Chapter 2

Literature Review

Image texture analysis has been an intensive area of research since 1960s as it has a great variety of applications and it is vital to understand how the human vision perception works on the different textures existing in the images [87]. Julesz demonstrated that the image texture can be described by statistics of order K, or the statistics of co-occurrence for intensities at K-tuples of pixels, in the year 1962. GLCM introduced by Haralick et. al [37, 88] is based on this theory presented by Julesz. Haralick et. al suggested that a total of fourteen texture features, each dependent on a group of spatial dependency probability distribution matrices produced for a specific image block in spatial domain, can be extracted from images. Extraction of these features from images is simple. The first order texton statistics are essential to human visual perception according to Julesz's texton theory, which was put forth in the early 1980s. The development of texture analysis techniques have been significantly affected by the extensive study of this texton theory. Laws proposed to use the 'texture energy' approach on monochrome images that uses few convolution with small masks and moving average techniques[89]. Bovik et. al [90] used two dimensional Gabor functions and represented the textures as irradiance patterns with a finite range of spatial frequencies, where the main characteristic frequencies of mutually exclusive textures differ dramatically. In [91] also, the authors take benefit from the concept of Gabor filtering, and from the filtered images energy around each pixel is computed after computing linear transformation. In [92], authors have proposed to consider the Gabor wavelet features for image retrieval based on image content. Gabor wavelet transform is applied to the input image, and mean and standard deviation of the transformed coefficients are calculated to extract crucial features for image retrieval. Malik and Perona [93] tried to visualize the human perception of texture with convolution of the image with a set of even symmetric linear filters. Statistical modeling-based texture features, such as Markov random fields [94–96] and fractal models, exist as well. In these techniques, texture images are modelled with random field probability distributions. By the turn of the century, numerous researchers had produced a substantial body of work using texton and its mathematical modeling. Along with these methods, works based on the Bag of Texton (BoT) were also included to the literature. Following the development of BoT, a number of works based on Bag of Words (BoW) were also created for texture analysis. These methods generate the Texton dictionary and the representation of the images is done using orderless histograms instead of Texton dictionary. However, it was apparent that a representation of texture that is independent of changes in scale, view point, illumination, etc. was needed. As a result, the sparse local descriptors Scale Invariant Feature Transform (SIFT) [97] and Speeded Up Robust Features (SURF), which are invariant to illumination, rotation, scale and also have faster computation time, came to light.

Ojala et al. presented the idea of LBP for texture analysis in [98]. It is a dense local texture descriptor employed in texture analysis and classification. The primary advantages of LBP are its better discriminating ability and insensitivity to monotonic grayscale variations. LBP was first introduced in 1996, although many researchers at the time viewed it as an ad hoc approach. After discussing the theoretical description of LBP, the research community offered a number of effective LBP variants that have been studied ever since. The LBP and its variations have drawn a lot of attention over the past ten years. Typically, they use a circular sampling pattern to encode the relationship between the center/reference pixel and its neighbors. The classical LBP encodes the local structural details around each pixel of image by considering the relationship between the centre and neighbors in a square neighborhood of 3×3. The centre and neighbor pixels are denoted as I_c and I_n respectively, where $n \in [0, 7]$ are the circular neighbors. The difference of I_c and I_n results into a binary code given by (2.1).

$$f(x) = \begin{cases} 1, & if \quad x \geq 0 \\ 0, & else \end{cases} \tag{2.1}$$

After achieving the binary code, pattern values are obtained using (2.2).

