Chapter 3

Feature descriptors based on NSST for remote sensing image retrieval

/

The current accessibility of remote sensing (RS) images has strengthened the progress of schemes to retrieve high resolution images with complicated exhaustive structures and varying spatial resolutions. In last one decade, image retrieval has received growing awareness in many real-world applications. The objective of remote sensing image retrieval (RSIR) is to identify relevant images based on their visual content. However correct retrieval of relevant RS images is not an easy task. In one RS image, there may exist several specific features that may elevate the complications of RSIR. For instance, within one RS image scene there may exist different objects such as 'Trees', 'Parking lot', 'Freeway', 'Building' and 'Vehicle'. And sometimes in a single scene one object may exists with different scales. Although tremendous efforts in RSIR has been made, it is still a tough task because of the presence of multiple complicated structures and spatial patterns.

The Shearlet transform (ST) overcomes the limitations of wavelet transforms and other directional transforms and delivers an optimal sparse description of an image at different scales and different orientations. Therefore, it becomes an effective alternative for analysing features of spatial images.

The transform domain statistical modelling based feature extraction tech-

niques have been quite popular in extracting global features of images. The low dimensional feature extraction is the main strength of such techniques. The performance of these techniques highly depends on the choice of accurate statistical models used for approximating the transform coefficients. The accurate statistical modelling of NSST coefficients using different distributions with reference to RSIR needs to be studied.

The techniques such as SIFT [97], BoVW [162], LBP [98], CLBP [103], multi-scale CLBP [104] and EMLBPs [123] etc. are local feature based methods capable of capturing micro structural details present in the images. The techniques such as GIST [134], Gabor [92], use of color histograms [112], DWT-GGD [144], global morphological texture descriptors [120], etc captures global or gross information from the images and are unable to describe dense local details. The local and global descriptors encode inter-dependent information present in the images and their blend could enhance the image retrieval performance. In [140], the both global and local features are combined in order to improve the discriminative power of features for retrieval of high resolution RS images. However, for an improved RSIR system, it is important to select efficient visual local and global features which can complement each other without putting much burden on the feature dimensions.

In this chapter, we investigate on statistical modeling of high resolution RS images with reference to RSIR applications. We also investigate on an appropriate combination of global and local features for improved RSIR without putting much burden on feature vector dimensions. We introduced the following two techniques in this chapter:-

1. RSIR via Symmetric normal inverse Gaussian (SNIG) modeling of NSST coefficients referred to as NSST-SNIG descriptor.
2. RSIR based on 3D-Local Ternary Pattern (3D-LTP) features and NSST domain statistical features (NSSTds) referred to as NSSTds-3DLTP descriptor.

3.1 RSIR via SNIG modeling of NSST coefficients

In this section, we have introduced one image feature descriptor for content based RSIR in NSST domain by modeling its coefficients using SNIG distribution. The

heavy-tailed non-Gaussian statistics of NSST detail subband coefficients of noise-free RS images can be well modeled using SNIG distribution. The four parameter SNIG distribution can be used to describe a variety of heavy tailed distributions. It is demonstrated through the Kolmogorov-Smirnov (KS) goodness of fit test, that the SNIG distribution most closely resembles the statistics of the detail NSST coefficients. An Expectation-Maximization (EM) kind of approach is utilized to estimate SNIG parameters in order to compute the Maximum Likelihood (ML) estimate. The NSST is first applied onto the input RS image and then SNIG parameters are computed from the image NSST detail subbands to form the feature vector.

3.1.1 Methodology

Before presenting a detailed discussion on proposed methodology, we present a brief overview of NSST.

3.1.1.1 Non subsampled shearlet transform (NSST)

ST possess great directional sensitivity and provides optimal sparse representation of higher-dimensional singularity compared to wavelets. Other multi geometric analysis approaches like curvelet and contourlet transform are introduced to overcome the constraints of wavelet. The continuous shearlet system [163] for any $\psi \in L^2(R^2)$ is given as-

$$ST(\psi) = \psi_{\lambda,\omega_o,\tau}(x) = \lambda^{-3/4}\psi(A_\lambda^{-1}S_{\omega_o}^{-1}(x-\tau)) \tag{3.1}$$

where $\lambda > 0$, $\omega_o \in R$, $\tau \in R^2$. The anisotropic dilation matrix A_λ and shear transformation matrix S_{ω_o} are responsible for controlling the resolution and orientation respectively and they can be represented as -

$$A_\lambda = \begin{bmatrix} \lambda & 0 \\ 0 & \sqrt{\lambda} \end{bmatrix} \tag{3.2}$$

and

$$S_{\omega_o} = \begin{bmatrix} 1 & \omega_o \\ 0 & 1 \end{bmatrix} \tag{3.3}$$

According to the dilation matrix A_λ, the x-directional scaling is equal to the square of the scaling in y-direction. A_λ is the dilation matrix and its general form is expressed as follows:

$$A_\lambda = diag(\lambda, \lambda^a) \tag{3.4}$$

where $\lambda \in (0, 1)$ is for controlling the degree of anisotropy. For discrete setting, $a = 1/2$, is considered. The shearing matrix S_{ω_o} controls the orientation with the help of variables related to slopes rather than the angles. For integer value of ω_o , the integer lattice remains invariant.

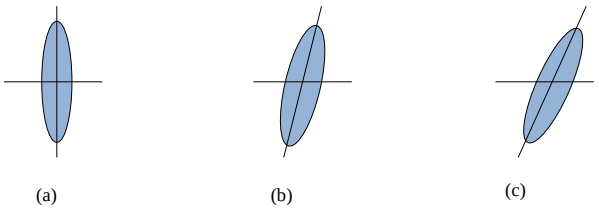


Figure. 3.1: The geometric impacts of parabolic scaling and shearing with fixed parameter λ and several parameter ω_o (a)$\omega_o = 0$ (b)$\omega_o = 1/4$ and(c) $\omega_o = 1/2$

Fig. 3.1 shows the geometric effects of parabolic scaling and shearing with fixed parameters λ and ω_o. The associated continuous ST of any $f \in L^2(R^2)$ is given by-

$$SH_\psi f(\lambda, \omega_o, \tau) =< f, \psi_{\lambda,\omega_o,\tau} > \tag{3.5}$$

SH_ψ maps the function f to the coefficients $SH_\psi f(\lambda, \omega_o, \tau)$ related with the variables scale $\lambda > 0$, the orientation $\omega_o \in R$ and location $\tau \in R^2$.

To construct continuous ST connected with a reconstruction formula, ψ

must be strongly localized and follow admissibility criterion given in [164] as -

$$\int_{R^2} \frac{|\hat{\psi}(\varrho_1, \varrho_2)|^2}{\varrho_1^2} d\varrho_2 d\varrho_1 \leq \infty \tag{3.6}$$

Each $f \in L^2(R^2)$ has the representation of the form given as,

$$f = \int_{R^n} \int_{-\infty}^{\infty} \int_0^{\infty} (f, \psi_{\lambda,\omega_o,\tau}) \psi_{\lambda,\omega_o,\tau} \frac{d\lambda}{\lambda^3} d\omega_o d\tau \tag{3.7}$$

An admissible shearlet is any function ψ, for which $\hat{\psi}$ is supported away from origin compactly. For conditions $\varrho = (\varrho_1, \varrho_2) \in \hat{R}^2, \varrho_1 \neq 0$ the classical shearlet ψ is given as-

$$\hat{\psi}(\varrho_1, \varrho_2) = \hat{\psi}_1(\varrho_1)\hat{\psi}_2(\frac{\varrho_2}{\varrho_1}) \tag{3.8}$$

According to [165], $\psi_1 \in L^2(R)$ follows the Calderon condition,

$$\sum_{j \in Z} |\hat{\psi}_1(2^{-j}\varrho)|^2 = 1 \quad \forall \varrho \in R \tag{3.9}$$

where supp $\hat{\psi}_1 \subset [-2, -1/2] \cup [1/2, 2]$ and $\psi_2 \in L^2(R)$ function is considered as bump with supp $\hat{\psi}_2 \subset [-1, 1]$ and

$$\sum_{k=-1}^{1} |\hat{\psi}_2(\varrho + k)|^2 = 1 \quad \forall \varrho \in [-1, 1] \tag{3.10}$$

The classical shearlet is a function ψ, that behaves wavelet-like along one axis and bump-like along another. At various scales λ, the frequency support of each member of classical shearlet $\psi_{\lambda,\omega_o,\tau}$,relies on a pair of trapezoids. These are aligned along a line of slope ω_o and are symmetric with respect to the origin.

The continuous shearlet function SH_ψ is said to be isometry if it follows

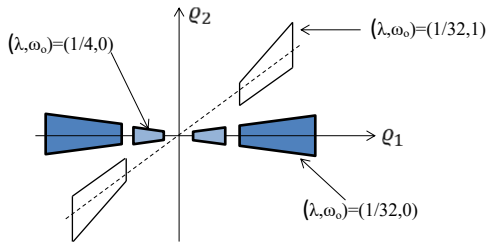


Figure. 3.2: The frequency support of the classical shearlets $\hat{\psi}_{\lambda,\omega_o,\tau}$ for different λ and ω_o

the condition given as-

$$H_{\psi}^{+} = H_{\psi}^{-} = 1 \tag{3.11}$$

where $H_{\psi}^{+} = \int_0^{\infty}\int_R \frac{|\hat{\psi}(\varrho_1,\varrho_2)|^2}{\varrho_1^2} d\varrho_2 d\varrho_1$ and $H_{\psi}^{-} = \int_{-\infty}^{0}\int_R \frac{|\hat{\psi}(\varrho_1,\varrho_2)|^2}{\varrho_1^2} d\varrho_2 d\varrho_1$.

The discrete shearlet system can be obtained by sampling the scale, shear and translation parameters properly as $(\lambda,\omega_o,\tau) = (2^{-p}, -k2^{-p/2}, S_{-k2^{-p/2}}A_2 - pr)$ and defined as

$$SH(\psi) = \psi_{p,q,r} = 2^{3/4p}\psi(S_q A_{2^p} - r), p,q \in Z, r \in Z^2 \tag{3.12}$$

In the same manner, the discrete ST can be defined as -

$$f \rightarrow SH_{\psi}f(p,q,r) =< f, \psi_{p,q,r} > \tag{3.13}$$

Classical shearlet function that shows fast decay both in the spatial and frequency domain is treated as a, well localized function. This property of classical shearlets ensure the formation of a frame by discrete shearlet system and hence results into optimal sparse approximations. The frequency tiling produced by discrete shearlets $\hat{\psi}_{p,q,r}$ is depicted in Fig. 3.3.

The Laplacian-pyramid algorithm and directional filter are used to obtain the discrete ST [166]. The directional filter is built with small sized support to reduce the impact of Gibbs phenomenon and computational complexity. The non-subsampled Laplacian-pyramid filter (NSLP) was introduced in [167]. NSST is

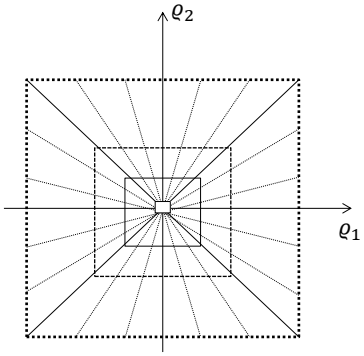


Figure. 3.3: The frequency tiling induced by the discrete shearlets $\hat{\psi}_{p,q,r}$

designed by replacing its Laplacian pyramid (LP) with NSLP to improve discrete ST [168]. In contrast to the ST, the NSST is a shift-invariant ST which can be expanded to multiple scale and direction. The pseudo-Gibbs situation as well as shearlet type artifacts are reduced in the subsequent filtering. The iterative process involved in NSLP is given as -

$$NSLP_{p+1}I = (Ph_p^1 \prod_{q=1}^{p-1} Ph_q^0)I \tag{3.14}$$

where I, $NSLP_{p+1}$ are the image and detail coefficients at scale $p+1$ respectively. Ph_q^0 and Ph_p^1 are the low and high pass filters used in NSLP at scale p and q respectively. The steps involved in NSST associated with NSLP is shown below-

Algorithm:

1. Step 1: NSLP is applied to decompose I_λ^{p-1} into low pass and high pass image I_λ^p and I_d^p respectively.
2. Step 2: Compute $\hat{I}_d^p$ in pseudo-polar grid and matrix MI_d^p is obtained.
3. Step 3: Band pass filters is applied to matrix MI_d^p to obtain $(\hat{I}_{d,q}^p)_{q=1}^{D_p}$
4. Step 4: Two dimensional inverse fast Fourier transform is applied to obtain the discrete ST coefficients $(I_{d,q}^p)_{q=1}^{D_p}$ in pseudo polar grid.

The approximation and detail subbands obtained after decomposing one RS image are shown in Fig. 3.5. High-frequency and low-frequency coefficients exist in the NSST detail and approximation subbands, respectively. High frequency detail coefficients are displayed in Fig. 3.5(d-e) and 3.5(f-i) at the finest scale (Scale 1) and the following coarsest scale, respectively.

