Chapter 1

Introduction

1.1 Introduction

Images are a key component of effective human communication among a variety of communication schemes. Human brains interpret visual information considerably more quickly than written words, hence images capture human attention more readily than other forms of communication like texts [1]. Images are a potent tool for communicating across boundaries, cultures, and languages because they enable connections between people who may be a million miles apart. Thus, the development of image acquisition technologies is regarded as a significant turning point in human development.

Earlier images were created to preserve priceless events, and image capturing was extremely expensive. However, the recent improvements in contemporary digital image capturing technologies and data storage, have made image acquisition much simpler, affordable, and easily reachable by masses . Digital images are therefore currently utilised in a variety of practical contexts, including face recognition, texture classification, object identification, surface surveillance, clinical decision-making etc [2–9]. As an outcome of the fast advancements in RS technology, the number of high resolution RS images has greatly escalated. These images have been used in the detection of deforestation, weather forecasting, disaster rescue and many military applications [10–13]. The proper examination and management of these high resolution RS images is challenging and still a hot issue. Searching a huge dataset to locate visually similar images that matches close to given input query image is an important problem in the proper utilization of enormous RS data. Also, due to advancements in medical imaging field, many medical image modalities, including computed tomography (CT), magnetic resonance imaging (MRI), and X-rays, are becoming increasingly important in the diagnosis and treatment of diseases [5, 14]. Number of these type of images are growing rapidly with growth in number of medical institutions. Different organizations today typically save records of previously gathered medical image data in their image libraries. The effective use of these databases highly depends on accurate management of these huge collection of images. For proper patient diagnosis and to retrieve medical images of interest, it is crucial to use this database effectively. It aids in the advancement of medical scientific research too. To fulfil this requirement, an effective image retrieval system for biomedical images is essential. Also, with development in imaging technology, the number of face images is rising and these images find applications in the field of image forensic, surveillance and criminology etc. The correct retrieval of face images of interest from these huge collection of face images is crucial in many situations. Hence there is a need of an effective image retrieval technique for face images also. Image retrieval is a systematic approach to find images of interest from these huge collection of images for different applications [15]. In an image retrieval system, the operator essentially submits a query image, and the system responds by archiving a set of related images from the database. Thus an image retrieval system browses, searches and achieves related images from the image database whenever a query is submitted to it.

In accordance with the search mechanism, there are two types of image retrieval systems. One is text-based image retrieval, in which the images are retrieved based on how similar the text data in the images are to each other. In this case, the images of the database are manually and intuitively annotated with meta data before being used. A set of images with the same annotation are displayed when a user searches for a specific image using the system. To do this, the user must enter the meta data or key words linked to the search into the system. Since assigning meta data is a progressive process and human perception frequently changes from person to person, these systems are time-consuming and frequently error prone [16]. The image retrieval system is based on image content where image similarity is searched depending on image content. The 'content' of an image refers to the colour, shape and texture, the basic visual attributes existing in an image [17]. Content based image retrieval (CBIR) systems are superior to text-based approaches in retrieving data because they are exclusively focused on the visual information obtained from the images and not on human judgement. CBIR techniques have been the subject of intensive research since 1990s, and as a result, the field has grown to an interesting research area [18–21].