$$LBP = \sum_{n=0}^{7} f(I_n - I_c) 2^n \tag{2.2}$$

The resultant histogram of the pattern map is regarded as the feature vector for image retrieval applications. Considering a neighborhood of size 3×3, a binary code of length 8 bit is obtained, which results into a feature dimension of size $2^8 = 256$. The key advantages of original LBP are: less complex computation, invariant towards illumination and easier implementation. Though original LBP possess these advantages, it has some disadvantages too. This increases the need for storage since it produces histograms with high dimension and poor uniqueness. LBP shows sensitiveness towards rotation in images and noise, it fails to capture large scale texture details too. Pietikainen et. al proposed a rotation-invariant version of LBP that maintains rotationally distinct patterns in order to considerably reduce the feature dimension and circumvent the restrictions of LBP. This rotation invariant version of LBP was introduced for classification of texture images [99]. Though it can provide rotation invariance upto some extent with less feature dimension, it has been observed that rotation invariant version of LBP does not perform well than original LBP for texture classification. However, in [98] Ojala et. al. showed that occurrence of few LBP patterns is more frequent compared to others. Depending on a uniformity measure, all the LBP patterns are divided into uniform and non uniform pattern. The patterns with more than two '0/1' or '1/0' transitions in a binary code are referred to as non uniform pattern and pattern with two or less than two transitions are considered as uniform patterns. However, Ojala et al. presented a rotation-invariant version of the uniform pattern for texture classification in order to provide improved rotation invariance and further lower the feature dimension [98].

After realising the potential of LBP in texture classification, researchers have found interest in local pattern based feature descriptor and its application in diverse fields. In the year 2006, Heikkila et. al proposed center symmetric local binary pattern (CSLBP)[100] by taking advantage of both SIFT and LBP descriptor. It is computationally efficient, robust to flat areas of image and change in illumination. As an alternative to comparing all of the nearby pixels to the central pixel, the CSLBP encodes the relationship between center symmetric neighbor pairs in local neighborhoods of the image. Authors presented the application of CSLBP for description of interest region. Again in 2007, Zhao and Pietkainen [101] proposed two three dimensional extension of LBP in the form of 3D-LBP and LBP in three orthogonal planes (LBP-TOP) for dynamic texture analysis for face images. Utilizing three orthogonal planes, this 3D-LBP is expanded to a 3D-LTP for the identification of human action[102].

In [103], a completed modeling of LBP (CLBP) is introduced where the discriminative power of the LBP descriptor is improved significantly via incorporation of sign, magnitude and central pixel details in a local neighborhood. The magnitude and central pixel information of a local neighborhood along with the sign details, improve the texture discrimination capability of the feature descriptor for effective texture classification. In [104], Chen et. al utilized CLBP for classification of RS images for the first time. The CLBP features are extracted in multiple scales to further enhance the discriminative power to represent the texture details of RS images.

Tan and Triggs introduced the LTP as a generalization of LBP that is more resilient to variations in illumination, more discriminant, and less susceptible to noise [105]. In the case of LBP operator, the center pixel was taken as threshold for further processing, however the LTP considers a user defined threshold value with the help of which the binary code is converted to three valued code. LTP quantizes gray levels in a zone of width surrounding the $+/-$ threshold to zero, to +1 above it, and to -1 below it. This way the feature descriptor has become less prone to noise but the resistant towards the gray scale transformations has reduced.

Murala et. al presented the idea of local tetra pattern (LTrP) which encodes the centre -neighbor relationship in a different way unlike LBP and LTP for texture image retrieval. It encodes the relationship of the centre pixel with the surrounding neighbors depending on the directions that are computed using the vertical and horizontal first-order derivatives[106]. Although LTrP is better than LBP in terms of performance, it is relatively expensive in terms of feature dimensions.

Nguyen and Caplier introduced the concept of elliptical local binary pattern (ELBP) in horizontal and vertical directions which is able to describe the micro structure details existing in face images [107]. The authors showed that the elliptical neighborhood is more suitable in case of face images, hence performs better than LBP which uses circular neighborhood.

Authors in [108] proposed local mesh pattern (LMeP) for retrieval of biomedical images. Unlike LBP, in a local neighborhood, LMeP encodes the association between neighboring pixels and the reference pixel. The relationship among the neighbors separated by a distance of 1, 2 and 3 are used for LMeP computation. Inspired by success of LMeP, the authors have introduced another local feature descriptor, local mesh peak valley edge patterns (LMePVEP). The peak/valley edges obtained by computing the first order derivatives are considered for encoding the relationship among the neighbors [109]. In these two techniques, the utilization of relationship of reference pixel with neighbors is not taken into consideration.

Local directional number pattern (LDN) was introduced in the literature by Rivera et. al in [110]. Kirsch compass masks and derivative Gaussian mask are used in LDN to obtain image maps in different directions. With a feature dimension of size 64, its performance in this case is superior to LBP and LTP.

