“The 19th century poet Emily Dickinson was reclusive to the point that she would allow a doctor to examine her only from a distance of several feet as she walked past an open door. If she were alive today, it’s likely that she would benefit from advances in medical imaging that could accommodate her standoffishness while still diagnosing the Bright’s disease that ended her life at age 55.” [1].

1.1 The Medical Image Computing Paradigm

Medical science – is one of the most important fields of natural science. It is the quest for understanding the structures and functions of the human body under all conditions of health, illness and injury [2]. This quest results in various models of human health which are immensely helpful in detecting and diagnosing illness and abnormalities, preventing diseases and disabilities, and designing therapies/treatments to alleviate the pain and suffering of the patients. Advancement toward these objectives has been so remarkable that nowadays, the average life span of humans in developed countries is almost twice its expected value a century ago [2,3].

The human body as shown by the schematic image in Fig. 1.1(a), is an incredibly complex system. In this regard, the major challenges to clinicians as well as researchers are the questions: how to acquire, process and display the massive amount of data regarding the static and dynamic properties of this human body.
Effective acquisition, processing and visualization of this data is extremely necessary, since this results in information that can be assimilated, interpreted and utilized in an efficient way to yield more useful diagnostic methods and treatment procedures [4]. Most of the times, the presentation of information as an ‘image’ is the most efficient approach to address these challenges [5, 6].

Unlike the ‘Glass Frog’, shown in Fig. 1.1(b), we cannot see through our skin to look at our organs. The skin that covers our bodies and protects our organs makes it difficult for doctors to see what is going on inside our bodies. Medical image computing tools allow doctors to see inside the human body in a non-invasive manner, so that they can diagnose and treat diseases. Different modalities (e.g. X-ray, Emission Computed Tomography (ECT), Positron Emission Tomography (PET), Single-Photon Emission Computed Tomography (SPECT), Magnetic Resonance Imaging (MRI), Ultrasonography (USG) and Computed Tomography (CT) etc.) of medical images reveal various characteristics of the human body. Some of these important characteristics are transmissivity, opacity, emissivity, reflectivity, conductivity, magnetizability etc. and changes in these characteristics with time [3, 4].

1Courtesy by Thitiya Mangprayool (downloaded from http://www.123rf.com/photo_13930940_human-anatomy.html)
2Courtesy by Steffen Reichle–The Nature Conservancy (downloaded from http://www.nature.org/cs/groups/webcontent/@web/@bolivia/documents/media/prd_021484.jpg)
This information is essential for improving human health through detection and diagnosis of illness and injury [2–4,7].

Since the discovery of X-rays by Wilhelm Conrad Rontgen in 1895, the continuous and remarkable development as well as advancement of existing and new image acquisition techniques provide different and much more important views of the anatomical, functional and molecular properties of a human body [2–4,8]. In the medical image computing paradigm, the main issue in the early years of research and development was regarding the storage and transfer of images due to limited computing capacity. The speed of pixel manipulation algorithms was an essential challenge [9–11]. At present, a method-driven modeling approach dominates and algorithms are developed on a methodological level in order to support diagnostic decisions or intervention planning. The characteristics of radiological work has gone through significant change: from the subjective interpretation of images towards diagnosis based on objective quantitative image parameters [11]. What is perhaps most remarkable about these advances is the fact that this required significant innovations in nearly all aspects of image processing, analysis and management domains [4,7]. The development of image computing and analysis systems for diagnostic support, operation planning and computer-aided surgery is a complex interdisciplinary process. Therefore, a multidisciplinary culture of research, including clinical practice, medical and biomedical research, image acquisition and image computing is needed [11]. Moreover, in an application-oriented integrative approach, advanced image processing, analysis and management methods have to be developed so that image-based medical diagnostics and patient treatment can be improved in the future.
1.2 Background and Related Works

The commonly used term ‘medical image computing’ means the utilization of digital image processing and analysis techniques for medical images used in diagnosis and treatment. In general, medical image computing paradigm covers the following five major domains as shown in Fig. 1.2 [12]. The first major domain namely image formation, includes all the steps from capturing the image to forming a digital image matrix. Generally, the output images obtained by different image formation schemes are of sub-optimal quality and not readily suitable for further use. In most cases, these images are undergone through some form of image enhancement. This refers to all types of manipulation of the obtained images, resulting in optimized and improved output images. The domain of image visualization indicates realistic visualization of the image data specifically in 2D, 3D and 4D format. The image analysis domain encompasses diverse sorts of techniques which are used for quantitative measurements and abstract interpretations of biomedical images. The last domain namely image management, sums up all
schemes that provide efficient and secure storage, communication, archiving, and access of medical images. Even though, in the last few decades, every aspect of the above mentioned domains of medical image computing paradigm has advanced significantly, but we still require substantial research activities for not only improving the existing schemes, but also to develop novel and intelligent techniques for known and hitherto unknown/unsolved problems.

Among the various domains of medical imaging paradigm described in Fig. 1.2, the present thesis contains some possible solutions to problems of four different sub-domains namely – MRI denoising, multimodal medical image fusion, classification/retrieval of medical images, effective and ethical management of digital medical images and related information through digital image watermarking techniques. The following section contains a brief review of some of the existing works related to these sub-domains.

1.2.1 Medical Image Enhancement: MRI Denoising

Generally, due to various sources of interference and other physical phenomena that affect the underlying measurement processes in imaging and data acquisition systems, medical images get deteriorated by different types of artifacts [3,13]. Some of the main artifacts in medical images are different types of noise (Rician in MRI, Speckle in USG and Poisson in SPECT etc.), limited contrast, motion artifacts and bias field etc. These artifacts limit the visual quality of the medical images. As a result, it becomes difficult for medical experts to analyze and interpret comprehensive and accurate information from the underlying medical images. Moreover, these artifacts also cause problems in subsequent high level image processing and analysis tasks. For example, the small differences that may exist between normal and abnormal tissues in a mammogram are confounded by
noise and other artifacts, often making it difficult to directly analyze and interpret the acquired images. Therefore, it is of paramount importance to improve the appearance and visual quality of the medical images by some image enhancement techniques [14–18].

In recent years, among the different modalities of medical images, MRI has gained popularity in diagnosis and treatment planning [19,20]. Over the years MRI technology has advanced significantly to be more cost-effective, providing improved spatio-temporal resolution as well as reducing the acquisition time. However, the quest of any of these objectives results in images with low signal-to-noise ratio (SNR) and exhibits significant artifacts (e.g, noise, partial voluming, background inhomogeneity and bias field etc.) which are undesirable [21,22]. These artifacts limit the accuracy of computer-aided diagnosis (CAD), clinical visual inspection and performance of subsequent high level post-processing procedures (e.g. segmentation, registration, tracking etc.). Therefore, pre/post-acquisition enhancement steps are essential for reducing these artifacts in MR images [19,23–25].

