



## Unconstrained texture classification using efficient jet texton learning

Swalpa Kumar Roy <sup>a</sup>, Dipak Kumar Ghosh <sup>b</sup>, Shiv Ram Dubey <sup>c</sup>,  
Siddhartha Bhattacharyya <sup>d,\*</sup>, Bidyut B. Chaudhuri <sup>e</sup>

<sup>a</sup> Department of Computer Science & Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, India

<sup>b</sup> Department of Electronics & Communication Engineering, Adamas University, Barasat, India

<sup>c</sup> Computer Vision Group, Indian Institute of Information Technology, Sri City, India

<sup>d</sup> RCC Institute of Information Technology, Kolkata, India

<sup>e</sup> Computer Vision & Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India



### ARTICLE INFO

#### Article history:

Received 26 August 2018

Received in revised form 30 October 2019

Accepted 31 October 2019

Available online 7 November 2019

#### Keywords:

Derivative of Gaussian (DtG)

Jet texton learning

Local jet vector (Ljv)

Texture classification

### ABSTRACT

This paper proposes a simple and effective texture recognition method that uses a new class of jet texton learning. In this approach, first a Jet space representation of the image is derived from a set of derivative of Gaussian (DtGs) filter responses upto 2nd order ( $\mathbb{R}^6$ ), so called local jet vector (Ljv), which satisfies the scale space properties, where the combinations of local jets preserve the intrinsic local structure of the image in a hierarchical way and are invariant to image translation, rotation and scaling. Next, the jet textons dictionary is learned using K-means clustering algorithm from DtGs responses, followed by a contrast Weber law normalization pre-processing step. Finally, the feature distribution of jet texton is considered as a model which is utilized to classify texture using a non-parametric nearest regularized subspace (Nrs) classifier. Extensive experiments on three large and well-known benchmark database for texture classification like KTH-TIPS, Brodatz and CURET show that the proposed method achieves state-of-the-art performance, especially when the number of available training samples is limited. The source code of complete system is made publicly available at <https://github.com/swalpa/JetTexton>.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

### 1.1. Research background and motivation

Texture is a recurrent basic primitive of visual pattern that reflects the distinctive ambiance between natural scenes. The texture region in an image can be characterized by the spatial distribution of image intensity or color if it follows a varying or nonuniform pattern. The non-uniformity is analogous to the variations in the captured scene due to the inherent pattern present in the scene. It can be visually perceived in images ranging from microscopic to multi-spectral remotely sensed images. It is very important to tackle the texture recognition problem as it is used in numerous applications that include industrial inspection, texture synthesis for animation and computer graphics, document image recognition, satellite and aerial imagery applications, and biomedical image analysis [1,2]. Thus, the classification of texture plays an important role in pattern recognition and

computer vision based solutions for many real-life problems of society and industry [3,4]. However, the capacity to distinguish between the textures characterizing the different local structures separated spatially is the main challenge of any texture recognition approach. Several texture classification methods have been introduced in the literature [5,6]. Most of the previously introduced descriptors are not having the required amount of discriminating power to work well for real-world textures. At the same time, most of these descriptors are also not computationally efficient for real-time applications due to the high dimensionality problem. The method of texture classification should be invariant to environmental changes, such as changes in viewpoint, rotation, scale, geometry of the underlying surface etc.

### 1.2. Related work

The Local Binary Pattern (LBP) [7], a simple and efficient statistical descriptor is used as a standard method for robust texture classification in many applications [8]. Some variants of LBP such as LBP variance [9], completed LBP [10], dominant LBP [11], binary rotation invariant and noise tolerant (BRINT) [12], scale selective local binary patterns (SSLBP) [13], multiresolution local binary pattern (MRLBP) [8], Complete dual-cross pattern (CDCP) [14], Local directional ZigZag pattern (LDZP) [15], Local morphological

\* Corresponding author.

E-mail addresses: [swalpa@cse.jgec.ac.in](mailto:swalpa@cse.jgec.ac.in) (S.K. Roy), [dipak@ieee.org](mailto:dipak@ieee.org) (D.K. Ghosh), [srdubey@iits.in](mailto:srdubey@iits.in) (S.R. Dubey), [dr.siddhartha.bhattacharyya@gmail.com](mailto:dr.siddhartha.bhattacharyya@gmail.com) (S. Bhattacharyya), [bbc@isical.ac.in](mailto:bbc@isical.ac.in) (B.B. Chaudhuri).

pattern (LMP) [16], Affine Differential Local Mean ZigZag Pattern (ADLMZP) [17], Local ZigZag max histograms of pooling pattern (LZHPP) [18] have been introduced to boost the robustness and accuracy of LBP descriptor.

In addition to LBP based methods, the texton learning based methods have also become popular for texture classification. These methods can be classified in terms of the following aspects: hard level assignment and histogram based representation; dense image sampling; a rear separating of the feature space; image filter response and patch feature. Leung and Malik [19] have proposed a method where the training images are enrolled and a vector of 48 dimensional response is generated for each class using a filtered operation. Then, the feature vectors are clustered using K-Means algorithm. The texture model is built from texton histogram for classification. Cula and Danna [20] have introduced 2-D texton learning by improving work of doing Leung and Malik [19] single texture and un-calibrated image classification. Recently, a texton dictionary is learnt in VZ-MR8 by Varma and Zisserman [21]. They considered the training dataset having the texture images with scale and rotation variations. Finally, the learned texton dictionary is used for the classification of the new unseen texture images. In 2009, Varma and Zisserman introduced another method called as VZ-Patch [22] to encode the feature distribution. The VZ-Patch is also based on the texton which is directly computed from raw image intensity values. The downside of these methods is the increased time complexity of feature extraction as well as feature matching. Varma and Garg [23] learned textons using the locally extracted fractal dimension and length signature from the MR8 filter responses to deal with large variations of scale and viewpoint.