In [111], by taking into account the similarity of the existence of 1 in the bit streams, the co-occurrence between the upper LTP and lower LTP binary codes acquired for both neighborhoods is computed. Local ternary co-occurrence pattern(LTCoP) is found to perform better in case of CT image retrieval with a fewer dimension however the selection of accurate threshold is a major challenge.

Sheng et. al [112] presented an sparse coding based LTP feature descriptor for classification of high resolution satellite scene images. Three different feature descriptors, SIFT, colour histogram and a rotation invariant version local ternary pattern histogram Fourier (LTP-HF) are used to extract complementary features from these high resolution scene images. Discrete Fourier transform based features extracted from the LTP histograms are considered as feature descriptor which provides invariance towards rotation. This approach shows substantially better results for high resolution scene image classification compared to original LTP features.

Pairwise rotation invariant co-occurrence LBP (PriCoLBP)[113] was introduced by Qi et. al for texture classification. It computes two variants of LBP, LBP^{u2} and LBP^{riu2} for a reference pixel and its adjacent pixels respectively. The order of sampling for the LBP^{u2} depends on the LBP^{riu2} computed for the reference one. Following that, the feature vector is determined by the co-occurrence of these two LBP codes. This approach is extended to multiple scales and orientations, increasing the feature dimension in the process of achieving greater improvement.

In Center symmetric local binary co-occurrence pattern (CSLBCoP) presented by [114] Verma and Raman, uses GLCM to compute the co-occurrence of CSLBP codes obtained in four different directions 0°, 45°, 90° and 135°. First the CSLBP pattern maps are obtained from the input image and then GLCM is obtained in four different directions at a distance ‘1’. The GLCM feature vectors obtained from the four matrices are concatenated to form the final feature vector.

This descriptor was shown to provide good retrieval performance for face, texture and biomedical images.

For the purpose of retrieving natural and texture-based images, Dubey et. al presented the bag of filter based LBP descriptor [115] in 2015. First, a bag of filters (BoF) is used to filter the input image. Finally, LBP features are concatenated to create the final feature vector after being generated from these filtered images. The authors state that as the filtered images contain more local information than the raw input image, they offer a considerably better discriminating feature vector than the raw input image.

In [116], local diagonal extrema pattern (LDEP) was introduced for CT image retrieval, which exploits the centre pixel and diagonal neighbor relationship in a local neighborhood of an image by exploring intensities of local diagonal extremas extracted from the first order diagonal derivatives. As LDEP considers the diagonal neighbors only, the overall feature dimension is less and shows better performance compared to LBP.

Verma and Raman introduced local tri-directional pattern (LTriDP)[117] which exploits the neighborhood pixel intensity values in three directions. Another pattern based on magnitude is also produced. The histograms obtained from all these pattern maps are concatenated to form the final feature vector. This descriptor was proposed for retrieval of texture images.

In [118], an image retrieval feature descriptor is proposed for retrieval of natural, biomedical and face images. Here the input image is first passed through multiscale Gaussian filters which results into three filtered images which constitute the 3D neighborhood for extraction of features. From these 3D neighborhood, spherically symmetric LTP features are extracted. In [119], authors have extracted features from images using the relationship between reference pixel and its neighbors in five different directions in a three dimensional cube produced from Gaussian filtered images obtained after passing the input image through multiscale Gaussian filters. The major issue or concern for these three dimensional approaches is their complex computation and higher feature dimension.

Aptoula demonstrated the application of global morphological features for retrieving RS images in [120]. The application of rotation-invariant point triplets and the circular covariance histogram was shown to provide encouraging results for RSIR. In addition, a few new descriptors, taking advantage of the Fourier power spectrum of the input's quasi-flat-zone-based scale space was also introduced by Aptoula. These approaches produced encouraging results with low feature dimension.

Bosilj et. al in [121] proposed an image retrieval technique using morphological descriptor called pattern spectra for RS images. They have considered the pattern spectra both locally and globally with a dense strategy. These approach showed encouraging results compared to other morphological approaches of RSIR.