MR images contain varying amounts of noise of diverse origins: noise from stochastic variation, numerous physiological processes, eddy currents, artifacts from the magnetic susceptibilities between neighboring tissues, rigid body motion, non-rigid motion, thermal noises from the patient and electronic noises from the MRI device [14,15]. One of the main noise component in MRI is due to thermal noise of the scanned object [19,25,26]. The variance of thermal noise can be described as the sum of noise variances from independent stochastic processes representing the body, the coil and the electronics [25,27]. The inverse discrete Fourier transform (IDFT) of the raw magnetic resonance (MR) data is normally used to construct the image in a single channel signal acquisition system. The signal component of the acquired measurement exists in both real and imaginary channels. Both of these orthogonal channels are affected by additive white Gaus-
sian noise (AWGN) [28]. Generally, the magnitude of the reconstructed MR image is utilized for visual inspection and subsequent computer aided analysis. In [29], it was shown that as the magnitude of the MRI signal is the square root of the sum of the squares of two independent Gaussian variables, it follows Rician distribution. In multichannel MRI the MR image is reconstructed by combining complex images and the noise distribution can be described by non-central Chi distribution. Based on the reconstruction mechanism, the noise amplitude varies according to the spatial location of the image and can follow Rician or Chi distribution in case of parallel imaging [30].

Over the years, many approaches have been proposed to address the difficult problem of MR image denoising. Generally, these existing schemes can be grouped into three different categories: filtering approach, transform domain approach and statistical approach [25]. In filtering approach, various linear and non-linear filters are used to denoise the MR images. Some of the remarkable techniques of this category are: spatial and temporal filtering, anisotropic diffusion filtering, combination of domain and range filtering and non-local means filtering etc. In [14], McVeigh et al. have proposed the spatial and temporal filter for reducing the Gaussian noise in MR images. This filtering scheme suffers from various problems: edge blurring, loss of spatial resolution and aliasing artifacts etc. To overcome these problems, Perona et al. have developed a multiscale smoothing and edge detection scheme called the anisotropic diffusion filter (ADF) [31]. ADF significantly improves the image quality by preserving object boundaries. ADF and its variants have been successfully applied to denoise MR images by various researchers [32–35]. As a non-iterative alternative to ADF, the bilateral filter was proposed by Tomasi et al. [36]. This filter is a combination of two Gaussian filters: domain and range filters. It does not involve the solution of partial differential equation (PDE) and can be implemented in single iteration. This filtering technique is applied for MRI
denoising by Walker et al. [37] and Xie et al. [38]. Wong et al. have proposed the trilateral filtering scheme for reducing noise in medical images [39, 40]. This filtering technique uses the local structural similarity along with the geometric and photometric similarities of bilateral filter. Recently, the popular non-local means (NLM) filter for denoising natural images has come up as an effective scheme for denoising MR images [25, 41–45]. The NLM filter exploits the redundancy of information contained within an image to remove the noise [46]. Several modifications of the original NLM algorithm have been proposed by various researchers for denoising MR images [25, 42, 43, 47–56].

In transform domain based denoising approaches various multiresolution analysis (MRA) and multiscale geometric analysis (MGA) tools are used to decompose the MR images into different frequency components that can be studied with a resolution matched to their scale. Wavelets, curvelet and contourlet transforms are some of the widely used transforms for image denoising. Wavelet transform (WT) and it’s variants have been widely used for MR image denoising by various researchers [16, 57–63]. The denoising methods based on WT, are not suitable for describing the signals which have high dimensional singularities such as edges and corners. To overcome the shortcomings of WT, and to detect, represent and process high dimensional data, some new advanced transforms have been proposed [64, 65]. Recently, these transforms have been used by several researchers for denoising MR images [66–68].

A few works have been reported in the literature for MR image denoising based on statistics/estimation methods [28, 69, 70]. The schemes of this category can be further divided into several sub-groups: maximum likelihood approach [71–75], linear minimum mean square error estimation approach [76, 77] and non-parametric estimation method [24, 78–82] etc.
1.2.2 Multimodal Medical Image Fusion

Even though, the visual qualities of different modalities of medical images are improved by various image enhancement techniques. It is often not possible for a single modality of medical image to provide the medical experts with comprehensive and accurate information. The rapid and extraordinary advances in various technology domains (e.g. sensor, electrical, electronics and communication etc.) and modern instrumentations have brought an urgency for processing techniques that can efficiently integrate information from these different sources into a single composite form for interpretation. Multiple imaging modalities can complement each other to provide more information to understand the real worlds of objects than the use of a single modality [83–86]. Fig. 1.3, shows an example describing the need of such image fusion (information integration). The T1-weighted MR image in Fig. 1.3(a), contains information about soft tissues and it also shows a lesion in the brain. But, the vascular nature of the lesion is not clear. The vascular nature of the lesion is evident in the magnetic resonance angiography image of Fig. 1.3(b), but the tissue information is low. Both the lesion and its vascular nature along with the soft tissues information are evident in the fused image of the
Fig. 1.3(c). Lower cost, less time, accurate as well as reliable information are some of the immediate potential advantages of image fusion (IF). Moreover, it enables features to be distinguished that are impossible to perceive with any individual sensor. These advantages correspond to the notions of redundant, complementary, more timely, and less costly information [84,85].

As a subset of IF techniques, multimodal medical image fusion (MIF) is the integration of complementary information from multimodal source medical images. The fused images are more suitable for human visual perception and further computer processing and analysis tasks [87,88]. The combination of images from different modalities, leads to additional clinical information which is not apparent in the separate imaging modalities. A successful fusion should extract complete information from source images into the result, without introducing any artifacts or inconsistencies [89–92]. Generally, depending on the merging scheme used, existing MIF schemes can be classified into three levels [93]: pixel level, feature level, and decision level. MIF usually employs the techniques at pixel level. According to whether multiscale decomposition (MSD) is applied, the pixel level fusion methods can be roughly classified into MSD-based [94–99] or non-MSD-based methods [100–103]. Compared to the latter, the former performs better, because it can capture salient image features at different scales, which are more suitable to the mechanism of human vision [91,98,104–108].

The MIF schemes based on the methodology of simplest pixel averaging methods, normally result in undesirable side effects such as reduced contrast, loss of image fine details and unwanted degradation etc. [100,101,109–111]. In more robust weighted averaging approaches for pixel level fusion, the fused pixel is estimated as the weighted average of the corresponding input pixels. However, the weight estimation usually requires a user-specific threshold. Many other MIF methods have been developed based on intensity-hue-saturation (IHS), principal
component analysis (PCA), and the Brovey transform etc. [102, 103, 112, 113]. These techniques are easy to understand and implement. But, these schemes suffer from spectral degradation; that is, they can yield high spatial resolution fused image, but they overlook important spectral information. Soft-computing techniques (e.g., artificial neural network (ANN), fuzzy logic (FL), genetic algorithm (GA) and evolutionary algorithm (EA) etc.) have also been introduced to produce better fused images [114–120]. However, the performance of these soft-computing schemes depends on the sample images and this is not an appealing characteristic.

