Computer Modeling in
Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2022.018339

ARTICLE

Remote Sensing Image Retrieval Based on 3D-Local Ternary Pattern (LTP) Features and Non-subsampled Shearlet Transform (NSST) Domain Statistical Features

Hilly Gohain Baruah*, **Vijay Kumar Nath and Deepika Hazarika**

Department of Electronics and Communication Engineering, Tezpur University, Assam, 784028, India

*Corresponding Author: Hilly Gohain Baruah. Email: hilly90@tezu.ernet.in

Received: 17 July 2021 Accepted: 22 October 2021

ABSTRACT

With the increasing popularity of high-resolution remote sensing images, the remote sensing image retrieval (RSIR) has always been a topic of major issue. A combined, global non-subsampled shearlet transform (NSST)-domain statistical features (NSSTds) and local three dimensional local ternary pattern (3D-LTP) features, is proposed for high-resolution remote sensing images. We model the NSST image coefficients of detail subbands using 2-state laplacian mixture (LM) distribution and its three parameters are estimated using Expectation-Maximization (EM) algorithm. We also calculate the statistical parameters such as subband kurtosis and skewness from detail subbands along with mean and standard deviation calculated from approximation subband, and concatenate all of them with the 2-state LM parameters to describe the global features of the image. The various properties of NSST such as multiscale, localization and flexible directional sensitivity make it a suitable choice to provide an effective approximation of an image. In order to extract the dense local features, a new 3D-LTP is proposed where dimension reduction is performed via selection of 'uniform' patterns. The 3D-LTP is calculated from spatial RGB planes of the input image. The proposed inter-channel 3D-LTP not only exploits the local texture information but the color information is captured too. Finally, a fused feature representation (NSSTds-3DLTP) is proposed using new global (NSSTds) and local (3D-LTP) features to enhance the discriminativeness of features. The retrieval performance of proposed NSSTds-3DLTP features are tested on three challenging remote sensing image datasets such as WHU-RS19, Aerial Image Dataset (AID) and PatternNet in terms of mean average precision (MAP), average normalized modified retrieval rank (ANMRR) and precision-recall (P-R) graph. The experimental results are encouraging and the NSSTds-3DLTP features leads to superior retrieval performance compared to many well known existing descriptors such as Gabor RGB, Granulometry, local binary pattern (LBP), Fisher vector (FV), vector of locally aggregated descriptors (VLAD) and median robust extended local binary pattern (MRELBP). For WHU-RS19 dataset, in terms of {MAP,ANMRR}, the NSSTds-3DLTP improves upon 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%,19.60%} and {15.60%,13.26%}, respectively. For AID, in terms of {MAP,ANMRR}, the NSSTds-3DLTP improves upon Gabor RGB, Granulometry, LBP, FV, VLAD and MRELBP descriptors by {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 PatternNet, the NSSTds-3DLTP respectively improves upon Gabor RGB, Granulometry, LBP, FV, VLAD and MRELBP descriptors by {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%} in terms of {MAP,ANMRR}. The moderate dimensionality of simple NSSTds-3DLTP allows the system to run in real-time.

KEYWORDS

Remote sensing image retrieval; laplacian mixture model; local ternary pattern; statistical modeling, KS test, texture, global features

1 Introduction

Due to advances in remote imaging sensors and earth observation technologies, the volume of high resolution remote sensing images have increased dramatically. Urbanization preparation, deforestation detection, weather prediction, farm monitoring, and military applications are only a few of the uses for remote sensing images. Hence, precise image retrieval techniques for remote sensing images are very important [1–4]. The increase in spatial resolution and database volume have led to difficulties in the manual annotation based image retrieval approaches. Hence it is very important to have a proper management framework to deal with this huge volume of remote sensing data. Image retrieval techniques based on image content plays crucial role to handle this data properly. Feature extraction and similarity measurement are the two key modules in content based image retrieval (CBIR). In the feature extraction module, features that describe visual content of an image are extracted, and then the similarity between the features extracted from the query and the database images is calculated. Remote sensing images consist of semantic objects of wide range as they cover quite large geographical area. The main challenge in case of remote sensing images is the presence of variations in appearance in semantic objects of same category [2]. Hence for retrieval of remote sensing images, the features extracted should be highly robust and descriptive. The time requirement for retrieval of images of interest out of the huge volume of remote sensing data is also to be considered while designing an efficient remote sensing image retrieval framework.

The texture, color and shape information are the primary visual attributes of any high resolution remote sensing images and describes important details for scene retrieval. The most of the earlier literature are based on these features. The various handcrafted features like scale invariant feature transform (SIFT) [5], color histogram [6], gist [7], histogram of oriented gradients (HOG) [8] and local binary pattern (LBP) [9], etc. exist in the literature. The color histogram and gist describes the global features whereas the SIFT, HOG and LBP describes the local features of an image. Global features denote the visual details of an image as a whole. The macrostructure informations in an image can be well captured with global features. In [1], Ferecatu and Boujemaa for image retrieval employed global image descriptors that are constructed using statistical representation of color, texture as well as shape features and demonstrated few search examples that exhibits the effectiveness of relevance feedback on remote sensing images. In [10], Ma et al. proposed a shape based descriptor that uses region and polygonal extraction. In an another approach, Yang et al. [11] showed the use of color layer based texture elements histogram along with color fuzzy correlogram for retrieval of remote sensing images. Few techniques employ both statistical model and multiresolution analysis to describe the global features of the images. Choy et al. [12] modeled the wavelet coefficients of images using three parameter generalized Gamma distribution and used its parameters to describe the texture feature. In [13], the image wavelet subband coefficients were modeled using finite mixtures of generalized Gaussian distribution. The parameters of this distribution were used as features to describe the subband images. Liu et al. [14] modeled the NSST coefficients of remote sensing images using Bessel K distribution and used its parameters to describe the texture feature. In [3], another remote sensing image retrieval approach based on statistical modeling was introduced where the symmetric normal inverse Gaussian (SNIG) distribution was used to model detail subbands of the NSST. The estimated SNIG parameters were used to construct the feature vector. The main strength of transform domain statistical modeling based techniques is that the texture discrimination here is treated as an issue of similarity measurement between statistical distributions, which is relatively easy to implement when compared to Markov random fields [15]. The global feature based techniques are usually effective on the categories that are largely texture based and carries image-scale details. The local feature based techniques however are effective in the categories which exhibits definite or perceptible structures whose presence/absence is used to discriminate the images. The local pattern based schemes such as LBP [16], local ternary pattern (LTP) [17], etc. captures microstructure information only and is highly appropriate for dense local feature extraction. In [18,19], a technique using patch based complete local binary pattern in multi-scale framework is introduced for remote sensing image scene classification. Bian et al. introduced extended multi-structure local binary pattern for scene classification of remote sensing images. In [20], the original Bag-of-words (BOW) model was improved by characterizing the images using local features that are extracted from base images for retrieval of remote sensing images. Sukhia et al. [21] proposed to use LTP in three different scales for extraction of features from remote sensing images. These features are then encoded with Fisher vector encoding scheme.

Both global and local features capture complementary informations and their combinations are observed to be effective in improving the retrieval and classification performance. In order to describe a high resolution image scene with high diversity, many techniques fail to supply discriminative details especially when some major structural information in the image usually dominate the image class. In such cases, the fusion of both local and global features are usually preferred to obtain improved performance [22,23]. In the last one decade, various schemes [24–31] have been introduced that combine both local and global features. In [24], Bian et al. fused the local features that are extracted using codebookless model and the global features that are exploited using saliency based multiscale multiresolution multistructure local binary pattern for classification of high resolution remote sensing scenes. In [25], Risojevic et al. extracted local features employing the SIFT and global features utilizing the enhanced Gabor texture descriptor, and were combined using a scheme to enhance the classification of remote sensing image scenes. In [26], Liu et al. introduced median robust extended LBP (MRELBP) which not only captures microstructure information but macrostructure too. In [27], Yang et al. extracted global features from high pass subband images of dual-tree complex wavelet and local features from LBP applied on all low pass subband images. The authors finally combined both these features for texture classification. In [28], Kabbai et al. combined both local and global features for image classification. Local features were extracted using speeded up robust feature descriptor and the global features were extracted through combination of wavelet transform based features with modified form of local ternary pattern (LTP). A multiple feature based regularized kernel is introduced for classification of hyperspectral images [29]. Various spatial features such as local feature, shape, spectral and global features are combined to supply more discriminative information. For local, global and shape features; LBP feature, Gabor feature and extended multiattribute profiles are exploited respectively.

It is discussed in [14] that the combination of features not everytime assures improved retrieval performance. For any images with high amount of details, it is essential to select efficient features that are supportive to each other in order to achieve improved retrieval results and to effectively blend them without increase in feature dimensions. Motivated from [3,14,24,27], we introduce a remote sensing image retrieval technique that uses an effective combination of new local 3D-LTP based features and novel global NSST domain statistical features. The 3D-LTP descriptor encodes both the colour cue information and local texture details. It is shown that the 2-state LM distribution best fits the statistics of detail NSST subband coefficients than BKF, Laplacian and SNIG distributions. Through accurate statistical modelling we calculate the discriminative global texture features from NSST subbands using the 2-state LM distribution parameters along with subband kurtosis and skewness. Since the local or global features describes complementary image informations and alone cannot provide discriminative description in many situations, we propose an effective blend of global and local features along with colour information to improve the discriminativeness of the features. The proposed NSSTds-3DLTP outperforms Gabor RGB, Granulometry, LBP, FV, VLAD and MRELBP descriptors in terms of MAP, ANMRR and P-R curve analysis for WHU-RS19,AID and PatternNet datasets. The NSSTds-3DLTP is highly suitable in retrieval of high resolution remote sensing images where accurate and fast search procedures are required in order to retrieve the most relevant images.

The main contributions of the paper are:

1. The image NSST detail subband coefficients are modeled using 2-state Laplacian mixture model. It is demonstrated that the 2-state Laplacian mixture model best fits the subband coefficients when compared to other highly non-Gaussian distributions such as Laplacian, Bessel K form and SNIG. The 2-state Laplacian mixture model parameters in addition with kurtosis and skewness are calculated from detail subbands, and the mean along with standard deviation are calculated from the approximation subband and are concatenated together to construct the feature vector, to represent the global features of the image. This global feature is referred to as NSSTds.
2. Since the classical LTP ignores the encoding of color feature which is also one of the crucial visual attribute, an extension to 3D-LTP is introduced in this paper in order to encode not only the local intensity variations across the planes but also the color information.
3. The proposed fusion of local and global features achieves highly discriminative feature representation with much less dimensions. This feature fusion is referred to as NSSTds-3DLTP.

The paper is organized with the following structure. Section 2 presents a brief review on NSST and detailed description on proposed framework in remote sensing image retrieval. Section 3 reflects the performance analysis on the experimental outcomes obtained. Section 4 concludes the paper.

2 Methodology

2.1 Nonsubsampled Shearlet Transform (NSST)

Though wavelet transform deals effectively with the point singularities of signals, it fails to capture the linear singularities that exist in the images [32]. To handle this problem, different multi-geometric analysis tools such as curvelet [33], contourlet [34] and shearlet transform [35] were introduced in the literature.

NSST inherits the benefits of classic theory of the affine systems, i.e., it is an extension of the wavelet theory. The important properties of NSST such as localiztion, multiscale, translational invariance and high directional sensitivity enables the NSST to provide a powerful image representation. Inspite of the significant developments that has been made, the effective texture description is still a demanding problem that needs attention. Therefore, in this paper we intend to develop a shearlet based texture descriptor to describe the texture information more effectively.

The continuous shearlet transform (ST) of image f in two dimension can be defined as follows:

$$ST_{\varphi,f}(i,s,t) = < f, \varphi_{i,s,t} > \tag{1}$$

where i, t and s represents scale, translation and orientation parameters, respectively [32]. The shearlet function $\varphi_{(}i,s,t)$ is defined as

$$\varphi_{i,s,t} = |detK_{i,s}|^{-1/2}\varphi(K_{i,s}^{-1}(a-t)); i > 0, s \in R, t \in R^2, \varphi \in L^2(R^2) \tag{2}$$

The notation L^2 denotes a vector space of square integrable functions on a 2-D euclidean space R^2.