A color image retrieval technique: multi channel decoded LBP (MDLBP) was put forward in [122]. Here, one 3:8 decoder block is utilized to compute the inter channel information of the encoded LBP maps after the R, G, and B channels of the color images have first been encoded using LBP. By combining the histograms of all the decoder output maps, the final features are constructed. The MDLBP shows a highly effective retrieval capability by outperforming different LBP versions in the retrieval of texture images. MDLBP, however, also suffers from a high dimensionality problem.

Bian et. al in [123] proposed an extended multi-structure local binary pattern (EMLBP) for classification of high resolution image scene. The images are first converted to YCbCr color space and features are extracted using EMLBP descriptor. Here, complementary features from local patches are extracted using three coupled descriptors with multi-structure sampling. Both isotropic and anisotropic image details are extracted from local image patches here and it shows improvement in performance compared to LBP and its different variants.

For the retrieval of face images, Dubey suggested a local descriptor termed as frequency decoded local binary pattern (FDLBP) [124], which was inspired by BoF-LBP. Here, input is first filtered with one low pass and four high pass filters in order to create several filtered images. Following filtering, LBP descriptors are used to encode the images. These LBP maps, which were created from low and high-pass filtered images, are then fed into decoders to produce inter-frequency data. Concatenating the features gleaned from the decoder outputs yield the final feature vector. The feature dimension of the descriptor is substantial despite the approach's good performance in the case of face image retrieval.

In [81], Dubey et. al proposed a bit-plane decomposition based local feature descriptor for retrieval of biomedical images. The 8 bit-plane slices obtained from the input image contains finer to coarser information. The relationship of bit-plane transformed values for a particular reference pixel of the image and the original reference pixel value is encoded with a binary pattern to form local bit-plane decoded pattern (LBDP) map. The histogram obtained from this is taken a feature descriptor for CT and MRI images.

Local bit-plane dissimilarity pattern (LBDISP), proposed in [82], provides a pattern map that is created by analyzing the degree of dissimilarity between adjacent pixels and the bit-plane's central pixel. When compared to LBP, both of these approaches perform better when retrieving CT images. These methods work well when they take advantage of the similarities and differences between the reference and its neighbors in each bit-plane. The retrieval rates can be increased if the differences between center-neighbor and neighbor-neighbor are carefully selected.

In [84] an another feature descriptor for biomedical image retrieval: local bit-plane adjacent neighborhood dissimilarity pattern (LBPANDP) was introduced. Here, the input image is first transformed into eight bit-planes. Then the information of dissimilarity between neighbors and its adjacent neighbors are computed for the first four significant bit-planes. This dissimilarity information is transformed into value ranging from 0 to 255. This transformed value is then compared with the reference pixel's intensity value using LBP and finally the LBPANDP feature descriptor is generated.

In [83], another bit-plane decomposition based feature descriptor for biomedical image retrieval was proposed. In each of most significant bit-planes, dissimilarity relationship of centre pixel and its neighbors and the neighbor and its adjacent neighbors in clockwise direction is computed first. Based on both of these dissimilarity information, a joint dissimilarity detail is computed using an adder. This joint dissimilarity details are then further encoded to three LBP-DAP pattern maps. Final feature vectors is constructed after concatenating the histograms obtained from this pattern maps.

Banerjee et. al suggested a feature descriptor called local neighborhood intensity pattern (LNIP) [125] for retrieval of face and texture images. This descriptor encodes the relative intensity difference of a specific pixel and center pixel of a local neighborhood with its adjacent neighbors. Both sign and magnitude is considered here to construct the pattern maps. The features extracted from both of these pattern maps are considered as a final feature vector. The consideration of mutual relationship between adjacent neighbors makes this descriptor robust towards change in illumination.

Motivated from LNIP, Ghose et. al proposed a different feature extraction technique called fractional LNIP [126]. Here, only one pattern map is obtained which represents both sign and magnitude information together. Multi-resolution information is obtained by extracting features from input image and Gaussian filtered images obtained from input image. These features have been utilized for retrieval of face, biomedical and texture images.