Deep convolution neural networks (CNN) are hierarchical representation of feature extraction block which directly generates feature from raw data and is predominately used to perform several challenging computer vision tasks to achieve state-of-the-art performance [24–26]. Cimpoi et al. [27] introduced FV-CNN where extraction of CNN features at multiple scales and orderless Fisher Vector pooling operation has been used to describe the texture. The scale and rotation invariant wavelet convolution scattering network (ScatNet) has been proposed by Mallat et al. [28,29] where wavelets are used convolution filters and learning is not needed. Chan et al. [30] proposed the PCANet which includes multistage principle component analysis (PCA) in cascading form, binary hashing and histogram pooling. Liu et al. proposed Genetic Algorithm based GANet [31] for global texture representation which encodes Fisher vector pooling of a CNN feature maps. However, CNNS require the powerful hardware, both for training and test because of complex architecture and the lack of geometric invariance of global CNN activation limits robustness of recognition.

### 1.3. Our contributions

Most of the texton learning oriented approaches deal with a large number of filter responses to achieve geometric invariance. However, it increases the computational cost due to many convolution operations. A novel jet texton learning based method is proposed in this work, which deals with only six filter response for efficient, yet robust texture classification. The contributions of this work are as follows: first the representation of an image in Jet space is obtained from the responses of derivative of Gaussian (DtGs) filters upto 2nd order in  $\mathbb{R}^6$  space where DtGs responses are preserved as an intrinsic local structure of the image in a hierarchical way, using the characteristics of Hermite polynomials and together with invariant to image scaling, rotation, and reflection. Next, the Jet textons are learned from DtGs responses followed by a contrast Weber law normalization based pre-processing

step, using the K-means clustering algorithm. The feature distribution of Jet texton is considered as a texture model, and the texture is classified using a non-parametric nearest regularized subspace (NRS) classifier [32].

The remaining part of the paper is organized as follows. Section 2 explains the theoretical background of local image decompositions, Jet and Jet space. Section 3 describes the details of the proposed Jet texton learning based method. In Section 4, the texture classification performance is analyzed and compared with the different variants of LBPs and also state-of-the-art bag-of-words methods. Finally, Section 5 draws the concluding remarks.

## 2. Intrinsic image structure, jet and jet-space

Analysis of image is a crucial job in the image processing area. Lindeberg [33] introduced a general framework of Scale Space representation to address the multiple resolution image structures. The “local deep structure” idea is connected with the concept of derivative defined by

$$f'(0) = \lim_{\Delta \rightarrow 0} \frac{f(0 + \Delta) - f(0)}{\Delta}.$$

So, it is not suitable in discrete domain, like images, that presence physical measurement. To measure the image derivative, the scale space analysis introduces the following two-step solution [33]:

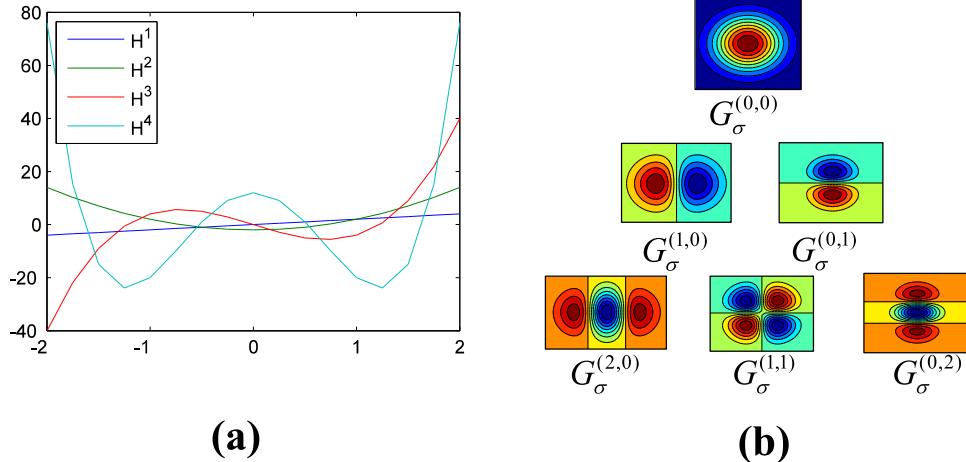
First, the image inner scale changing is measured by convolving (symbolize as \*) Gaussian kernels with image. The 1D Gaussian kernel at scale  $\sigma \in \mathbb{R}^+$  is represented by

$$G_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$$

The 2D convolution of an image patch with Gaussian function can be derived by computing the convolution of 1D Gaussian function with image patch sequentially in the horizontal and vertical directions by utilizing the 2D Gaussian function separability property ( $G_\sigma(x, y) = G_\sigma(x)G_\sigma(y)$ ) [34]. The image scaling, i.e.,  $I_\sigma = G_\sigma * I$  is efficiently computed using magnificent characteristic of the Gaussian function in space and frequency localization [35] even if the input image ( $I$ ) is the responses of physical measurement and a directly sampled function. Secondly, In scale space approach, the image derivative of up-scaled/down-scaled image can be calculated by convolving the DtGs ( $I'_\sigma = G'_\sigma * I$ ) with the original image in alternative way, as proposed by Young [36]. He observed that in our receptive fields, the derivatives of Gaussian perceive the image structure more accurately compared to other functions like Gabor transform does. The 1-D DtGs with  $\sigma > 0$  is represented as  $G_\sigma(x) = G_\sigma^0(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$  and