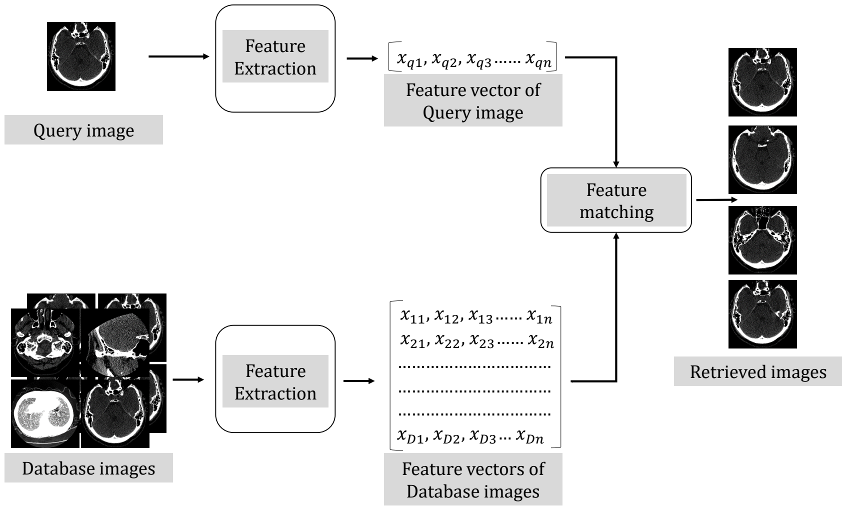

Figure. 1.1: Block diagram of a CBIR framewrok

Essentially, there are two main components that make up a CBIR system. One is the extraction of discriminative features from the images, and the other is the matching of features extracted from query and database images. During extraction of features, the visual contents of the query and database images are first analysed and feature vector is constructed. Whereas in comparison of features or similarity measurement, the distance between the query image feature vector and each database image's feature vector are computed with some distance measures such as Chi-square, Euclidean etc and finally the images are retrieved based on the closest matching distances [16, 17]. The main focus of this dissertation is to develop discriminative feature extraction techniques for CBIR applications. The schematic block diagram of the CBIR system is shown in Fig. 1.1. In the Fig. 1.1 the feature vector for query image is represented as $x_{q1}, x_{q2}, ...x_{qn}$ and for database images, the feature vector is represented as $x_{11}, x_{12}, ...x_{1n} : x_{21}, x_{22}, ...x_{2n} : ...... : x_{D1}, x_{D2}, ...x_{Dn}$ where n is the length of the feature vector and D is the number of images present in the database.

The performance of a CBIR system highly relies on discriminativeness of features for retrieval of images. This is due to the fact that feature space image representation is more effective than pixel space in terms of distinctness, storage, and computation. Typically, a feature vector is low-dimensional yet discriminative description of an image itself, with each feature containing details on a certain visual aspect of the image. If these features fail to understand the visual contents of an image properly, the accurate image retrieval will not be possible. Even if the human brain is capable of deciphering an image's visual contents, the challenge is getting a machine to comprehend that information and create a feature vector so that it can correctly retrieve or classify various images. Different works have been done to overcome these challenges and an ideal feature descriptor should possess a few qualities such as

- It should be able to extract distinct and discriminative features from images.
- It should be low dimensional for fast image matching.
- It should be invariant towards different variations e.g. rotation, illumination, blur, noise, scaling, translation and view point variation etc.

An important phase in the development of a CBIR system is the extraction of useful characteristics from the images because these features have a significant impact on how well one image retrieval system can retrieve information. Typically, based on intensity and geometry, there are two types of feature vectors. Color and texture falls into intensity based category and shape based feature vectors fall into geometry based category. These features can be extracted either from spatial domain images or from images in transform domain. These extracted features can be either local or global in nature. Low, middle and high level features are the three categories in which the features fall into [22]. Color, texture and shape, these three fundamental visual attributes of an image fall into the first category and these are extracted from images directly without any a priori knowledge. Middle level features are the logical features which require external information about the particular objects present in the images. Whereas the high level of features necessitate a deep comprehension of the significance of the objects that appear in the images. The middle and high level features are dependent on low level features and the system performance highly depends on its effectiveness. In a CBIR system, the challenge is in bridging the semantic gap issue, which is the gap between fundamental low-level features and high-level logical features.

A CBIR system extract the visual attributes of an images' in terms of low level contents, such as color, shape, or texture. These attributes carry unique characteristics of their own as well as limitations too which are very much crucial for development of an effective CBIR framework [15]. The fundamental visual characteristic of an image that is most frequently used is color as it is invariant towards changes in image size, rotation, orientation, scale and translation [23]. Different color models are usually utilized to extract color features. Three dimensional color spaces such as RGB(Red-Green-Blue) and HSV(Hue-Saturation-Value) etc. are frequently used color models. Each color model has it's own characteristics and unique property. Several categories of color feature extraction techniques exist, including histogram-based techniques (such as RGB, Hue, and Opponent histograms), moment-based techniques (such as Color moments and Color moment invariants), and color scale invariant feature transform(SIFT) techniques (such as HSV, Hue, and Opponent SIFT) etc. There are certain limitations to color features as well, including noise sensitivity, susceptibility to gray scale changes, and illumination change. Color features are unable to differentiate two images having similar color distribution. Another important visual attribute of an image is the shape information. Shape descriptors must be independent of scale, orientation and viewpoint [24, 25]. Shape descriptors can be either region-based or contour-based [8]. Region-based shape descriptors utilise both the interior region of the shape and the object shape boundary, and contour-based shape descriptors use the object shape border information only. Some examples of frequently used shape descriptors are Fourier descriptors [26–28], curvature scale spaces [29], and moment-based approaches [5, 30–34]. For human eyes, texture is one of the easily distinct attribute of image. Although there is no agreed-upon definition of texture, a few authors claim that it can be described as a measurement of the coarseness, directionality, contrast, regularity, line-likeness and roughness of an image's content. Tamura features, local binary pattern (LBP) and histogram of gradient magnitudes (HoG) are some of the examples of widely used texture features. Compared to color and shape descriptors, texture attributes have advantages. There are various texture features which represent various aspects of an image, each with some advantages and drawbacks of their own [35]. Texture features are exploited by researchers from different fields and applied in various application such as pattern recognition, object detection, image classification, image inpainting, image segmentation and image retrieval etc. Researchers have worked on development of numerous texture features to extract distinct information from images. The review on these broad range of texture feature extraction methods have been presented by different researchers earlier and in [36] recently. Based on the type of basic principle involved, the existing texture feature extraction techniques are grouped into seven different perceptions. They are [36]-