The parameter $K_{i,s} = \begin{bmatrix} i & -\sqrt{i}s \\ 0 & \sqrt{i} \end{bmatrix}$ for $i > 0,\ s \in R,\ t \in R^2$. The $K_{i,s}$ matrix can be factorized as $K_{i,s} = \begin{bmatrix} 1 & -s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} i & 0 \\ 0 & \sqrt{i} \end{bmatrix} = B_s A_i$ where B_s and A_i denotes shear and diagonal matrices. It is important to note that anisotropic dilation is carried out by A_i (for multiscale partitions) and shearing is done by B_s matrix (for directional analysis).

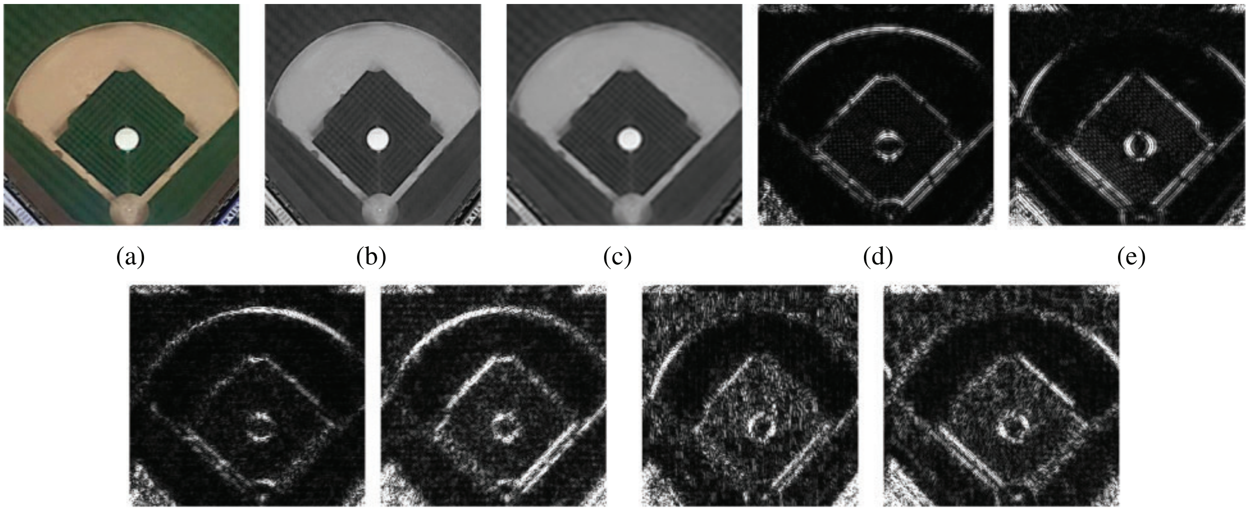
In NSST, nonsubsampled Laplacian pyramid (NSLP) filtering results in low and high frequency components and directional filtering with different shearing matrices lead to shift invariant form of shearlet transform. The NSST removes the up sampling and down sampling operations unlike shearlet transform and therefore is completely invariant [36,37]. Also NSST is multi-scale and has got high directional selectivity. Therefore the use of NSST in image retrieval applications could do justice to these powerful features of NSST in effectively describing the features of input image.

In Fig. 1, a visual example of image NSST approximation and detail subbands for one remote sensing image is shown. The NSST approximation and detail subbands (Fig. 1) refer to the subbands that contains low-frequency coefficients and the high-frequency coefficients respectively. Figs. 1d–1e and 1f–1i respectively shows high frequency detail coefficients at the finest scale/Scale 1 and at next coarsest scale, i.e., Scale 2.

2.2 The Proposed Remote Sensing Image Retrieval Framework

This subsection describes the proposed NSSTds-3DLTP feature in a RSIR framework in details. The framework consists of two major modules. The first module calculates the global NSSTds features using NSST-domain statistical parameters and the second module calculates the local 3D-LTP features from RGB channels. The proposed global NSSTds features are combined with proposed local 3D-LTP features to generate a fused representation NSSTds-3DLTP leading to an enhanced feature description.

The block diagram of NSSTds-3DLTP based framework is presented in Fig. 2.



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

(f) (g) (h) (i)

Figure 1: Example of NSST 2-level decomposition of a remote sensing image from PatternNet dataset. (a) Original image; (b) Its grayscale form; (c) Approximation subband; (d) 1st Detail subband (Scale 1); (e) 2nd Detail subband (Scale 1); (f) 1st Detail subband (Scale 2); (g) 2nd Detail subband (Scale 2); (h) 3rd Detail subband (Scale 2); (i) 4th Detail subband (Scale 2)

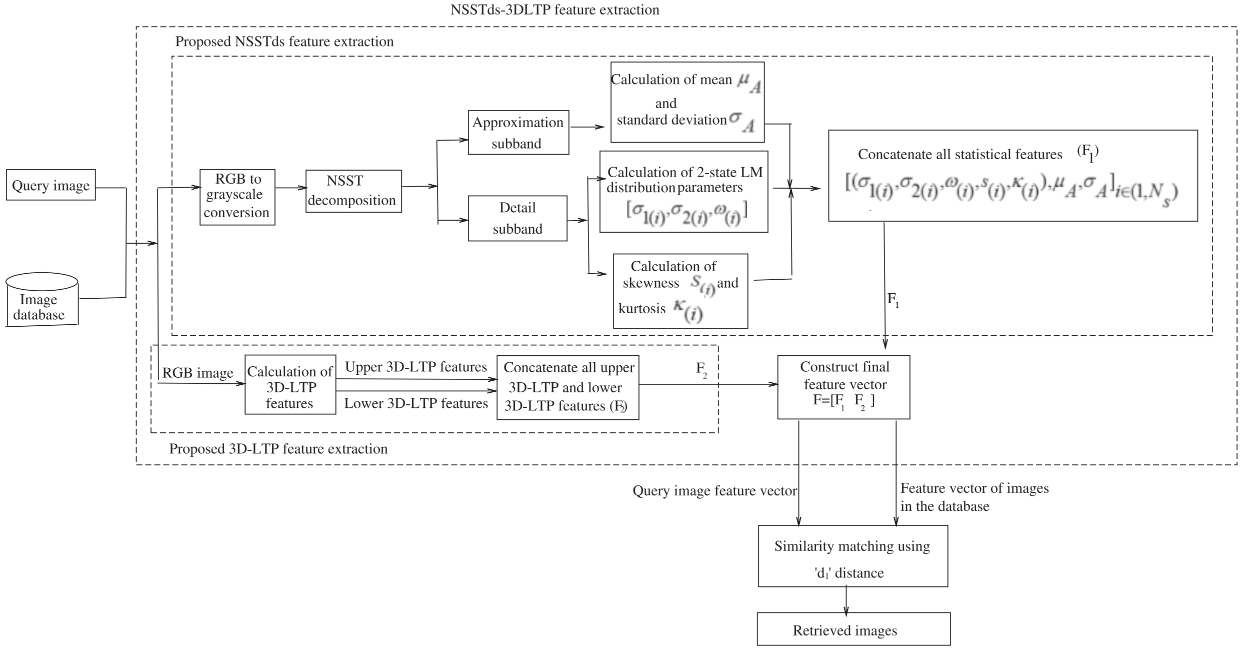

NSSTds-3DLTP feature extraction
Proposed NSSTds feature extraction
Query image
Image database
RGB to grayscale conversion
NSST decomposition
Approximation subband
Detail subband
Calculation of mean μ_A and standard deviation σ_A
Calculation of 2-state LM distribution parameters $[\sigma_{1(i)},\sigma_{2(i)},\omega_{(i)}]$
Calculation of skewness $S_{(i)}$ and kurtosis $\kappa_{(i)}$
Concatenate all statistical features (F_1) $[(\sigma_{1(i)},\sigma_{2(i)},\omega_{(i)},s_{(i)},\kappa_{(i)}),\mu_A,\sigma_A]_{i\in(1,N_s)}$
F_1
RGB image
Calculation of 3D-LTP features
Upper 3D-LTP features
Lower 3D-LTP features
Concatenate all upper 3D-LTP and lower 3D-LTP features (F_2)
F_2
Construct final feature vector $F=[F_1\ F_2]$
Proposed 3D-LTP feature extraction
Query image feature vector
Feature vector of images in the database
Similarity matching using 'd_1' distance
Retrieved images

Figure 2: The block diagram of proposed NSSTds-3DLTP descriptor in an image retrieval framework

2.2.1 *Computation of NSST Domain Statistical Features (NSSTds)*

With statistical modeling of transform coefficients, the texture discrimination problem can be solved with much less dimensions by simply measuring the similarity between the statistical models. This technique is relatively easier to implement and is highly effective. The parametric distributions have been employed to model the image transform coefficients distribution, for retrieval of images [3,14]. In the literature, wavelet transform domain statistical modeling of images have been quite popular. The non-Gaussian distributions such as generalized Gaussian, finite mixture of generalized Gaussian [13] and generalized Gamma [12] based models have been successfully used in image retrieval applications. The discrete wavelet transform are not capable of describing the linear singularities present in images. As also discussed in previous subsection, the multiscale geometric analysis tools such as shearlet provides solution to the above problem as this transform provides excellent sparse representation for higher dimensional singularities. Very recently in [3], it was demonstrated that the image NSST coefficients obey highly non-Gaussian statistics and the symmetric normal inverse Gaussian (SNIG) distribution was shown to be more appropriate than Laplacian and BKF models [14] in modeling the NSST detail coefficients of remote sensing images. Laplacian mixture (LM) model has been known for its good ability to capture very heavy tails of highly non-Gaussian empirical data. It should be noted that the tails of mixture of two Laplacian distributions decays slower than the tail of one Laplacian distribution. The mixture of three or more Laplacian distributions may give more heavy tails than a single Laplacian distribution or mixture of two Laplacian distributions, however as the number of parameters to be estimated increases the potential to estimate them accurately gets decreased. Therefore, we propose to model the NSST subband coefficients of remote sensing images using 2-state LM model [38,39].

In this paper, a mixture of two individual Laplacian distributions is referred to as 2-state LM distribution or model. Let $P_{x(j)}(x(j))$ (where $j = 1, 2 \ldots N_T$ and N_T is the no. of set of coefficients) denote a 2-state LM model for modeling the image NSST high-frequency detail coefficients $x(j)$ which is expressed as [38]:

$$P_{x(j)}(x(j)) = \omega(j) P_1(x(j)) + (1 - \omega(j)) P_2(x(j)) \tag{3}$$

where $\omega(j)$ and $(1 - \omega(j))$ are the weights to two individual 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 $P_2(x(j)) = \frac{1}{\sigma_2(j)\sqrt{2}} e^{\frac{-\sqrt{2}|x(j)|}{\sigma_2(j)}}$, Eq. (3) can be expressed as:

$$P_{x(j)}(x(j)) = \omega(j) \frac{1}{\sigma_1(j)\sqrt{2}} e^{\frac{-\sqrt{2}|x(j)|}{\sigma_1(j)}} + (1 - \omega(j)) \frac{1}{\sigma_2(j)\sqrt{2}} e^{\frac{-\sqrt{2}|x(j)|}{\sigma_2(j)}} \tag{4}$$

The $\sigma_1(j)$ and $\sigma_2(j)$ respectively are the standard deviations of individual pdf's $P_1(x(j))$ and $P_2(x(j))$. The parameters $\sigma_1(j)$, $\sigma_2(j)$ and $\omega(j)$ are first initialized and then estimated using Expectation-Maximization (EM) algorithm [38,39].

To estimate the parameters of 2-state LM distribution, the parameters $[\sigma_1, \sigma_2, \omega]$ are first initialized and subsequently the Expectation Maximization computation procedures are iteratively carried out till the condition of convergence is reached.

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

$$r_1(j) \leftarrow \frac{\omega(j) P_1(x(j))}{\omega(j) P_1(x(j)) + (1 - \omega(j)) P_2(x(j))} \tag{5}$$

and

$$r_2(j) \leftarrow (1 - r_1(j)) \tag{6}$$

The responsibility elements must assure $r_1(j) + r_2(j) = 1$.