$$G_\sigma^m(x) = \frac{d^m}{dx^m} G_\sigma(x) = \left(\frac{-1}{\sigma\sqrt{2}}\right)^m H^m\left(\frac{x}{\sigma\sqrt{2}}\right) G_\sigma(x), \quad m \in \mathbb{Z}^+ \quad (1)$$

where  $m$  is a positive integer.  $G_\sigma^0(x)$ , and  $H^m(x)$  represent basic of Gaussian kernel and Hermite polynomial (Fig. 1(a)) with  $m$ th order [37] respectively. The normalization of DtGs can be performed using the condition  $\int |G_\sigma^k(x)| dx = 1$ . Here,  $G_\sigma^1(x)$  and  $G_\sigma^0(x)$  represent  $\ell^1$ -normalized differentiating and blurring filters, respectively. According to Hermite polynomials, for odd order the neighborhood functions are anti-symmetrical ( $G_\sigma^m(-x) = -G_\sigma^m(x)$ ), and for even order are symmetrical ( $G_\sigma^m(-x) = G_\sigma^m(x)$ ). We consider the Fourier domain to compute the consequences of neighborhood operators (that is, the  $G_\sigma^m(x)$ ) throughout the image. In Fourier domain, the  $n$ th order operator convolution represents as multiplication of  $n$ th times with envelope of Gaussian



**Fig. 1.** (a) Hermite polynomials. (b) Upto 2nd order 2- $\mathbb{D}$  derivatives of the Gaussian (DtGs).

frequency which extends the implementation efficiency. In 2- $\mathbb{D}$  (Fig. 1(b)), the DtGs at  $\sigma$  scale are expressed by

$$\begin{aligned} G_\sigma^{(m,n)}(x, y) &= G_\sigma^m(x)G_\sigma^n(y), \quad m, n \in \mathbb{Z}^+ \\ G_\sigma(x, y) &= G_\sigma^{(0,0)}(x, y) \end{aligned} \quad (2)$$

hence, the scale-space technique permits computing derivative of image at any scale with any order. The convolution formalism can be used to compute the entire image derivative whereas for a single location the mechanism of inner product ( $\langle \cdot, \cdot \rangle$ ) is widely accepted. The image derivatives w.r.t. the origin at scale  $\sigma$  is defined by

$$\begin{aligned} J_{(m,n)} &= (-1)^{(m+n)} \langle G_\sigma^{(m,n)} | I \rangle \\ &= (-1)^{(m+n)} \int_{x,y \in \mathbb{R}} G_\sigma^{(m,n)}(x, y) I(x, y) dx dy. \end{aligned} \quad (3)$$

However, measurement of the image derivative  $J_{(m,n)}$  (Fig. 2) depends with inner scale  $\sigma$  variation. Therefore, the responses of scale normalized DtGs ( $J_{(m,n)}^s$ ) are computed by multiplying  $\sigma^{n+m}$  with  $J_{(m,n)}$ , and expressed as,

$$J_{(m,n)}^s = \sigma^{n+m} J_{(m,n)}. \quad (4)$$

To avoid incorrect presumption, we should be noted that DtGs are not orthonormal kernel (e.g.,  $\langle G_\sigma^{(2,0)} | G_\sigma^{(0,2)} \rangle = 1/(16\pi\sigma^6)$ ). The vector of DtG responses  $\{G_\sigma^{(m,n)} | 0 \leq m + n \leq k\}$  up to  $k$ th order structure,  $J_{(m,n)}^k = \langle G_\sigma^{m,n} | I \rangle$  is denoted to a local  $\mathcal{L}$ -jet, whereas  $J_{(m,n)}^k \in \mathbb{R}^{\mathcal{L}}$  ( $\mathcal{L} = \frac{(k+2)!}{2*k!}$ ) is referred as an jet space element [38]. In this work, we are considering the measurements up to 2nd order structure which require kernel of six DtGs represented as,  $\{\vec{G} = (G_\sigma^{(0,0)}, G_\sigma^{(1,0)}, G_\sigma^{(0,1)}, G_\sigma^{(2,0)}, G_\sigma^{(1,1)}, G_\sigma^{(0,2)})\}$ , and referred as the DtG family (Fig. 1(b)). The responses of DtGs calculated by Eq. (4) are concerned as  $\vec{J} = (J_{(0,0)}^s, J_{(1,0)}^s, J_{(0,1)}^s, J_{(2,0)}^s, J_{(1,1)}^s, J_{(0,2)}^s)$  is known a 6-jet. Fig. 3 presents the responses for a sample image patch of Fig. 2 with higher orders local structure (up to 2nd order).

### 3. Proposed method

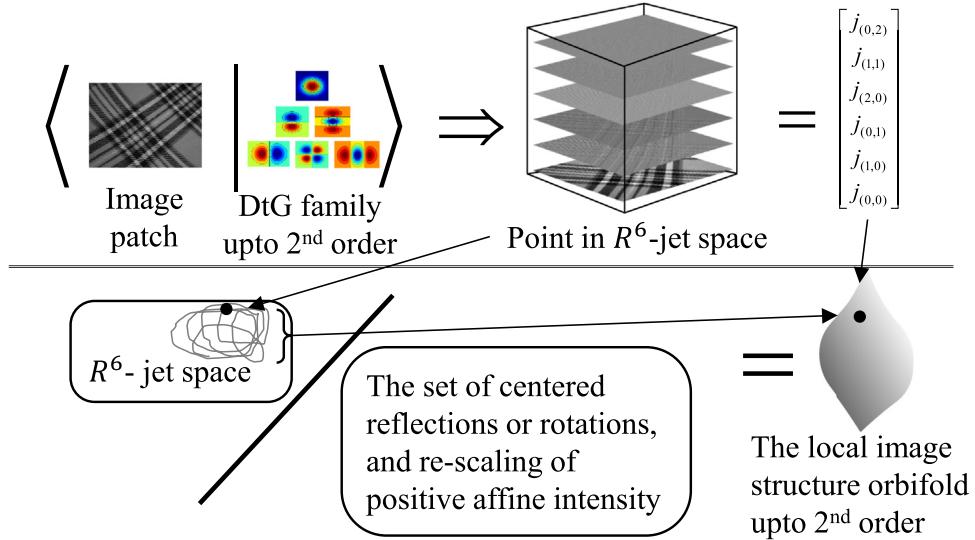
This work proposes a framework (shown in Fig. 4), which consists of two phases: jet texton learning and matching of feature distribution for texture classification. The inside information of the proposed framework are describe as follows:

#### 3.1. Jet texton learning

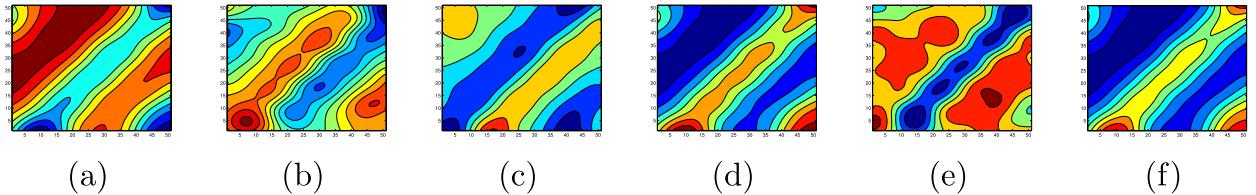
There is no preferable scale which should be justified a real-world natural texture. The representation abstract feature of the whole image at different scales can be detected by searching. Therefore, the desired image representation can be represented by considering scale-space relations in order to facilitate the consideration of different class of resolutions. In this work, the morphology of the image object, at a particular scale is explored with the help of spatial derivatives of texture image with different Gaussian structures (anisotropic and isotropic). Florack et al. [38] formed a spatial derivatives using multi-scale  $\mathcal{L}$ -jet image representation, as well as derivatives with respect to the scale ( $\sigma$ ) based on the principle visual cortex system. A set of responds for derivative of Gaussian (DtGs) filters upto 2nd order ( $\mathbb{R}^6$ ) which stacks to form a vector in jet space representation of an image called local jet vector (Ljv), satisfying the statistical in variance characteristics of scale space [39,40]. In other words, the jet can also be understood as isolating an image patch with a Gaussian window and then probing it with *Hermite* function which is not unlike a windowed Fourier transform. To learn the Jet textons, at first the given texture image is transformed into local  $\mathcal{L}$ -jet ( $\mathcal{L} = 6$ ) according to the Eq. (3), where the elements of 6-jet demonstrate the DtGs response upto 2nd order (shown in Fig. 4(a)). The scale normalized derivative responses of 6-jet for a given image  $I$ ,  $\{\vec{J} = (J_{(0,0)}^s, J_{(1,0)}^s, J_{(0,1)}^s, J_{(2,0)}^s, J_{(1,1)}^s, J_{(0,2)}^s)\}$  are represented as a vector  $\vec{J} = (J^1, J^2, \dots, J^{L-1}, J^L)$  which we specify as the local jet vector (Ljv) (shown in Fig. 4(b)). The filters in the filter bank are translated to make it the zero-mean and motivated by Weber's law [22] a contrast normalization pre-processing step is used in this work to normalize the DtG responses. Let  $\|\mathbf{J}^{(i,j)}\|_2$  be the  $\ell_2$  norm of the DtG responses at pixel  $(i, j)$ . The normalized values of the DtGs responses are computed as follows,

$$\mathbf{J}^{(i,j)} \leftarrow \mathbf{J}^{(i,j)} \times \frac{\log(1 + \frac{L^{(i,j)}}{0.03})}{L^{(i,j)}}, \quad (5)$$