Roy et. al proposed local directional zigzag pattern (LDZP)[127] which uses Kirsch masks in six different directions and obtain local directional edge maps. Local zigzag pattern is then applied in local directional edge maps and finally uniform pattern histograms are obtained from the edge maps. LDZP features are insensitive towards change in illumination.

In [128], local directional relation pattern (LDRP) is proposed for retrieval of face images. Here the directional neighbors are encoded in different radius. This descriptor makes use of the reference pixel's relationship with the encoded directional neighbors. This descriptor also possess robustness towards illumination variation in face images.

Median robust extended LBP (MRELBP) is an another local feature descriptor which takes into account of both micro and macro information existing in images [129]. Here the comparison of the median values of adjacent local neighborhood is performed unlike simple LBP. Information from a larger neighborhood is considered for extraction of texture details present in images. Using MRELBP, three different types of pattern maps are obtained and histograms from these pattern maps are finally combined to produce the final feature vector. This descriptor is invariant towards noise, rotation and gray scale variations. This descriptor was shown with an application in RSIR [130].

A local feature descriptor called local directional mask maximum edge pattern (LDMaMEP) was introduced by Vipparthi et. al in [9] for image retrieval and classification. Here, the maximum edge pattern (MEP) and maximum edge position patterns (MEPP) are computed based on the magnitude and position of the maximum directional edge after computing the directional edge information of the image. Multi-resolution Gaussian filters are integrated here to boost the discriminative power of the descriptor. This descriptor showed encouraging retrieval performance for texture and biomedical images along with good recognition results for face images.

Chakraborty et. al in [131],proposed a local gradient hexa pattern (LGHP) for face image retrieval and recognition. The relationship of a reference pixel with its neighbors is utilized here for different derivative directions at various radial width. This descriptor showed improved performance for face image recognition and retrieval in the presence of varied pose, illumination , expression and lighting condition.

Vipparthi and Nagar proposed an another local feature descriptor for color image retrieval called as color directional local quinary patttern (CDLQP)[132]. The difference in gray scale between the reference pixel and its neighbors in various directions is used to calculate the directional edge information. To obtain color texture characteristics for color image retrieval, these features are individually extracted from the R, G, and B channels.

Sukhia et. al in [133] proposed another LTP based image retrieval approach for RS images. The input image is first enhanced using regularized histogram equalization and discrete cosine transform (RHE-DCT) which helps to extract discriminative features from the high resolution RS images.

An extensive volume of works are done for the development of local feature descriptors for extraction of distinct and discriminative information from different types of images such as RS, face, texture and biomedical images for effective image retrieval. However, a global approach of extracting features from whole image also exist in the literature. With global features, the macro-structural details existing in images can be captured effectively.

Authors in [134] presented a low-dimensional image scene representation known as the 'spatial envelope'. The major spatial structure of a scene is represented by a few number of perceptual dimensions (naturalness, openness, roughness, expansion, ruggedness). These spatial envelope properties make it simple to identify the GIST, an abstract representation of the image scene. The spatial envelope properties organize image scenes as human eyes do, and is able to retrieve images that share the same semantic category.

It was illustrated how effectively and robustly the color histogram could represent the objects in multicolored images in [135]. To calculate the color histograms, the image colors are discretized first, then the number of appearance of this discretized color is counted. This color histogram shows invariance towards rotation and translation.

Authors of [136] presented another approach for representing image textures globally. Here the statistical representation of shape, texture and color existing in images are considered as global features. The histogram based on Hough transform was considered to extract the shape information of the images whereas the spectral power density based on Fourier transform was used as texture features. The integration of both these descriptors showed improved performance for retrieval of RS images.

Ma and Sethi in [137] presented one approach for retrieval of infrared satellite images by utilizing shape details existing in images. The images are initially pre-processed by grouping them into blocks that are each represented by the mean and standard deviation of the intensity values. The second step is to extract regions with region growing segmentation. After that, polygonal approximation is used to characterize the extracted regions. These resultant polygons are used as features for infrared image retrieval applications.

Yang and Hao in [138] proposed another descriptor for image retrieval applications. This approach first converts the images to HSV color space. From each of the color layer, one statistical histogram by counting the number of texture elements is obtained. Color fuzzy correlogram is extracted to represent the global color feature from images. Both these features are then concatenated to construct the final feature descriptor. This descriptor showed encouraging results for texture image retrieval.