Because, real-world objects usually contain structures at many different scales or resolutions and multiresolution or multiscale approaches can provide a means to exploit this fact. The multiresolution techniques have attracted more and more interest in IF. WT and its variants are used in MIF extensively by various researchers [94, 121–125]. Recently, several novel MGA tools, like, curvelet, contourlet, ripplet etc. are found to offer better advantages of directionality, localization, anisotropy and multiscale properties, which cannot be perfectly achieved by traditional multiscale analysis like WT [64, 65, 126]. As a consequence, MIF schemes based on these transform domains result in superior fused images than produced by WT [95–98, 127–129].

1.2.3 Medical Image Classification and Retrieval

Nowadays, modern hospitals and health-care centers are producing large number (sometimes, 100, 000 images a day; this is about 100 GB of data) of medical images of diverse modalities. This generates huge repositories of valuable information, which in many cases is difficult to process and manage appropriately [12,130–135]. Development of automated diagnostic tools to draw quicker and easier inferences from these huge databases has become an important area of research in biomedical
engineering [136–145]. Very often, an important initial step in developing such tools is the search (classification/retrieval) of potential subjects from different categories of medical images stored in databases, before an adequate course of action can be suggested for pathological subjects.

At present, medical images can be searched in three different ways: text based search (TBS), content based search (CBS) and semantic based search (SBS) methods [136, 146, 147]. Sometimes, a combination of these above mentioned schemes are also used for medical image search purpose [148–151]. In TBS schemes, images are classified/retrieved by manually annotated and previously stored text descriptions and traditional database techniques. Even though, TBS systems are fast and easy to develop, but these have several inherent shortcomings: incompleteness or inability to search due to ambiguous or missing descriptions in the file headers, dependency on time-consuming and subjective annotation procedures, and ineffectiveness of natural language of diagnostic reports in representing the true content of the medical images [140, 147, 152–156]. Thus, to support effective image searching, various methods based on image content have been developed.

In CBS systems, images are indexed and retrieved from databases based on their visual content (low level image features) such as color, texture, shape, salient points, patches and visual words etc. [131, 136, 153, 157–163]. Initially, medical images were included in the content based image retrieval (CBIR) field as a subdomain for trials [142, 164–167]. These efforts have been embodied in several content-based medical image retrieval (CBMIR) systems [131, 168–172]. Some of the remarkable CBMIR systems are: Automatic Search and Selection Engine with Retrieval Tools (ASSERT) [173], Image Indexing by Content (I²C) [164], COntent-Based Retrieval Architecture (COBRA) [174], GNU Image Finding Tool for Medical images (MedGIFT) [175], Medical Image Access and Presentation System (MIAPS) [176], Spine Pathology and Image Retrieval System (SPIRS) [177]
and Image Retrieval for Medical Applications (IRMA) [178–180] etc. There are promising CBIR-based CAD systems, mostly specialized on a particular application domain, e.g. MR brain volumes [181,182], lung cancer [183], mammography [184], chest CT [185], or bone age assessment [186], etc. Flexible image retrieval engine (FIRE) system handles different kinds of medical data as well as non-medical data like photographic databases [187]. Some very good reviews of the works done in this domain can be found in [131,136,146,149,152,153,166,188–191].

One of the major problems in CBS for image data is ‘semantic gap’. The ‘semantic gap’ is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation. User seeks semantic similarity, but the database can only provide similarity by data processing [147,149,192,193]. An alternative to CBMIR is the semantics based medical image retrieval, where the main goal is to obtain the semantics of the images, by means of automatic image annotation [147,149,150,192–195]. The current state-of-the-art in medical image search approaches has been presented in [131,190].

Among the different modalities of medical images, the non-invasive nature of MRI together with its rich information provision, makes it the widely used method for diagnosis and treatment planning [20,196]. Various researchers are not only trying to improve the MR image quality, but also seeking novel methods for easier and quicker inference from the images. In recent years, MRI has emerged as one of the popular choice to study the human brain [197,198]. This necessitates the requirement of developing automated diagnosis tools to draw quicker and easier inferences from the brain MR images [135,197,199–203]. In recent years, various approaches of brain MR image classification [204–215] and retrieval [216–225] have been proposed by different researchers.

A lot of research works are going on for developing effective classification and
retrieval systems for medical images. Various research groups are increasingly attracted to medical image classification and retrieval problem. Many international competitions are now emerging to assist in the benchmark of feature sets, retrieval and classification schemes. One such annual competition is known as Cross Language Evaluation Forum of Image (ImageCLEF\(^1\)). Since 2004, the ImageCLEF competition has conducted text-based as well as image-based retrieval each year. The tasks increasingly involved more data, a higher number of classes, and a more complicated class structure. In 2005, it has also included a medical image annotation task. Competitions are mainly based on the IRMA project X-ray library [153, 170, 178, 226], which consists of medical radiographs taken from clinical routine at the Department of Diagnostic Radiology, Aachen University Hospital, Germany. Images are classified by medical experts according to the imaging modality, the examined region, the image orientation with respect to the body and the biological system under evaluation [137, 180, 226–234]. Some of the remarkable medical image search systems based on this IRMA radiographic images can be found in [137, 154, 170, 194, 226, 235–244].

1.2.4 Medical Image Watermarking

Due to its importance in clinical diagnosis, treatment, research and other commercial and non-commercial applications, medical information is highly valuable and critical. The modern integrated health-care delivery systems (such as hospital information systems (HISs), medical information system (MISs), Picture Archiving and Communication Systems (PACS) etc.) provide easier access, effective manipulation and efficient distribution of medical information [245–251]. On the other hand, these advances have introduced new risks for inappropriate use of medical information, given the ease with which digital form of data could be

\(^1\)http://imageclef.org
manipulated [252–256]. Consequently, there is a need to design a system for effective storage, access controlling and manipulation restriction of medical images and related information, keeping the authenticity, integrity and confidentiality requirements of medical data intact, for effective management purpose [255, 257, 258].

Recently, studies show that digital watermarking (DWM) is a promising way to facilitate sharing and remote handling of information in the digital world in a secure and private manner [259–266]. Medical image watermarking (MIW) is a subset of digital watermarking, where medical information is embedded in medical images and/or videos. When digital watermarking techniques are used in medical domain, these schemes must follow some specific requirements [257]: imperceptibility, robustness, capacity, authenticity, reversibility, intactness of region of interests (ROIs) and complexity etc. MIW schemes have several applications in medical data management domain: saving digital data storage and network bandwidth, avoid detachment of medical images and related information, confidentiality and security of patient record, non-repudiation, integrity control, authentication, indexing, access control and captioning etc. [252, 267].

In the last few decades, researchers have developed a number of MIW schemes both in spatial [268–278] as well as in transform domain [?, 279–286]. Various researchers have classified these existing MIW techniques differently based on their applicability or working methodology. For example in [255], G. Coatrieux et al. have mentioned that three different kinds of MIW schemes are available. The first class groups methods that embed information within region of non-interest (RONI) in order not to compromise the diagnosis capability [253, 269, 273, 280, 287, 288]. The second approach corresponds to reversible watermarking [269, 284, 285, 289, 290]. The third approach consists in using classical watermarking methods while minimizing the distortion [246, 291, 292].