Maximization procedure: The $\omega(j)$ is calculated using:

$$\omega(j) \leftarrow \frac{1}{N_m} \sum_{i \in N_m(j)} r_1(i) \tag{7}$$

where $N_m(j)$ denotes a square shaped local window with N_m coefficients inside it and is positioned at the $x(j)$ as center. The $\sigma_1(j)$ and $\sigma_2(j)$ are calculated using:

$$\sigma_1^2(j) = \frac{\sum\limits_{i \in N_m(j)} r_1(i)|x(j)|^2}{\sum\limits_{i \in N_m(j)} r_1(i)} \tag{8}$$

$$\sigma_2^2(j) = \frac{\sum\limits_{i \in N_m(j)} r_2(i)|x(j)|^2}{\sum\limits_{i \in N_m(j)} r_2(i)} \tag{9}$$

In order to defend the use of 2-state LM model in modeling the statistics of NSST coefficients, we perform Kolomogrov-Smirnov (KS) goodness of fit test [38] considering Laplacian, BKF and SNIG as probable models. The KS test statistic supplies the information on distance between empirical CDF (ECDF) and the CDF of a probable distribution. In this test, while calculating the distance information between ECDF and each probable distributions CDF, the one which give minimum KS statistic value is declared as the best fit for the given empirical data.

Mathematically, the KS test can be expressed as [38]:

$$KS_v = max_{x \in N} |F(x) - \hat{F}(x)| \tag{10}$$

where $\hat{F}(x)$ and $F(x)$ are the ECDF and CDF of the model respectively. The parameter x in Eq. (10) denotes the emprical data, i.e., the high frequency NSST detail coefficients whereas N denotes the total number of coefficients in the set of data.

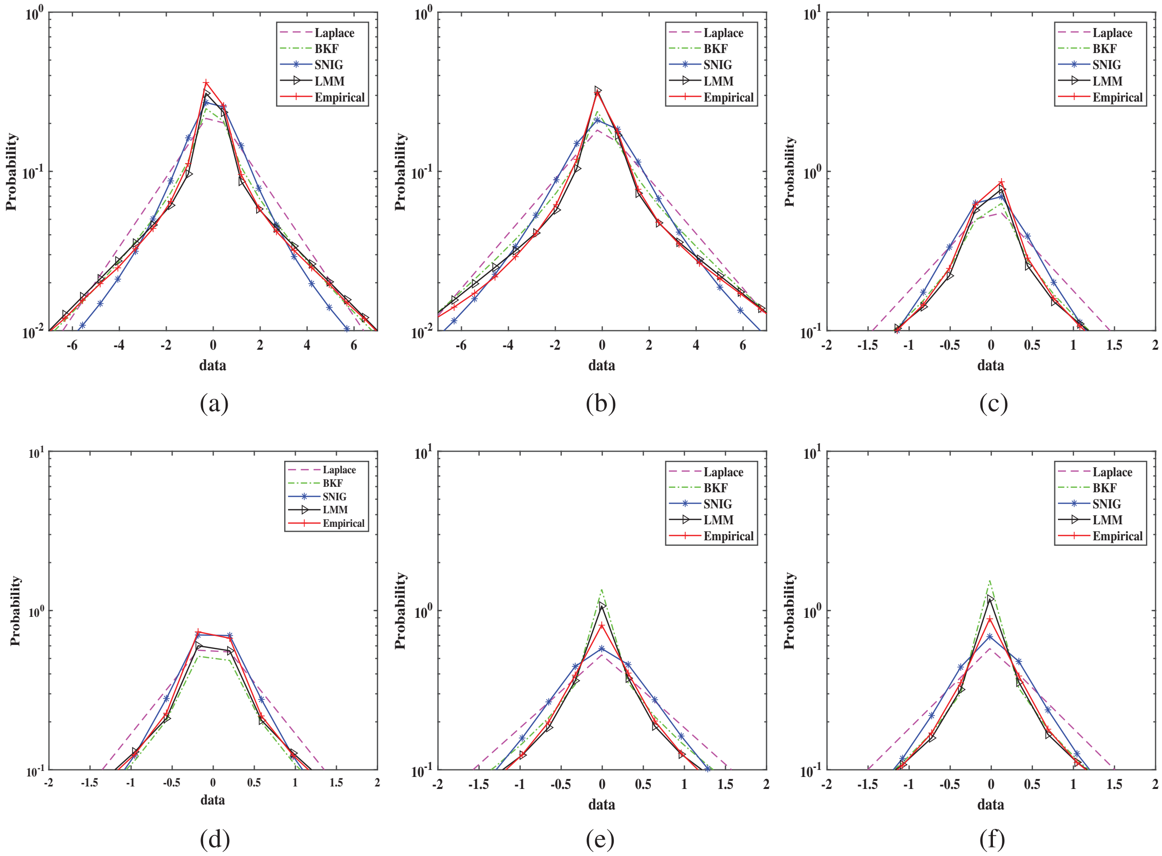
Table 1 exhibits the average KS test statistics for remote sensing images taken from well known WHU-RS19 image dataset.

Table 1: Average KS test values for WHU-RS19 dataset

Dataset	pdf	Level 1								Level 2				Level 3	
		Subband								Subband				Subband	
		1	2	3	4	5	6	7	8	1	2	3	4	1	2
WHU-RS19	Laplacian	0.042	0.041	0.038	0.036	0.041	0.047	0.042	0.034	0.049	0.040	0.046	0.050	0.051	0.055
	BKF	0.080	0.084	0.085	0.099	0.087	**0.014**	0.083	0.088	0.062	0.077	0.064	0.063	0.062	0.051
	SNIG	0.029	0.026	0.028	0.028	0.028	0.027	0.028	0.028	0.027	0.030	0.027	0.028	0.033	0.031
	LMM	**0.013**	**0.014**	**0.013**	**0.013**	**0.016**	**0.014**	**0.016**	**0.015**	**0.016**	**0.019**	**0.017**	**0.018**	**0.013**	**0.012**

For the purpose of performing KS test, we consider a 3-level NSST decomposition (with 1,2,3 directions) that yields one approximation subband and a total of 14 detail subbands. The KS test was performed on 20 random images taken from diverse classes such as ‘Airport’, ‘Beach’, ‘bridge’, ‘commercial’, ‘desert’, ’farmland’, ‘footballfield’, ‘forest’, ‘Industrial, ‘Meadow’, ‘Park’, ‘River’, ‘Pond’, ‘Railway’, ‘Port’ and ‘Residential’ of WHU-RS19 dataset and finally averaged in order to find the most appropriate distribution that approximates the statistics of high-frequency NSST detail coefficients, considering Laplacian, BKF and SNIG distributions. It is clearly visible from Table 1 that for most of the subbands, the KS test value for 2-state LM model is the smallest that reveals clearly that it is able to approximate the detail subband coefficients better than Laplacian, Bessel K form (BKF) and SNIG distributions.

In addition to KS test and to further demonstrate the suitability of 2-state LM distribution in modeling the image NSST detail coefficients, we plotted the histogram plots (Fig. 3) of various detail subbands of NSST in logarithmic domain where Laplacian, BKF, SNIG and LM model pdf’s are fitted in log domain. Fig. 3 clearly demonstrates the superiority of 2-state LM distribution in approximating the statistics of high frequency detail coefficients as compared to other statistical models. Both Fig. 3 (through log histogram plots) and Table 1 (through KS test statistic) confirms that the 2-state LM model provides best fit compared to Laplacian, BKF and SNIG distributions.



(a) (b) (c)

(d) (e) (f)

Figure 3: The log histogram plot for six NSST subbands of one example image from WHU-RS19 dataset where Laplacian, BKF, SNIG and LM pdfs are fitted to empirical histogram in log domain. (a) Subband 1 (Scale 1); (b) Subband 2 (Scale 1); (c) Subband 1 (Scale 2); (d) Subband 2 (Scale 2); (e) Subband 3 (Scale 2); (f) Subband 4 (Scale 2)

Given the 2-state LM model, the probability density function (pdf) of NSST coefficients in each subband can be fully described using three parameters $\sigma_1, \sigma_2, \omega$. We use two more statistical parameters namely skewness (s) and kurtosis (κ) to describe the detail NSST subband coefficients. The skewness and kurtosis indicates about the distribution symmetry and distribution peakedness respectively. For a given sample of n values, we calculate the sample kurtosis and skewness using following expressions:

$$s = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right]^{3/2}} \tag{11}$$

$$\kappa = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right]^{2}} \tag{12}$$

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

We use simple statistical mean and standard deviation features to describe the statistics of approximation subband.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{13}$$

$$\mu = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) \tag{14}$$

Finally to describe the image NSST subbands using statistical features, we calculate the 2-state LM distribution parameters $(\sigma_1, \sigma_2, \omega)$, skewness ($s$) and kurtosis ($\kappa$) from each detail subband along with the mean (μ_A) and standard deviation σ_A estimated from the approximation subband, and concatenate them to construct the NSSTds feature vector 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 number of detail subbands).

2.2.2 *Computation of Inter-Channel 3D-Local Ternary Pattern (3D-LTP) Features*

The NSSTds proposed in previous subsection are regarded as global features of an image. The global feature based description however fails to describe the detailed arrangement and perceptible objects present in an image which usually can be best described using local features. For instance, few land use and land cover based categories are illustrated largely by discrete objects such as baseball fields and storage tanks. In order to address this issue we propose a new 3D-LTP based technique where it is directly applied on the spatial RGB color channels [40].

The 2-fold motivation of extension of LTP to 3D-LTP are:

1. Since the RGB planes have high inter-plane correlation, the 3D-LTP exploits the relationship between a pixel intensity in one plane with respect to the neighbors in the next plane in reference to the same spatial position, thus capturing the color cue information too.
2. Since the 3D-LTP can capture the above local inter-plane relationship, the process behaves like a high-pass kind of filter which catches the local intensity variations in an orientation.

The traditional and popular LBP technique [9] describes the texture feature by computing a LBP value where a center pixel is compared to all its neighbors in a circular neighborhood and a 0/1 is assigned to each neighbor based on the center pixel and neighboring pixel difference as follows:

$$LBP_{R,T} = \sum_{i=1}^{T} 2^{i-1} f\left(I(p_i) - I(p_c)\right) \tag{15}$$

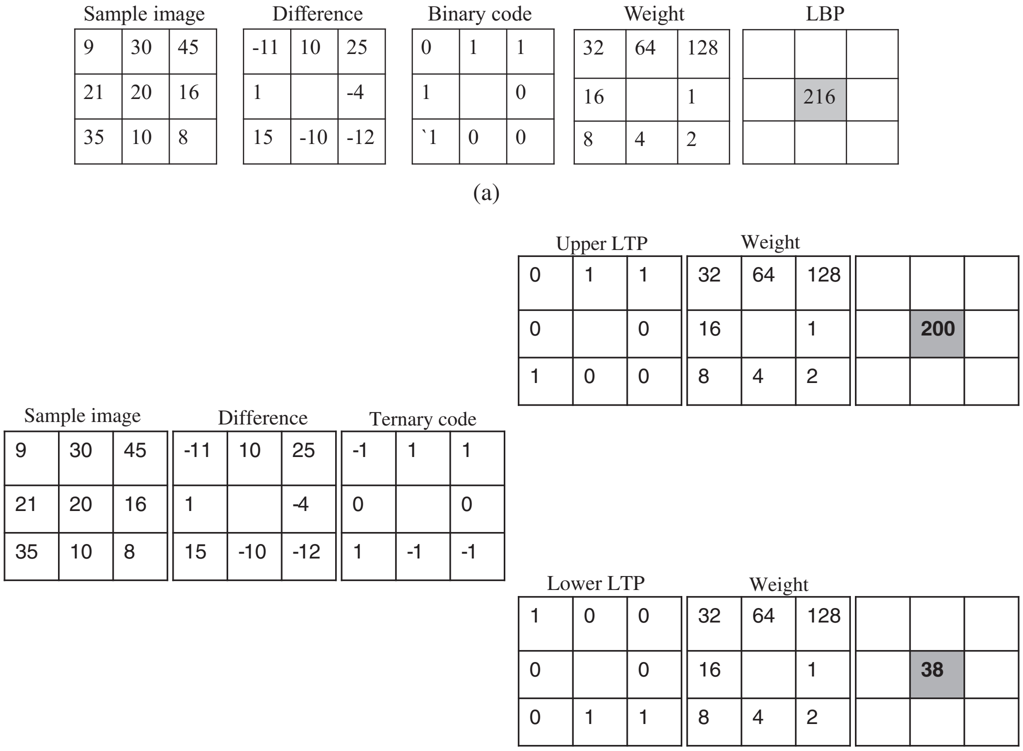
$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & else \end{cases} \tag{16}$$

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 a 3-valued code called local ternary pattern (LTP) [17] which is an extension to LBP where the pixel values in the range of ± threshold (th) around $I(p_c)$ are quantized to 0, for the values above $(I(p_c) + th)$ are quantized to $+1$ and the values less than $(I(p_c) - th)$ are quantized to -1. The function $f(x)$ is thus modified as follows:

$$f(x, I(p_c), th) = \begin{cases} +1 & x \geq I(p_c) + th \\ 0 & |x - I(p_c)| < th \\ -1 & x \leq I(p_c) - th \end{cases} \tag{17}$$

where $x = (I(p_i) - I(p_c))$. A sample example of LBP and LTP calculation is shown in Fig. 4.