Sucharitha and Senapati in [24] suggested one Zernike moment (ZM) based CBIR technique. ZM is a global shape descriptor and is the set of orthogonal complex moment and insensitive to image size, translation and orientation. However magnitude of this descriptor does not change with image rotation and scaling. This ZM based descriptor performed better compared to other shape based descriptor due to these properties. However the higher order ZM computation is computationally complex and sensitive towards image noise. Hence low order ZM are computed and utilized for retrieval of texture images with different objects. Kumar et. al also proposed a biomedical image retrieval approach based on lower order ZM in[25].

In [139], Aggarwal et. al proposed to use orthogonal Fourier Mellin moments (OFMM) for retrieval of CT and MRI images. Due to outstanding information representation capabilities, OFMMs can represent the full image's information with a small number of coefficients. This approach also achieved outstanding results for biomedical image retrieval when tested for CT and MRI images.

Local and global details of one image carry complementary information to each other and it is observed that the fusion of these two descriptors increase discriminative power of the final feature vector formed. Authors in [140] presented one RS image classification approach based on feature descriptor which is the fusion of both local and global details of an image. The global details of these images are extracted by enhanced Gabor texture descriptor (EGTD). The local information are extracted using SIFT. The fusion of these EGTD and SIFT based feature descriptors showed encouraging classification results for RS imagery as both encode complementary details of an image.

In [141], Bian et. al proposed to fuse both local and global features for classification of high resolution scene images. Here, the global information of the image are extracted with two different descriptors based on pixel intensities and differences. The local attributes are extracted utilising local codebookless model (CLM) features. This fusion of local and global features attained superior performance for classification of high resolution scene images.

Sucharitha and Senapati in [85] introduced one feature descriptor which utilizes both local and global details existing in biomedical images for image retrieval. Here, one local feature descriptor 'local directional edge binary pattern' (LDEBP) with which the local edge information in four different directions for each pixel of an image are proposed. The global shape details are obtained with low order ZM features. It is observed that the blending of multiple features such as global shape and texture features is highly effective in improving the retrieval performance of biomedical images.

Zeng et. al in [142] proposed another RSIR technique by incorporating both local and global details existing in images. This approach extracted features using global context features and local object level features. The concatenation of these two types of features showed improved classification performance for scene images.

The spatial domain feature extraction schemes suffer from the drawback of analysing the signal only in one scale. The spatial-frequency domain feature extraction schemes [74–76] are much efficient because of their ability to analyse the visual details in a multiscale and variable orientation states. The multiscale analysis enables concurrent localization support both in the spatial as well as in the frequency state/domain.

In some of the algorithms the parameters of the distributions that are used to model the transform coefficients are used as features for image retrieval applications. The advantage of this parametric modelling is that only a few parameters are needed to characterize the texture [79]. Buccigrossi et. al [80] mentioned that the histogram of the wavelet coefficients are highly non Gaussian and these coefficients can be modeled using a variety of distributions, including modified Gaussian, Gamma, Weibull, and mixtures of Gaussian, among others.

Authors in [143] proposed to utilize both first and second order details extracted from subbands obtained after decomposing image with discrete wavelet transform (DWT). To capture first order texture information, the parameters of exponential distribution, which is used to model the wavelet subband coefficients are utilized. To obtain second order details, the eight features from the co-occurrence matrices obtained from DWT subands are considered. Both these features together, further improves the discriminating power of the final feature descriptor for texture characterization.

Do and Vetterli in [144] proposed to model the marginal distribution of the wavelet coefficients with generalized Gaussian density (GGD) distribution. These parameters are estimated with maximum likelihood estimator and used to construct the final feature vector for image retrieval of texture images.

Kokare et. al designed two different approaches for retrieval of texture images using dual tree complex wavelet transform (DTCWT) and dual-tree rotated complex wavelet filter (DT-RCWF) jointly [145]. After applying DTCWT and DT-RCWF image decomposition, the energy and standard deviation of the subbands were computed and employed as feature vectors. The features obtained from DTCWT and DT-RCWF, subbands carry complementary details and performs well for retrieval of texture images. However this approach lacks shift invariance and possess less discrimination capacity.