- Statistical type: The statistical characteristics of the spatial distribution of grey levels in an image are taken into account as texture descriptors in statistical approaches. The pairwise pixel spatial relationship which is a second order statistics of an image is taken as features in Gray level co-occurrence matrix (GLCM)[37, 38]. GLCM works better in conditions where the image textures are distinctly different from each other but due to large dimension, it's computational time is high. Although GLCM performs quite well in terms of processing time and complexity for document images, the memory need is also relatively high. GLCM is unable to work effectively under the presence of noise. Grey level run length matrix (GLRLM) extracts statistical information of higher order, from images as texture features. This texture feature is based on the concept that both coarse and fine texture details can be characterized by a high or low number of pixels with nearby neighbors that have identical gray values. These features have shown relatively less robust result compared to other texture features [39]. Another statistical approach of extracting texture feature is autocorrelation based. Here the dot product of the image with its shifted copies is computed. The information about periodic and similar patterns are obtained through this feature. Depending on primitives, the autocorrelation function either decreases slowly or rapidly with increase in distance which provides the measure of coarseness. Though this feature provide image coarseness details, it is not sufficient enough to be used broadly for texture feature extraction [40]. Due to the presence of rotational invariance and simpler computation property, the histogram of oriented gradients (HoG), a statistical texture descriptor which is widely utilized in image segmentation and classification of texture [41]. HoG feature is rotation invariant as it ignores the gradient orientation whereas computes the histogram over gradient magnitude of pixels in an image. Local mapped pattern-based techniques are another statistical technique for extracting texture features, and some of them are employed as feature descriptors because of their robustness to image rotation [42]. This texture descriptor is an updated version of method called local fuzzy pattern [42]. In local mapped patterns, each local square neighborhood is mapped to a histogram bin. Local energy pattern (LEP) [43], variogram [44]and Tamura [45], these three feature descriptors extract distinct information from images out of which LEP retain local texture of images significantly, variograms are easy to implement and Tamura features carry significant visual understanding of the image textures. LEP features remain unaffected by the image conditions. Based on their individual benefits, each of these three textural features are utilized in a variety of applications. Local binary pattern (LBP) [46] and its different variants have shown excellence in capturing texture details in an image. Its implementation is simple and computationally less complex, hence it has been used in multiple applications [47–50]. In the case of a shape index-based histogram technique, the shape index of an image captures the image structural information of order two [51]. Another statistical approach for capturing texture details of an image is local Weber descriptor which is based on Weber's law [52] and is utilized in gender recognition applications from face images [53].

- Structural type: Structural approach for texture feature extraction deals with the repetition of primitives or basic textural element in regular or approximately regular manner. The regions of an image with consistent gray level, edges in different directions, regular line arrangements etc. are some of the examples of image primitives. As these primitives play significant role, the identification of different primitives is an important task. For this task Laplacian of Gaussian (LoG) or difference of Gaussian (DoG) filters are considered [54]. The structural approach works well for images with regular textures and is capable of providing symbolic representation of an image. The texture which is highly random, however, is not adequately represented by structural techniques.