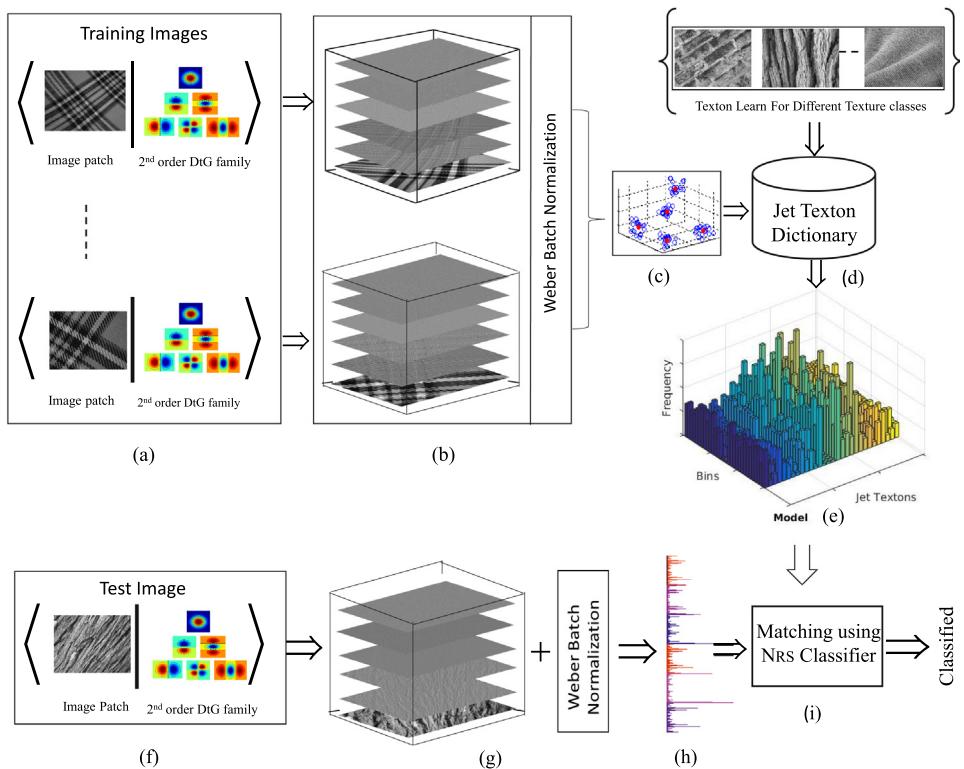
where  $L^{(i,j)} = \|\mathbf{J}^{(i,j)}\|_2$  is the magnitude of the DtG output at coordinate  $(i, j)$ . To form jet texton dictionary of each class, the normalized DtG responses of randomly selected training image set are combined and clustered using  $K$ -means clustering algorithm (shown in Fig. 4(c)). Let  $L$  be the number of randomly selected training image for the first class which is represented as  $X_1 = [x_{11}, x_{12}, \dots, x_{1L}] \in \mathbb{R}^{M \times N \times L}$ , where  $M \times N$  denotes the size of a texture image. So, the total number of normalized DtG responses of  $L$  selected images will be  $L \times \mathcal{L}$  and can be



**Fig. 2.** Diagrammatically, top-left indicates the image structure measurement using up to 2nd order DtG filters bank and top-right shows the resulting local jet vector.



**Fig. 3.** (a)–(e). Examples of responses for image patch shown Fig. 2 for upto 2nd order local structure.



**Fig. 4.** Proposed texture classification framework: (a) randomly selected training images and DtGs structure upto 2nd order; (b) the scale normalized derivative responses, 6-jet known as local jet vector (Ljv) for training images (a); (c) K-means clustering; (d) jet texons dictionary for training images of all class; (e) distribution of learned models; (f)–(g) test image and its jet representation; (h) distribution of test model; (i) matching using Nrs classifier.

represented as  $J_{X_1} = [J_{x_{11}}, J_{x_{12}}, \dots, J_{x_{1L}}, \dots, J_{x_{1(L \times C)}}] \in R^{M \times N \times L \times C}$ , where  $J_{x_{1l}}$  represents  $l$ th element of  $J_{X_1}$ . The dictionary of  $K$  jet texton for first class, represented by  $D_1 = [d_{11}, d_{12}, \dots, d_{1K}] \in R^{M \times N \times K}$ , is to be learned from the training set  $X_1$ , where  $d_{1j}$ ,  $j \in 1, \dots, K$  is a jet texton. The conventional clustering algorithm,  $K$ -means is applied to determine the  $K$  number of textons by finding the solution of following optimization equation [21].

$$\arg \min_{\Omega_1} \sum_{j=1}^K \sum_{J_{x_{1l}} \in \Omega_{1j}} \|J_{x_{1l}} - d_{1j}\|_2^2 \quad (6)$$

In  $K$ -means clustering, the dataset  $X_1$  is partitioned into  $K$  different groups  $\Omega_1 = \{\Omega_{11}, \Omega_{12}, \dots, \Omega_{1j}, \dots, \Omega_{1K}\}$  so as to minimize the within-cluster sum of squares (Wcss), where  $d_{1j}$  defines the cluster mean of group  $\Omega_{1j}$ . The dictionary  $D_1$  is more compact compared to  $X_1$  (i.e  $K \ll L$ ). However, it is expected that the learned texton in dictionary  $D_1$  could represent all the samples in  $X_1$ . In this way the dictionary  $D = [D_1, D_2, \dots, D_C]$  is formed, which contains a total of  $C \times K$  textons for all  $C$  classes (shown in Fig. 4(d)). Finally, the models of training images are formed by extracting the histogram of learned jet textons (shown in Fig. 4(e)) and these models are utilized to recognize the class of test image during the classification phase.

### 3.2. Feature distribution matching

In pattern classification problems, different types of classifiers such as nearest neighbor classifier (Nnc) [14], support vector machine (Svm) [11], multiple fuzzy-classifier [41], ensemble extreme learning machine (Eelm) [42], and nearest regularized subspace (Nrs) classifier [32] have been used to recognize the test samples. In this work, the class of test image is recognized by matching the model histogram of test image (build up procedure is shown in Fig. 4(f)-(h)) with the learned models of training images (shown in Fig. 4(i)) using Nrs classifier [32]. In Nrs classifier, the class of the test sample  $y \in R^d$  is referred to one of the  $C$  classes. Let  $x_i$  be the  $i$ th training sample and a vector  $\mathbf{X} = \{x_i\}_{i=1}^n \in R^d$  is formed by stacking all training samples with labels of categories  $\omega_c \in \{1, 2, \dots, C\}$ , where  $n$  and  $C$  represent the total number of training examples and the number of categories, respectively. Let  $n_c$  be the number training examples for  $c$ th category, where  $\sum_{c=1}^C n_c = n$ . If the total number of training samples of  $c$ th category is  $\mathbf{X}_c$ , then the category of test example  $y$ ,  $\bar{y}_c$  is approximated using a linear combination of class-wise available training samples as

$$\bar{y}_c = \mathbf{X}_c \alpha_c \quad (7)$$

where  $\mathbf{X}_c$  is of size  $d \times n_c$  and  $\alpha_c$  is the weighted coefficients of  $n_c \times 1$  size vector. The classification for a test feature  $y$  is done by assigning appropriate class label for which the residual is minimized [32], i.e.,