According to Navas et al. [257], MIW algorithms reported in literature can be
divided into two categories: The algorithms that focus on tamper detection and authentication [272, 280, 287–289, 293–296]. The tamper detection algorithms use such watermarks which are able to localize the changes or alterations where the tampering was done. The second category of algorithms focus on electronic health record (EHR) data hiding [246, 254, 269, 280, 286, 292, 297–299]. These techniques give more importance in hiding higher payload in image, keeping the imperceptibility very high.

Researchers proposed watermarking techniques and reported findings in the literature to satisfy both integrity and confidentiality requirements [247, 300–307], while hiding EHR in medical image to make it more usable. Both fragile and robust watermarking techniques are used for integrity control and EHR hiding. Anand et al. have used least significant bit (LSB) plane technique in spatial domain to insert patient information encrypted using a log function in medical images [246]. In [254], Cho et al. have applied non-blind watermarking method to medical images both in spatial and transform domain. Zhou et al. have presented a MIW scheme that attaches digital signature and EHR into the mammographic images using the LSB replacing technique [291]. Chao et al. have proposed a secure data hiding technique based on bipolar multiple-base conversion to allow a variety of EPR data to be hidden within the same mark image [300]. In [292], Acharya et al. have adapted MIW for interleaving patient information and graphical signals with medical images to reduce storage and transmission overheads. Nayak et al. [298] have extended the work of Acharya et al. [292] in transform domain (discrete Fourier transform (DFT), discrete cosine transform (DCT) and discrete wavelet transform (DWT)). In this scheme, text files are encrypted using Rijndael algorithm and Electrocardiographic (ECG) signal is encrypted by differential pulse code modulation technique, prior to interleaving.

Shih et al. [280], have proposed a MIW technique embedding a fragile water-
mark and textual data around the ROI of a medical image based on GA in DCT domain. In [288], Giakoumaki et al. have presented a wavelet based multiple MIW approach for addressing the confidentiality protection and both origin and data authentication problems. Luo et al. have presented a lossless scheme for medical image processing [308]. Their method provides relatively high data embedding rate and original image can be recovered distortion free. Memon et al. have proposed a method to embed the watermark information in RONI. Encryption of the embedded data is also done to provide additional security using Bose-Chaudhuri-Hocquengham (BCH) in their method [303]. In [285], a robust, reversible, blind and double watermarking scheme that embeds/extracts watermarks using recursive dither modulation (RDM) scheme and differential evolution (DE) optimization in DWT domain based on singular value decomposition (SVD) is proposed. Extensive reviews on different MIW schemes can be found in [255, 257, 258, 309].

1.3 Theoretical Preliminaries

In this thesis work, several mathematical tools/techniques have been used to develop different solutions to various problems of medical image computing sub-domains. The following section begins with the theoretical preliminaries of the different tools/techniques used in the thesis. A brief description of the different quantitative measures used to objectively evaluate the performances of the proposed solutions is also included in this section.

1.3.1 Multiscale Image Transforms

Efficient representation of visual information lies at the foundation of many image processing and analysis tasks, including compression, filtering, and feature extraction etc. Efficiency of a representation refers to the ability to capture significant
information of an object of interest in a small description [310,311]. Moreover, for human visual system (HVS), it is well-known that the receptive fields in the visual cortex are characterized as being localized, oriented, and bandpass [312]. Various experiments in searching for the sparse components of (both still and time-varying) natural images produced basis images that closely resemble the aforementioned characteristics of the visual cortex [313]. More importantly, the results suggest that for a computational image representation to be efficient, it should have the following properties [310,311]: multiresolution, localization, critical sampling, directionality and anisotropy. Among these desiderata, the first three are successfully provided by the separable wavelet system. However, the last two require new challenging nonseparable constructions.

A key distinguishing feature of natural images is that they have intrinsic geometric structures. In particular, visual information is mainly contained in the geometry of object boundaries or edges. Edges or boundaries of objects cause discontinuities or singularities in image intensity. How to efficiently represent singularities in images poses a great challenge in image processing. The well known Fourier transform (FT) can only provide an efficient representation for smooth images but not for images that contain edges. It is also well known that 1D singularities in a function (which has finite duration or is periodic) destroy the sparsity of Fourier series representation of the function, which is known as Gibbs phenomenon. In contrast, WT is able to efficiently represent a function with 1D singularities [314,315]. In particular, wavelets are good at isolating the discontinuities at edge points. However, as a result of their construction by separable extension from 1D bases, wavelets in 2D cannot “see” the smoothness along the contours. In addition, separable wavelets can capture only limited directional information, which is an important and unique feature of multi-dimensional signals.

Ridgelet transform overcomes the limitations of WT [316]. It can resolve 1D
singularities along an arbitrary direction (including horizontal and vertical direction). But, since ridgelet transform is not able to resolve 2D singularities, Candes and Donoho have proposed the first generation curvelet transform based on multiscale ridgelet [64]. Curvelet transform can resolve 2D singularities along smooth curves. Similar to curvelet, contourlet [65] and ripplet [126] transforms were proposed to resolve 2D singularities. In the following section brief descriptions are given about some of MRA/MGA tools used in this thesis.

1.3.1.1 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 [65] 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 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 1.4. 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 [317]:

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

(1.1)

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 $\mu_{j,n}(t)_{n \in Z^2}$ 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 [318]. 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 1.5. As shown in Figure 1.5, corresponding to the subbands indexed, the equivalent analysis and synthesis filters are denoted using $H_k$ and $G_k$, $0 \leq k < 2^m$.

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:

$$g_k^{(l)}[n - s_k^{(l)} n]_{0 \leq k < 2^l, n \in Z^2}$$

(1.2)
\[ g[n] = 2\Pi \frac{n_1(l+1) + n_2}{N} - \psi\left(\frac{n_1l}{N} + n_2\right) \]  
(1.3)

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

\[ \psi(x) = \frac{1 - \cos(\pi x)}{\pi x} \]  
(1.4)

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

\[ W_j = \bigoplus_{k=0}^{2^l-1} W_{j,k} \]  
(1.5)

A DFB is designed to capture the high frequency content like smooth contours and directional edges. Fig. 1.6, shows the frequency partition of CNT, and Fig. 1.7,
Figure 1.7: Contourlet transform of *Barbara* image.

presents an image along with its subbands after decomposition by CNT.

1.3.1.2 Nonsubsampled Contourlet Transform

In 2006, Arthur L. da Cunha et al. have proposed an overcomplete transform called the NSCT [319]. NSCT is a fully shift-invariant, multiscale and multi-direction expansion that has a fast implementation. The CNT is not shift invariant due to the presence of the down-samplers and up-samplers in both the LP and DFB stages of CNT [319]. NSCT achieves shift-invariance property by using the Non-subsampled pyramid filter bank (NSPFB) and the Non-subsampled DFB (NSDFB).

