman and T. F. Shipley, "A theory of visual interpolation in object perception," *Cognitive Psychol.*, vol. 23, no. 2, pp. 141–221, 1991.

[269] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *Proc. of CVPR*, pp. 1–8, 2008.

[270] P. Ou and H. Wang, "Prediction of stock market index movement by ten data mining techniques," *Modern Appl. Sci.*, vol. 3, no. 12, pp. 28–42, 2009.

[271] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. of CVPR*, vol. 2, pp. 2169–2178, 2006.

[272] T. Kobayashi, A. Hidaka, and T. Kurita, "Selection of histograms of oriented gradients features for pedestrian detection," in *Proc. of ICONIP*, pp. 598–607, 2008.

[273] H. M*ü*ller and T. M. Deserno, "Content-based medical image retrieval," in *Biomedical Image Processing*, Biological and Medical Physics, Biomedical Engineering, pp. 471–494, 2011.

[274] R. O. Chàvez, M. Montes, and E. Sucar, "Image re-ranking based on relevance feedback combining internal and external similarities," in *Proc. of FLAIRS*, pp. 140–141, 2010.

[275] L. Yang and A. Hanjalic, "Supervised reranking for web image search," in *Proc. of MM*, pp. 183–192, 2010.

[276] F. L*ó*pez, J. Valiente, R. Baldrich, and M. Vanrell, "Fast surface grading using color statistics in the CIE lab space," in *Pattern Recognition and Image Analysis*, vol. 3523 of *Lecture Notes in Computer Science*, pp. 666–673, 2005.

[277] N. Doghmane, Z. Baarir, N. Terki, and A. Ouafi, "Study of effect of filters and decomposition level in wavelet image compression," *Courrier du Savoir*, vol. 3, pp. 41–45, 2003.

[278] E. Mortensen and W. Barrett, "Interactive segmentation with intelligent scissors," *Graph. Mod. in Image Process.*, vol. 60, no. 5, pp. 349–384, 1998.

[279] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, 2006.

[280] O. Ikeda, "Segmentation of faces in video footage using controlled weightson hsv color," in *Proc of 13th Scandinavian Conf. on Image Analysis*, vol. 2749 of *Lecture Notes in Computer Science*, pp. 163–170, 2003.

[281] C. H. Lee and M. F. Lin, "Ego-similarity measurement for relevance feedback," *Expert Syst. Appl.*, vol. 37, no. 1, pp. 871–877, 2010.

[282] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, and Y. Pan, "A multimedia retrieval framework based on semi-supervised ranking and relevance feedback," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 723–742, 2012.

[283] M. E. ElAlami, "A novel image retrieval model based on the most relevant features," *Knowl-Based Syst.*, vol. 24, pp. 23–32, 2011.

[284] E. Yildizer, A. M. Balci, M. Hassan, and R. Alhajj, "Efficient content-based image retrieval using multiple support vector machines ensemble," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2385–2396, 2012.

[285] X. Y. Wang, H. Y. Yang, Y. W. Li, and F. Y. Yang, "Robust color image retrieval using visual interest point features of significant bit-planes," *Digital Signal Process.*, vol. 23, no. 4, pp. 1136–1153, 2013.

[286] X. Zhou, M. Pitkanen, A. Depeursinge, and H. Müller, "A medical image retrieval application using grid technologies to speed up feature extraction in medical image retrieval," *Philippine J. of Inf. Tech.*, 2009.

[287] T. Weyand and T. Deselaers, "Combining content-based image retrieval with textual information retrieval," *Research Project, RWTH Aachen*, pp. 1–23, 2005.

[288] M. S. Simpson, D. You, M. M. Rahman, Z. Xue, D. Demner-Fushman, and *et al*, "Literature-based biomedical image classification and retrieval," *Comp. Med. Imag. and Graph.*, vol. 39, pp. 3–13, 2015.

[289] H. D. Tagare, C. C. Jaffe, and J. Duncan, "Medical image databases: a content-based retrieval approach," *J. Am. Med. Inform. Assoc.*, vol. 4, no. 3, pp. 184–198, 1997.

[290] T. M. Lehmann, M. O. Guld, C. Thies, B. Fischer, D. Keysers, and M. Kohnen, "Content-based image retrieval in medical applications for picture archiving and communication systems," in *Proc. of SPIE: Med. Img.: PACS and Integrated Medical Inf. Syst: Design and Evaluation*, vol. 5033, pp. 109–117, 2003.