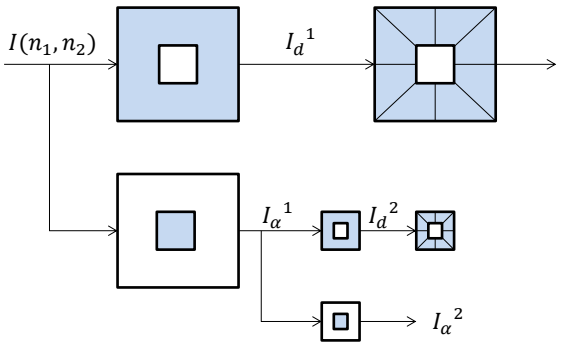
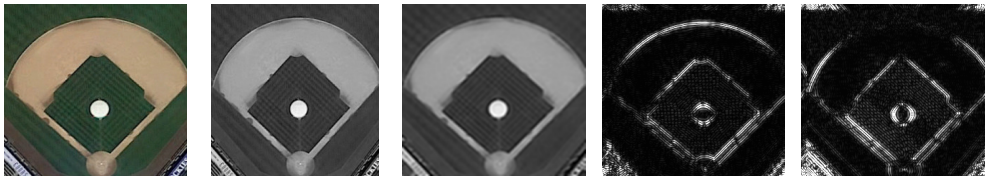
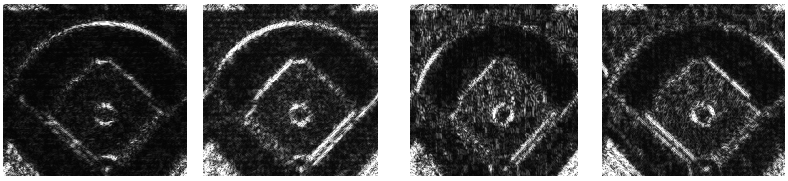


Figure. 3.4: The illustration of the Laplcian filter and directional filtering



(a) Original image (b) Its grayscale form (c) Approximation subband (d) 1^{st} Detail subband (Scale 1) (e) 2^{nd} Detail subband (Scale 1)



(f) 1^{st} Detail subband (Scale 2) (g) 2^{nd} Detail subband (Scale 2) (h) 3^{rd} Detail subband (Scale 2) (i) 4^{th} Detail subband (Scale 2)

Figure. 3.5: Subbands obtained through two-level NSST decomposition of a sample RS image

3.1.1.2 Proposed methodology

Histogram plots of two RS images' NSST detail subbands are shown in Fig. 3.6

From Fig. 3.6, it is seen that the statistics of image NSST detail subband coefficients have a sharp peak near zero with quite heavy tails. This indicates that most of the NSST coefficients in a subband are equal to zero and the subband coefficients can be best modeled using some highly non-Gaussian distributions.

We present a RSIR scheme called NSST-SNIG which is based on statistical modeling of NSST coefficients using SNIG distribution, in this subsection.

The block diagram of RSIR technique with the NSST-SNIG feature de-

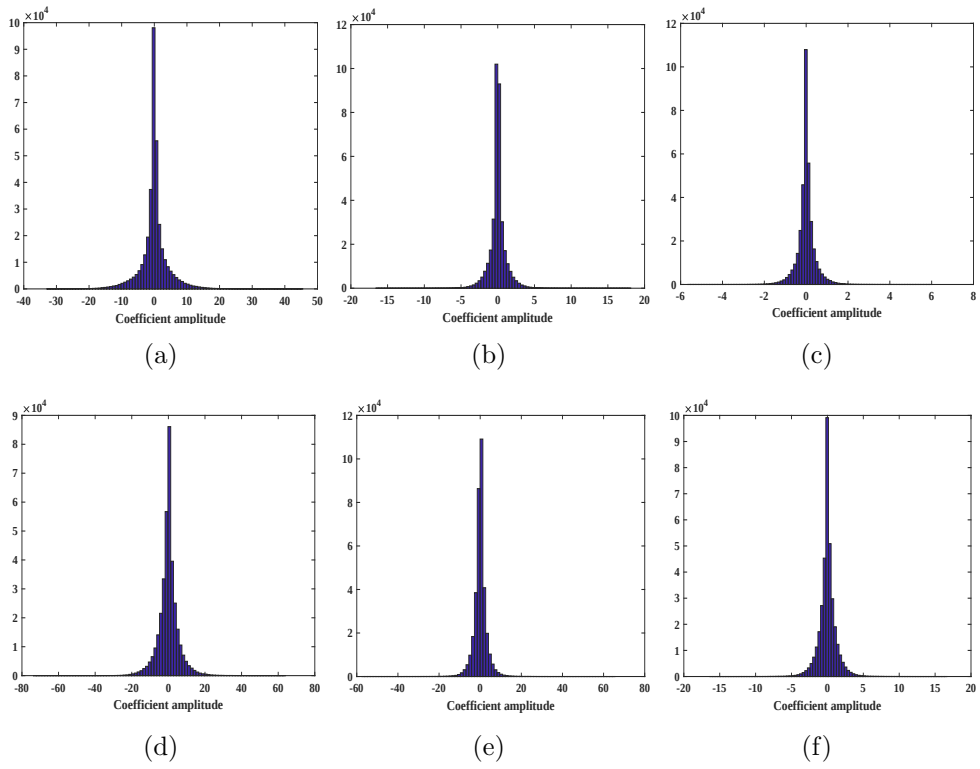


Figure. 3.6: (a-c) and (d-f) shows the plots corresponding to the NSST detail subbands of image 1 and image 2 from WHU-RS19 respectively

scriptor is presented in Fig. 3.7.

The NSST coefficients' heavy tail nature can be modelled by the SNIG [169]. This distribution has four parameters and is closed-form flexible [170]. Considering the normal distribution $N(\mu_m, \sigma^2)$ with variance σ^2, mean μ_m and Inverse Gaussian distribution $I_G(\kappa, \delta)$ with variance $V(y) = \frac{\delta}{\kappa^3}$ and $E(y) = \frac{\delta}{\kappa}$ as mean, the probability density function (pdf) of inverse Gaussian (IG) can be expressed as [170],

$$f_{IG}(y) = \frac{\delta}{\sqrt{2\pi}} e^{(\delta\kappa)} y^{-3/2} e^{-1/2(\frac{\delta^2}{y} + \kappa^2 y)} \tag{3.15}$$

If $N(\mu_m + \beta y, y)$ denotes the conditional distribution of c given y and if y obeys itself as an $I_G(\kappa, \delta)$,then the combined distribution that results is the $SNIG(\alpha, \beta, \mu_m, \delta)$ with $\alpha = \sqrt{\kappa^2 + \beta^2}$. The pdf of $SNIG$ is [170]

$$h(c; \alpha, \beta, \mu_m, \delta) = \frac{\alpha}{\pi} e^{(\delta\sqrt{\alpha^2-\beta^2} - \beta\mu_m)} \phi(c)^{-1/2} K_1(\delta\alpha\phi(c)^{1/2}) e^{(\beta c)} \tag{3.16}$$

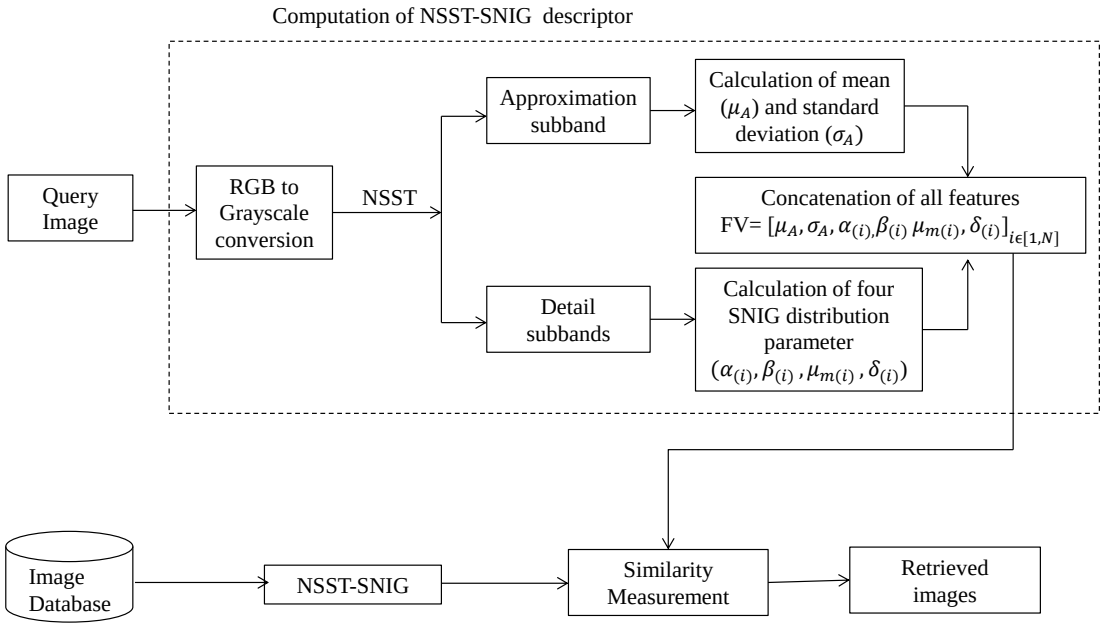


Figure. 3.7: Block diagram of the NSST-SNIG based feature extraction approach

with

$$\phi(c) = 1 + [\frac{c - \mu_m}{\delta}]^2 \tag{3.17}$$

The $K_b(c)$ represents the 3^{rd} kind modified Bessel function of order b.

The KS goodness of fit test is taken into account to support the usage of SNIG in the modeling of NSST coefficients. For the ML estimate of SNIG parameters, [170], an EM kind of method is utilized as follows:

In EM type of algorithm, there are two main steps. First one is E step followed by M step. In case of E step, the set of parameters to be estimated are considered as $\varphi = (\alpha, \beta, \mu_m, \delta)$. These $\varphi^{(k)}$ are the values of the parameters after the k^{th} iteration, then the pseudovalues S_i and w_i are calculated as

$$S_i = E(y_i|c_i, \varphi^{(k)}) = \frac{\delta^k \phi^{(k)}(c_i)^{1/2} K_o(\delta^{(k)}\alpha^{(k)}\phi^{(k)}(c_i)^{1/2})}{\alpha^{(k)} K_1(\delta^{(k)}\alpha^{(k)}\phi^{(k)}(c_i)^{1/2})}$$

$$w_i = E(y_i^{-1}|c_i, \varphi^{(k)}) = \frac{\alpha^{(k)} K_{-2}(\delta^{(k)}\alpha^{(k)}\phi^{(k)}(c_i)^{1/2})}{\delta^{(k)}\phi^{(k)}(c_i)^{1/2} K_{-1}(\delta^{(k)}\alpha^{(k)}\phi^{(k)}(c_i)^{1/2})}$$

for $i = 1, 2, ...n.$ where $\phi^{(k)}(c) = 1 + [\frac{(c - \mu_m^{(k)})}{\delta^{(k)}}]^2$

Using the derived pseudovalues from the E-step, the M step adjusts the parameters. Compute $\hat{M} = \sum_{i=1}^{n} \frac{S_i}{n}$and $\hat{\wedge} = n(\sum_{i=1}^{n}(w_i - \hat{M}^{-1}))^{-1}$. The following equations are utilized to update the parameters:

$$\delta^{(k+1)} = \hat{\wedge}^{1/2}$$

$$\gamma^{(k+1)} = \frac{\delta^{(k+1)}}{\hat{M}}$$

$$\beta^{(k+1)} = \frac{\sum_{i=1}^{n} c_i w_i - \bar{c}\sum_{i=1}^{n} w_i}{n - \bar{S}\sum_{i=1}^{n} w_i}$$

$$\mu_m^{(k+1)} = \bar{c} - \beta^{(k+1)}\bar{S}$$

$$\alpha^{(k+1)} = [(\gamma^{(k+1)})^2 + (\beta^{(k+1)})^2]^{1/2}$$

where $\bar{S} = \sum_{i=1}^{n} \frac{S_i}{n}$.

This algorithm's initial values are calculated employing the moment-based approach ([171]). The EM operation is applied repeatedly until the parameters converges.

The empirical cumulative distribution function (ECDF) for a given dataset and the specified model cumulative distribution function (CDF) are compared using KS test statistics [171]. The KS test is defined as[171–173]

$$KS_{res} = \max_{c \in N} |F(c) - \tilde{F}(c)| \tag{3.18}$$

where the model CDF and ECDF are represented by $F(c)$ and $\tilde{F}(c)$, respectively. The distribution whose value of KS_{res} is lowest is regarded as having the best fit to the data. The average KS statistics obtained from few RS image NSST subband are presented in Table 3.1. In Fig. 3.8, the log histogram of one of the finest NSST detail subbands for six images from three different databases are shown. The distributions such as Bessel K form (BKF), laplacian and SNIG pdf's are fitted to this empirical histogram in logarithmic domain. Both Table 3.1 and Fig. 3.8 make it evident that SNIG is preferable to all other pdfs like Laplacian and

Table 3.1: Average KS statistics for a few RS images

NSST decomposition	pdf	Laplacian	BKF	SNIG
	Level 1	0.0417	0.0798	**0.0298**
		0.0408	0.0844	**0.0260**
		0.0381	0.0845	**0.0285**
		0.0360	0.0986	**0.0282**
		0.0406	0.0873	**0.0284**
		0.0467	**0.0140**	0.0277
		0.0418	0.0836	**0.0285**
		0.0343	0.0883	**0.0286**
	2	0.0489	0.0625	**0.0271**
		0.0402	0.0767	**0.0296**
		0.0458	0.0643	**0.0267**
		0.0499	0.0632	**0.0280**
	3	0.0514	0.0619	**0.0326**
		0.0548	0.0506	**0.0314**

BKF for estimating the NSST coefficient statistics.

Features from one approximation subband and N_1 total detail subbands are combined to generate the feature vector as follows:

$$FV = [\alpha_1, \beta_1, \mu_{m_1}, \delta_1, ...\alpha_{N_1}, \beta_{N_1}, \mu_{m_{N_1}}, \delta_{N_1}, \mu_A, \sigma_A] \tag{3.19}$$

where α_1, β_1, μ_{m_1}, and δ_1 are the SNIG parameters from the 1^{st} NSST detail subband; similarly, α_{N_1}, β_{N_1}, $\mu_{m_{N_1}}$, and δ_{N_1} are parameters of N_1^{th} subband. The μ_A is the mean of NSST approximation subband and σ_A is its standard deviation.

After extracting the features, the similarity measurement is used to detect similar images related to a particular query. The query and database image features are compared for similarity. In the NSST-SNIG framework, d_1 distance is used to provide the best results when compared to other distance measures for determining similarity.