- Transform type: In transform type approach, the images are first transformed into a scale space whose coordinate system can closely interpret the characteristics of the textures of an image. For the purpose of extracting texture features, there are various transform based techniques available, including the texture energy approach, locally encoded transform features, and techniques based on various image transforms, such as Fourier, Gabor, wavelet, shearlet, and contourlet [36]. Both the micro and macro structural details of an image can be captured by the texture energy technique. The locally encoded transform feature histogram (LETRIST) captures discriminative texture data that is unaffected by changes in Gaussian noise, rotation, scale, and other factors [55]. Various Fourier transform based techniques are available in the literature. The translational invariance can be achieved by using the Fourier spectrum. The main limitation of Fourier transform is its inability to analyze the local texture differences. Further it comprise of details localized in the frequency domain only but not in the space. By investigating the spectrum one cannot analyse on the structural localization of texture, the frequency response of which can be seen in the spectrum. The Fourier transforms are employed in applications such as finger print identification [56] etc. In order to overcome these limitations, the Gabor or wavelet based techniques are introduced in the past. The Gabor filters exhibit improved spatial localization as compared to Fourier transforms. The Gabor transform based approaches perform multi-resolution decomposition and extracts both the frequency and orientation information. The Gabor transforms are used in biomedical applications, but these feature extraction techniques sometime capture redundant information due to non orthogonality of the Gabor filters. The image content are analysed in both frequency and spatial domain in case of Gabor transform. The wavelet transforms provide localization via dilations in the scale and in the spatial domain through translations of a function known as mother wavelet. The wavelet transforms are also well capable of examining the image content in both spatial and frequency domains. However wavelet transform does not possess translation and rotation invariant property. The limitation of wavelet transform are overcome by multiscale geometric analysis tools such as shearlet, curvelet and contourlet transform etc. The texture descriptors based on these transforms have found multiple applications including texture classification [36].

- Model type: In case of model based texture feature extraction approaches, the textures of images are represented with some mathematical models. The various types of model based texture feature extraction techniques use complex network [57], mosaic model [58], random field [59–62], fractal measures [63], gravitation [64] and Wold decomposition [65]. Complex network based method utilizes the theory of complex network to characterize the texture of an image. Complex network based techniques perform well for texture classification and it is rotation invariant upto some extent. The selection of various parameters is a tedious procedure that may be impacted by noise when a complex network-based feature is computed. The mosaic models study the geometrical processes involved in the formation of the visual attributes of an image. Under mosaic models there are two types of features: cell structure and coverage models. Random field model based approaches are simple and they enable the modelling of both isotropic and non-isotropic textures existing in images. Autoregressive, moving average, Markov random field and generalized long correlation model, these four are the members of fractal field model based approaches. Autoregressive model based approach works on the principle that there exist interaction between local pixels. For moving average models, the image is considered to be the outcome of circular convolution of a single stationary input process with a point spread function. For Markov random field models, a pixel's intensity is believed to rely on previous pixels in a chain, which creates a graphical model. The generalized correlation model describes textural images with correlations that have small order models and span significant distances. The analysis of the fractal textures of an image is less expensive in computational time. It is used to characterize natural textures. The gravitational model based approach transforms image into different states with gravitational collapse technique and features are obtained from this states. Though this model based approach finds application in classification purpose, the cost of computation is high here. The Wold decomposition based approach can capture both macro and micro textures of an image. These approaches deal with measurement of randomness, periodicity and directionality.

- Graph type: The graph based techniques transform an image into graph structure and extracts the graph structure attributes as texture features. The graph based approaches are categorized into local graph structures [65], graph of tourist walk approach [66] and shortest path base techniques [67]. Local graph based approach characterizes textures present in a local graph neighborhood. These approach takes less computation time and are invariant towards illumination, shifting and scaling. Graph of tourist walk approach works better with textures having similar visual characteristics. Statistical moments derived from the shortest paths in the landscape between pairs of points are taken as textures in shortest paths in graphs based techniques. These methods based on graphs are able to extract information from both micro and macro patterns and hence are used in different applications.

- Learning type: In learning based approaches, the features that are extracted from sample data is used to train the system which is subsequently considered for classification purpose. Deep learning [68], extreme learning [69] and vocabulary learning [70] are the three prime learning approaches for extraction of texture features. The deep learning based techniques can represent image textures significantly. However these techniques require significant amount of training data and are expensive to compute. Extreme learning machine based approach performs well for classification of textures and is computationally fast. Vocabulary learning based approaches rely on dataset and the dictionary learning is computationally complex. These different learning methods have advantages of their own and their performance is also encouraging however their computational complexity is comparatively high.