[291] H. Greenspan and A. T. Pinhas, "Medical image categorization and retrieval for PACS using the GMM-KL framework," *IEEE Trans. Inf. Technol. Biomed*, vol. 11, pp. 190–202, March 2007.

[292] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level using patch-based visual words," *IEEE Trans. Med. Imaging*, vol. 30, no. 3, pp. 733–746, 2011.

[293] C. B. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: Current status and future directions," *J Digit Imaging*, vol. 2, no. 24, pp. 208–222, 2011.

[294] M. Chowdhury, S. Das, and M. K. Kundu, "Effective classification of radiographic medical images using LS-SVM and NSCT based retrieval system," in *Proc. 5th Int. Conf. on Computers & Devices for Communication*, pp. 1–4, 2012.

[295] M. O. Guld, M. Kohnen, D. Keysers, H. Schubert, B. Wein, and et al., "Quality of DICOM header information for image categorization," in *Proc. of SPIE: Int. Symp. Med. Img.*, vol. 4685, pp. 280–287, 2002.

[296] L. R. Long, S. Antani, D.-J. Lee, D. M. Krainak, and G. R. Thoma, "Biomedical information from a national collection of spine X-rays: Film to content-based retrieval," in *Proc. of SPIE: Med. Img.: PACS and Integrated Medical Inf. Syst.: Design and Evaluation*, vol. 5033, pp. 70–84, 2003.

[297] H. Müller, A. Rosser, J. P. Vallee, and A. Geissbuhler, "Comparing feature sets for content-based image retrieval in a medical-case database," in *Proc. SPIE: Med. Img.*, vol. 5371, pp. 99–109, 2004.

[298] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications – clinical benefits and future directions," *Int. J. Medical Inf.*, vol. 73, no. 1, pp. 1–23, 2004.

[299] M. Rahman, B. C. Desai, and P. Bhattacharya, "Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion," *Comput. Med. Imag. Graph.*, vol. 32, no. 2, pp. 95–108, 2008.

[300] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, "Histopathology image classification using bag of features and kernel functions," in *Proc of the 12th Conf on Artificial Intelligence in Medicine: Artificial Intelligence in Medicine*, pp. 126–135, 2009.

[301] M. Rahman, S. K. Antani, and G. R. Thoma, "A learning-based similarity fusion and filtering approach for biomedical image retrieval using SVM classification and relevance feedback," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 4, pp. 640–646, 2011.

[302] N. J. Fesharaki and H. Pourghassem, "Medical x-ray image hierarchical classification using a merging and splitting scheme in feature space," *J. Med. Signals Sens.*, vol. 3, no. 3, pp. 150–163, 2013.

[303] B. C. Ko, S. H. Kim, and J.-Y. Nam, "X-ray image classification using random forests with local wavelet-based cs-local binary patterns," *J. Digit. Imaging*, vol. 24, no. 6, pp. 1141–1151, 2011.

[304] L.-j. Zhi, S.-m. Zhang, D.-z. Zhao, H. Zhao, and S.-k. Lin, "A new two-step method for medical image retrieval," in *Proc. of Int. Conf. on Inf. Tech. and Computer Sci.*, pp. 305–309, IEEE, 2009.

[305] K. Iftekharuddin, M. Prajna, S. Samanth, and M. Indhukuri, "Mega voltage x-ray image segmentation and ambient noise removal," in *Proc. of 2nd Joint EMBS/BMES*, vol. 2, pp. 1111–1113, 2002.

[306] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," in *Proc. of SPIE: Media Forensics and Security II*, vol. 7541, pp. 754110–754121, 2010.

[307] Y. Chen, S.-K. Park, Y. Ma, and R. Ala, "A new automatic parameter setting method of a simplified PCNN for image segmentation," *IEEE Trans. Neural Netw.*, vol. 22, no. 6, pp. 880–892, 2011.

[308] Z. Cui and G. Zhang, "A novel medical image dynamic fuzzy classification model based on ridgelet transform," *J. Softw.*, vol. 5, no. 5, pp. 458–465, 2010.

[309] M. Rahman, P. Bhattacharya, and B. C. Desai, "A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 1, pp. 58–69, 2007.

[310] X. Xu, D. Zhang, and X. Zhang, "An efficient method for human face recognition

using nonsubsampled contourlet transform and support vector machine," *Opt. Appl.*, vol. XXXIX,, no. 3, pp. 601–615, 2009.

[311] K.-H. Yap and K. Wu, "Fuzzy relevance feedback in content-based image retrieval systems using radial basis function network," in *Proc. of the Int. Conf. on Multimedia and Expo*, vol. 3, pp. 177–180, 2005.