Sample image

9	30	45
21	20	16
35	10	8

Difference

-11	10	25
1		-4
15	-10	-12

Binary code

0	1	1
1		0
`1	0	0

Weight

32	64	128
16		1
8	4	2

LBP

	216	

(a)

Sample image

9	30	45
21	20	16
35	10	8

Difference

-11	10	25
1		-4
15	-10	-12

Ternary code

-1	1	1
0		0
1	-1	-1

Upper LTP

0	1	1
0		0
1	0	0

Weight

32	64	128
16		1
8	4	2

	200	

Lower LTP

1	0	0
0		0
0	1	1

Weight

32	64	128
16		1
8	4	2

	38	

(b)

Figure 4: Example of LBP and LTP calculation for a sample image. (a) LBP; (b) LTP (for threshold 5)

LTP can capture image details better than LBP, as LTP provides 3 valued code to the difference between centre pixel and its neighbouring pixels.

The extension to 3D-LTP encodes not only the color cue information but also the local texture information in a color image. Given a RGB color image, the proposed inter-channel 3D-LTP produces six new images as shown in sample example computation in Fig. 5. The encoding of R channel considers the center/reference pixel in R and consider neighbors from G channel. Similarly, the encoding of G channel considers center/reference pixel in G channel and neighbors from B channel and the center/reference pixel in B channel consider neighbors from R channel for encoding B channel. The pattern images formed with 3D-LTP are presented in Fig. 6. Inter channel LTP when calculated for each of R-G, G-B and B-R combination provides one upper and one lower LTP. Therefore the R-G, G-B and B-R combinations produces a total of six pattern images, i.e., three upper LTPs and three lower LTPs.

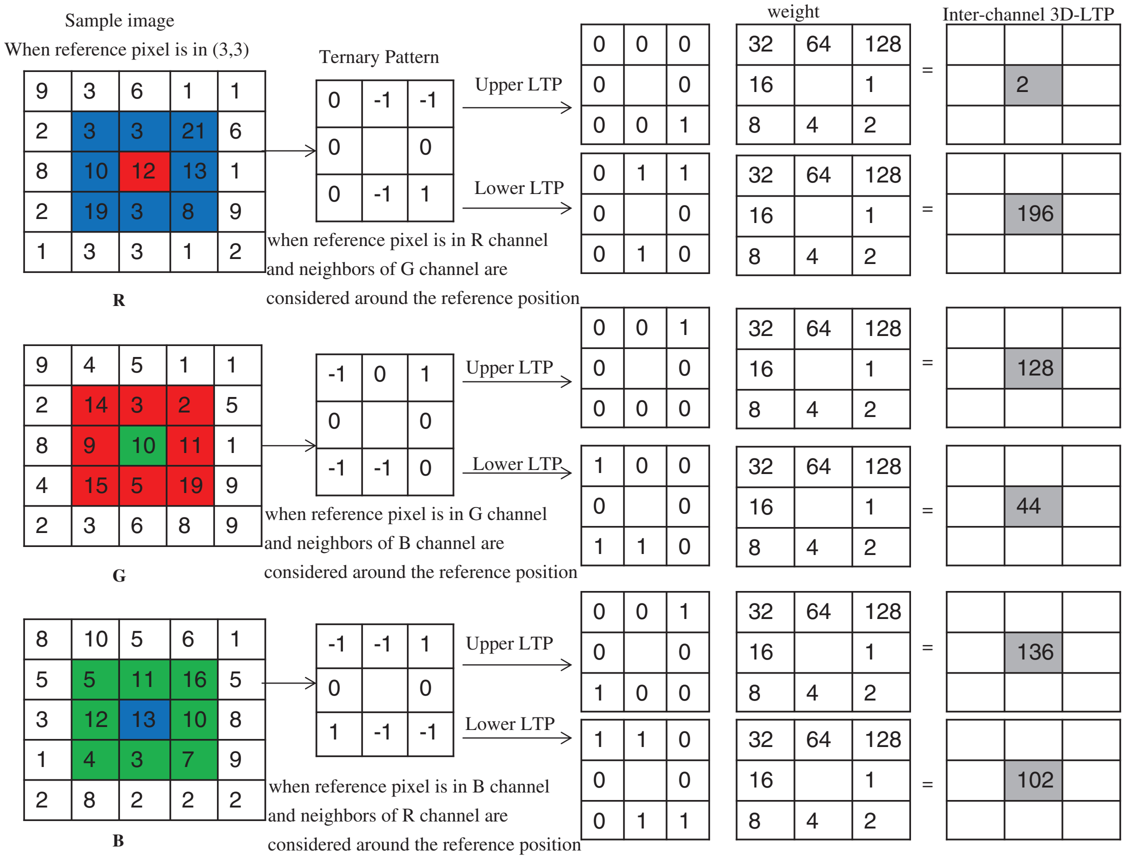


Figure 5: A sample example for 3D-LTP calculation

To reduce the feature dimensions, we identify and consider only the 'uniform' patterns in 3D-LTP. In this paper, the 'uniform' refers to uniform appearance of 3D local ternary patterns which means the patterns that have two or less number of discontinuities in circular binary representation and rest are referred to as non-'uniform' [16]. For example, 00010000_2 is an uniform pattern as it has only 2 bitwise 0/1 transitions and 00101001 is non-'uniform' pattern with more than 2 spatial transitions. It is observed that these ‘uniform’ patterns constitutes a majority of patterns that corresponds to important features like edges, textures, sharp corners, etc. For an image with R=1 and T=8, the unique ‘uniform’ patterns would be 58.

2.2.3 Fusion of NSST-Domain Statistical Features (NSSTds) and 3D-LTP Based Features

For describing a complex scene having complicated patterns and spatial structures, the blend of complementary features such as local and global features are usually preferred to achieve good results. In this subsection, we introduce a fused feature description using NSSTds and 3D-LTP based features for retrieval of remote sensing images.

If feature vector $F_1 = [(\sigma_{p_1(i)}, \sigma_{p_2(i)}, \omega_{(i)}, s_{(i)}, \kappa_{(i)}), \mu_A, \sigma_A]_{i \in (1,N_s)}$ denotes the NSSTds calculated from detail and approximation subbands and F_2 is the feature vector obtained from 3D-LTP ‘uniform’ histograms, then the final feature vector is described by $F = [F_1, F_2]$.

For example, if an input image is decomposed using 4-level NSST with 3,3,4,4 directions (coarser to finest scale), we obtain a total of 49 subbands. Among these 49 subbands, one is low-frequency approximation subband and rest 48 are high-frequency detail subbands, i.e., 8(2^3), 8(2^3), 16(2^4) and 16(2^4) exists in Scale 4 (most coarsest), Scale 3, Scale 2 and Scale 1 (finest) respectively. With NSSTds, each high-frequency detail subband is represented by a 5 dimensional feature vector, so a total of 48 detail subbands will provide a feature vector of dimension (48 × 5) = 240. Since only standard deviation and mean parameters are used to describe an approximation subband, a total of $(240 + 2) = 242$ dimension is required by NSSTds descriptor to describe a total of 49 subbands (both approx.+detail subbands). From Fig. 5 it can be clearly seen that with 3D-LTP technique (using uniform features), a total of 6, i.e., 3 upper LTP and 3 lower LTP feature maps are obtained. As 59 uniform features are obtained as a result of encoding a single feature map, a total of $(59 \times 6) = 354$ features are therefore required to Encode 6, i.e., 3 upper and 3 lower LTP feature maps. For the given NSST decomposition setting, finally with proposed NSSTds-3DLTP a total of $242 + 354 = 596$ features are obtained.

2.2.4 Steps of NSSTds-3DLTP Feature Extraction Methodology

The algorithm for the proposed feature extraction technique is as follows:

Input-Image; Output-Feature vector

1. Convert the color remote sensing image into gray scale image.
2. Apply the NSST on the gray scale image.
3. Extract the 2-state LMM parameters, kurtosis and skewness from each NSST detail subband and concatenate it with the mean (μ_A) and standard deviation (σ_A) calculated from the approximation subband to form the feature vector F_1.
4. Calculate the 3D-LTP based features from the R,G,B color channels of original remote sensing image to form the feature vector F_2.
5. The NSST-domain statistical features (NSSTds) and the 3D-LTP based features are finally concatenated to form the final feature vector set $F = [F_1, F_2]$ after normalization.

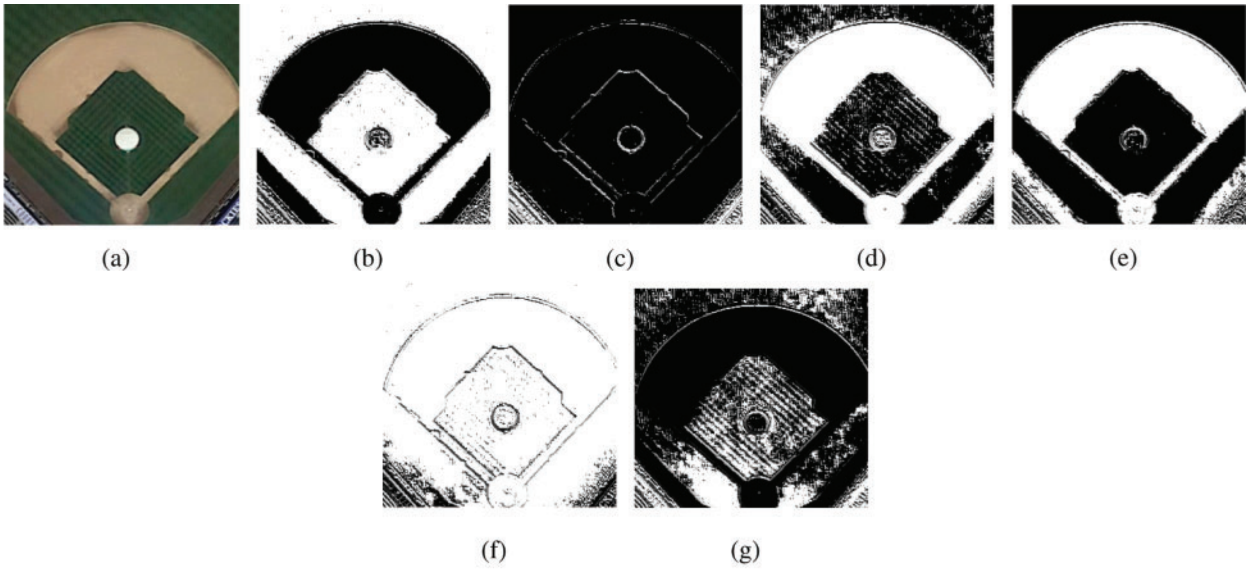


Figure 6: The pattern images obtained for an image from PatternNet dataset with inter channel 3D-LTP (a) Original image (b) Upper LTP pattern image after encoding R channel, (c) Lower LTP pattern image after encoding R channel, (d) Upper LTP pattern image after encoding G channel, (e) Lower LTP pattern image after encoding G channel, (f) Upper LTP pattern image after encoding B channel, (g) Lower LTP pattern image after encoding B channel

2.2.5 Similarity Measure

The extracted features of query image and database images are matched using a similarity metric. In the experiments, the NSSTds-3DLTP features has been evaluated using ‘d1’, Euclidean, Manhattan, Canberra and Chi-square similarity measures.