- Entropy type: Entropy based approaches assess the complexity or irregularity of visual textures in an image in terms of entropy and these approaches are the bi-dimensional expansion of well known one dimensional entropy measures [71]. Their implementation is easy and entropy is extracted directly from images. These entropy based techniques are divided into two dimensional sample [72], distribution [73] and multi-scale entropies [73]. The two-dimensional sample entropy is used to assess the uncertainty in pixel arrangements in an image. These are automated technique and invariant towards change in rotation. The two dimensional distribution entropy overcomes the limitations of two dimensional sample entropy approaches, have faster implementation time and rotation invariant. The two dimensional multi-scale entropy based techniques are better than the other two as they are robust to the size of image and takes less time to implement.

Table 1.1: Types of texture features

Sl. no.	Types	Discussion
1	Statistical	Few examples of statistical texture features are: GLCM [37, 38],GLRLM [39],HoG [41],LEP [43], variogram [44]and Tamura [45],LBP [46].
2	Structural	Some examples of image primitives include regions of an image with a consistent grayscale value, edges in different directions, and regular line arrangements, among others.
3	Transform	There are numerous transform-based techniques for extracting texture features, including the texture energy approach, locally encoded transform features, and techniques based on various image transforms, such as Fourier, Gabor, wavelet, shearlet, and contourlet [36].
4	Model	Mathematical models are used to depict the textures of an image in this type of feature. The various types of model based texture feature extraction techniques use complex network [57], mosaic model [58], random field [59–62], fractal measures [63], gravitation [64] and Wold decomposition [65].
5	Graph	The image is converted into a graph structure, and the properties of the graph structure are employed to describe the texture. The graph based approaches are categorized into local graph structures [65], graph of tourist walk approach [66] and shortest path base techniques [67].
6	Learning	The features extracted from sample data is used to train the system which is subsequently considered for classification purpose. Deep learning [68], extreme learning [69] and vocabulary learning [70] are the three prime learning approaches for extraction of texture features.
7	Entropy	It evaluates the complexity or irregularity of an image's visual textures in terms of entropy, and these methods represent the bi-dimensional expansion of well-known one-dimensional entropy measures [71]. Their implementation is easy and entropy is extracted directly from images.

In Table 1.1, various popular types of texture features are summarized. Although a wide variety of image feature descriptors have been introduced in literature so far, the deep learning based schemes have attained very good results on biomedical and RS images till now. However these techniques have an inherent drawback of complex pre-training procedure to modify the parameters. The carefully designed hand crafted features can perform at par with many basic deep learning based techniques without such pre-training constraints and are relatively simple too.

The multiscale geometric analysis tools such as curvelet, shearlet etc. have attracted considerable attention because of its powerful multiresolution and directional descriptions which is compatible with human based discrimination of images. Despite of a vast literature available on spatial domain feature descriptors, not much statistical and non-statistical concept based studies have been attempted using directional transforms.

1.2 Motivation

The characterization of textures is an important step in many image processing tasks including image retrieval, classification etc. As discussed in previous subsection, a significant amount of progress has been made in texture feature extraction in various applications, however effective and robust texture description is still demanding. Although several wavelet transform based texture description models were introduced in the literature, a few techniques using multi-scale geometrical analysis tools such as contourlet [74], curvelet [75], shearlet [76] etc. also have been proposed as an alternative over wavelets to overcome its serious drawbacks. The multi-scale geometric analysis tools allow optimal description via multiresolution as well as geometric directional analysis. In curvelet transform to incorporate directionality, the series of functions offered at various scales and locations are rotated in a variety of directions. However curvelets do not adhere to the Cartesian grid, hence rotations are one of their key drawbacks. As a result the curvelets cannot be directly implemented numerically [77]. The contourlet transform is capable of capturing the smooth curves and contours of an image effectively but it has limitation of restricted directionality [74]. The shearlets are thoroughly localized, supplies more directional sensitivity and provides the orientations that doubles at every finer scale. In this thesis, we focus on non subsampled shearlet transform (NSST) based image feature descriptors for retrieval of remote sensing (RS) and biomedical images. The NSST is a translational invariant form of shearlet transform [77]. The statistical modelling of transform domain coefficients using appropriate probability distribution function (pdf) allows to characterize the feature with only fewer parameters [78]. The performance of such techniques highly depends on the choice of accurate statistical models that best approximates the empirical pdf [79]. Several pdf models are introduced that estimate the visual details in a multiple orientation and multiple scale domain by employing wavelet transform [80].It is observed that till now no definite study has been carried out in the statistical modeling of NSST coefficients of RS images. The high resolution RS images comprise of complex texture and spatial patterns along with sometimes important isolated details. Such complicated scenes are often described by both local and global features. Therefore, it is necessary to select an appropriate combination of complementary features in the form of local and global features in order to obtain the improved results. Identifying the dominant spatial features in order to characterize the global structures and extraction of discriminative local fine details in a high resolution complex scene image is challenging and requires more investigation.