Selvan and Ramakrishnan [146] proposed to extract features from wavelet subbands by estimating the parameters of exponential function considered for modelling the singular values obtained after applying singular value decomposition. The parameters are estimated using maximum likelihood estimator. This type of approach is computationally simple.

Kwitt and Uhl in [147] proposed another statistical model based approach for extraction of features from texture images. Here they have used several different statistical models such as Laplace, GGD, Weibull, Reyleigh and Gamma density, to approximate the DTCWT subband coefficients. For measurement of similarity they have used Kullback-Leibler distance(KLD). This technique with statistical model parameters as features have obtained improved texture image retrieval performance.

Authors in [148] proposed to model the wavelet detail coefficients using genaralized Gamma density distribution. The three parameters obtained from each of the detail subbands are concatenated together to construct the final feature descriptor. Generalized Gamma density provides better control over shape of the model with an extra index shape parameter. Symmetrized KLD is utilized here to measure the similarity between features of query and database images.

In [149], authors introduced a facial expression recognition method by extracting LBP features from selected curvelet subbands obtained after decomposing an image with curvelet transform. The features for different classes of face expression are computed first and their average for each class is considered as the feature vector. This descriptor shows encouraging results for recognition of facial expression.

Allili in [150] put forwarded another texture feature for image retrieval by modelling the marginal distribution of wavelet coefficients with finite mixture of generalized Gaussian distribution (MoGG). This distribution extracts more discriminative information from wavelet subbands compared to single probability distribution function. The measurement of similarity is performed using KLD with Monte carlo sampling methods.

In [151], another approach for representing facial expression in curvelet transform was proposed. The input image is decomposed using curvelet transform and complete LBP (CLBP) features are extracted from these subbands to construct the final feature vector. These CLBP features consist of three components and features from all these three components are encoded to obtain the final feature description of the facial expression.

Authors in [152] introduced a wavelet transform based feature descriptor for CT images where the relationship among the neighbours in a local neighborhood is considered first, next to that the relationship with centre is encoded. Before encoding the centre-neighbour relationship, the range of centre pixel is transformed to match with the local wavelet decomposed values. The main limitation of wavelet transforms is its lack of capability in approximating the line and curve singularities existing in images. The multiscale geometrical analysis tools such as curvelet, shearlet, contourlet etc overcomes the limitations of wavelet. The contourlet transform has less directional characteristics than curvelets. The shearlets when compared to other directional transforms has no restrictions on the number of directions. The shift invariant kind of shearlets .i.e. NSST has multiscale, multidirectional or flexible directionality, localization and shift invariance features. Limited efforts have been made in the development of image feature descriptors using these directional transforms.

Shinde et. al proposed one biomedical image retrieval technique based on fast curvelet transform [153]. First of all, the input image is decomposed with curvelet transform. The directional energies obtained from these subbands are used to construct the feature vector for retrieval of biomedical images.

Srivastava and Khare [154] proposed another feature descriptor for retrieval of natural images in curvelet transform domain. LBP is applied on the subbands obtained after decomposing the image with curvelet transform. Afterwards GLCM features are extracted from these LBP pattern maps. Both these complementary features get benefited from local pattern and co-occurrence details present in images and gets encouraging performance results for retrieval of biomedical images.

Yang and Yang in [155] introduced a feature extraction approach for texture images by extracting parameters of generalized Gamma density models and LBP from subbands obtained after applying DTCWT on images. They have used generalized Gamma density model to approximate the DTCWT detail subband coefficients and its parameters as global feature descriptor. The LBP features are extracted from the approximation subband to extract the local details. Concatenation of both of these features constitute the final feature vector and obtains better performance of retrieval for texture images.

Bessel K form (BKF) distribution was utilized by Liu et. al to approximate the NSST detail coefficients as BKF distribution can effectively model the heavy tailed distributions [79]. The estimated parameters of BKF was used as texture feature for RSIR. The effective selection of accurate model for approximating the NSST coefficients is important for design of discriminative texture descriptor with less dimension.