$$\text{class}(y) = \arg \min_{c=1, \dots, C} (r_c) \quad (8)$$

where  $r_c = \|\bar{y}_c - y\|_2$  represents residual between  $\bar{y}_c$  and  $y$  which is to be categorized. In Nrs [32],  $\alpha_c$  represents the class specific weight vector which is determined mathematically as follows

$$\alpha_c = \arg \min_{\alpha^*} \|y - \mathbf{X}_c \alpha^*\|_2 + \tau \|\Gamma_{c,y} \alpha^*\|_2 \quad (9)$$

where  $\Gamma_{c,y}$  is the biasing of Tikhonov regularization matrix for  $c$ th category and test example  $y$ ,  $\tau$  represents regularization parameter used globally (best-approximated result is achieved when  $\tau$  is set to be  $5 \times 10^{-2}$ ) which balances the minimization of regularization terms with the residual, and  $\alpha^*$  is the different

value of  $\alpha_c$ . Mathematically, a diagonal depiction of  $\Gamma_{c,y}$  [32] is in the following form

$$\Gamma_{c,y} = \begin{bmatrix} \|y - x_{c,1}\|_2 & 0 \\ & \ddots \\ 0 & \|y - x_{c,n_c}\|_2 \end{bmatrix}$$

where the individual elements i.e,  $x_{c,1}, x_{c,2}, \dots, x_{c,n_c}$  are columns of the matrix  $\mathbf{X}_c$  for the  $c$ th category. As stated by the Eq. (9), the class specific weight vector  $\alpha_c$  is calculated as follows [32]

$$\alpha_c = (\mathbf{X}_c^T \mathbf{X}_c + \tau^2 \Gamma_{c,y}^T \Gamma_{c,y})^{-1} \mathbf{X}_c^T y. \quad (10)$$

## 4. Results and discussions

### 4.1. Texture databases

The performance of the introduced Jet Texton Learning approach is evaluated on three standard sets of texture databases: KTH-TIPS [43], Brodatz album [44], and CUReT [45] texture databases, respectively. Due to the insufficient intra-class variation and huge number of texture categories with limited number of examples per category make the experiments running on whole database challenging. The descriptions of mentioned databases are as follow:

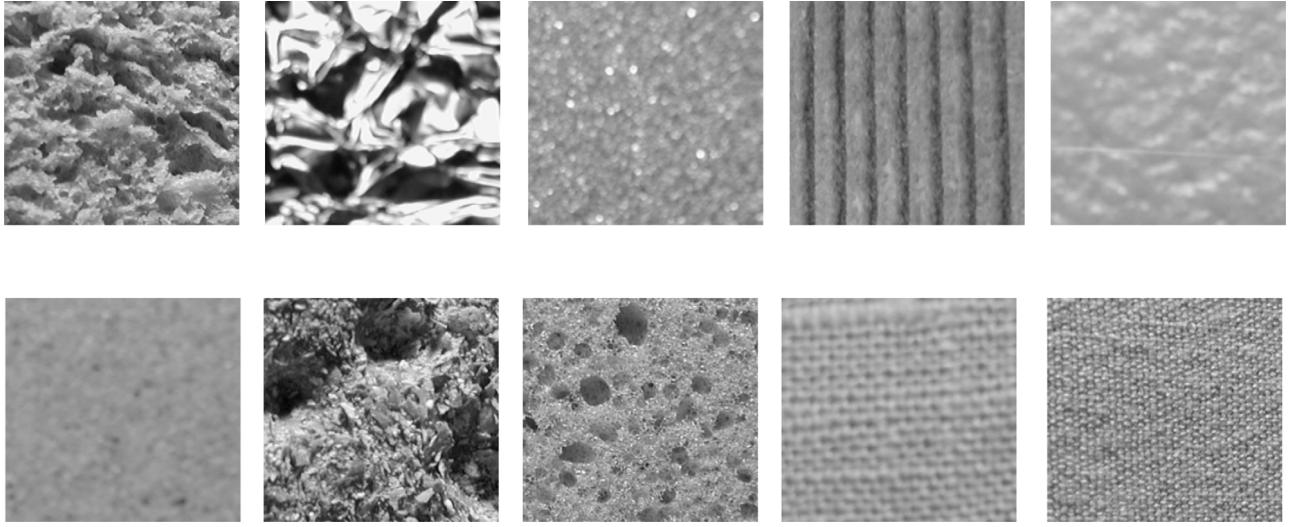
The **KTH-TIPS** database [43] is enlarged from CUReT by imaging new samples of 10 textures as shown in Fig. 5. It contains texture images with three different geometric poses, four illuminations, and nine different variations of scales with size  $200 \times 200$ .

**Brodatz** [44] album consists of 32 homogeneous texture classes where 25 sub-images of size  $128 \times 128$  is obtained by non-overlapping partitioning of each image, and each sub-image is sampled down to  $64 \times 64$  pixels shown in 6.