It has been observed that NSSTds-3DLTP exhibits best retrieval performance with ‘d1’ distance measure in comparison to Euclidean, Manhattan, Canberra and Chi-square distance measures. Therefore, the NSSTds-3DLTP descriptor employs ‘d1’ distance measure for feature matching in the image retrieval framework.

The $d1$ distance measure is given by:

$$D(d_k,q)=\sum_{j=1}^{L_f}\left|\frac{f_k(j)-f_q(j)}{1+f_k(j)+f_q(j)}\right| \tag{18}$$

where $D(d_k,q)$ denotes the distance between d_k and q where d_k is the k^{th} database image and q denotes the query image. The length of the feature vector is L_f. The parameter f_k denotes the k^{th} feature vector in the database of features and f_q denotes the query feature vector. The least distance value indicates the best match of the image in the database.

2.2.6 *Feature Vector Matching*

The algorithm of the feature vector matching is as follows:

Input: Query image; Output: Most similar retrieved images

1. Calculate the feature vector of each image using NSSTds-3DLTP in the database.
2. Calculate the feature vector of query image using NSSTds-3DLTP.
3. Calculate the similarity between each database image feature vector and the feature vector of query image using d_1 distance.
4. Sort the similarity values obtained from Step (3).
5. The final sorted result are the most similar retrieved images from the database.

3 Experimental Results and Discussion

This section presents experimental results to validate the performance of the proposed fused features. The experiments were carried out on a system with Intel core i5-7200U CPU, 2.50 GHz and 8GB RAM using MATLAB computing platform. First the experimental settings are described which includes the database details and performance evaluation criteria. Next, the experimental test results and analysis are presented where the proposed descriptor is compared with many well known existing descriptors.

3.1 Experimental Settings

3.1.1 *Description of Datasets*

Three publicly available popular remote sensing image databases namely WHU-RS19 [41], Aerial Image Dataset (AID) [42,43] and PatternNet [44,45] are utilized in the experiments, the details of which are given below:

1. WHU-RS19 Dataset
   WHU-RS19 dataset [41] consists of a total of 1005 images of 19 different classes with a high spatial resolution of 600 × 600 pixels (Table 2). All the images are taken from huge satellite images (from Google earth imagery) where the light, emergence of objects and their positions changes notably with repeated occlusions and is therefore known to be a challenging dataset. The different image classes in this dataset are-airport, beach, bridge, commercial, desert, farmland, football field, forest, industrial, meadow, mountain, park, parking, pond, port, railway station, residential, river and viaduct. Sample images from each image class are shown in Fig. 7.

Table 2: Databases used in the experiments

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

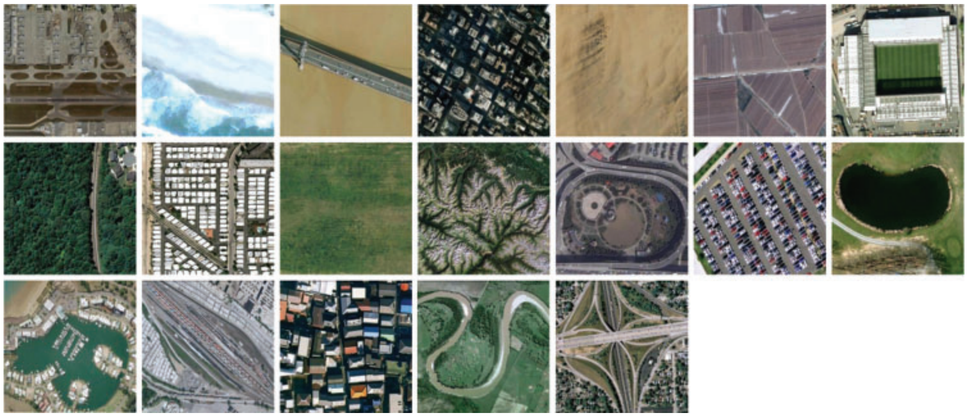


Figure 7: Example image from each class of WHU-RS19 dataset

2. Aerial Image Dataset (AID)

   AID dataset [42,43] is known to be one of the largest annotated aerial image dataset which is composed of a total of 30 classes with 10000 images (Table 2). The remote sensing images in this dataset are obtained using dissimilar imaging sensors that are used at separate time periods under diverse imaging situations which reduces the inter-class variations and escalates the intra-scale variations, therefore bringing in more difficulties in correct retrieval of similar images. Each class has around 220–420 images of size 600 × 600 pixels. The images of this dataset are classified into the following classes-airport, bare land, baseball field, beach, bridge, center, church, commercial, dense residential, desert, farmland, forest, industrial, meadow, medium residential, mountain, park, parking, playground, pond, port, railway station, resort, river, school, sparse residential, square, stadium, storage tanks and viaduct. The images of this large scale aerial dataset are collected selectively from Google Earth imagery. Example image of each class of AID is presented in Fig. 8.

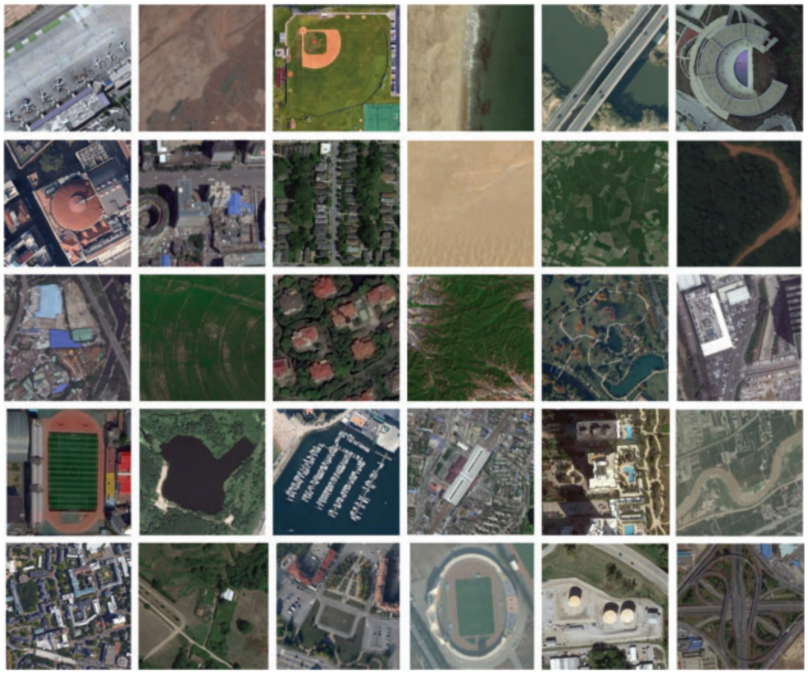


Figure 8: Example image from each class of AID dataset

3. PatternNet

   PatternNet [44,45] is the largest high resolution remote sensing image dataset (Table 2). The images in this dataset are collected from Google Earth imagery or via Google MAP API of US cities. This dataset consists of 38 image classes, each with 800 images of dimension 256 × 256. The main advantage of this dataset is the presence of very less background compared to other remote sensing image datasets. PatternNet is composed of the following classes- airport, baseball field, basketball court, beach, bridge, cemetery, chaparral, christmas tree farm, closed road, coastal mansion, cross walk, dense residential, ferry terminal, football field, forest, freeway, golf course, harbor, intersection, mobile home park, nursing home, oil gas field, oil well, overpass, parking lot, parking space, railway, river, runway, runway marking, shipping yard, solar panel, sparse residential, storage tank, swimming pool, tennis court, transformer station and waste water treatment plant. The sample image from each class are presented in Fig. 9.

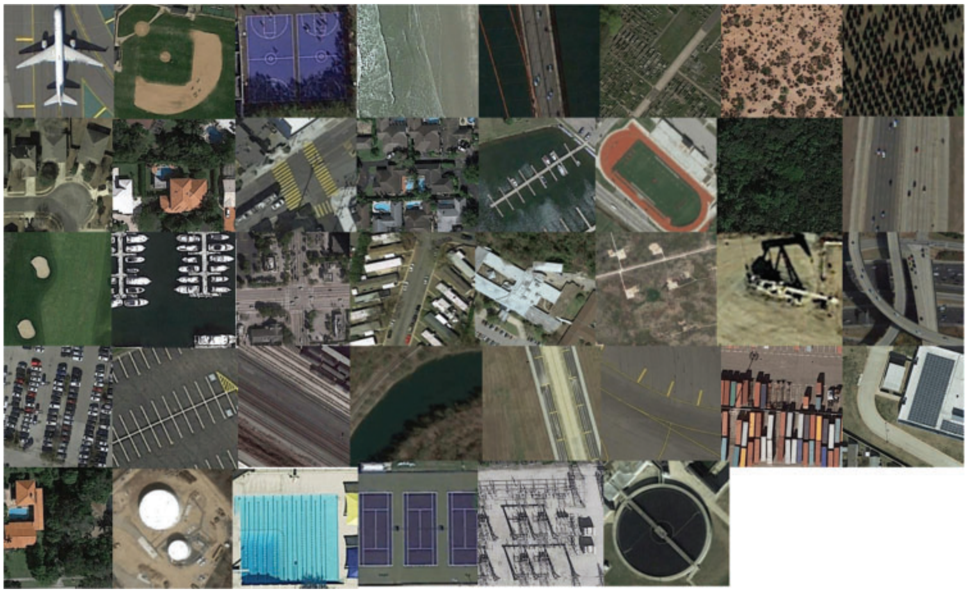


Figure 9: Example image from each class of PatternNet dataset

3.1.2 Performance Evaluation Criteria

The retrieval performance of the proposed descriptor is evaluated using average normalized modified retrieval rank (ANMRR), mean average precision (MAP) and precision-recall (P-R) graph, the details of which are given below:

1. Average normalized modified retrieval rank (ANMRR)
   It is often employed to assess the retrieval performance of MPEG-7 standard and is quite popular in the field of remote sensing image retrieval. The value of ANMRR ranges between 0 and 1. Smaller the value, higher the retrieval efficacy [3,46]. For any query image (q), $Gr(q)$ denotes the size of ground truth images and let the ground truth image at k^{th} position is retrieved at the location $Rank(k)$. Subsequently, the image ranks that are treated as acceptable from retrieval point of view are expressed as $K(q)$ which is two times that

of $Gr(q)$ and the images belonging to higher ranks are given a penalty as:

$$Rank(m) = \begin{cases} Rank(m) & \text{if } Rank(m) < K(q) \\ 1.25K(q) & \text{if } Rank(m) > K(q) \end{cases} \tag{19}$$

Therefore the average rank (A_r) for q is given as:

$$A_r(q) = \frac{1}{Gr(q)} \sum_{m=1}^{Gr(q)} Rank(m) \tag{20}$$

To control the effect of variable no. of ground truths of query image, the normalization is done and averaged for all query images N_Q to compute ANMRR:

$$ANMRR = \frac{1}{N_Q} \sum_{q=1}^{N_Q} \frac{A_r(q) - 0.5[1 + Gr(q)]}{1.25K(q) - 0.5[1 + Gr(q)]} \tag{21}$$

2. Mean Average Precision (MAP)
The MAP is one way to congregate Precision-Recall curve into a single value that assess the rank place of all ground truth. Let $Pr_{ave}(q)$ is the average precision for each query image q which is simply the average of precision values of each relevant item:

$$Pr_{ave}(q) = \frac{\sum_{k=1}^{n} (Pr(k) * rel(k))}{No.\ of\ relevant\ items} \tag{22}$$

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:

$$MAP = \frac{\sum_{q=1}^{N_Q} Pr_{ave}(q)}{N_Q} \tag{23}$$

The range of MAP value lies between 0 and 100. Higher value of MAP signifies better retrieval performance of the descriptor [3,46].

3. Precision-Recall (P-R) curve
Precision and recall, both are popularly used for image retrieval performance assessment. The ratio of number of relevant images retrieved to the number of images retrieved gives the precision value whereas recall is the ratio of number of relevant images retrieved to the number of relevant images in a database. The descriptor that shows largest area under the curve indicates high precision and high recall which exhibits better results relevancy and improved correct relevant image retrieval [46].