It was found that approaches for retrieving biomedical images using local patterns, such as LBP, are unable to capture the extremely minute details found in the images. In the last several years, several local bit-plane-based image retrieval approaches have been developed as a remedy to this issue [81–84]. The purpose of these types of techniques is to split the input image into a number of bit-planes based on its bit depth. The bit-planes with the highest significance carry the coarsest image details, whereas those with the least significance carry the finest image details. Effective bit-plane encoding captures coarse to fine image features. Only the most significant bit-planes have been used in a few recently developed techniques in effort to minimize the feature dimensions, which restricts the potential of the technique to capture a greater amount of intricate details [83, 84]. Further it is noticed that all the existing bit-plane based descriptors are introduced in spatial domain which show limitations in capturing the anisotropic information at different scales. As directional transforms such as NSST etc. provides multiresolution and oriented descriptions which is consistent with human observation of images, it is expected that the bit-plane decomposition and encoding concept when extended to such directional transforms can improve the results significantly. However no such approaches have been exploited in the literature so far and needs further investigation. As the biomedical images usually comprise of view of various body portions and important unique structures, textures as well as shape features, the gross or overall aspect of these biomedical images plays an important role in correct recognition, something that local features cannot do alone. In the literature, many popular local texture feature descriptors are introduced for retrieval of biomedical images because of their ability to catch fine details of the image. However , a recent few techniques have demonstrated the importance of combining both global shape and local texture features [85, 86]. In addition to the essential local texture details, the global shape features enables clear discrimination between the images for effective retrieval of biomedical images. Such multi-feature combination is not always successful in providing improved results, it is the blend of effective visual features that are mutually supportive to each other (without increase in feature length) is important in providing important results. The development of such multi-features frameworks which provides improved results without increase in feature lengths is challenging and needs more investigation. The feature vector dimension is one of the important parameter that influences the analysis of any image retrieval framework in terms of run time complexity. The run time complexity of a system consists of two important parts i.e. total feature extraction time and the total similarity matching time. The total feature extraction time is the time required for extracting the features of all the database images which usually is done in advance on any dataset. The total similarity matching time is the time required for similarity matching of query image features with the features of each image in the database. In any CBIR, especially when the database is quite huge, the total similarity matching time is more crucial than total feature extraction time which generally is performed in advance. It should be noted that the total similarity matching time completely depends upon feature dimensions while total feature extraction time relies on the algorithmic complexity.

From the above discussion, we can observe that there is a scope for further study of NSST based image retrieval techniques that supplies better image retrieval performance without increasing the feature dimension substantially. While for enhancing the retrieval performance of RS images, it is crucial to use an accurate statistical model for approximating the NSST coefficients and also it is important to choose effective local and global features with reduced feature dimensions. While improving the retrieval results of biomedical images, the bit-plane decomposition and encoding concept can be suitably exploited in the NSST domain, also an effective multi-feature set comprising of shape and texture features that are complementary to each other can be exploited in NSST domain with reduced feature dimensions. All these aspects, should hence , be investigated in developing effective retrieval frameworks for RS and biomedical images.

1.3 Thesis Contribution

The following provides an overview of the key contributions of the thesis::

1. Two NSST-based image feature descriptors that use statistical modeling of NSST coefficients for remote sensing image retrieval(RSIR) are proposed.

- The first descriptor for RSIR is based on statistical modelling of NSST coefficients using symmetric normal inverse Gaussian (SNIG) model. We have shown that the statistics of NSST coefficients of RS images is highly non-Gaussian. It has been demonstrated, via the use of the Kolmogorov-Smirnov (KS) goodness of fit test, that the four parameter SNIG distribution when compared to other probable non-Gaussian distributions provide the best fit to the image NSST coefficients. An Expectation-Maximization (EM) type of technique is employed for Maximum-Likelihood (ML) estimation of SNIG parameters. The final feature vector is constructed using the SNIG parameters estimated from the NSST detail subbands, and the mean and standard deviation estimated from the NSST approximation subband. [1]