NSPFB is a shift-invariant filtering structure accounting for the multiscale property of NSCT. This is achieved by using two-channel non-subsampled 2D filter banks. It has no downsampling or upsampling and hence shift-invariant. Perfect
reconstruction is achieved provided the filters satisfy the following identity

$$H_0(z)G_0(z) + H_1(z)G_1(z) = 1 \quad (1.6)$$

where, $H_0(z)$ is the lowpass decomposition filter, $H_1(z)$ is the highpass decomposition filter, $G_0(z)$ is the lowpass reconstruction filter, and $G_1(z)$ is the highpass reconstruction filter.

In order to obtain the multiscale decomposition, NSPFB are constructed by iterated non-subsampled filter banks. For the next level all filters are upsampled by 2 in both dimensions. Therefore, they also satisfy the perfect reconstruction identity. The equivalent filters of a $k$-th level cascading NSPFB are given by

$$H_n^{eq}(z) = \begin{cases} H_1(z^{2^{n-1}}) \prod_{j=0}^{n-2} H_0(z^{2^j}), & 1 \leq n < 2^k \\ \prod_{j=0}^{n-1} H_0(z^{2^j}), & n = 2^k \end{cases} \quad (1.7)$$

where, $z^j$ stands for $[z_1^j, z_2^j]$.

The NSDFB is constructed by eliminating the downsamplers and upsamplers of the DFB by switching off the downsamplers/upsamplers in each two channel filter bank in the DFB tree structure and upsampling the filters accordingly [319]. The outputs of the first level and second level filters are combined to get the four directional frequency decomposition. The synthesis filter bank is obtained similarly. All filter banks in the NSDFB tree structure are obtained from a single non-subsampled filter bank (NSFB) with fan filters. To obtain multidirectional decomposition, NSDFBs are iterated and to get the next level decomposition all filters are up sampled by a quincunx matrix given by

$$QM = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (1.8)$$
NSCT is obtained by combining NSPFB and NSDFB as described by the Fig. 1.8(a). The resulting filtering structure approximates the ideal partition of the frequency plane displayed in Fig. 1.8(b). It must be noted that different from the contourlet expansion NSCT has a redundancy given by $R = \sum_{j=0}^{J} 2^{4j}$, where $2^{4j}$ is the number of directions at scale $j$.

1.3.1.3 Type-I 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 [126]. 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 [64]. 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 [126].

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 [126]. 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 $\theta$ are sampled at equal-spaced intervals. $a_j$, $b_k$ and $\theta_l$ substitute $a$, $b$ and $\theta$, respectively, and satisfy that $a_j = 2^{-j}$, $b_k = [c \cdot 2^{-j} \cdot k_1, 2^{-j/d} \cdot k_2]^T$ and $\theta_l = \frac{2\pi}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot l$, where $k = [k_1, k_2]^T$, 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 $n, m \in \mathbb{N}$ and $n, m$ are both primes. In the frequency domain, the corresponding frequency response of ripplet function is in the form

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

(1.9)

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

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

(1.10)

$$\sum_{l=-\infty}^{+\infty} |V\left(\frac{1}{c} \cdot 2^{-\lceil j(1-1/d) \rceil} \cdot \omega - l\right)|^2 = 1$$

(1.11)

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

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

(1.12)

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

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

(1.13)

where, $R_{j,k,l}$ 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 localization in both spatial and frequency domain, it provides a more efficient and effective representation for images or 2D signals.

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

Jun Xu et al. have showed that RT can provide a more effective representation for images with singularities along smooth curves [126]. 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 [107]. All these experiments show that RT based image coding is suitable for representing texture or edges in images.

1.3.2 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 [320]. 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 [321]. Because these models provide a simple, effective way for studying synchronous pulse dynamics in networks, these were recognized as being very potential in image processing [322–326]. Johnson et al. carried on a number of modifications and variations to tailor this model’s performance in image processing algorithms [323,327]. This modified neural model is called 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. 1.9. 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 [322, 325]:

\[
F_{m,n}[t] = e^{-\alpha_F} F_{m,n}[t-1] + V_F \sum_{k,l} W_1^{m,n,k,l} Y_{m,n}[t-1] + S_{m,n} \quad (1.14)
\]

\[
L_{m,n}[t] = e^{-\alpha_L} L_{m,n}[t-1] + V_L \sum_{k,l} W_2^{m,n,k,l} Y_{m,n}[t-1] \quad (1.15)
\]

\[
U_{m,n}[t] = F_{m,n}[t](1 + \beta L_{m,n}[t]) \quad (1.16)
\]
In Eq.(1.14) to Eq.(1.18), the indexes $m$ and $n$ refer to the pixel location in the image, $k$ and $l$ refer to the dislocation in a symmetric neighborhood around one pixel, and $t$ denotes the current iteration (discrete time step). Here $t$ varies from 1 to $T$ (total number of iterations). The dendritic tree is given by Eqs.(1.14)–(1.15).

The two main components $F$ and $L$ are called feeding and linking, respectively. $W_{1m,n,k,l}^{1}$ and $W_{2m,n,k,l}^{2}$ are the synaptic weight coefficients and $S$ is the external stimulus. $V_F$ and $V_L$ are normalizing constants. $\alpha_F$ and $\alpha_L$ are the time constants; generally, $\alpha_F < \alpha_L$. The linking modulation is given in Eq.(1.16), where $U_{m,n}[t]$ is the internal state of the neuron and $\beta$ is the linking parameter. The pulse generator determines the firing events in the model in Eq.(1.17). $Y_{m,n}[t]$ depends on the internal state and threshold. The dynamic threshold of the neuron is Eq.(1.18), where $V_\theta$ and $\alpha_\theta$ are normalized constant and time constant, respectively. Generally, the firing time output (time matrix) of PCNN is used for different image processing and analysis applications. The time matrix is computed as follows:

$$G(m, n) = \sum_{t=1}^{T} Y_{m,n}[t]$$

The time matrix $G$ of PCNN includes the information of the image intensity distribution as well as the spatial geometrical structures of the image, which makes it suitable for various image processing and analysis applications.

The working mechanism of PCNN can be described as follows: the input stimulus (pixel intensity) is received by the feeding element and the internal activation element combines the feeding element with the linking element. The value of $Y_{m,n}[t]$ is given by:

$$Y_{m,n}[t] = \begin{cases} 
1, & U_{m,n}[t] > \theta_{m,n}[t] \\
0, & \text{otherwise}
\end{cases}$$

(1.17)

$$\theta_{m,n}[t] = e^{-\alpha_\theta \theta_{m,n}[t-1]} + V_\theta Y_{m,n}[t]$$

(1.18)
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. More information about PCNN can be found in [325,326].