The following is NSST-SNIG feature extraction algorithm in an image retrieval framework:

Algorithm: RSIR approach with NSST detail subbands modeling using SNIG.
Require: Input: Query image: Output: n_T number of images retrieved

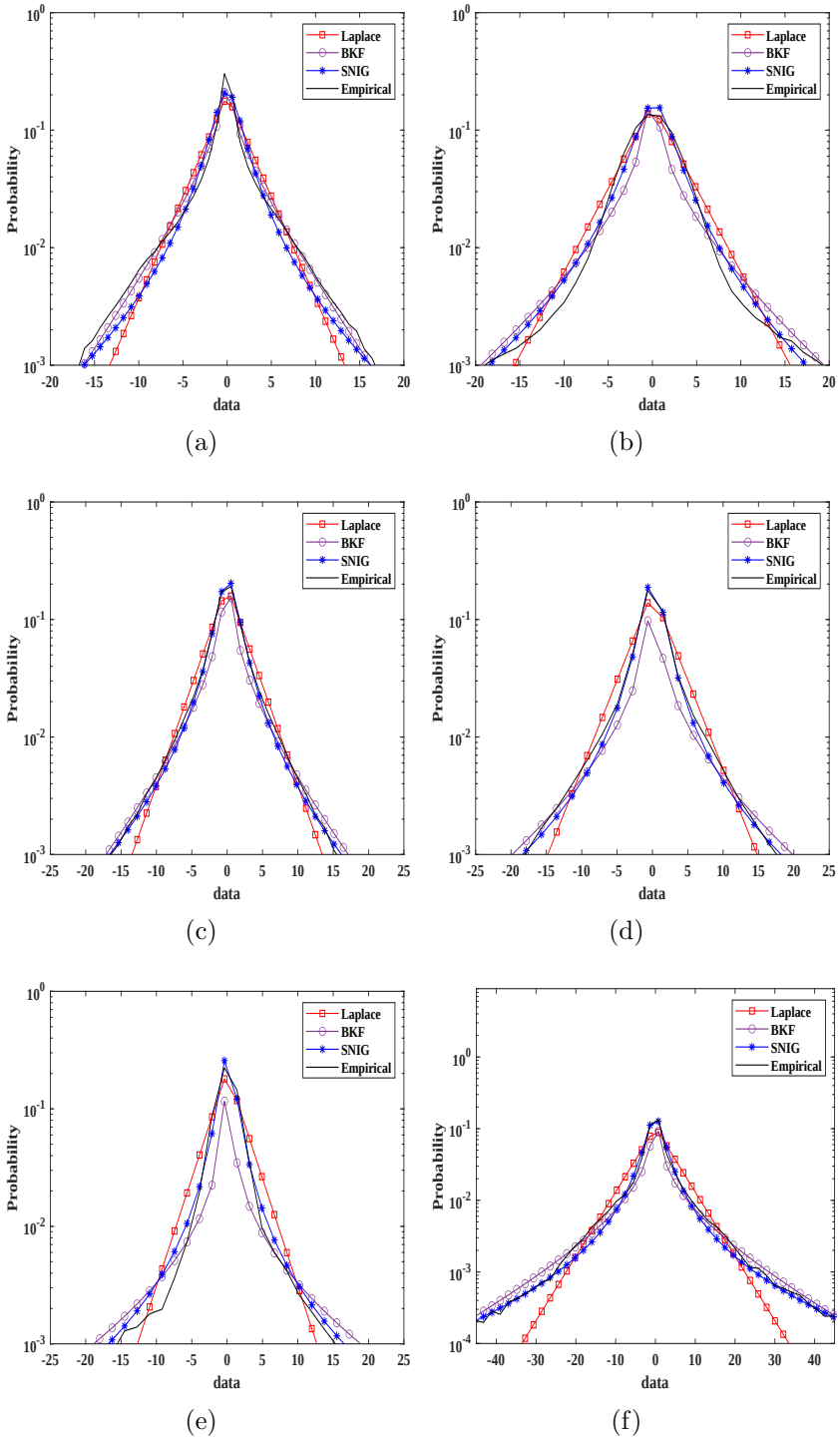


Figure. 3.8: The Laplacian, BKF and SNIG pdf's fitted to the empirical histogram in log state for one finest NSST subband of six different images (Two each from WHU-RS19, AID and PatternNet)

1. Load the input RS image and convert it to grayscale if it is a color image.
2. Decompose the RS images with NSST to obtain different subbands.
3. Model the NSST detail coefficients with SNIG distribution and estimate the SNIG parameters using EM type of approach
4. Extract the μ_A and σ_A from approximation subband and combine it with features obtained from step 3 to construct the final feature vector.
5. Compute d_1 distance for similarity measurement between features of query and database images.
6. Retrieve the final n_T number of images based on the shortest matching distance calculated.

3.1.2 Experimental results

In this subsection, the dataset used for the experiments, the performance assessment metrics employed, and the similarity distance metrics employed are elaborated before the discussion of the experimental outcomes achieved.

3.1.2.1 Datasets considered

For the experiments we have considered three publicly available RS image datasets- WHU-RS19, Aerial image dataset (AID) and PatternNet dataset. The details about these three datasets are presented in Table 3.2.

S.No.	Database	Total images	No. of classes	Image size
1.	WHU-RS19	1005	19	600×600
2.	AID	10000	30	600×600
3.	PatternNet	30400	38	256×256

Table 3.2: Databases used in the experiments

- WHU-RS19 dataset : This dataset is comprised of Google Earth satellite images. These images have a resolution of up to 0.5 meters. This collection includes a total of 19 distinct image classes taken from various locations

utilizing satellite images with varying resolutions, scale, orientation, and illumination [174, 175]. In the Fig. 3.9, example image of each class of the dataset is presented.

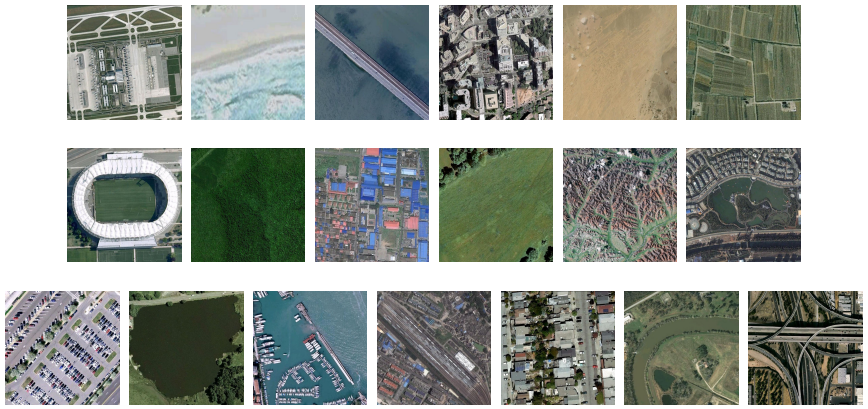


Figure. 3.9: The sample image taken from individual classes of WHU-RS19 dataset

- Aerial Image Dataset (AID): One of the most extensive collections of annotated aerial images, AID [176, 177] contains 10000 images in 30 classes. The RS images in this dataset were created using multiple imaging sensors at different times and under different imaging conditions. This results in less intra-scale variance and more inter-class variation, while also making it more difficult to correctly retrieve identical images. Each class has roughly 220-420 images with size 600×600. This extensive aerial dataset's images are taken from Google Earth imagery in a chosen manner. Fig. 3.10 presents an example image for each AID class.

Figure. 3.10: The sample image taken from individual classes of AID dataset

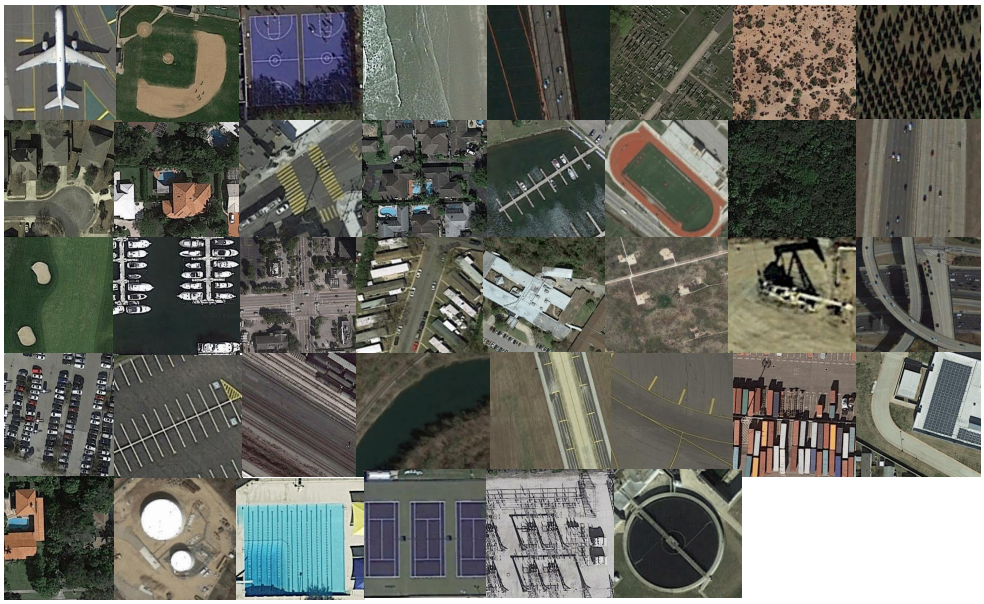


Figure. 3.11: The sample image taken from individual classes of PattenNet dataset

- PatternNet : PatternNet [178, 179] is the largest high resolution RS dataset available, with a total of 30400 photos evenly distributed into 38 different image classes. These images have a size of 256×256. This dataset's images were compiled from Google Earth imagery or the Google MAP API of several US cities. In Fig. 3.11 examples of each class in this dataset are displayed.

3.1.2.2 Performance evaluation measures

The NSST-SNIG's retrieval performance is measured using three performance assessment metrics: average normalized modified retrieval rank (ANMRR), mean average precision (MAP), and precision-recall (P-R) graph. The details of these evaluation measures are discussed below.

1. ANMRR:

   ANMRR is a popular performance evaluation measure for evaluating the effectiveness of retrieval performance in MPEG-7 standard, particularly in the field of RSIR techniques. ANMRR values vary from $0 to 1$. A lower value for the ANMRR is indicative of a high level of retrieval accuracy [180].

   $Gr(q_1)$ specifies the dimensions of the ground truth images for each query image (q_1), and the ground truth image at the k^{th} position is retrieved at $Rank(k)$. The admissible image ranks are then stated as $K(q_1)$, which is

twice $Gr(q_1)$, and images with higher ranks are penalized as follows:

$$Rank(o) = \begin{cases} Rank(o) & \text{if } Rank(o) < K(q_1) \\ 1.25K(q_1) & \text{if } Rank(o) > K(q_1) \end{cases} \tag{3.20}$$

For q_1, the average rank (A_r) is given as:

$$A_r(q_1) = \frac{1}{Gr(q_1)} \sum_{o=1}^{Gr(q_1)} Rank(o) \tag{3.21}$$

The normalization is computed to control the effect of varying number of ground truths of query image and further averaged averaged for all query images N_Q to compute ANMRR:

$$ANMRR = \frac{1}{N_Q} \sum_{q_1=1}^{N_Q} \frac{A_r(q_1) - 0.5[1 + Gr(q_1)]}{1.25K(q_1) - 0.5[1 + Gr(q_1)]} \tag{3.22}$$

2. MAP:

The MAP is one method for aggregating the P-R curve into a single number that determines the rank position of every ground truth. Let us consider,$Pr_{ave}(q_1)$ represents the average precision for every query image q_1, which is simply the average of the precision values for all of the relevant items:

$$Pr_{ave}(q_1) = \frac{\sum_{k=1}^{n} (Pr(k) * rel(k))}{No. \quad of \quad relevant \quad images} \tag{3.23}$$

where $rel(k)$ denotes a function which outputs 1 if the item at k^{th} rank is valid or relevant else outputs 0. The $Pr(k)$ denotes the precision at k. The Pr_{ave} values over all query items lastly provides the MAP:

$$MAP = \frac{\sum_{q_1=1}^{N_Q} Pr_{ave}(q_1)}{N_Q} \tag{3.24}$$

The MAP value ranges between 0 and 100. A higher MAP value indicates that the descriptor has greater retrieval performance[180,181]. The example computation of MAP for two queries for top 10 retrieved cases are shown in Fig. 3.12.

3. P-R curve

Precision and recall are both commonly employed for assessing the effectiveness of image retrieval systems. Precision is the ratio of relevant images

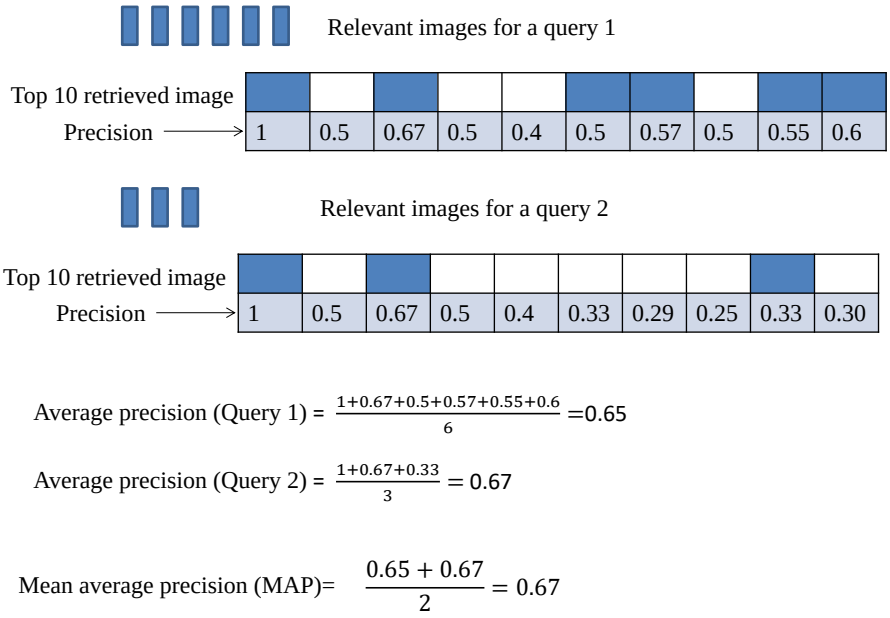


Figure. 3.12: Example computation of MAP considering two query images with different number of relevant images in dataset

retrieved to the total number of images retrieved, whereas the term 'recall' refers to the ratio of relevant images that were successfully retrieved to the total number of relevant images contained in the database. The descriptor with the biggest area under the curve implies high precision and strong recall, which demonstrates more results relevancy and enhanced retrieval of proper relevant images [180].