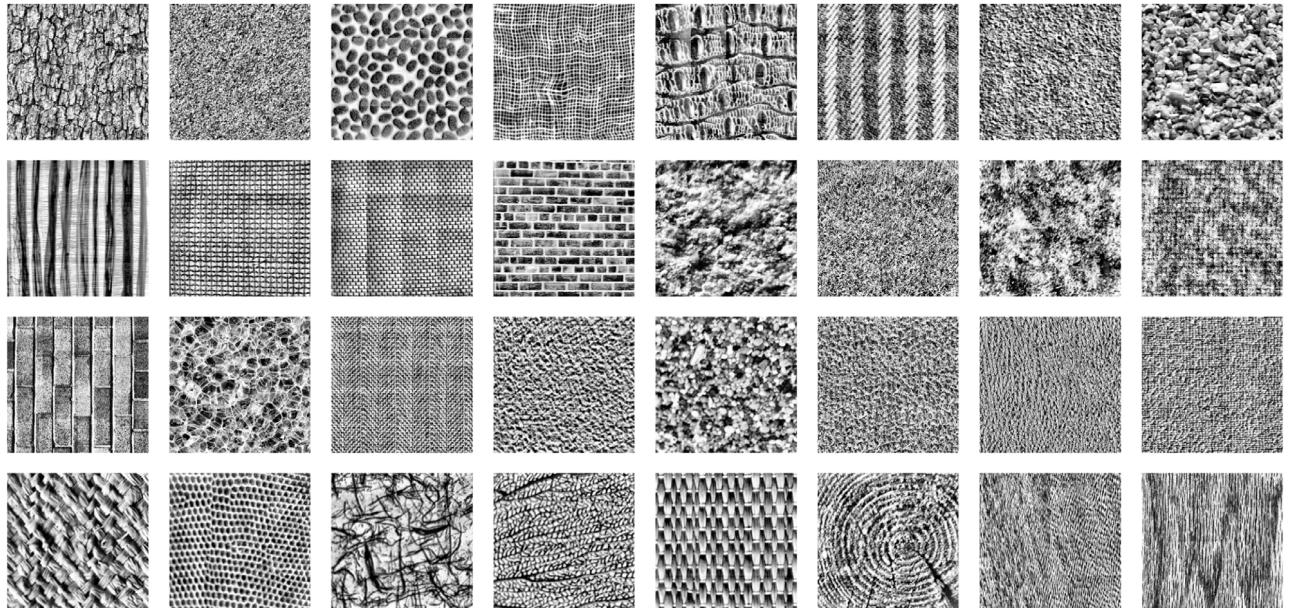
The 92 texture images per class with total 61 classes has been included in **CUReT** database [45]. This database is organized to attain large intra-class variation where images are collected in varying lighting and view-point conditions with uniform scale. The 61 texture classes each containing 92 images are cropped into  $200 \times 200$  spatial neighborhood and transformed to gray scale [21] as shown in Fig. 7. The **KTH-TIPS**, **Brodatz**, and **CUReT**, database are outlined in Table 1.

### 4.2. Results of experiment

In order to judge the texture classification outcome and robustness of the proposed Jet Texton Learning approach, the experiments are carried out on three well-known texture databases (KTH-TIPS, Brodatz, and CUReT) which consist of images with different pose, scale, rotation and illumination conditions as mentioned in Table 1. The samples are captured in uncontrolled environments. In this work, where  $L$  the number of samples for each class are randomly selected as training samples to learn the jet texton dictionary and subsequently train the Nrs classifier. The remaining  $N - L$  samples are used as test samples for performance analysis of the proposed model, where  $N$  represents total number of samples per class. In KTH-TIPS, to learn the jet texton dictionary, the DtGs responses upto 2nd order of  $L = 40$  randomly selected images from each texture category are arranged and clustered using  $K$ -means algorithms. So,  $K = 12$  jet textons are learned from each of the  $C = 10$  texture categories, resulting in a jet texton dictionary of  $12 \times 10 = 120$  ( $K \times C$ ) textons. Similarly, jet texton dictionaries of  $32 \times 15 = 480$  textons, and  $61 \times 19 = 1159$  textons are learned for brodatz and CUReT databases, respectively. Here, we have chosen the same value of  $K$  for the different texture classes by putting equal importance to all different classes. The performance of proposed model depends



**Fig. 5.** KTH-TIPS database from where ten texture images from each class has been taken haphazardly.



**Fig. 6.** In experiment thirty two texture images taken from each class randomly of Brodatz textures.

**Table 1**

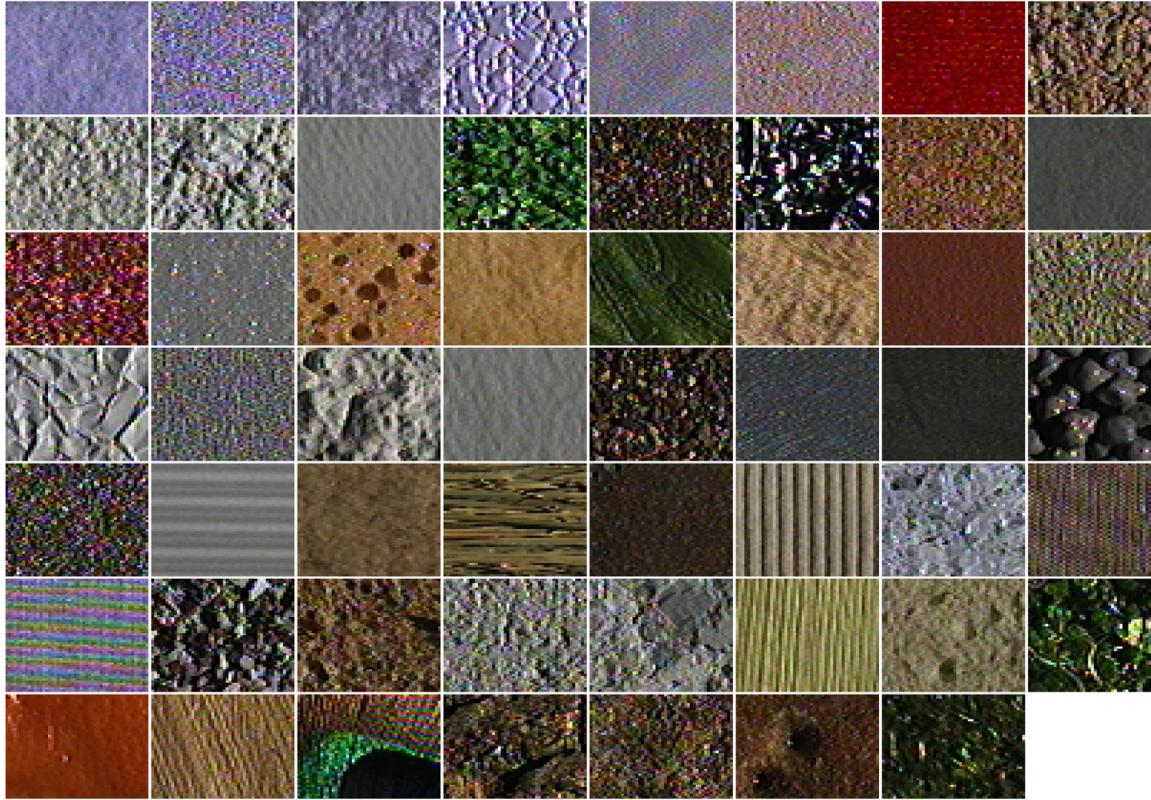
Summary sheet of experimental texture database.