1.3.3 Least Square-Support Vector Machine

Recently, support Vector Machines (SVM) have been shown to be effective for many classification problems [328]. 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 least square version of SVM (LS-SVM) 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 obtain a linear set of equations instead of a QP problem in the dual space [329,330]. Given a training set:

\[
\{(x_i, y_i)\}_{i=1}^{N} \text{ and } y_i = \{+1, -1\}
\]  

(1.20)
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:

$$\min_{w, b, e} J(w, b, e) = \frac{1}{2} w'w + \frac{1}{2} C \sum_{i=1}^{n} e_i^2$$  \hspace{1cm} (1.21)$$

subject to the equality constraints

$$y_i[w'\varphi(x_i) + b] = 1 - e_i$$  \hspace{1cm} (1.22)$$

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.(1.21) is defined as follows:

$$\mathcal{L}(w, e_i, b, \alpha_i) = \min_{w, b, e} J(w, b, e) - \sum_{i=1}^{n} \alpha_i \{ y_i[w'\varphi(x_i) + b] - 1 + e_i \}$$  \hspace{1cm} (1.23)$$

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

$$\begin{bmatrix} 0 & -Y \\ Y & \varphi' \varphi + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix}$$  \hspace{1cm} (1.24)$$

where, $\varphi = [\varphi(x_1)'y_1, ..., \varphi(x_n)'y_n]$, $Y = [y_1, ..., y_n]$, $\mathbf{1} = [1, ..., 1]$, 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
\[ f(x) = \text{sgn}\left[\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b\right] \] (1.25)

### 1.3.4 Earth Mover’s Distance

In computer vision applications, feature distributions are often used to summarize the content of an image. Consequently, it becomes important to define a distance between two distributions. This requires in turn a notion of distance between the basic elements that appear in the distributions. EMD is such a consistent measure of distance, or dissimilarity, between two distributions of points in a space for which a ‘ground distance’ (distance measure between single features) is given [331].

Intuitively, given two distributions, one of them can be seen as a mass of earth properly spread in space, the other as a collection of holes in that same space. The EMD measures the least amount of work needed to fill the holes with earth. Here, a unit of work corresponds to transporting a unit of earth by a unit of ground truth. A distribution can be represented by a set of clusters where each cluster is represented by its cluster center (e.g., mean, mode etc.), and by the fraction (weight) of the distribution that belongs to that cluster. Such a representation is called signature of the distribution. Signatures can be of variable or fixed size. EMD reflects the minimal cost that must be paid to transform one signature into the other. The EMD has many desirable properties: it is more robust in comparison to other histogram matching techniques, in that it suffers from no arbitrary quantization problems due to the fixed binning of the latter. It allows for partial matching, and it can be applied to signatures with different sizes. When used to compare distributions that have the same overall mass, the EMD is a true metric.
Computing the EMD is based on a solution to the well known transportation problem. Assuming, that several suppliers, each with a given amount of goods, are required to supply several consumers, each with a given limited capacity. For each supplier-consumer pair, the cost of transporting a single unit of goods is given. The transportation problem is then to find a least expensive flow of goods from the suppliers to the consumers that satisfies the consumer’s demand. Matching signatures can be naturally cast as transportation problem by defining one signature as the supplier and the other as the consumer, and by setting the cost for a supplier-consumer pair to equal the ground distance between an element in the first signature and an element in the second. Intuitively, the solution is then the minimum amount of work required to transform one signature into the other.

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

Computing EMD thus becomes finding a flow \( F = [f_{ij}] \) with \( f_{ij} \) representing the flow between \( p_i \) and \( q_j \) which minimizes the overall cost.

\[
\text{WORK}(P, Q, F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d(p_i, q_j) f_{ij} \tag{1.26}
\]

subject to the constraints:

\[
f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n \tag{1.27}
\]

\[
\sum_{j=1}^{n} f_{ij} \leq w_{p_i}, 1 \leq i \leq m \tag{1.28}
\]
\[
\sum_{i=1}^{m} f_{ij} \leq w_{qj}, 1 \leq j \leq n \quad (1.29)
\]
\[
\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = \min(\sum_{i=1}^{m} w_{pi}, \sum_{j=1}^{n} w_{qj}) \quad (1.30)
\]
Constraint Eq.(1.27) ensures movement of goods from suppliers to consumers and not the other way. Constraint Eq.(1.28) defines the upper bound on the capacity of the suppliers while Eq.(1.29) represents the upper bound on the capacity of the consumers. Constraint Eq.(1.30) ensures that maximum possible supplies to be moved from suppliers (\(P\)) to consumers (\(Q\)), called the total flow. Once the solution to optimal flow is obtained EMD is defined as the work normalized by the total flow:
\[
EMD(P, Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d(p_i, q_j) f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \quad (1.31)
\]
EMD by its definition extends to distance between sets or distributions of elements, thereby facilitating partial matches. It can be shown that, EMD is a metric, if the ground distance is a metric and the total weights of two signatures are equal.

### 1.3.5 Image Quality and Quantitative Performance Measures

In this thesis, the performance effectiveness of various proposed solutions, overcoming different problems of medical image processing, analysis and management sub-domains have been evaluated through extensive experiments and comparisons. Both subjective (qualitative) as well as objective (quantitative) measures have been used for this purpose. All the solutions reported in subsequent parts of this thesis have been implemented in MATLAB, and experiments have been carried out on a PC with 2.10 GHz CPU and 2 GB RAM. In the following section, brief descriptions of the different image quality and quantitative performance measures
are given, which have been used to objectively evaluate the effectiveness of the proposed solutions.

For evaluating the performance effectiveness of the proposed medical image enhancement scheme, the image quality and quantitative performance measures used are:

**Root-Mean-Square-Error (RMSE):** RMSE is computed by the following formula:

\[
RMSE = \sqrt{\frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} (X(m,n) - Y(m,n))^2}
\]  

(1.32)

where, \(X\) and \(Y\) are two images of size \(M \times N\) and \(X(m,n)\) indicates the gray-value of the pixel of image \(X\) at position \((m,n)\). In case of denoising applications, one of the two images is the denoised image and the other one is the noise-free (reference) image. Lower value of RMSE indicates better denoised results.

**Peak-Signal-to-Noise-Ratio (PSNR):** PSNR is an approximation to human perception of reconstruction quality. Higher value of PSNR represents better output image and it is computed as follows:

\[
PSNR = 10 \times \log_{10} \left( \frac{MAX^2_L}{MSE} \right)
\]  

(1.33)

where, \(MAX_L\) indicates the maximum possible gray value of the images (in case of 8-bit gray image this is 255) and MSE represents the mean-square-error between two images \((X\) and \(Y)\) of size \(M \times N\) and is computed by the following formula:

\[
MSE = \frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} (X(m,n) - Y(m,n))^2
\]  

(1.34)
Mean Structural Similarity Index (MSSIM): It is a method for measuring the similarity between two images $X$ and $Y$ [332].

$$\text{MSSIM}(X,Y) = \frac{1}{B} \sum_{b=1}^{B} \text{SSIM}(x_j, y_j)$$ (1.35)

where, $x_j$ and $y_j$ are the image contents at the $j$-th local window; and $B$ is the number of local windows in the image. The SSIM metric is calculated on various windows of an image. The measure between two windows $x$ and $y$ of common size $N \times N$ is:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$ (1.36)

where, $\mu_p$ and $\sigma_p^2$ represents the average and variance of $p$, respectively; $p = \{x, y\}$. The term $\sigma_{xy}$ indicates the covariance of $x$ and $y$. $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ are two variables to stabilize the division with weak denominator and $L$ indicates the dynamic range of the pixel-values with $k_1 = 0.01$ and $k_2 = 0.03$ by default. Higher value of MSSIM represents better output.