3.1.2.3 Distance measures used for similarity assessment

The distance between features extracted from query and database images is determined by similarity measurement, which is essential in a CBIR system. After the measurements are calculated, the distance values are organized in ascending order. Different distance metrics, such as d_1, Euclidean, Manhattan, Chi-square, and Canberra, are used to determine how similar two images are.

- d_1 distance

$$D(I_q, DB_k) = \sum_{j=1}^{F_l} |\frac{F_{DB_{kj}} - F_{I_{qj}}}{1 + F_{DB_{kj}} + F_{I_{qj}}}| \tag{3.25}$$

- *Euclidean* distance

$$D(I_q, DB_k) = \sqrt{\sum_{j=1}^{F_l} (F_{DB_{kj}} - F_{I_{qj}})^2} \tag{3.26}$$

- *Manhattan* distance

$$D(I_q, DB_k) = \sum_{j=1}^{F_l} |F_{DB_{kj}} - F_{I_{qj}}| \tag{3.27}$$

- $Chi - square$ distance

$$D(I_q, DB_k) = \frac{1}{2} \sum_{j=1}^{F_l} \frac{(F_{DB_{kj}} - F_{I_{qj}})^2}{F_{DB_{kj}} + F_{I_{qj}}} \tag{3.28}$$

- *Canberra* distance

$$D(I_q, DB_k) = \sum_{j=1}^{F_l} \frac{|F_{DB_{kj}} - F_{I_{qj}}|}{|F_{DB_{kj}}| + |F_{I_{qj}}|} \tag{3.29}$$

$D(I_q, DB_k)$ measures the distance between the query image I_q and the database's k^{th} image. The feature vector has a length of F_l. The j^{th} feature of the k^{th} image of the database and the query image, respectively, are $F_{DB_{kj}}$ and $F_{I_{qj}}$.

3.1.2.4 Results obtained

In the experiments, 3 levels of decomposition with NSST in the 1,2,3 directions are used, which yields 1 approximation and 14 detail subbands. In order to evaluate how effectively the NSST-SNIG descriptor retrieves images, it is compared to six other well-known low-dimensional local and global hand-crafted descriptors. The WHU-RS19, AID and PatternNet datasets are considered to perform all the experiments.

Table 3.3 presents the comparison of the performance of NSST-SNIG with that of LBP[182, 183], Granulometry[184], Gabor L[185], Gabor RGB[186], and FV[187] for WHU-RS19 dataset. It is observed that the NSST-SNIG outperforms the other methods for WHU-RS19 both in terms of MAP and ANMRR. The Table 3.4 presents the ANMRR and MAP values obtained for AID and PatternNet

Table 3.3: Retrieval performance comparison of ANMRR and MAP in WHU-RS19

Dataset	Methods	MAP	ANMRR	FD
WHU-RS19	LBP [182, 183]	23.55	0.674	256
	Granulometry [184]	21.41	0.717	78
	Gabor L [185]	25.47	0.658	32
	Gabor RGB [186]	27.00	0.649	96
	FV [187]	19.33	0.726	512
	NSST-SNIG	**28.35**	**0.584**	58

Table 3.4: Retrieval performance comparison for AID and PatternNet datasets in terms of ANMRR and MAP

Dataset	**Methods**	**MAP**	**ANMRR**	**Dataset**	**Methods**	**MAP**	**ANMRR**
AID	LBP [182, 183]	10.36	0.857	PatternNet	LBP [182, 183]	29.99	0.686
	Granulometry [184]	9.04	0.877		Granulometry [184]	14.60	0.817
	Gabor L [185]	10.79	0.916		Gabor L [185]	17.43	0.851
	Gabor RGB [186]	11.69	0.843		Gabor RGB [186]	26.53	0.686
	FV [187]	**16.40**	0.748		FV [187]	19.23	0.760
	NSST-SNIG	12.40	**0.707**		NSST-SNIG	**31.52**	**0.638**

dataset. It is observed that for PatternNet, NSST-SNIG obtains clearly outperforming values compared to all other approaches taken for comparison. However for AID dataset, the NSST-SNIG, shows inferior results compared to FV and shows better values compared to other techniques. FV is very good at capturing local details and AID dataset contains a great number of images that contain more of local details compared to global textures which is directly evident from the results presented in Table 3.4 and Fig. 3.14. In addition, Table 3.3 also displays the feature vector dimensions of the NSST-SNIG and all other methods compared in this study. It is observed that NSST-SNIG outperforms LBP, Granulometry, Gabor L, Gabor RGB, and FV in terms of all performance evaluation parameters while having a lesser dimensionality of feature vectors. Even though the proposed method has a slightly larger number of feature vectors than Gabor L, it works much better than Gabor L.

The Fig.s 3.13, 3.14 and 3.15 presents the average precision values obtained for each image classes of WHU-RS19, AID and PatternNet dataset respectively. These average precision values are calculated considering 20 trials of 20 randomly selected images for top 20 retrieved images. The Fig.s 3.13-3.15 show the superiority of proposed NSST-SNIG over other feature descriptors for most of the individual classes. NSST-SNIG shows better performance for individual classes beach, desert, football field, forest, meadow of WHU-RS19 dataset. For AID dataset, the NSST-SNIG showed better average precision for image classes such as bareland, beach, forest, meadow, medium residential, railway station and sparse residential. For PatternNet dataset, the average precision value is found to

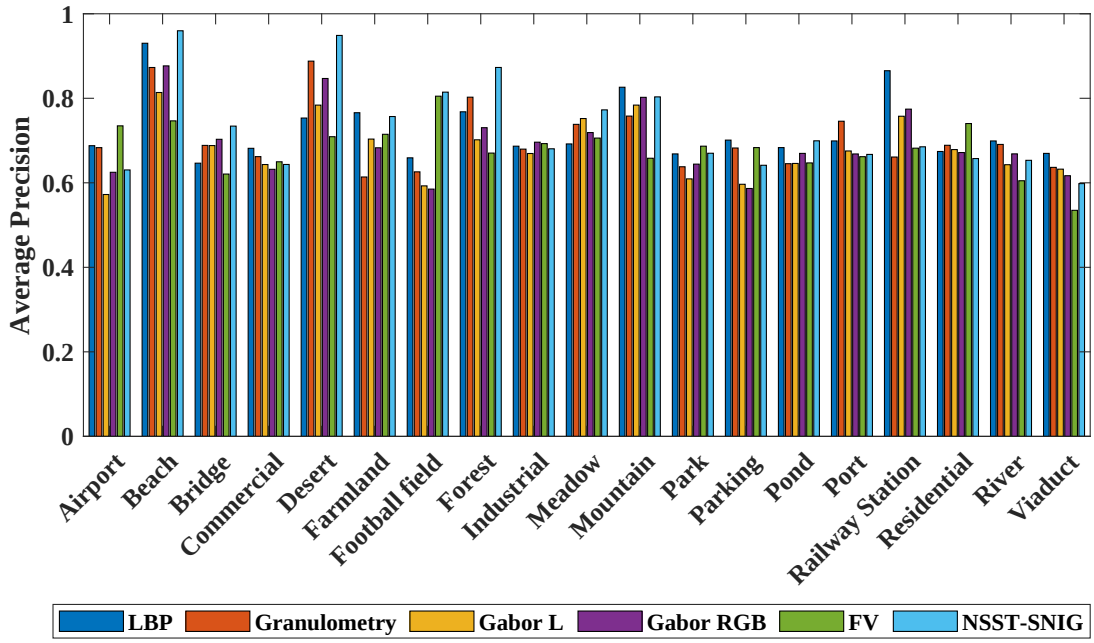


Figure. 3.13: The per class average precision comparison of various schemes for WHU-RS19

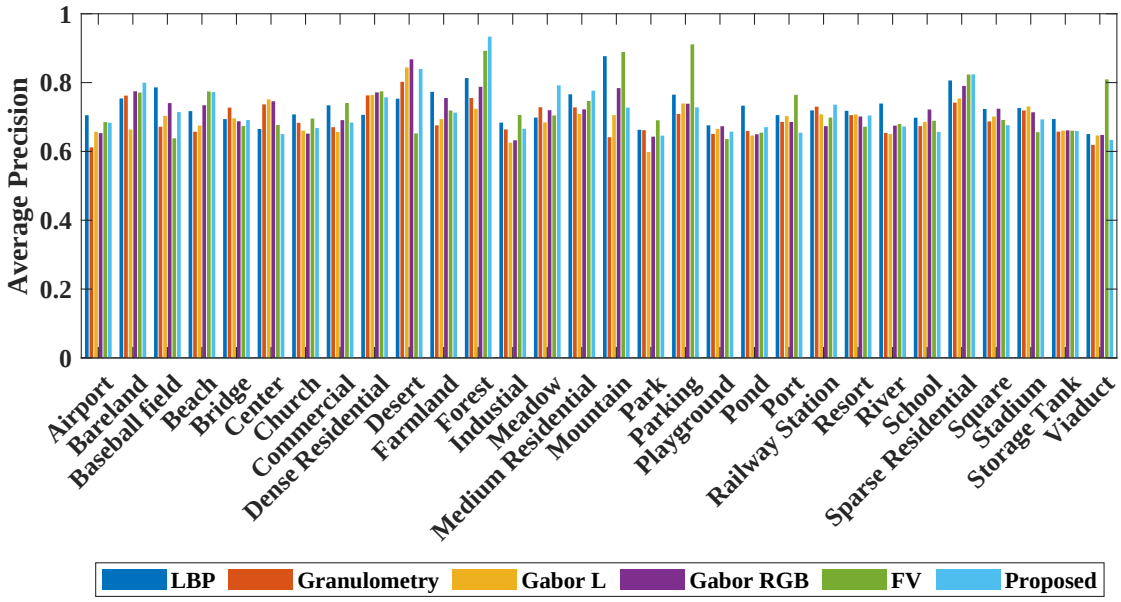


Figure. 3.14: The per class average precision comparison of various schemes for AID dataset

be better for image classes such as airplane, beach, forest, harbor, cemetary, chaparral, coastal mansion, christmas tree farm, dense residential, runway, shipping yard, ferry terminal, mobile homepark, parking lot, sparse residential, swimming pool and wastewater treatment plant. It is observed that the NSST-SNIG in most of the cases performs better for the image classes having more amount of textures as well as high scale attributes such as forest, dense residential, beach, bareland

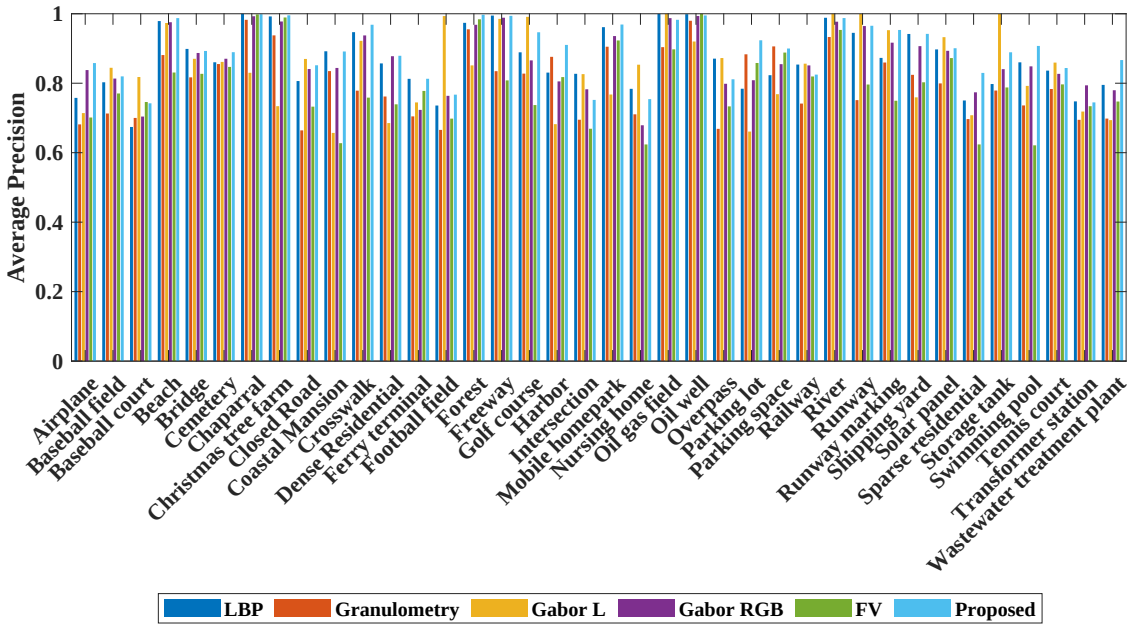


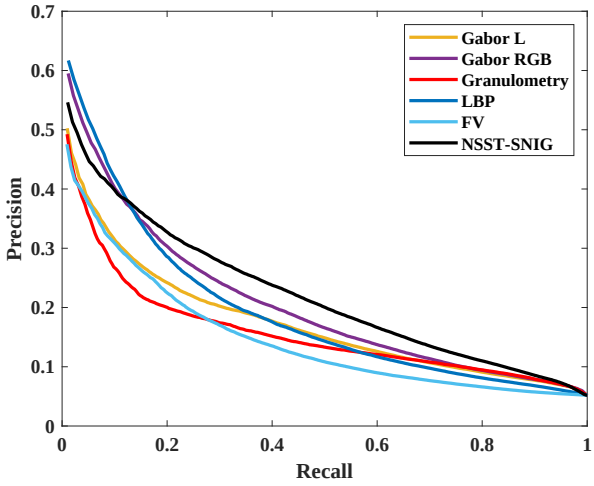
Figure. 3.15: The per class average precision comparison of various schemes for PatternNet dataset

etc. and shows inferior results for the classes such as intersection, baseball, railway, storage tank, parking, golf course, church etc. This shows that the NSST-SNIG has limitations in capturing local image features.