- In the second method, we effectively incorporated the statistical characteristics of the global NSST domain (NSSTds) with the features of the local three-dimensional local ternary pattern (3D-LTP). Two-state laplacian mixture distribution (LM) models the image NSST detail subband coefficients, and EM method estimates its three parameters. Using log histogram plots and KS test we have demonstrated that the 2-state LM distribution is the most suitable distribution in comparison to other probable distributions for approximating the statistics of NSST detail subband coefficients. In addition to 2-state LM distribution parameters we also extract skewness and kurtosis statistics from the detail subbands. From NSST approximation subband, we extract only mean and standard deviation parameters. The statistical parameters from NSST detail and approximation subbands are concatenated to construct the discriminative global features. We extend the classical LTP concept to 3D-LTP by utilizing the spatial RGB planes with the intention to encode both the color feature along with local intensity variations across RGB planes. Finally a fused feature representation (NSSTds-3DLTP) is suggested to improve the discriminativeness of features by combining the effective global (NSSTds) and local (3D-LTP) features. [2]

[1] Baruah, H. G., Nath, V. K., and Hazarika, D. (2019). Remote sensing image retrieval via symmetric normal inverse Gaussian modeling of nonsubsampled shearlet transform coefficients. In International Conference on Pattern Recognition and Machine Intelligence (pp. 359-368). Springer, Cham.

[2] Baruah, H. G., Nath, V. K.,and Hazarika, D. (2022) A Remote Sensing Image Retrieval Based on 3D-Local Ternary Pattern (LTP) features and Non-subsampled Shearlet Transform (NSST) Domain Statistical Features, Computer Modeling in Engineering and Sciences, 131(1), 137-164(SCIE)

2. Two NSST based texture feature descriptors that use bit-plane decomposition in transform domain are introduced for the retrieval of biomedical images.

   - The LBP and its variants show limitations in capturing very fine details present in the images. Some descriptors decompose the input image into a no. of bit-planes in order to catch the coarse to very fine image details and each bit-plane are further encoded in an attempt to capture all the coarse to very fine image details as possible. As the spatial domain image representation lacks both multi-scale information and anisotropic details, we attempt to investigate the bit-plane decomposition and encoding concept in NSST domain for biomedical image retrieval in order to improve its discriminative power of features. The bit-plane decomposition cannot be applied directly to the NSST subband coefficients because this structure is not sufficient for the texture cues. Before feature extraction the sensitivity of NSST coefficients with regard to local variations is required to be reduced. Therefore, we perform non-linearity followed by smoothing over the coefficients prior to feature extraction. Non-linearity addition to NSST coefficients is required in order to detect the distinction between texture regions with similar average brightness and 2^{nd}-order details. With an intention to investigate the efficacy of bit-plane encoding in NSST domain, we extend the existing spatial domain local bit-plane dissimilarity adder pattern (LBPDAP) technique to NSST domain. Following the incorporation of non-linearity and smoothing over the NSST coefficients, we normalize it in the range of [0-255] values. Then we apply bit-plane decomposition to split it into eight bit-planes. In each bit-plane, we consider 'centre-neighbor' and 'neighbor-neighbor' dissimilarity association and they are combined using an 'adder' before being encoded to a value. The NSST-LBPDAP is then computed by comparing this encoded value to the corresponding centre pixel energy reference value. The NSST-LBPDAP effectively captures more fine details than existing spatial LBPDAP and exhibits encouraging retrieval results over the spatial version.

   - Motivated from the study on NSST-LBPDAP, in the second work we have proposed a novel NSST-domain local bit-plane neighbor dissimilarity pattern known as NSST-LBNDP for retrieval of CT and MRI images. We first introduce non-linearity, then apply smoothing to all NSST detail and approximation subband coefficients, and finally normalize it to 8-bit values. These normalized values are then disintegrated into an 8 number of bit-planes. In each bit-plane, with respect to each reference a powerful dissimilarity association between each neighbor and all its adjacent neighbors is exploited to obtain a dissimilarity matrix which is further thresholded and multiplied with suitable weights followed by summation to provide an encoded bit-plane value. Next, all the obtained 8 no. of bit-plane encoded values are then compared to the corresponding reference local energy feature value to get the binary values which are then suitably weighted and added to supply the ultimate NSST-LBNDP value.[3]

3. Two NSST-based feature descriptors that combine shape and texture features are proposed for biomedical image retrieval.