Quality Index based on Local Variance (QILV): Based on the assumption that a great amount of the structural information of an image is coded in its local variance distribution, QILV for two images $X$ and $Y$ is defined as follows [333]:

$$\text{QILV}(X,Y) = \frac{2\mu_{V_X}\mu_{V_Y}}{\mu^2_{V_X} + \mu^2_{V_Y}} \cdot \frac{2\sigma_{V_X}\sigma_{V_Y}}{\sigma^2_{V_X} + \sigma^2_{V_Y}} \cdot \frac{\sigma_{V_XV_Y}}{\sigma_{V_X}\sigma_{V_Y}}$$ (1.37)

where, $\sigma_{V_XV_Y}$ represents the covariance between the variances of $X$ and $Y$, $\sigma_{V_Y}$ denotes the standard deviation of the local variance and $\mu_{V_X}$ indicates the mean of the local variance with locality defined as a window of size $B$.
(default $B = 11$). High value of $QILV$ means better results.

The selected image quality and quantitative performance measures used in the objective analysis for evaluating the proposed MIF schemes are as follows:

**Standard Deviation (STD):** It measures the contrast of the fused image. An image with high contrast would have a high standard deviation. It is calculated as follows:

$$STD = \sqrt{\frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} (X(m,n) - \bar{X})^2}$$  \hspace{1cm} (1.38)

where, $M \times N$ denotes the size of the image $X$ and

$$\bar{X} = \frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} |X(m,n)|$$  \hspace{1cm} (1.39)

**Entropy (EN):** The entropy of an image is a measure of its information content. It is the average number of bits needed to quantize the intensities in the image. It is defined as

$$EN = -\sum_{g=0}^{L-1} p(g) \log_2 p(g)$$  \hspace{1cm} (1.40)

where $p(g)$ is the probability of grey-level $g$, and the range of $g$ is $[0,.....,L-1]$. An image with high information content would have high entropy.

**Spatial Frequency (SF):** Spatial frequency can be used to measure the overall activity and clarity level of an image. Larger SF value denotes better fusion result and it is calculated as follows:

$$SF = \sqrt{RF^2 + CF^2}$$  \hspace{1cm} (1.41)
where, RF is the row frequency and CF is the column frequency:
\[
RF = \frac{1}{M \times (N - 1)} \sum_{m=0}^{M-1} \sum_{n=0}^{N-2} (X(m, n + 1) - X(m, n))^2
\]
(1.42)
and
\[
CF = \frac{1}{(M - 1) \times N} \sum_{m=0}^{M-2} \sum_{n=0}^{N-1} (X(m + 1, n) - X(m, n))^2
\]
(1.43)

**Mutual Information (MI):** It measures the degree of dependence between two images. A larger measure implies better quality. In case of MIF, given the fused image \(Z\) and two source images \(X\) and \(Y\) of size \(M \times N\) each, MI is defined as [334]:
\[
MI = MI(X, Z) + MI(Y, Z),
\]
(1.44)
where,
\[
MI(X, Z) = \sum_{u=1}^{L} \sum_{v=1}^{L} h_{X,Z}(u, v) \log_2 \frac{h_{X,Z}(u, v)}{h_X(u)h_Z(v)}
\]
(1.45)
where, \(h_X, h_Z\) are the normalized gray level histograms of \(X\) and \(Z\), respectively. \(h_{X,Z}\) is the joint gray level histogram of \(X\) and \(Z\), and \(L\) is the number of bins. \(MI(X, Z)\) indicates how much information the fused image \(Z\) conveys about the reference \(X\). Thus, the higher the mutual information between \(Z\) and \(X\), the more likely \(Z\) resembles the ideal \(X\).

\(Q_{XY}^{Z}\): Given the fused image \(Z\) and two source images \(X\) and \(Y\) of size \(M \times N\) each, C. S. Xydeas et al. proposed an objective image fusion performance measure \(Q_{XY}^{Z}\) as follows [335]:
\[
Q_{XY}^{Z} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} (Q_X(m, n)w_X(m, n) + Q_Y(m, n)w_Y(m, n))}{\sum_{m=1}^{M} \sum_{n=1}^{N} (w_X(m, n) + w_Y(m, n))}
\]
(1.46)
where, $Q^{XZ}(m,n) = Q^{g}_{XZ}(m,n)Q^{XZ}_{a}(m,n)$. $Q^{g}_{XZ}(m,n)$ and $Q^{XZ}_{a}(m,n)$ are the edge strength and orientation preservation values, respectively. $Q^{YZ}(m,n)$ is similarly computed. $w^{X}(m,n)$ and $w^{Y}(m,n)$ reflect the importance of $Q^{XZ}(m,n)$ and $Q^{YZ}(m,n)$, respectively. The dynamic range of $Q^{XY/Z}$ is $[0, 1]$, and it should be as close to 1 as possible.

Q$_0$: It is a universal image quality index proposed by Wang et al. [335]. $Q_0$, between the source image $X$ and the fused image $Z$ is defined as follows:

$$Q_0(X,Z) = \frac{2\sigma_{XZ} \cdot 2\bar{X}\bar{Z}}{(\sigma^2_{X} + \sigma^2_{Z}) \cdot (X^2 + Z^2)} \quad (1.47)$$

where, $\sigma_{XZ}$ represents the covariance between $X$ and $Z$. $\sigma_X$, $\sigma_Z$ indicate the standard deviations of $X$ and $Z$; and $\bar{X}$, $\bar{Z}$ represent the mean value of $X$ and $Z$, respectively. $Q_0(X,Y,Z)$ is the average between $Q_0(X,Z)$ and $Q_0(Y,Z)$:

$$Q_0(X,Y,Z) = \frac{Q_0(X,Z) + Q_0(Y,Z)}{2} \quad (1.48)$$

Note that $-1 \leq Q_0 \leq 1$, and it should be also as close to 1 as possible.

Quantitative evaluation of the proposed image classification and retrieval systems and their performance comparison with other state-of-the-art techniques have been analyzed using the following statistical measures:

**Sensitivity (true positive fraction):** is the probability that a diagnostic test is positive, given that the person has the disease.

$$Sensitivity = Recall = \frac{TP}{TP + FN} \quad (1.49)$$

**Specificity (true negative fraction):** is the probability that a diagnostic test
is negative, given that the person does not have the disease.

\[ Specificity = \frac{TN}{TN + FP} \] (1.50)

**Precision (positive predictive value):** is the probability that a person truly has the disease given that his diagnostic test was positive.

\[ Precision = \frac{TP}{TP + FP} \] (1.51)

**Accuracy:** is the probability that a diagnostic test is correctly performed.

\[ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \] (1.52)

where,

TP (True Positive) - correctly classified positive cases,

TN (True Negative) - correctly classified negative cases,

FP (False Positive) - incorrectly classified negative cases, and

FN (False Negative) - incorrectly classified positive cases.