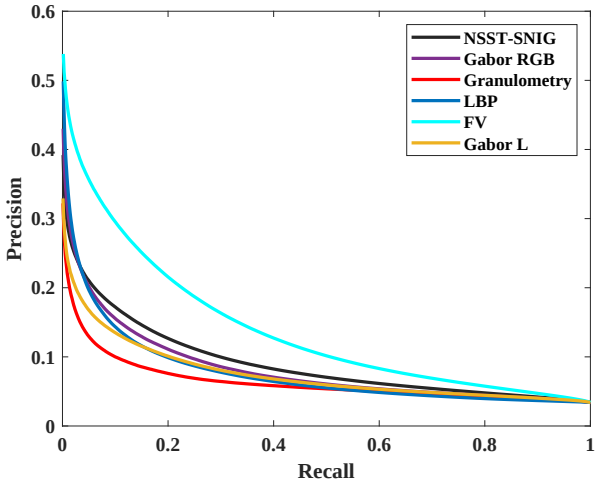
Fig. 3.16(a-c) presents the P-R curve for WHU-RS19, AID and PatternNet dataset. From these curves, it is observed that the P-R curve for NSST-SNIG shows the superiority over LBP, Granulometry, Gabor L, Gabor RGB, and FV for both WHU-RS19 and PatternNet datasets. In case of AID, except for FV, the NSST-SNIG shows similar trend over all other descriptors.

In Table 3.5, the total retrieval time required for the three considered RS datasets are presented. The total retrieval time for a certain dataset is calculated by estimating how long it would take to match the query image with each image in the dataset. The feature dimensions play a major role in this calculation. It has been observed that the total retrieval time for NSST-SNIG is less than LBP, Granulometry, Gabor RGB and FV, as these techniques posses higher feature dimension compared to NSST-SNIG. Similar trend is observed for all the datasets.

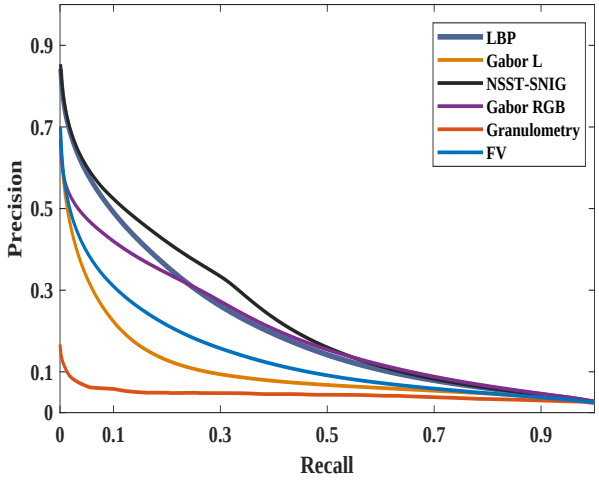
A few instances of image queries from various classes and their related retrieved results utilizing the proposed NSST-SNIG descriptor (for WHU-RS19) are shown in Fig. 3.17.The proposed descriptor is able to correctly retrieve images from certain classes (beach, bridge, desert, farmland, river, and pond) when given an image from that class as a query with the exception of the classes in-



(a) WHU-RS19



(b) AID



(c) PatternNet

Figure. 3.16: The P-R curve for (a)WHU-RS19, (b)AID and (c) PatternNet dataset

Table 3.5: Comparison of total retrieval time (in seconds) of the NSST-SNIG with other techniques

Method	LBP	Granulometry	Gabor L	Gabor RGB	FV	NSST-SNIG
WHU-RS19	2.74	1.77	1.42	2.19	4.59	1.54
AID	8.71	3.49	2.51	6.41	11.01	4.32
PatternNet	10.45	7.12	4.28	7.86	19.47	6.08

Figure. 3.17: Images retrieved for some of the classes of WHU-RS19 dataset(The black-bordered query image is on the left, while the top five retrieval results are on the right.)

dustrial, viaduct and mountain, which provides one incorrect retrieved result. For 'industrial' class, one image from 'residential' class is wrongly retrieved. For the image class 'viaduct', one image from 'Railway' class is wrongly retrieved. Again, one image from 'forest' gets retrieved incorrectly for query image from 'mountain' class. It is observed that for image classes, beach, bridge, desert, farmland, river, and pond, the NSST-SNIG descriptor retrieves images correctly, however, the NSST-SNIG performs poorly on the classes industrial, viaduct, and mountain, with at least one incorrect retrieval result observed in each of these classes (Fig. 3.17).

3.2 RSIR based on 3D-LTP features and NSST domain statistical features

The combination of suitable local and global features is found to be successful in enhancing the retrieval/classification results as they both carry complementary details. Some techniques fail to capture discriminative information from high resolution images where the presence of few major structural details dominate the image class. In these circumstances, blending of local and global features has been observed to work better. Different approaches are presented where combination of local and global features are utilized for extraction of discriminative details from images [141, 142]. In [188], it is discussed that the retrieval performance improvement is not always guaranteed by the combination of visual attributes. Complementary visual attributes should be extracted from high resolution images in order to improve retrieval efficiency without increasing the feature dimension.

In this work, the features of high-resolution RS images are extracted using a combination of global NSST - domain statistical features (NSSTds) and local three dimensional local ternary pattern (3D-LTP) features. We use a 2-state LM distribution to model the statistics of NSST detail subband coefficients and the EM approach is employed to compute its three parameters. In order to include more statistical information of NSST coefficients, in addition to 2-state LM parameters we also compute kurtosis and skewness from detail subbands and mean as well as standard deviation from approximation subband. We then concatenate all of these statistical parameters together to form NSSTds features. The many characteristics of NSST, including multiscale, localization, and adjustable directional sensitivity, make it an appropriate option for providing an accurate approximation of an image. A new 3D-LTP is introduced for the purpose of capturing the

dense local details. Spatial RGB planes of the input image are used to calculate the 3D-LTP. The suggested inter-channel 3D-LTP captures both color and local texture information. At last, a blended feature description i.e. NSSTds-3DLTP is suggested to improve the discriminative power of features by combining global (NSSTds) and local (3D-LTP) features.

3.2.1 Methodology

This subsection elaborates on the proposed NSSTds-3DLTP framework for RSIR application. The framework comprise of two important parts: The first module creates global NSSTds features using statistical parameters from the NSST domain, while the second module computes local 3D-LTP features using RGB channels.

The block diagram of NSSTds-3DLTP based framework is presented in Fig. 3.18.

3.2.1.1 Calculation of NSST domain statistical features (NSSTds)

By comparing the similarity between statistical models, the texture discrimination problem can be addressed with substantially less dimensions when statistics of transform coefficients are statistically modelled using statistical distributions. This method is highly successful and relatively simpler to use. For image retrieval, the distribution of image transform coefficients has been modelled using parametric distributions [181, 188]. Frequent use of statistical modeling of images in the wavelet transform domain is observed in the literature. Applications for image retrieval have effectively used various non-Gaussian distributions like generalized Gamma [189], Laplacian, Cauchy etc. Images having linear singularities cannot be described well by the discrete wavelet transform. The MGA tools like shearlet, which was also discussed in previous section, give a solution for the aforementioned issue because these transforms offer an excellent sparse representation for singularities of higher dimensions.

In earlier research, it was shown that the RS images' NSST detail coefficients conform to strongly non-Gaussian statistics and the SNIG distribution is preferable to other non- Gaussian distributions like laplacian and BKF models for modeling the NSST detail coefficients. The ability of LM model to capture the large tails of non-Gaussian empirical data is well established [171]. When

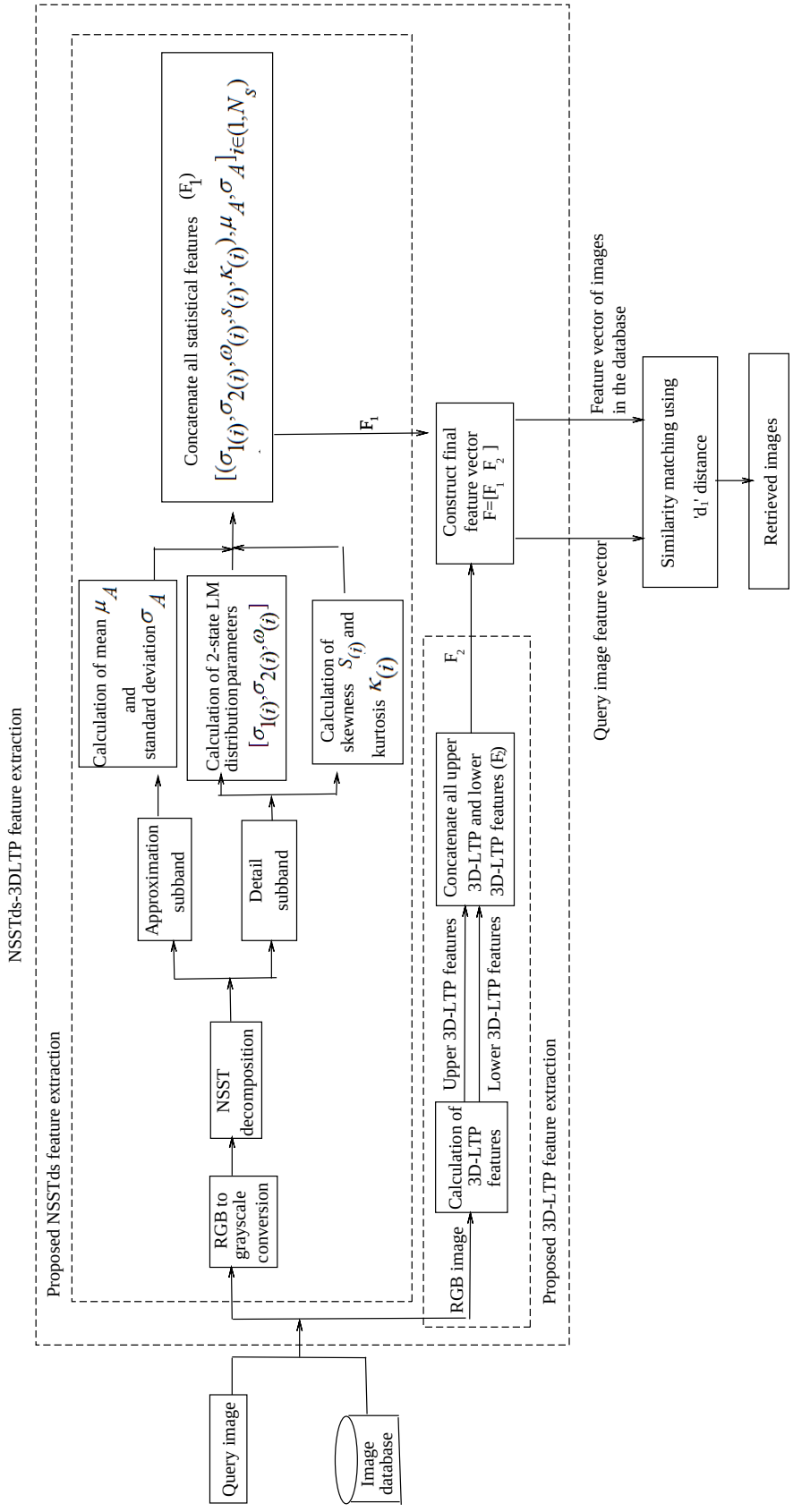


Figure. 3.18: The NSSTds-3DLTP descriptor's functional block diagram in an image retrieval framework

two Laplacian distributions are combined, the tails degenerate more slowly than the tail of a single Laplacian distribution. The likelihood of correctly estimating each parameter decreases with the increase in number of parameters to be evaluated, even though a mixture of three or more Laplacian distributions may produce heavier tails than either a single Laplacian distribution or a mixture of two Laplacian distributions. As a result, we considered about using a 2-state LM model to simulate the statistics of NSST subband coefficients of RS images[171, 190].

The 2-state LM distribution or model refers to the mixing of two independent Laplacian distributions. Let $P_{x(j)}(x(j))$ indicate a 2-state LM model for modeling the image NSST detail coefficients $x(j)$, where j=1,2.....N_T and N_T is the no. of set of coefficients [171]:

$$P_{x(j)}(x(j)) = \omega(j)P_1(x(j)) + (1-\omega(j))P_2(x(j)) \tag{3.30}$$

where $\omega(j)$ and $(1-\omega(j))$ represent the weights of two distinct Laplacian pdf's $P_1(x(j))$ and $P_2(x(j))$ respectively.

When $P_1(x(j)) = \frac{1}{\sigma_1(j)\sqrt{2}} e^{\frac{-\sqrt{2}|x(j)|}{\sigma_1(j)}}$ and $P_2(x(j)) = \frac{1}{\sigma_2(j)\sqrt{2}} e^{\frac{-\sqrt{2}|x(j)|}{\sigma_2(j)}}$, (3.30) can be expressed as:

$$P_{x(j)}(x(j)) = \omega(j)\frac{1}{\sigma_1(j)\sqrt{2}} e^{\frac{-\sqrt{2}|x(j)|}{\sigma_1(j)}} + (1-\omega(j))\frac{1}{\sigma_2(j)\sqrt{2}} e^{\frac{-\sqrt{2}|x(j)|}{\sigma_2(j)}} \tag{3.31}$$

To estimate the parameters of the 2-state LM distribution, the parameters σ_1, σ_2, and ω are first initialized where σ_1, σ_2 are the standard deviations of P_1 and P_2 respectively. The EM computation procedures are then iteratively performed until the condition of convergence is satisfied [171, 190].