    - The gross perspective of biomedical images is highly crucial in assisting more clear discrimination between the images which the local texture features cannot do alone. We introduce an effective combination of spatial Zernike moment (ZM) based global shape features and a powerful NSST domain maximum of subbands local directional edge pattern based texture features called ZM-NSST-MSLDEP for biomedical image retrieval. The biomedical images consists of variety of texture and shape details. The ZMs with very less features are quite effective in describing these shape information. The influence of noise on the magnitude of ZM coefficients is also less, because ZMs are obtained through summation procedure. ZMs are orthogonal, they exhibit almost no redundancy between the coefficients so that the moments at various orders describe individual and distinct features of an image. Before extracting the local texture features from the NSST subbands, the non-linearity procedure followed by smoothing process is inserted to the NSST subband coefficients and subsequently are normalized to an 8 bit value. the local texture information is then computed from a set of NSST detail subbands available in a scale using directional maximum edge idea. In a given scale, with respect to each reference, first the edges in a particular direction is computed for each NSST detail subband and then based on the magnitude of maximum to minimum edge distributions of all subbands, the NSST domain local texture feature value in particular direction is obtained. At each scale, the NSST domain

[3]Baruah, H. G., Nath, V. K., Hazarika, D., and Hatibaruah, R. (2022). Local bit-plane neighbour dissimilarity pattern in non-subsampled shearlet transform domain for bio-medical image retrieval. Mathematical Biosciences and Engineering, 19(2), 1609-1632(SCIE).

local texture feature values at 0°,45°,90°,135°,180°,225°,270° and 315° directions are computed. The local textures from NSST approximation subband is computed via simple yet effective 'uniform' LBP patterns (LBP^{u2}). Finally the features from NSST approximation and NSST detail subbands are concatenated to form the final feature vector. The proposed global shape and local texture features are complementary to each other and exhibits improved performance than many well known local texture feature descriptors and ZM alone.[4]

- Motivated from the combination of shape and texture features, we attempt to investigate a similar hybrid framework in a low dimensional form. We propose to combine low order ZM based shape features and low dimensional but effective NSST domain texture features. The texture features are computed from singular value decomposition (SVD) of the NSST coefficients after inserion of non-linearity followed by smoothing process. We model the pdf of these singular values using Weibull distribution whose parameters are estimated using ML technique. It is demonstrated that the Weibull distribution best fits the singular values as compared to generalized Gaussian(GG) and exponential distributions. The LBP^{u2} features are extracted from NSST approximation subband. The final feature vector concatenate the features obtained from spatial ZMs, the NSST approximation and NSST detail subbands. [5]

1.4 Thesis Organization

This section provides a summary of how the chapters in this thesis are structured.

- Chapter 1: ***Introduction***

  This chapter outlines CBIR and its different components, as well as motivation for the work completed, the thesis's contribution, and the thesis' organisation.

- Chapter 2: ***Literature review***

[4] Baruah, H.G., Nath V.K. and Hazarika, D., CT image retrieval via blend of zernike moment based global shape features and non-subsampled shearlet transform (NSST) domain local texture features. (Manuscript under preparation)

[5] Baruah, H.G., Nath V.K. and Hazarika, D., Biomedical image retrieval using ZM and SVD based statistical modeling in NSST domain. (Manuscript under preparation)

In this chapter, a review of the existing literature on CBIR methods for biomedical and RS images is presented.

- Chapter 3: ***Feature descriptors based on NSST for remote sensing image retrieval***

  In the first part of the chapter, a brief discussion on NSST is presented. In the later part, the two proposed descriptors for RSIR are described in detail. Experimental results are briefly discussed in order to demonstrate the viability of the proposed descriptors.

- Chapter 4: ***NSST domain Feature descriptors that use bit-plane decomposition for biomedical image retrieval***

  Two new descriptors based on local texture features are presented in this chapter for biomedical image retrieval. Results from experiments performed on publicly accessible biomedical datasets demonstrate the efficiency of the proposed descriptors in comparison to other well-known existing techniques.

- Chapter 5: ***Biomedical image retrieval in NSST domain using shape and texture features***

  This chapter presents two new descriptors for biomedical image retrieval that utilize both shape and texture information. The experiments conducted on publicly available datasets show the efficacy of using both types of features for biomedical image retrieval.

- Chapter 6: ***Conclusions and Future scope***

  This chapter provides an overview of the conclusions reached from the completed work. Discussion of the work's future scope is also included in this chapter.