3.2 Results and Analysis

Here for the experiments, the remote sensing images are decomposed using NSST with 3,3,4,4 directions to extract the NSST-domain statistical features. It yields one approximation subband along with 48 detail subbands. For each database, the performance of proposed descriptor is compared with Gabor RGB [47], Granulometry [48], LBP [16], FV [49], VLAD [50] and MRELBP [51] in terms of MAP and ANMRR (Tables 3 and 4).

Table 3: Performance comparison of the proposed descriptor for WHU-RS19 dataset in terms of ANMRR and MAP

Dataset	Methods	MAP	ANMRR
WHU-RS19	Gabor RGB [47]	31.69	0.570
	Granulometry [48]	23.39	0.670
	LBP [16]	24.06	0.663
	FV [49]	38.06	0.532
	VLAD [50]	41.28	0.561
	MRELBP [51]	38.91	0.520
	NSSTds	38.71	0.499
	3D-LTP	32.85	0.576
	NSSTds-3DLTP	**44.98**	**0.451**

Table 4: Performance comparison of the proposed descriptor for AID and PatternNet datasets in terms of ANMRR and MAP

Dataset	Methods	MAP	ANMRR	Dataset	Methods	MAP	ANMRR
AID	Gabor RGB [47]	11.69	0.843	PatternNet	Gabor RGB [47]	26.53	0.686
	Granulometry [48]	9.04	0.877		Granulometry [48]	14.6	0.817
	LBP [16]	10.36	0.857		LBP[16]	29.99	0.686
	FV [49]	16.40	0.748		FV [49]	19.23	0.760
	VLAD [50]	19.61	0.734		VLAD [50]	28.98	0.638
	MRELBP [51]	18.19	0.679		MRELBP [51]	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**

From Tables 3 and 4, it is observed that the global NSSTds features performs better than the local 3D-LTP features both in terms of ANMRR and MAP for WHU-RS19 and AID databases, and both NSSTds and 3D-LTP performs quite close for Patternnet. The good performance of proposed NSSTds features is due to its suitability of capturing texture features for retrieval of remote sensing images and can effectively describe features mainly over multiple scales and multiple orientations. For each database, the proposed fusion NSSTds-3DLTP outperforms all the existing methods including MRELBP which is also known for its ability to capture both global and local features (Tables 3 and 4). The proposed method shows good improvement over other methods both in terms of MAP and ANMRR which verifies the efficacy of combining proposed NSST domain statistical feature and 3D-LTP. In terms of MAP,ANMRR, the NSSTds-3DLTP improves upon 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%,19.60% and 15.60%,13.26% respectively for WHU-RS19 dataset. For AID, in terms of {MAP,ANMRR}, the NSSTds-3DLTP improves upon Gabor RGB, Granulometry, LBP, FV, VLAD and MRELBP descriptors by {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 and for PatternNet dataset the NSSTds-3DLTP respectively improves upon Gabor RGB, Granulometry, LBP, FV, VLAD and MRELBP descriptors by {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%} in terms of {MAP,ANMRR}. Unlike most of the other methods, which are either local or global, the subfeature 3D-LTP not only encodes the color cue but the local texture information is extracted too. Further, the subfeature NSSTds is capable of capturing the image information at multiple scales and orientations. The complementary characteristics of NSSTds and 3D-LTP are combined in NSSTds-3DLTP to produce a highly discriminative representation.

In Tables 5–7, the average precision of all descriptors including proposed descriptors (NSSTds, 3D-LTP, NSSTds-3DLTP) for individual classes and for each database is shown. The average precision value is calculated for 20 top retrieved images found in 20 trials for 20 images selected randomly from each image class. Tables 5–7 show that the NSSTds-3DLTP features performs best in most of the individual classes compared to other techniques including global NSSTds and local 3D-LTP. From Tables 5–7, it can be clearly seen that the global NSSTds features alone when compared to local 3D-LTP show good performance on the specific classes like Forest, river, residential, bareland, school, mountain, parking etc. which are more texture based and have image-scale attributes (Fig. 10). However, the local 3D-LTP alone when compared to global NSSTds show good performance on the classes like intersection, railway, baseball field, freeway, storage tank, golf course,church, commercial, pond, medium residential, bridge etc. that contains unique arrangement of structures in the absence of which the images cannot be retrieved correctly (Fig. 10). These results confirms that the local and global features contain mutually supportive details and their combination is expected to improve the discriminativeness of features. The retrieval of challenging images such as tennis court, dense residential, sparse residential, stadium, playground etc. are significantly improved using fused NSSTds-3DLTP descriptor. In Fig. 12, a few image query examples from different classes and its corresponding retrieved results using proposed NSSTds-3DLTP descriptor (for all the three databases) are shown. From Fig. 12a, it is observed that the images from the classes beach, bridge, desert, farmland, industrial and river from WHU-RS19 dataset when given as input query images exhibit correct retrieval results except for viaduct class where it provides one incorrect retrieved result, i.e., an image from Railway class is wrongly retrieved here. Likewise, from Fig. 12b, it is observed that the images from the classes airport, bareland, church, dense residential, desert, medium residential and school from AID dataset when given as input query images exhibit correct retrieval results except for Beach class where it provides one incorrect retrieved result i.e., an image from bridge class is wrongly retrieved here. And, from Fig. 12c it is observed that the images from the classes baseball field, beach, cemetry, chaparral, closed road, coastal mansion, waste water treatment plant from PatternNet dataset when given as input query images exhibit correct retrieval results except for Airplane class where it provides one incorrect retrieved result i.e., an image from waste water treatment plant class is wrongly retrieved here. From Tables 3 and 4, Figs. 11 and 12, we can conclude that the NSSTds-3DLTP is able to achieve encouraging results for most of the images over other techniques, however it shows a few cases of incorrect retrieval too in relatively simple class such as ‘beach’.

Table 5: Average precision per class for WHU-RS19 dataset

Sl. No.	Class	Average precision								
		Gabor RGB	Granulo.	LBP	FV	VLAD	MRELBP	NSSTds	3D-LTP	NSSTds-3DLTP
1	Airport	65.62	69.31	69.37	77.05	73.71	70.24	68.01	67.34	68.89
2	Beach	89.24	87.14	92.49	87.40	93.87	92.88	93.24	91.19	90.89
3	Bridge	75.18	68.75	65.31	69.12	71.39	74.58	75.42	75.48	70.78
4	Commercial	67.75	65.77	65.96	78.94	79.19	85.11	67.31	72.55	72.10
5	Desert	89.92	89.93	76.22	76.26	85.38	84.57	90.22	98.03	98.35
6	Farmland	74.02	62.14	79.43	76.96	79.17	78.85	80.02	87.54	80.89
7	Football field	61.59	64.15	65.98	70.04	74.66	75.34	76.25	73.73	79.06
8	Forest	83.27	81.21	77.29	89.84	97.09	86.56	87.34	85.02	95.48
9	Industrial	68.75	68.32	68.51	78.07	74.23	66.52	66.92	66.57	71.27
10	Meadow	80.46	75.36	68.10	76.96	83.37	88.93	82.68	80.80	88.06
11	Mountain	80.74	76.04	84.02	95.91	96.46	90.68	88.44	67.60	90.26
12	Park	72.01	66.04	68.53	81.94	76.78	76.31	66.78	80.69	79.28
13	Parking	69.29	68.12	70.03	84.13	88.68	80.28	74.63	72.19	78.78
14	Pond	71.48	64.78	69.36	68.00	67.83	79.41	70.23	71.35	84.41
15	Port	62.79	75.69	69.00	71.29	72.30	77.42	74.03	73.67	76.49
16	Railway station	77.09	66.06	86.52	85.08	86.92	84.67	77.47	83.06	82.93
17	Residential	68.29	68.39	67.22	71.46	75.99	80.30	76.45	67.94	78.35
18	River	73.43	65.36	71.64	72.46	67.44	77.01	80.77	73.89	82.30
19	Viaduct	62.68	63.61	66.87	78.12	76.31	76.64	77.63	69.72	77.65

Table 6: Average precision per class for AID dataset

Sl. No.	Class	Average precision								
		Gabor RGB	Granulo.	LBP	FV	VLAD	MRELBP	NSSTds	3D-LTP	NSSTds-3DLTP
1	Airport	64.81	61.08	71.28	68.50	67.70	71.79	98.79	69.89	98.69
2	Bareland	76.52	74.71	76.08	77.51	76.40	69.72	86.15	82.14	91.61
3	Baseball field	73.88	68.78	77.60	63.66	71.03	69.88	72.49	78.84	82.92
4	Beach	74.82	67.32	70.07	77.92	87.27	74.25	73.81	82.40	81.37
5	Bridge	70.44	73.82	69.01	68.38	72.90	77.82	66.44	68.91	67.08
6	Center	74.56	73.60	69.13	70.42	71.30	75.47	70.93	64.80	68.68
7	Church	65.56	70.47	69.63	70.75	72.04	73.60	67.99	72.87	75.75
8	Commercial	68.88	67.33	71.87	74.62	75.36	73.21	69.37	79.04	79.65
9	Dense residential	76.09	76.69	70.33	80.16	84.05	71.16	83.28	78.39	87.76
10	Desert	88.64	80.20	73.77	66.77	70.81	85.79	85.75	89.71	96.50
11	Farmland	73.53	70.61	75.78	71.94	74.92	67.89	72.75	73.18	75.11
12	Forest	77.13	74.60	83.66	91.01	93.74	77.61	92.95	86.50	94.94
13	Industrial	63.32	66.77	70.21	71.67	74.12	81.46	67.69	66.33	69.64
14	Meadow	71.86	71.00	69.52	70.58	74.52	75.40	78.66	89.23	93.46
15	Medium residential	70.93	72.39	73.53	72.03	73.34	67.15	74.74	84.29	86.00
16	Mountain	79.25	65.82	86.53	89.94	90.39	69.09	88.43	71.19	93.25
17	Park	63.74	64.73	62.00	68.53	70.87	74.04	66.39	68.77	71.88
18	Parking	73.50	70.63	75.21	90.05	92.07	67.58	85.51	77.12	85.03
19	Playground	68.77	66.08	67.56	63.88	70.82	70.60	69.25	69.40	76.60
20	Pond	65.43	65.22	73.57	66.09	67.27	75.55	73.19	72.69	80.40

(Continued)

Table 6 (continued)

Sl. No.	Class	Average precision								
		Gabor RGB	Granulo.	LBP	FV	VLAD	MRELBP	NSSTds	3D-LTP	NSSTds-3DLTP
21	Port	69.51	70.45	70.35	78.97	81.71	64.25	77.86	77.52	84.30
22	Railway station	67.26	73.54	72.46	69.49	71.61	67.56	70.76	68.51	69.14
23	Resort	68.73	68.02	71.05	64.85	72.98	74.20	67.58	72.53	73.94
24	River	66.75	66.69	72.57	67.26	73.46	68.20	72.88	68.15	80.47
25	School	73.96	66.36	70.14	68.26	67.31	69.08	77.04	65.11	76.41
26	Sparse residential	78.90	75.02	80.20	81.37	84.51	71.27	90.04	88.51	96.25
27	Square	70.64	71.43	71.79	68.18	70.56	69.39	65.79	65.72	65.36
28	Stadium	70.22	73.47	71.32	67.09	71.82	68.47	71.42	69.51	75.61
29	Storage tank	66.00	64.93	66.96	68.29	74.17	67.49	75.59	69.95	80.54
30	Viaduct	62.71	65.36	63.21	80.54	80.85	71.21	73.43	64.36	77.61