Expectation procedure: Here, the responsibility element $r_1(j)$ is calculated for each iteration using:

$$r_1(j) \leftarrow \frac{\omega(j)P_1(x(j))}{\omega(j)P_1(x(j)) + (1-\omega(j))P_2(x(j))} \tag{3.32}$$

and

$$r_2(j) \leftarrow (1 - r_1(j)) \tag{3.33}$$

Table 3.6: Average KS test values for a few RS images

NSST decomposition	pdf	Laplacian	BKF	SNIG	LMM
	Level 1	0.041	0.079	0.029	**0.013**
		0.048	0.084	0.026	**0.014**
		0.038	0.084	0.028	**0.013**
		0.036	0.098	0.028	**0.013**
		0.040	0.087	0.028	**0.016**
		0.046	**0.014**	0.027	**0.014**
		0.041	0.083	0.028	**0.016**
		0.034	0.088	0.028	**0.015**
	2	0.048	0.062	0.027	**0.016**
		0.040	0.076	0.029	**0.019**
		0.045	0.064	0.026	**0.017**
		0.049	0.063	0.028	**0.018**
	3	0.051	0.061	0.032	**0.013**
		0.054	0.050	0.031	**0.012**

The elements of responsibility must guarantee that $r_1(j) + r_2(j) = 1$.

Maximization operation: Here, $\omega(j)$ is calculated using:

$$\omega(j) \leftarrow \frac{1}{N_m} \sum_{i \in N_m(j)} r_1(i) \tag{3.34}$$

where $N_m(j)$ stands for a square-shaped local window that contains N_m coefficients and is positioned at $x(j)$. The $\sigma_1(j)$ and $\sigma_2(j)$ values are derived using:

$$\sigma_1^2(j) = \frac{\sum\limits_{i \in N_m(j)} r_1(i)|x(j)|^2}{\sum\limits_{i \in N_m(j)} r_1(i)} \tag{3.35}$$

$$\sigma_2^2(j) = \frac{\sum\limits_{i \in N_m(j)} r_2(i)|x(j)|^2}{\sum\limits_{i \in N_m(j)} r_2(i)} \tag{3.36}$$

With the Laplacian, BKF, and SNIG as viable models, we conduct a KS goodness-of-fit test to support the usage of the 2-state LM model for modelling the statistics of NSST coefficients.

The average KS test statistics for RS images are shown in Table 3.6. We use a 3-level NSST decomposition (with 1, 2, and 3 directions) to execute the KS test. Twenty random images from different classifications were used in the KS

test, including 'Airport','Beach','Farmland','Bridge','Commercial','Footballfield', 'Desert','Industrial','Forest', 'Park','Meadow','River','Pond' and 'Railway'. The 2-state LM model has the smallest KS test result for the majority of subbands, which suggests that it more closely resembles the actual subband coefficients than the Laplacian, BKF and SNIG distributions.

In addition to the KS test, we used the histogram plots (Fig. 3.19) of different detail subbands of the NSST, where Laplacian, BKF, SNIG, and LM model pdf's are fitted in log domain, to show that the 2-state LM distribution is a good fit for modelling the image NSST detail coefficients. As shown in Fig. 3.19, the 2-state LM distribution performs better than other statistical models in approximating the statistics of high frequency detail coefficients. In comparison to the Laplacian, BKF, and SNIG distributions, the 2-state LM model provides the best match, as demonstrated by both Fig. 3.19(through log histogram plots) and Table 3.6(through KS test statistic).

Three parameters (σ_1, σ_2, and ω) of 2-state LM model can be used to fully characterize the pdf of NSST coefficients in each subband. To explain the specific NSST subband coefficients, we employ two additional statistical parameters: skewness (s) and kurtosis (κ). The skewness and kurtosis give information about the symmetry and peakedness of the distribution, respectively. We use the following equations to get the sample skewness and kurtosis for a given sample of n values:

$$s = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right]^{3/2}} \tag{3.37}$$

$$\kappa = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right]^{2}} \tag{3.38}$$

where x_i and $\overline{x}$ denotes the i^{th} value of x and the sample mean respectively.

To describe the statistics of an approximation subband, we use the mean ((μ_A) and the standard deviation (σ_A).

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \tag{3.39}$$

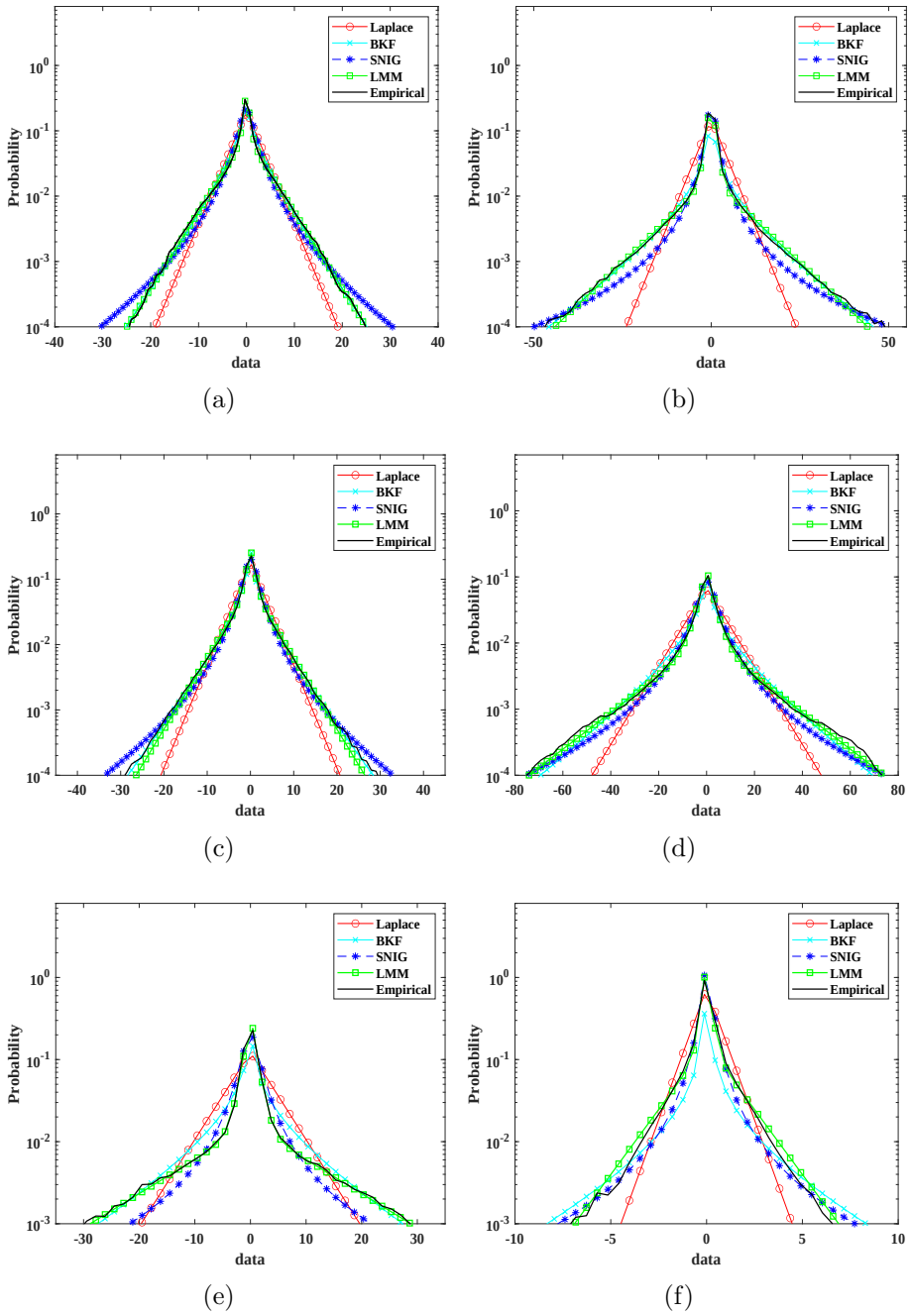


Figure. 3.19: Log histogram of one finest NSST subband of six different images (Two each from WHU-RS19, AID and PatternNet) where Laplacian, BKF, SNIG, and LM pdfs fitted to empirical histogram in log domain

$$\mu = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) \tag{3.40}$$

We compute skewness (s), kurtosis (κ) and 2-state LM distribution parameters from detail subbands along with the estimated mean and standard deviation from the approximation subband. Finally we concatenate them together to form the NSSTds descriptor as : $F_1 = [(\sigma_{1(i)}, \sigma_{2(i)}, \omega(i), s_{(i)}, \kappa_{(i)}), \mu_A, \sigma_A]_{i\in(1,N_s)}$ (N_s is the total no. of detail subbands.

3.2.1.2 Computation of Inter-channel 3D-LTP features

The NSST introduced in the previous section captures the global information. The image's fine arrangement and observable components, which are normally best described using local features, are overlooked by the global feature-based description. For example, a few land use and land cover classes are predominantly represented by discrete objects like baseball fields and storage tanks. To address this problem, we introduce a novel 3D-LTP-based method [191], which is directly applied to the spatial RGB color channels.

The following benefits motivated us to extend LTP to 3D-LTP:

1. Given R, G and B planes of a color image, the 3D-LTP uses the association between a reference intensity value in a given plane and its neighboring intensities in the next neighboring plane w.r.t the same reference position in order to capture the color cue details.

2. As 3D-LTP is able to extract the above mentioned local associationship between the inter R,G and B planes, this procedure acts as a high pass filter which enables the catching of local intensity differences in a particular direction.

The widely used and traditional LBP technique, as described in [192], computes an LBP value by comparing a center pixel to all of its neighbors in a circular neighborhood, and then assigning a 0/1 to each neighbor based on the difference between the center pixel and neighboring pixels as follows:

$$LBP_{R,T} = \sum_{i=1}^{T} 2^{i-1} f\left(I(p_i) - I(p_c)\right) \tag{3.41}$$

$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & else \end{cases} \tag{3.42}$$

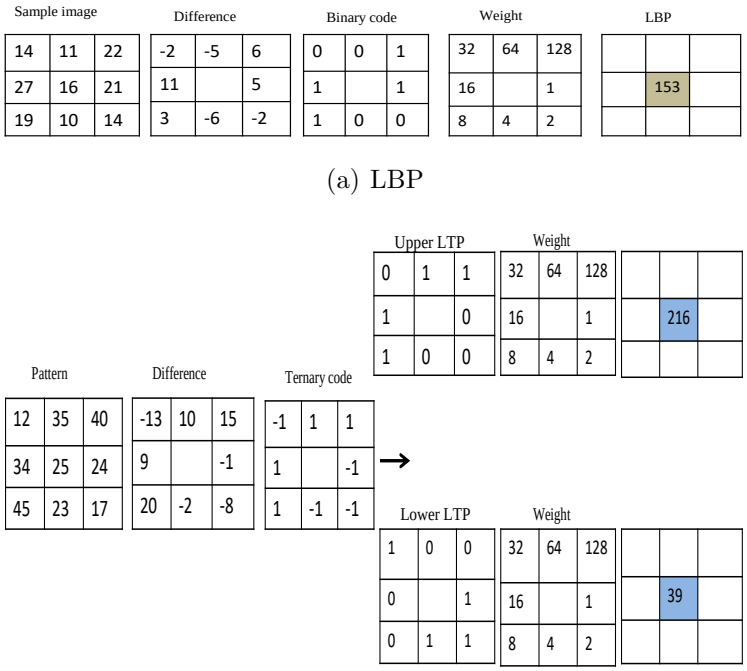
where $I(p_c)$ is the center pixel value, $I(p_i)$ are the neighboring values, T denotes the total no. of neighbors and R is the neighborhood radius.

Tan and Triggs proposed LTP [105],in which the function $f(x)$ is obtained as follows:

$$f(x, I(p_c), th) = \begin{cases} +1 & x \geq I(p_c) + th \\ 0 & |x - I(p_c)| < th \\ -1 & x \leq I(p_{c)} - th \end{cases} \tag{3.43}$$

where $x = (I(p_i) - I(p_c))$.

Fig. 3.20 presents the example computation of LBP and LTP.



(a) LBP

(b) LTP (for threshold 5)

Figure. 3.20: Illustration of LBP and LTP computation for an example image

Since LTP provides a three-valued code to the difference between the center pixel and its surrounding pixels, it is able to capture image details better than LBP.

The extension to 3D-LTP encodes the local texture information of a color image in addition to the color cue information. As seen in the example computation in Fig. 3.21, the inter-channel 3D-LTP generates six pattern images from an RGB color image. The center/reference pixel in the R channel and its G channel neighbors are taken into account during the R channel's encoding. The center/reference pixel in the G channel is examined along with its neighbors from the B channel, whereas the center/reference pixel in the B channel is examined along with its neighbors from the R channel. The 3D-LTP-formed pattern images are shown in Fig. 3.22. An upper and a lower LTP are obtained by calculating the inter channel LTP for each R-G, G-B, and B-R combination. As a result, three upper LTPs and three lower LTPs are produced by the R-G, G-B, and B-R combinations, for a total of six pattern representations.