The image quality and quantitative performance measures used to evaluate the performance of the proposed MIW schemes are:

**Weighted Peak Signal to Noise Ratio (WPSNR):** It uses the principle of redundancy of HVS toward high frequency components in images. WPSNR is nothing but PSNR weighted by HVS parameter and is expressed as:

\[ WPSNR = 10 \times \log_{10}(\frac{MAX_{L}^2}{NVF \times MSE}) \] (1.53)
where, \( NVF \) is the noise visibility function defined as

\[
NVF = NORM\left( \frac{1}{1 + \delta_{\text{block}}^2} \right)
\]

(1.54)

where, \( \delta_{\text{block}} \) is the standard deviation of block of pixels having a specific size \((8 \times 8)\) and \( NORM \) is normalization function used to normalize the \( NVF \) value in the range from zero to unity.

**Total Perceptual Error (TPE):** In this thesis, visual degradation due to watermark embedding is quantitatively measured using the TPE measurement calculated from the Watson Metric [336, 337]. Lower the value of TPE the better the result.

### 1.4 Scope and Contributions of the Thesis

In this thesis work, several novel solutions have been proposed to solve some diverse problems from four different fields of medical image computing paradigm: medical image enhancement (MRI denoising), multimodal MIF, classification and retrieval of medical images and effective and ethical management of digital medical images and related information through digital watermarking technique.

The guiding motivation behind this thesis work can be summarized as follows: generally, the visual quality of medical images produced by different imaging instruments get deteriorated by various types of noise. This is due to various sources of interference and other physical phenomena that affect the underlying measurement processes in imaging and data acquisition systems. Therefore, it is of paramount importance to improve the visual quality of medical images by some image enhancement techniques. Even though, assuming a medical personnel has high quality medical images for diagnosis and treatment planning. It is
often not possible for a single modality of medical image to provide the medical personnel comprehensive and accurate information. Most of the time, the medical personnel has to observe and analyze multiple medical images of different and same modalities (with different parameter settings), simultaneously, to get the required information. The use of multiple imaging modalities on a single patient, for example MRI and CT or PET requires sophisticated image fusion algorithms to get the complete and quite often complementary information. Moreover, because of the huge amount of imaging data, the existing manual methods of analysis and interpretation of medical images are tedious, time consuming, costly and subject to experience of human observer. This necessitates the requirement of developing automatic classification and retrieval techniques to draw quicker and easier inferences from the medical images. Finally, the widespread and convenient availability of digital data has spurred the search for efficient and ethical management technique for medical images and related information. Consequently, there is a need to design a system for effective storage, access controlling and manipulation restriction of medical images and related information, keeping the authenticity, integrity and confidentiality requirements of medical data intact for effective and ethical management purposes.

1.4.1 Problem Definitions

Within the previously described scope of the thesis, few problems in various medical image computing sub-domains have been identified. Specifically, the problems for which some novel efficient solutions have been proposed in this thesis can be listed as follows:

1. This problem is related to MR image denoising by the popular NLM filtering technique. Specifically, the problem is how to reduce the computational
1. Novel denoising technique for MR images, which is not only computationally efficient but also at the same time capable of producing better/comparable denoised results.

2. What could be an effective saliency measure to identify the important information (pixel/coefficients/features etc.) from multimodal source medical images? Moreover, how to incorporate intrinsic subtle details of the source medical images into the fused image to improve the result? Furthermore, how to automatically estimate the proper values of the parameters used in MIF techniques?

3. How to construct compact and effective feature vector representations by capturing the intrinsic subtle details of the medical images for efficient classification and retrieval purposes?

4. How can MIW technique be used as an all-in-one solution tool to effectively and ethically manage various critical issues of medical images and its related information? Moreover, to increase the applicability of MIW scheme, how could we make it robust to high compression and various common watermarking attacks?

1.4.2 Contributions

The research work presented in this thesis, advances knowledge in certain areas of medical image processing, analysis and management domains. Some potential contributions of the research outcomes are noted below:

1. Novel denoising technique for MR images, which is not only computationally efficient but also at the same time capable of producing better/comparable denoised results.
2. Development of novel automatic MIF schemes to integrate complementary and contrasting information from multimodal source medical images for the purpose of better diagnosis and treatment planning.

3. Designing of novel robust classification and retrieval systems for diverse modalities of medical images.


1.5 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. 1.10 and is described briefly next:

- **Chapter 1: Introduction**
The present chapter is the first one in this thesis work. This chapter contains the introductory discussion, brief review of related works and overview of rest of the following thesis. A brief theoretical exposition of various techniques/tools used in this thesis is also included here.

- **Chapter 2: Medical Image Enhancement: MRI Denoising**

  The first contributory chapter of this thesis work presents a novel noise removal approach based on PCNN adaptive improved unbiased NLM filter to overcome some of the shortcomings of existing NLM based MR image denoising techniques.

  *Some portions of the material in this chapter are from the articles [338,339] published by the author.*

- **Chapter 3: Multimodal Medical Image Fusion**

  This chapter presents two novel multimodal MIF schemes to overcome some of the shortcomings of state-of-the-art MIF methods. The first proposed method is based on MGA of NSCT and modified spatial frequency (MSF) motivated PCNN. The second MIF scheme is also based on NSCT and a reduced PCNN (RPCNN). This approach explores the use of fuzzy logic for building an efficient MIF technique based on HVS response model.

  *Some portions of the material in this chapter are from the articles [340,341] published by the author.*

- **Chapter 4: Medical Image Classification and Retrieval**

  In this chapter, two different solutions are proposed for two different medical imaging problems: brain MRI classification and medical image retrieval. The first solution is for classifying normal and abnormal brain MR images based on RT. This method is then improved by making it robust against common
MRI artifacts (rotation, varying dynamic range etc.). The second part of this chapter contains a possible solution to overcome certain shortcomings of the existing CBMIR systems. Here, a novel CBMIR scheme is proposed based on NSCT, clustering mechanism and EMD to retrieve medical images similar to the given query image.

Some portions of the material in this chapter are from the articles [342–344] published by the author.

- **Chapter 5: Medical Image Watermarking**
  Chapter 5 of this thesis work contains two different solutions to problems regarding effective and ethical management of digital medical images and related information. This results in two novel MIW approaches: the first one is based on spatial domain and the other on transform domain.

  Some portions of the material in this chapter are from the articles [345,346] published by the author.

- **Chapter 6: Conclusion and Future Work**
  The final chapter of the present thesis contains the concluding remarks and direction of future scopes of research.