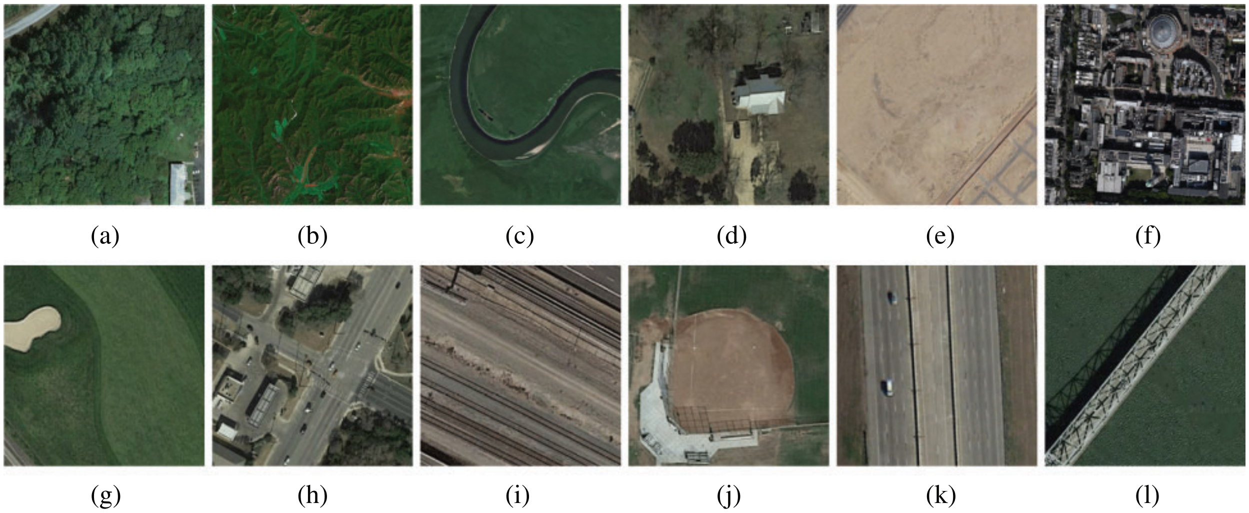
Table 7: Average precision per class for PatternNet dataset

Sl. No.	Class	Average precision								
		Gabor RGB	Granulo.	LBP	FV	VLAD	MRELBP	NSSTds	3D-LTP	NSSTds-3DLTP
1	Airplane	85.10	66.64	75.70	70.62	73.13	71.23	79.34	94.08	91.20
2	Baseball field	83.28	72.95	80.03	76.89	76.26	76.48	68.64	89.44	84.91
3	Baseball court	69.96	71.10	69.01	75.98	76.91	89.48	78.81	75.97	75.20
4	Beach	98.03	89.60	97.43	83.84	92.65	91.56	99.57	99.44	99.90
5	Bridge	88.06	79.84	88.45	83.16	82.47	85.67	90.69	91.81	95.00
6	Cemetery	84.53	85.21	86.84	84.86	81.48	88.98	93.99	93.43	97.15
7	Chaparral	99.31	98.19	99.98	99.76	99.96	75.86	99.86	100.00	100.00
8	Christmas tree farm	97.97	94.00	99.27	98.60	99.66	85.24	99.93	99.06	99.81
9	Closed road	84.74	66.31	80.94	72.66	76.78	71.18	86.38	90.97	93.27
10	Coastal mansion	85.34	82.88	87.25	62.85	65.16	74.93	85.70	92.23	94.20
11	Crosswalk	94.19	80.08	94.11	76.30	77.43	82.95	97.25	98.42	99.15
12	Dense residential	87.22	77.17	86.07	69.57	73.57	80.03	90.42	91.15	92.80
13	Ferry terminal	72.49	71.25	79.66	75.64	73.39	79.32	75.86	84.87	84.40
14	Football field	77.22	65.02	71.01	69.04	79.04	82.29	82.22	82.90	87.65
15	Forest	98.02	95.26	97.31	97.27	98.37	89.22	99.87	99.13	99.94
16	Freeway	99.20	81.83	99.51	81.61	97.88	82.39	99.19	99.96	99.97
17	Golf course	87.38	84.70	86.77	74.69	66.83	85.30	87.39	94.81	94.00
18	Harbor	81.79	87.37	83.96	82.70	93.18	81.46	94.98	90.76	95.22
19	Intersection	79.82	70.28	82.04	66.67	69.19	81.29	82.50	86.98	89.11
20	Mobile home park	93.97	92.05	96.19	92.99	97.10	74.61	98.93	98.48	99.22
21	Nursing home	67.54	68.71	76.34	63.68	69.37	83.53	73.50	77.06	81.65
22	Oil gas field	98.63	90.27	99.89	89.98	95.42	86.47	99.47	99.58	99.96
23	Oil well	99.12	97.21	99.82	99.98	99.84	79.27	98.20	99.88	99.76
24	Overpass	80.45	68.03	88.83	71.97	78.92	83.22	76.40	86.41	82.65
25	Parking lot	82.80	88.57	78.04	87.62	97.58	86.73	97.36	93.62	97.25
26	Parking space	86.10	89.39	85.34	89.26	93.12	81.08	88.19	89.84	90.06
27	Railway	83.67	72.42	84.59	79.89	78.84	78.07	78.90	89.73	89.49
28	River	97.16	94.01	98.75	95.25	94.46	76.42	99.60	99.87	99.97
29	Runway	95.69	74.88	94.43	79.71	83.88	77.93	96.86	96.79	98.57

(Continued)

Table 7 (continued)

Sl. No.	Class	Average precision								
		Gabor RGB	Granulo.	LBP	FV	VLAD	MRELBP	NSSTds	3D-LTP	NSSTds-3DLTP
30	Runway marking	90.26	86.16	86.73	74.68	75.22	80.58	89.45	95.00	94.71
31	Shipping yard	89.07	83.34	93.13	78.32	87.91	78.27	93.92	96.70	97.29
32	Solar panel	88.39	81.81	91.20	87.72	91.77	84.75	90.74	92.54	93.16
33	Sparse residential	76.29	70.01	75.42	61.45	66.63	95.14	80.03	79.99	84.51
34	Storage tank	84.97	77.88	78.02	76.60	84.46	80.78	80.94	86.90	88.49
35	Swimming pool	82.82	71.32	85.64	63.77	69.71	76.75	86.96	92.51	91.63
36	Tennis court	81.99	78.92	83.79	80.75	89.17	77.11	87.04	84.88	87.94
37	Transformer station	80.50	69.75	74.16	75.94	77.22	84.59	77.98	78.34	80.55
38	Wastewater treat. plant	79.22	72.15	78.45	75.96	78.31	77.44	88.46	85.06	91.64

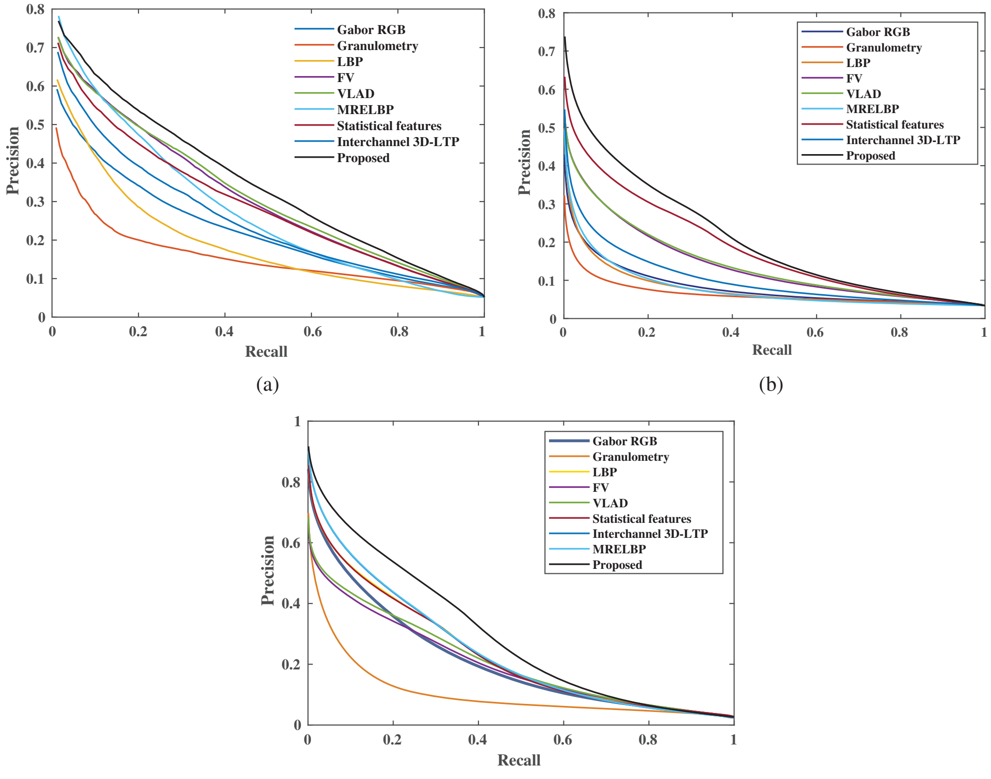


(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

Figure 10: Image classes on which improved results are achieved using global NSSTds alone (1st row) and local 3D-LTP alone (2nd Row) when compared to each other. (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’

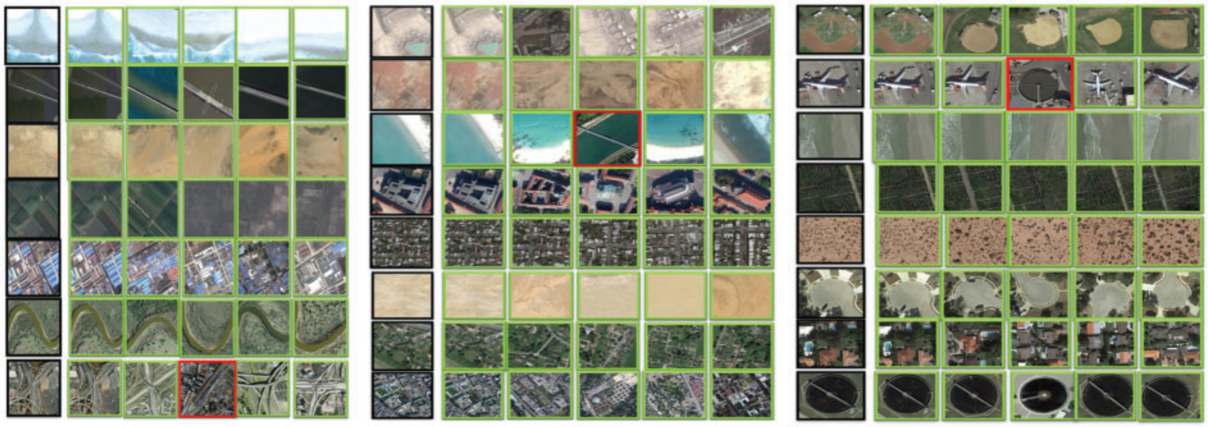
In order to further show the superiority of proposed fused features over other techniques including NSSTds and 3D-LTP, the P-R curves for all the techniques are shown in Figs. 11a–11c for WHU-RS19, AID and PatternNet databases respectively. Precision is defined as the ratio of number of relevant images retrieved to the total number of retrieved images, however recall is defined as the ratio of number of relevant images retrieved to the total number of relevant images present in the database. Precision indicates the accuracy of retrieval and recall indicates about the efficacy of the retrieval performance. The precision-recall graph describes about the inherent trade-off between these two parameters and is an important performance indicator in retrieval systems. The descriptor who has the largest area enclosed by its precision-recall curve (high precision and high recall) is considered as the the superior one.



(a) (b)

(c)

Figure 11: The precision-recall curve for (a) WHU-RS19, (b) AID and (c) PatternNet dataset



(a) (b) (c)

Figure 12: Few cases of input query examples taken from different classes and the corresponding retrieval results 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). (a) WHU-RS19, (b) AID, (c) PatternNet

In Fig. 11a, i.e., for WHU-RS19 dataset, the P-R curve obtained using NSSTds-3DLTP encloses the largest area and exhibits the best results followed by VLAD, FV, NSSTds, MRELBP, 3D-LTP, Gabor RGB, LBP and Granulometery descriptors. For AID dataset, the NSSTds-3DLTP exhibits the superior results followed by NSSTds, VLAD, FV, 3D-LTP, MRELBP, Gabor RGB, LBP and Granulometry (Fig. 11b). Similarly, for PatternNet too, the NSSTds-3DLTP shows the best results followed by MRELBP, 3D-LTP, LBP, NSSTds, VLAD, FV, Gabor RGB and Granulometry (Fig. 11c). The P-R curve results are observed to be consistent with Tables 2 and 3 results.

From Table 8, it can be seen that the feature dimensions of NSSTds-3DLTP is less than MRELBP and higher than other techniques. The Gabor RGB, granulometery, LBP, FV and VLAD techniques have comparatively less feature dimensions than NSSTds-3DLTP, but their performance is also well less than NSSTds-3DLTP. The NSSTds-3DLTP outperforms state of the art MRELBP with relatively less feature dimensions.