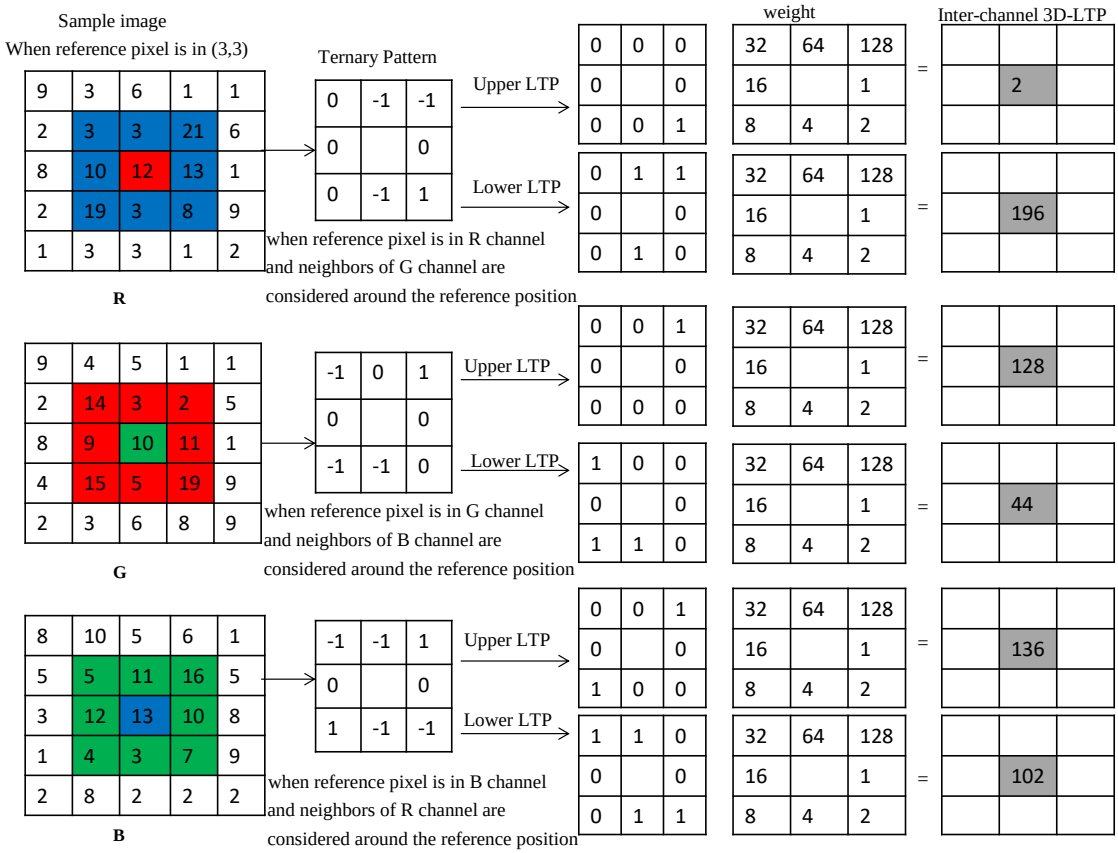
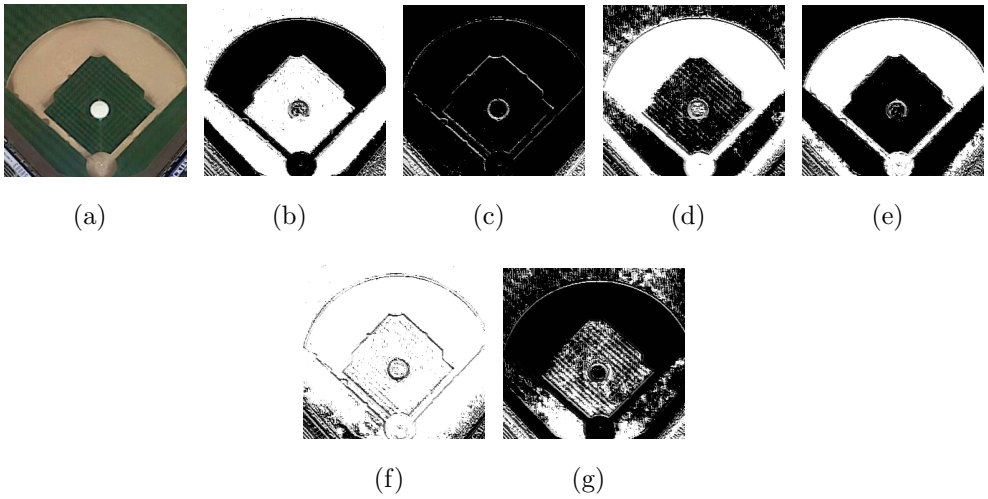


Figure. 3.21: A sample example for 3D-LTP calculation

In 3D-LTP, in order to reduce the feature dimensions, we simply recognize and examine the'uniform' patterns. The term 'uniform' in this context refers to the uniform appearance of 3D LTP, that is, patterns with two or fewer discontinuities in circular binary representation; other patterns are referred to as 'non-uniform' [182, 183]. Given that it only contains two bitwise 0/1 transitions, 00010000 is an example of a uniform pattern, whereas 00101001 is a non-uniform pattern with

several spatial transitions. These 'uniform' patterns have been shown to represent the majority of patterns that correspond to important characteristics like edges, textures, etc. There are 58 different 'uniform' patterns for an image with R=1 and T=8.



(a) (b) (c) (d) (e)

(f) (g)

Figure. 3.22: The resultant pattern maps generated from an example image using 3D-LTP operation (a) Original image (b) Upper-LTP pattern map generated from R-channel, (c)Lower-LTP pattern map generated from R-channel,(d) Upper-LTP pattern map generated from G-channel,(e) Lower-LTP pattern image generated from G-channel,(f) Upper-LTP pattern map generated from B-channel,(g)Lower-LTP pattern map generated from B-channel

3.2.1.3 Fusion of NSSTds and 3D-LTP features

To achieve the best results when describing a scene with intricate patterns and spatial structures, it is common to combine complementary aspects such as local and global features. We provide a blended feature representation based on NSSTds and 3D-LTP for retrieving RS images.

$F_1 = [(\sigma_{p_1(i)}, \sigma_{p_2(i)}, \omega_{(i)}, s_{(i)}, \kappa_{(i)}), \mu_A, \sigma_A]_{i\in(1,N_s)}$ and F_2 are the statistical features (NSSTds) extracted from NSST subbands and 'uniform' 3DLTP features extracted from RGB channels. This two feature vectors are concatenated together to form the final feature vector as $F = [F_1, F_2]$.

By decomposing the RS images with 4-level NSST with 3,3,4,4 directions, a total of 48 detail and 1 approximation subband is obtained. From the detail subbands, $5 \times 48 = 240$ features are extracted. Hence a total 242 (240+2) features are obtained from both detail and approximation subband. From 3DLTP, six pattern maps are obtained (3.22). From this six pattern images, a total of (59 $\times$

6) = 354 features are generated. Hence, the final fused NSSTds-3DLTP descriptor possess a feature dimension of length 242 + 354 = 596 .

The steps of NSSTds-3DLTP feature extraction operation in an image retrieval structure are summarized as follows :

Algorithm: RSIR with modeling the NSST detail coefficients with 2-state LMM distribution **Require:** Input: Query image; Output: n_T‘ number of images retrieved.

1. Load the RS image and convert it into grayscale image if the input image is a colored one.
2. Decompose the gray scale query image using NSST
3. Concatenate the extracted 2-state LMM parameters, kurtosis, and skewness from each NSST detail subband with the mean and standard deviation from the approximation subband to create the feature vector F_1.
4. Calculate the 3D-LTP features based on the R,G,B color channels of the original RS image to produce the feature vector F_2.
5. Combine NSST-domain statistical features (NSSTds) and the 3D-LTP based features to create the final feature vector set. $F = [F_1, F_2]$ after normalization
6. Retrieve similar images using d_1 as the similarity measurement distance

---

3.2.2 Experimental results and discussion

This section offers experimental results that show how effectively the NSSTds-3DLTP perform in a image retrieval framework.Prior to providing the database data, the experimental circumstances are presented. The experimental findings and analyses are then provided, in which the NSSTds-3DLTP is compared with a number of well-known descriptors. The experiments make use of three datasets of RS images that are openly accessible. The datasets' specifics are shown in Table 3.2.

The effectiveness of the NSSTds-3DLTP is assessed for each database in terms of MAP and ANMRR in comparison to Gabor RGB [193], Granulometry [184], LBP [98], FV [194], VLAD [195], and MRELBP [129] (Table 3.7-3.8). .

Table 3.7: Retrieval performance comparison of NSSTds-3DLTP with other schemes for WHU-RS19

Dataset	Descriptors	MAP	ANMRR
WHU-RS19	Gabor RGB [186]	31.69	0.570
	Granulometry [184]	23.39	0.670
	LBP [182, 183]	24.06	0.663
	FV [187]	38.06	0.532
	VLAD [195]	41.28	0.561
	MRELBP [129]	38.91	0.520
	NSSTds	38.71	0.499
	3D-LTP	32.85	0.576
	NSSTds-3DLTP	**44.98**	**0.451**

Table 3.8: Retrieval performance comparison of NSSTds-3DLTP with other schemes for AID and PatternNet

Dataset	Descriptors	MAP	ANMRR	Dataset	Descriptors	MAP	ANMRR
AID	Gabor RGB [186]	11.69	0.843	PatternNet	Gabor RGB [186]	26.53	0.686
	Granulometry [184]	9.04	0.877		Granulometry [184]	14.60	0.817
	LBP [182, 183]	10.36	0.857		LBP[182, 183]	29.99	0.686
	FV [187]	16.40	0.748		FV [187]	19.23	0.760
	VLAD [195]	19.61	0.734		VLAD [195]	28.98	0.638
	MRELBP [129]	18.19	0.679		MRELBP [129]	29.53	0.618
	NSSTds	24.62	0.680		NSSTds	29.93	0.672
	3D-LTP	17.54	0.775		3D-LTP	32.32	0.653
	NSSTds-3DLTP	**29.53**	**0.657**		NSSTds-3DLTP	**35.23**	**0.615**

According to Tables 3.7-3.8, for the WHU-RS19 and AID databases, the global NSSTds features outperform the local 3D-LTP features in terms of ANMRR and MAP, whereas for the PatternNet dataset, both NSSTds and 3D-LTP features perform fairly similarly. The proposed NSSTds features' fine performance is due to the use of accurate statistical model which further allows better capturing of texture features of RS images and their ability to capture crucial features, particularly over a range of different scales and orientations. The proposed fused NSSTds-3DLTP surpasses all other approach for each database, including MRELBP, which is a recent technique and is able to for its capacity to capture both global and local properties (Table 3.7-3.8). The combined NSSTds-3DLTP show notable improvement over NSSTds and 3DLTP alone both in terms of ANMRR and MAP. In terms of [MAP,ANMRR] ,the NSSTds-3DLTP shows improvement over Gabor RGB, Granulometry, LBP, FV, VLAD, and MRELBP descriptors by ([41.93,20.87]%, [92.30,32.68]%,[86.14,31.97]%,[18.18,15.22]%,[8.96,19]%, [15.60,13.26]% respectively)(for WHU-RS19 dataset), ([152.60, 22.06]%, [226.65,25.08]%, [185.03, 23.33]%, [80.06, 12.16]%, [50.58, 10.49]%, and [62.34,3.24]%) respectively (for AID dataset), ([32.79, 10.34]%, [141.30, 24.72]%, [17.47, 10.34]%, [83.20, 19.07]%,

[21.56, 3.60]%, and [19.30,0.48]%) respectively (for PatternNet dataset). The subfeature 3D-LTP not only encodes the color cue but also extracts the local texture data, in contrast to the most of other methods, which are either local or global. Additionally, the NSSTds subfeature may record image data at various sizes and orientations. The NSSTds-3DLTP creates a highly discriminative representation by combining the complementing traits of NSSTds and 3D-LTP.

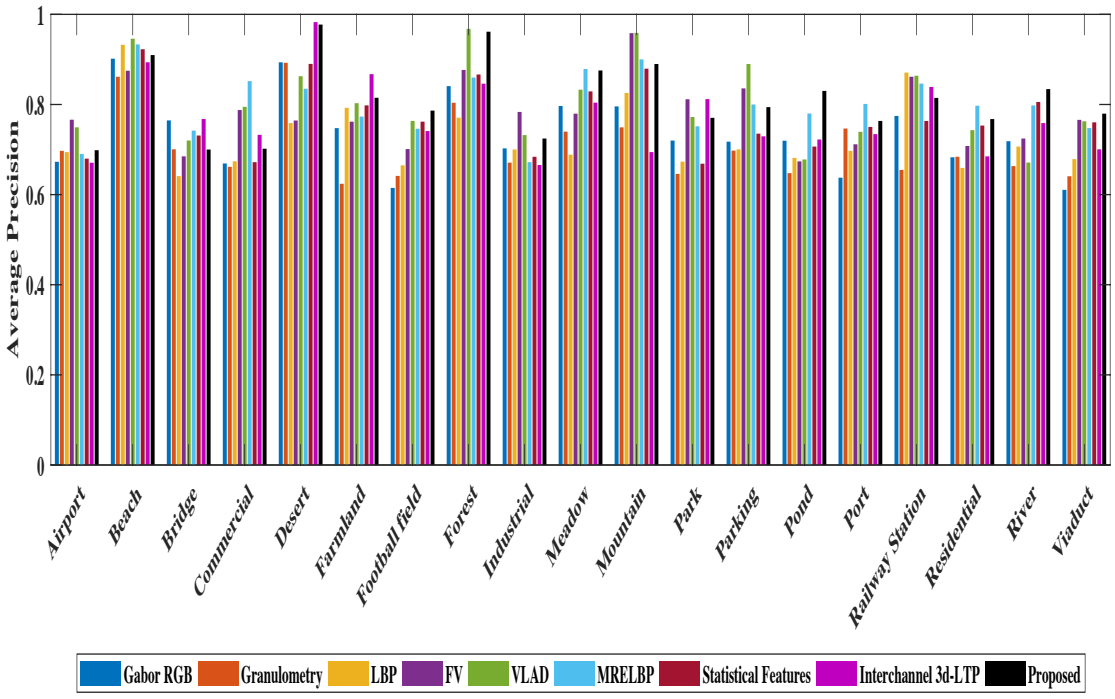


Figure. 3.23: The per class average precision comparison of the NSSTds-3DLTP with other techniques for WHU-RS19 dataset

The average precision of all descriptors, including NSSTds, 3D-LTP, NSSTds-3DLTP, for various classes and databases is shown in Figs. 3.23, 3.24 and 3.25. The average precision value is computed using the top 20 images that were correctly retrieved after 20 trials, with 20 images being randomly chosen from each image class. When compared to other techniques, such as global NSSTds and local 3D-LTP, the NSSTds-3DLTP features perform the best in the majority of the individual classes, as shown in Figs. 3.23, 3.24 and 3.25. As observed in Fig. 3.23-3.25, the global NSSTds features alone perform well on classes like 'forest','river','bareland','residential','school','mountain', and 'parking', which are more texture-based and have image-scale properties. However, the local 3D-LTP alone exhibits good performance when compared to the global NSSTds on classes like an intersection, railroad, baseball diamond, freeway, storage tank, golf course,medium residential, church, pond, bridge, commercial etc. that contain a distinctive arrangement of structures without which the images cannot be cor-