In [156], authors suggested to consider both local and global information extracted from DTCWT subbands for retrieval of images with textures. One local descriptor called local eight direction pattern (LEDP) is proposed to extract local features from DTCWT subbands and LBP features are extracted from spatial images. However the global information are extracted by modelling both relative phase and magnitude of DTCWT subbands coefficients using wrapped Cauchy (WC) distribution. This fusion of statistical global features and local pattern based features, achieved better texture image retrieval performance. The performance of transform domain statistical modelling based image descriptors highly depends on the selection of accurate statistical model that approximates the statistics of transform coefficients and hence must be selected carefully.

Shinde et. al in [157] proposed another biomedical image retrieval technique called as local neighborhood based wavelet feature descriptor (LNFWD). The image is decomposed with triplet half band filter bank (THFB). The relationship among these wavelet subband coefficients in 3×3 neighborhood is exploited to construct the LNFWD pattern. When tested for biomedical images, it performed significantly well with a fewer dimension.

Kumar and Nagarajan in [158] proposed one local feature descriptor based on contourlet transform. They introduced contourlet tetra pattern (Cont.-TrP) which utilizes the multidirectional information provided by contourlet subbands. This Cont.-TrP encodes more directional details compared to traditional LBP. Cont.-TrP extracts both local and global information present in images. This descriptor showed significantly good performance in case of texture images.

He et. al presented one shearlet transform based rotation invariant feature descriptor for image classification in [159]. Here, energy-based features from shearlet subbands are first produced, and then these features are quantized and encoded to produce a form of rotation invariant feature for classification for texture images. The final feature vector is created by concatenating the energy histograms.

Purkait et. al in [160] proposed another local descriptor in shearlet transform domain for retrieval of texture images. This descriptor produces feature vector with less feature dimension and extracts local edge details from images. In local shearlet energy gammodian pattern (LSEGP), first the image is decomposed using shearlet transform. As the range of subband coefficients is very broad, local energy is computed for each subband. After the computation of local energy, local gammodian binary pattern (LGBP) is computed for all the subbands, the final feature vector is constructed by concatenating all the LGBP features together.

In [161], Alinsaif and Lang proposed one shearlet transform based histopathological image classification technique. Four different types of local and global texture features are extracted in complex shearlet transform domain. To tackle the issue of higher dimensionality principal component analysis (PCA) is utilized. These feature performs well for classification of histopathological images for different combination of the features considered here for feature extraction.

2.1 Summary of the literature review and research gap found

Based on the literature survey of different well known spatial and transform domain image feature descriptors, the following gaps were noticed and requires to be explored:

1. Although shearlets are well localized, provides high directional sensitivity and supplies optimal description by multiscale and geometric directional analysis, not much work related to NSST-based image retrieval have been reported.

2. In transform domain techniques, the statistical modelling based image feature descriptors are quite popular because of their low dimensionality and effectiveness. However no definite study on statistical modelling of image NSST coefficients is available in the literature. This point we consider it to be a research gap because the selection of an appropriate statistical model that approximates the image transform coefficients is highly crucial and directly affects the image retrieval performance and thus needs attention.

3. The spatial domain local pattern descriptors such as LBP cannot sufficiently capture the fine image details. To overcome such problems the recently introduced bit-plane decomposition based approaches are very helpful and have demonstrated convincing results. In spite of such encouraging alternative, a few spatial domain schemes exists in the literature and so far no related transform domain i.e NSST-based work has been attempted. This is observed to be a research gap and should be addressed, as the NSST based image description exhibits many advantages.

4. The biomedical images are generally grouped into various classes based on the part of the body or a particular portion/area they illustrate or based on the visual structural similarities of the image content. The overall shape characteristics of the biomedical image too contribute a crucial role in discriminating between various shapes and textures of the image content. Therefore due to the presence of such complex geometrical structures, patterns and textures in biomedical images, a combination of multiple type of features or a formation of effective multi-feature set consisting of shape-texture information without much elevating the dimensions is required. Not much work on low dimensional NSST based multi feature set for biomedical images are available. This is identified as a research gap where effective

NSST-based multi-feature sets consisting of shape and texture information can be explored for improved biomedical image retrieval.