Table 8: Comparison of feature dimensions of various techniques

Methods	Gabor RGB	Granulometry	LBP	FV	VLAD	MRELBP	Proposed
Dimension	96	78	256	512	512	800	596

4 Conclusions

This paper combines global feature based on NSSTds and local feature based on 3D-LTP to generate a combined representation, i.e., NSSTds-3DLTP for retrieval of high-resolution remote sensing image. The complementary characteristics of local and global texture features along with colour information are utilized in NSSTds-3DLTP to produce a highly discriminative representation. Through KS test and log histogram plots we have shown that the 2-state LM distribution is the most appropriate distribution that approximates the statistics of high-frequency detail subband coefficients. Five statistical parameters are extracted from each NSST subband to form the feature vector of NSSTds. The 3D-LTP exploits both local texture details and colour information whereas the NSSTds exploits only global texture information. The image retrieval experiments using WHU-RS19, AID and PatternNet datasets validate the superior performance of NSSTds-3DLTP over many existing techniques with an encouraging margin. The NSSTds-3DLTP achieves the performance without any requirement for a pre-training and without parameter tuning and with less dimensions which is important from real-time implementation point of view.

In the future work, more effective local/global combination will be investigated and the shape based features will be exploited too.

Acknowledgement: The research fellowship support through Visvesvaraya PhD Scheme for Electronics and IT by Digital India Corporation, Ministry of Electronics and Information Technology (MeitY), Government of India is highly acknowledged by the authors.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Ferecatu, M., Boujemaa, N. (2007). Interactive remote-sensing image retrieval using active relevance feedback. *IEEE Transactions on Geoscience and Remote Sensing, 45(4),* 818–826. DOI 10.1109/TGRS.2007.892007.
2. Imbriaco, R., Sebastian, C., Bondarev, E., de With, P. H. N. (2019). Aggregated deep local features for remote sensing image retrieval. *Remote Sensing, 11(5),* 1–23. DOI 10.3390/rs11050493.
3. Baruah, H. G., Nath, V. K., Hazarika, D. (2019). Remote sensing image retrieval via symmetric normal inverse Gaussian modeling of nonsubsampled shearlet transform coefficients. *International Conference on Pattern Recognition and Machine Intelligence*, pp. 359–368. Tezpur, India.
4. Xiong, W., Lv, Y., Cui, Y., Zhang, X., Gu, X. (2019). A discriminative feature learning approach for remote sensing image retrieval. *MDPI Remote Sensing, 11(3),* 1–19. DOI 10.3390/rs11030281.
5. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60,* 91–110. DOI 10.1023/B:VISI.0000029664.99615.94.
6. Swain, M. J., Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision, 7,* 11–32. DOI 10.1007/BF00130487.
7. Oliva, A., Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36. DOI 10.1016/S0079-6123(06)55002-2.
8. Navneet, D., Bill, T. (2005). *Histograms of oriented gradients for human detection. IEEE Computer Vision and Pattern Recognition,* pp. 886–893. San Diego, CA, USA.
9. Ojala, T., Matti, P., Topi, M. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 24(7),* 971–987. DOI 10.1109/TPAMI.2002.1017623.
10. Ma, A., Sethi, I. K. (2005). Local shape association based retrieval of infrared satellite images. *7th IEEE International Symposium on Multimedia*, pp. 551–557. Irvine, CA, USA.
11. Yang, F. P., Hao, M. L. (2017). Effective image retrieval using texture elements and color fuzzy correlogram. *Information, 8(1),* 1–11. DOI 10.3390/info8010027.
12. Choy, S. K., Tong, C. S. (2010). Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval. *IEEE Transactions on Image Processing, 19(2),* 281–289. DOI 10.1109/TIP.2009.2033400.
13. Allili, M. S. (2012). Wavelet modeling using finite mixtures of generalized Gaussian distributions: Application to texture discrimination and retrieval. *IEEE Transactions on Image Processing, 21(4),* 1452–1464. DOI 10.1109/TIP.2011.2170701.
14. Liu, Z., Zhu, L. (2018). A novel retrieval method for remote sensing image based on statistical model. *Multimedia Tools and Applications, 77,* 24643–24662. DOI 10.1007/s11042-018-5649-6.
15. Haindl, M., Vacha, P. (2006). Illumination invariant texture retrieval. *18th International Conference on Pattern Recognition*, pp. 276–279. Hong Kong, China.
16. Ojala, T., Pietikainen, M., Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7),* 971–987. DOI 10.1109/TPAMI.2002.1017623.
17. Tan, X., Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing, 19(6),* 1635–1650. DOI 10.1109/TIP.2010.2042645.
18. Chen, C., Zhang, B., Su, H., Li, W., Wang, L. (2016). Land-use scene classification using multiscale completed local binary patterns. *Signal, Image and Video Processing, 10,* 745–752. DOI 10.1007/s11760-015-0804-2.
19. Huang, L., Chen, C., Li, W., Du, Q. (2016). Remote sensing image scene classification using multi-scale completed local binary patterns and fisher vectors. *MDPI Remote Sensing, 8(6),* 1–17. DOI 10.3390/rs8060483.
20. Yang, J., Liu, J., Dai, Q. (2015). An improved bag-of-words framework for remote sensing image retrieval in large-scale image databases. *International Journal of Digital Earth, 8(4),* 273–292. DOI 10.1080/17538947.2014.882420.

21. Sukhia, K. N., Riaz, M. M., Ghafoor, A., Ali, S. S. (2020). Content-based remote sensing image retrieval using multi-scale local ternary pattern. *Digital Signal Processing, 104,* 102765. DOI 10.1016/j.dsp.2020.102765.
22. Wang, Y., Zhang, L., Tong, X., Zhang, L., Zhang, Z. et al. (2016). A three-layered graph-based learning approach for remote sensing image retrieval. *IEEE Transactions on Geoscience and Remote Sensing, 54(10),* 6020–6034. DOI 10.1109/TGRS.2016.2579648.
23. Bosilj, P., Aptoula, E., Lefèvre, S., Kijak, E. (2016). Retrieval of remote sensing images with pattern spectra descriptors. *ISPRS International Journal of Geo-Information, 5(12),* 1–16. DOI 10.3390/ijgi5120228.
24. Bian, X., Chen, C., Tian, L., Du, Q. (2017). Fusing local and global features forhigh-resolution scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(6),* 2889–2901. DOI 10.1109/JSTARS.4609443.
25. Risojevic, V., Babic, Z. (2013). Fusion of global and local descriptors for remote sensing image classification. *IEEE Geoscience and Remote Sensing Letters, 10(4),* 836–840. DOI 10.1109/LGRS.2012.2225596.
26. Liu, L., Lao, S., Fieguth, P. W., Guo, Y., Wang, X. et al. (2016). Median robust extended local binary patternfor texture classification. *IEEE Transactions on Image Processing, 25(3),* 1368–1381. DOI 10.1109/TIP.2016.2522378.
27. Yang, P., Yang, G. (2018). Statistical model and local binary pattern based texture feature extraction in dual-tree complex wavelet transform domain. *Multidimensional Systems Signal Processing, 29,* 851–865. DOI 10.1007/s11045-017-0474-z.
28. Kabbai, L., Abdellaoui, M., Douik, A. (2019). Image classification by combining local and global features. *The Visual Computer, 35,* 679–693. DOI 10.1007/s00371-018-1503-0.
29. Yan, X., Jiangtao, P., Qian, D. (2020). Multiple feature regularized kernel for hyperspectral imagery classification. *APSIPA Transactions on Signal and Information Processing, 9,* 1–8. DOI 10.1017/ATSIP.2020.8.
30. Zeng, D., Chen, S., Chen, B., Li, S. (2018). Improving remote sensing scene classification by integrating global-context and local-object features. *MDPI Remote Sensing, 10(5),* 1–19. DOI 10.3390/rs10050734.
31. Lv, Y., Zhang, X., Xiong, W., Cui, Y., Cai, M. (2019). An end-to-end local-global-fusion feature extraction network for remote sensing imagescene classification. *MDPI Remote Sensing, 11, 3006(24),* 1–20. DOI 10.3390/rs11243006.
32. Wang, X., Tao, J., Shen, Y., Bai, S., Song, C. (2019). A nsst pansharpening method based on directional neighborhood correlation and tree structure matching. *Multimedia Tools and Applications, 78(18),* 26787–26806. DOI 10.1007/s11042-019-07841-5.
33. Candes, E., Demanet, L., Donoho, D., Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling & Simulation, 5(3),* 861–899. DOI 10.1137/05064182X.
34. Do, M. N., Vetterli, M. (2005). The contourlet transform: An efficient directional multiresolution image representation. *IEEE Transactions on Image Processing, 14(12),* 2091–2106. DOI 10.1109/TIP.2005.859376.
35. Easley, G., Labate, D., Lim, W. Q. (2008). Sparse directional image representations using the discrete shearlet transform. *Applied and Computational Harmonic Analysis, 25(1),* 25–46. DOI 10.1016/j.acha.2007.09.003.
36. Hou, B., Zhang, X., Bu, X., Feng, H. (2012). Sar image despeckling based on nonsubsampled shearlet transform. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(3),* 809–823. DOI 10.1109/JSTARS.4609443.
37. Farhangi, N., Ghofrani, S. (2018). Using bayesshrink, bishrink, weighted bayesshrink, and weighted bishrink in nsst and swt for despeckling sar images. *EURASIP Journal on Image and Video Processing, 2018(4),* 1–18. DOI 10.1186/s13640-018-0244-3.
38. Hazarika, D. (2017). *Despeckling of synthetic aperture radar (SAR) images in the lapped transform domain (Ph.D. Thesis)*. Tezpur University, India.
39. Rabbani, H., Vafadust, M., Abolmaesumi, P., Gazor, S. (2008). Speckle noise reduction of medical ultrasound images in complex wavelet domain using mixture priors. *IEEE Transactions on Biomedical Engineering, 55(9),* 2152–2160. DOI 10.1109/TBME.10.
40. Banerji, S., Sinha, A., Liu, C. (2013). New image descriptors based on color, texture, shape, and wavelets for object and scene image classification. *Neurocomputing, 117,* 173–185. DOI 10.1016/j.neucom.2013.02.014.

41. WHU-RS19 (2018). http://dsp.whu.edu.cn/cn/staff/yw/hrsscene.html.
42. Xia, G. S., Hu, J., Hu, F., Shi, B., Bai, X. et al. (2017). Aid: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing, 55(7),* 3965–3981. DOI 10.1109/TGRS.2017.2685945.
43. AID (2018). http://www.lmars.whu.edu.cn/xia/aid-project.html.
44. Zhou, W., Newsam, S., Li, C., Shao, Z. (2018). Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing, 145,* 197–209. DOI 10.1016/j.isprsjprs.2018.01.004.
45. PatternNet (2018). https://sites.google.com/view/zhouwx/dataset.
46. Napoletano, P. (2018). Visual descriptors for content-based retrieval of remote sensing images. *International Journal of Remote Sensing, 39(5),* 1343–1376. DOI 10.1080/01431161.2017.1399472.
47. Bianconi, F., Fernández, A. (2007). Evaluation of the effects of gabor filter parameters on texture classification. *Pattern Recognition, 40(12),* 3325–3335. DOI 10.1016/j.patcog.2007.04.023.
48. Hanbury, A., Kandaswamy, U., Adjeroh, D. A. (2005). Illumination-invariant morphological texture classification. In: *Mathematical morphology: 40 years on,* pp. 377–386. DOI 10.1007/1-4020-3443-1.
49. Florent, P., Jorge, S., Thomas, M. (2010). Improving the fisher kernel for large-scale image classification. *European Conference on Computer Vision*, pp. 143–156. Heraklion, Crete, Greece.
50. Hervé, J., Matthijs, D., Cordelia, S., Patrick, P. (2010). Aggregating local descriptors into a compact image representation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3304–3311. San Francisco, CA, USA.
51. Liu, L., Lao, S., Fieguth, P. W., Guo, Y., Wang, X. et al. (2016). Median robust extended local binary pattern for texture classification. *IEEE Transactions on Image Processing, 25(3),* 1368–1381. DOI 10.1109/TIP.2016.2522378.