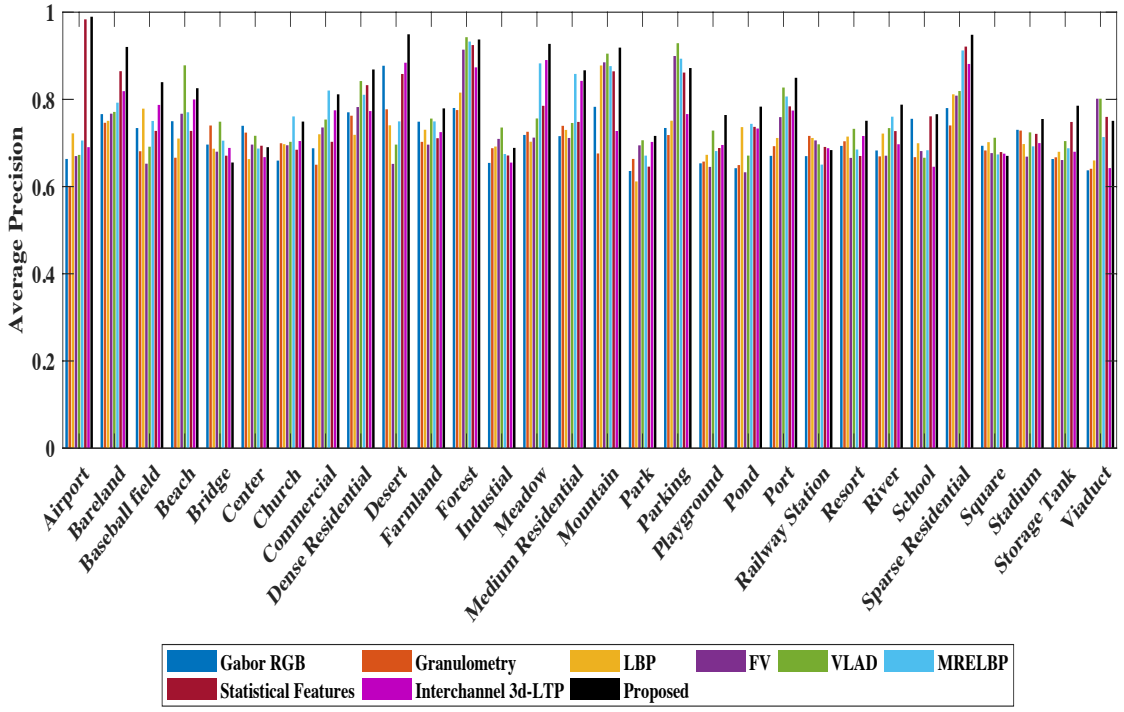


Figure. 3.24: The per class average precision comparison of the NSSTds-3DLTP with other techniques for AID

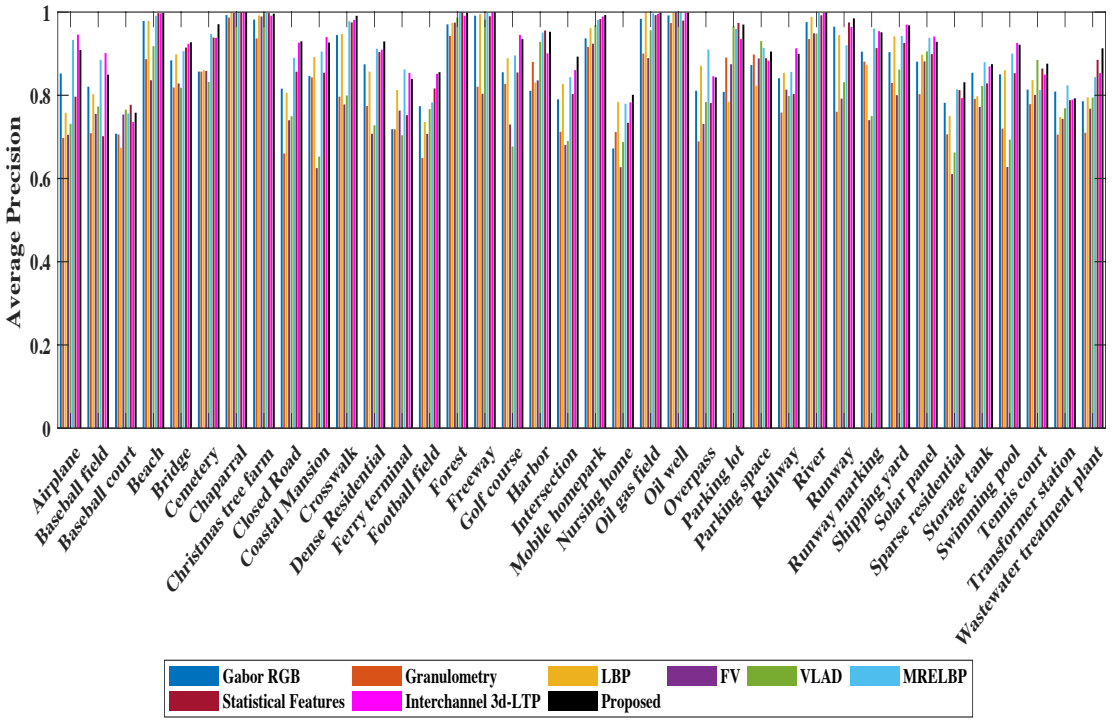


Figure. 3.25: The per class average precision comparison of the NSSTds-3DLTP with other techniques for PatternNet dataset

rectly retrieved. These findings demonstrate that local and global features both contain information that is mutually beneficial, and their combination is predicted to increase the discriminative power of features. Using fused NSSTds-3DLTP descriptors, retrieval of difficult images like tennis court, dense residential, sparse residential, stadium, playground, etc. is greatly enhanced.

(a) Forest (b) Mountain (c) River (d) Residential (e) Bareland (f) School

(g) Golf Course (h) Intersection (i) Railway (j) Baseball field (k) Freeway (l) Bridge

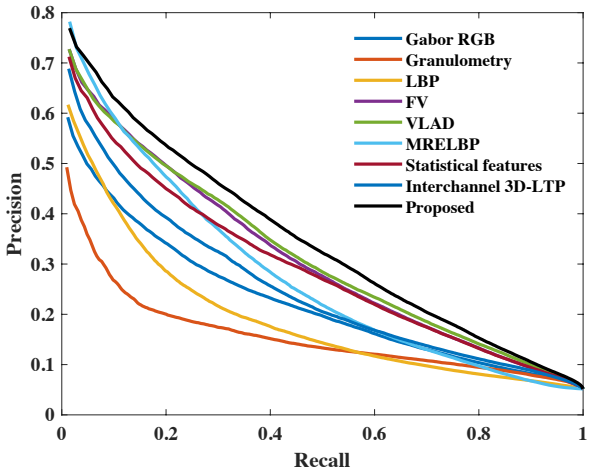
Figure. 3.26: Image classes for which global NSSTds alone (first-row) and local 3D-LTP alone (second-row) produce superior results when compared to one other.

Table 3.9: Comparison of feature dimensions of various techniques

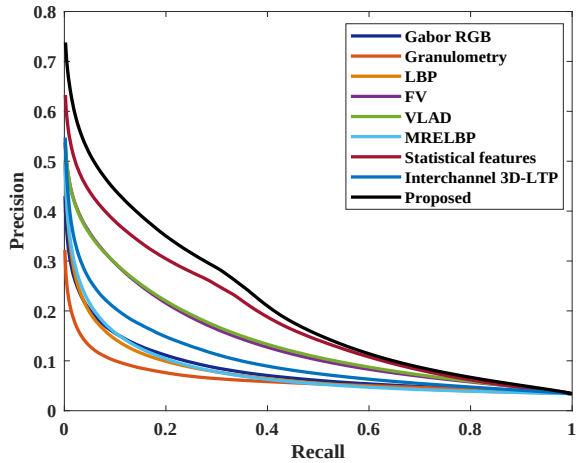
Methods	Gabor RGB	Granulometry	LBP	FV	VLAD	MRELBP	NSSTds-3DLTP
Dimension	96	78	256	512	512	800	596

Table 3.9 lists all other methods and the proposed method's feature dimension. It is clear from Table 3.9 that the feature dimensions of NSSTds-3DLTP are lower than MRELBP and larger than those of other approaches. In comparison to NSSTds-3DLTP, the Gabor RGB, granulometery, LBP, FV, and VLAD algorithms have smaller feature dimensions, yet they perform significantly worse. With relatively fewer feature dimensions, the NSSTds-3DLTP surpasses the most recent MRELBP.

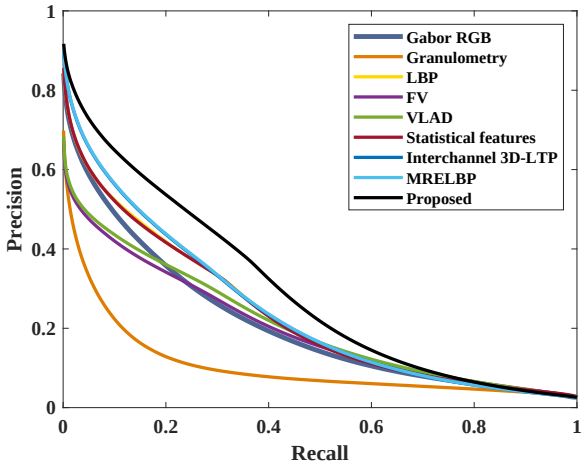
In Table 3.10, the total retrieval time is presented for NSSTds-3DLTP as well as other feature descriptors considered for comparison. The NSSTds-3DLTP is much faster than MRELBP but slower than other descriptors. However, the performance of all other descriptors is much inferior to the NSSTds-3DLTP scheme.



(a) WHU-RS19



(b) AID



(c) PatternNet

Figure. 3.27: The P-R curve obtained for different datasets

Table 3.10: Total retrieval time (in seconds) comparison of the NSST-SNIG with other techniques

Dataset	Gabor RGB	Granulometry	LBP	FV	VLAD	MRELBP	NSSTds-3DLTP
WHU-RS19	2.19	1.77	2.74	4.59	5.10	10.23	5.85
AID	6.41	3.49	8.71	11.01	11.86	15.31	12.10
PatternNet	7.86	7.12	10.28	19.47	20.12	25.12	21.05

The P-R curves for each technique are shown in Fig. 3.27(a-c) for the WHU-RS19, AID, and PatternNet databases, respectively, to further highlight the superiority of the proposed fused features over all other existing techniques, including NSSTds and 3D-LTP alone. According to Fig. 3.27(a), the P-R curve derived using NSSTds-3DLTP for the WHU-RS19 dataset encompasses the largest area and shows the best results, followed by VLAD, FV, NSSTds, MRELBP, 3D-LTP, Gabor RGB, LBP, and Granulometery descriptors. NSSTDs-3DLTP displays the best results for the AID dataset, followed by NSSTds, VLAD, FV, 3D-LTP, MRELBP, Gabor RGB, LBP, and Granulometry (Fig. 3.27(b)). In the same way, NSSTds-3DLTP yields the best results for PatternNet, followed by MRELBP, 3D-LTP, LBP, NSSTds, VLAD, FV, Gabor RGB, and Granulometry (Fig. 3.27(c)). Results from Table 3.7 and Table 3.8 of the referenced literature are noted to be consistent with the P-R curve results.

A few instances of image queries from various classes and their related retrieved results utilizing the proposed NSSTds-3DLTP descriptor (for WHU-RS19) are shown in Fig. 3.28. The images from the classes beach, bridge, desert, farmland, industrial,river, mountain and pond from the WHU-RS19 dataset, when given as input query images, exhibit correct retrieval results, with the exception of the viaduct class, which provides one incorrect retrieved result, i.e. an image from Railway class is wrongly retrieved here, as can be seen in Fig. 3.28.

3.3 Summary

In this chapter, two NSST-based feature descriptors are proposed for RSIR. The first one is a low dimensional global feature based approach that uses NSST domain statistical parameters for construction of feature vector. We have demonstrated

Figure. 3.28: Retrieved images obtained for few input query images taken from multiple classes, using NSSTds-3DLTP (Input query image, correct retrieved results and wrong retrieved results are enclosed in Black, Green and Red coloured boxes respectively for more clarity.)

that the NSST detail subbands coefficients can be best approximated using SNIG distribution when compared to other probable non-Gaussian distribution such as BKF and Laplacian. We have extracted SNIG parameters and the simple mean along with standard deviation are extracted from the approximation subband to form the final feature vector. The parameter estimation of SNIG distribution is carried out using an EM type algorithm. Due to the use of NSST and an accurate statistical distribution that approximates the NSST coefficient statistics along with other effective statistical features, the proposed NSST-SNIG outperforms many low dimensional well known feature descriptors with less retrieval time.

The high resolution RS images have quite complex and highly varied scene contents where many schemes cannot provide the discriminative information in many situations as the crucial structural details present in the image sometimes influences the corresponding image class. The blend of local and global features solves this problem to achieve desired performance. We introduce a technique that blends the global NSST domain statistical features and spatial domain local texture features for improved retrieval of RS images. As the performance of low dimensional transform domain statistical modelling based methods relies mostly on the choice of accurate statistical models, in the next work we demonstrate that the 2 -state LM model as the most appropriate statistical distribution when compared to BKF, Laplacian and also SNIG distribution. We extract 2 -state LM model parameters along with kurtosis and skewness to form the final feature vector which we refer ro as global NSSTds features. We introduce an extension of LTP which is referred as 3D-LTP in order to extract both local intensity variation information across the RGB planes and the color information too. We fuse both the local 3D-LTP and global NSSTds features in order to encode the wide variations in an image scene. From the extensive experiments, we observe that the fused features i.e. NSSTds-3DLTP provides powerful discriminative features and outperform many state of the art schemes including MRELBP which is well known example of fusion of global and local features.