[312] D. Keysers, H. Ney, B. B. Wein, and T. M. Lehmann, "Statistical framework for model-based image retrieval in medical applications," *J. Electron. Imaging*, vol. 12, no. 1, pp. 59–68, 2003.

[313] D. K. Iakovidis, N. Pelekis, E. E. Kotsifakos, I. Kopanakis, H. Karanikas, and et al., "A pattern similarity scheme for medical image retrieval," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 442–450, 2009.

[314] J. R. Logan, T. McCashland, and D. A. Lieberman, "Evaluation of reliability in the use of endoscopic terminology," *Medinfo.*, vol. 11, pp. 396–400, 2004.

[315] A. Depeursinge, T. Zrimec, S. Busayarat, and H. M*ü*ller, "3D lung image retrieval using localized features," in *Proc. of SPIE: Medical Imaging: Computer-Aided Diagnosis*, vol. 7963, pp. 79632E–79632E–14, 2011.

List of Publications of the Author

Journals:

1. M. K. Kundu, M. Chowdhury and S. Rota Bul, "A graph-based relevance feedback mechanism in content-based image retrieval", Knowledge Based System (Elsevier), Vol. 73, pp 254–264, 2015 (**IF: 3.325**).

2. M.Chowdhury and M. K. Kundu, "Comparative assessment of efficiency for content based image retrieval systems using different wavelet features and pre-classifier", Multimed. Tools Appl.(Springer), Vol. 72, No. 3, pp. 1–36, 2014(**IF:1.331**).

3. 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 11(I):1-13; 2012.

4. M. K. Kundu, M. Chowdhury and M. Banerjee, "Interactive Image Retrieval using M-band wavelet, Earth Mover's Distance and Fuzzy Relevance Feedback", International Journal of Machine Learning and Cybernetics (Springer), Vol. 3, No. 4, pp.285–296, 2012 (**IF:1.110**).

Conference:

1. M. K. Kundu and M. Chowdhury, "Image retrieval using NN based pre-classification and fuzzy relevance feedback", INDICON 2010, pp.1-4, Jadavpur University, India, 2010.

2. M. K. Kundu, M. Chowdhury and M. Banerjee, "Interactive Image Retrieval with Wavelet Features", Proc. 4th Int. Conf. on Pattern Recognition and Machine Intelligence (PREMI 2011), Lecture Notes in Computer Science series (LNCS 6744), pp 167-172, Higher School of Economics, National Research University, Moscow, Russia, June 2011.

3. M. Chowdhury, S. Das and M. K. Kundu, “Interactive Content Based Image Retrieval using Ripplet Transform and Fuzzy Relevance Feedback”, Proc. Indo-Japan Int. Conf. on Perception and Machine Intelligence (PerMIn-12), pp. 243-251, SINP Auditorium, Salt Lake City, Kolkata organized by CDAC, Kolkata, India, January 2012.

4. M. Chowdhury, S. Das and M. K. Kundu,“Effective Classification of Radiographic Medical Images Using LS-SVM and NSCT based Retrieval System”, Proc. 5th Int. Conf. on Computers And Devices for Communication (CODEC 2012), IEEE Press, pp.1-4, Hotel Hyatt Regency, Kolkata, India, December 2012.

5. M. Chowdhury, S. Das and M. K. Kundu,“Compact Image Signature Generation: An Application in Image Retrieval”, Proc. 5th Int. Conf. on Computer Science and Inf. Tech. (CSIT 2013), IEEE Press, pp.1-7, Applied Science University, Amman, Jordan, March 27-28, 2013.

6. M. Chowdhury, S. Das and M. K. Kundu, “A ripplet Transform Based Statistical Framework for Natural Color Image Retrieval Interactive”, Proc.17th Int. Conf. Image Analysis and Processing (ICIAP 2013), Lecture Notes in Computer Science series (LNCS 8156), Vol-I, pp.492-502, Castel dell’Ove, Naples, Italy, September 9-13, 2013.

7. M. Chowdhury and M. K. Kundu, “Endoscopic Image Retrieval System Using Multiscale Image Features”,Proc. of the 2nd Int. Conf. on Perception and Machine Intelligence (PerMIn ’15), pp.64-70, SINP Auditorium, Salt Lake City, Kolkata organized by CDAC, Kolkata, India, February 2015.

Journal Communicated:

1. M. K. Kundu, M. Chowdhury and S. Das, “Interactive Radiographic Image Retrieval System”, Computer Methods and Programs in Biomedicine (Major Revision Submitted), 2016.