Texture database	Image rotation	Illumination variation	Scale variation	Texture classes	Sample size (pixels)	Samples per class	Total samples
KTH-TIPS	✓	✓	✓	10	200 × 200	81	810
Brodatz	✓		✓	32	64 × 64	64	2048
CUReT	✓	✓		61	200 × 200	92	5612

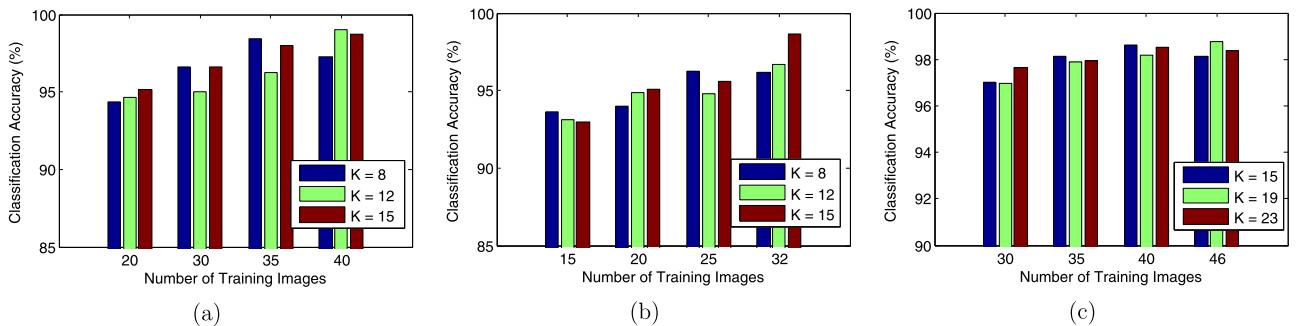
on the aforementioned parameters ( $K$ , number of cluster centers and  $L$ , number of training examples per category). Therefore these parameters are selected through evaluating the results of the Jet Texton Learning approach by varying the value of  $L$  and  $K$ . The classification results with different  $L$  and  $K$  are shown in Fig. 8. It is observed that the best performance of the proposed method on KTH-TIPS, Brodatz, and CUReT databases are achieved as 99.00% 98.63% and 98.75% for  $(L, K) = (40, 12)$ ,  $(L, K) = (32, 15)$ , and  $(L, K) = (46, 19)$ , respectively. In the literature, usually LBP variants are used for texture classification task. We compared the results of the Jet Texton Learning approach with some powerful variants of LBP, which includes CLBP [10], LBPV [9], LTP [46],

DLBP [11],  $Lbp_{R,N}^{sri\_su2}$  [47], PRICOLBP [48], BRINT [12], COALBP [49] and LWP [3]. The comparative results on KTH-TIPS, Brodatz, and CUReT are shown in Table 2.

The observations noted from Table 2 are as follows: The scale-invariant  $Lbp_{R,N}^{sri\_su2}$  descriptor performs better than  $Lbp_{R,N}^{riu2}$ . However, the multi-resolution  $Lbp_{R,N}^{riu2}$  descriptor, and  $CLBP\_S_{R,N}^{riu2}/M_{R,N}^{riu2}/C$  descriptor is better than the given performance of  $Lbp_{R,N}^{sri\_su2}$  descriptor; and much worse than the proposed descriptor. As a result the unvarying extraction and accurate scaling for each pixel is crucial. However,  $Lbp_{R(i,j),8}^{sri\_su2}$  contributes valuable performance in controlled environment [47], but losses status over complex databases. CDP [14] achieves improved performance



**Fig. 7.** Sample images of the Columbia–Utrecht database. In experiments, all the color images are mapped to the corresponding gray version.



**Fig. 8.** The proposed method's texture classification performances with different number of training samples ( $L$ ) and cluster centers ( $K$ ) for (a) KTH-TIPS, (b) Brodatz and (c) CUReT databases.

over CLBP since the CDCP patterns are extracted at the component and holistic levels. DLBP, when combined with Gabor features, achieves a classification performance higher than the conventional LBP with NNC. However, its performance is quite less than the proposed method, as it does not consider changes in scale. The proposed method achieves remarkable classification performance compared to CLBP, LBPF, LBPHF\_S and LBPHF\_S\_M on KTH-TIPS, Brodatz and CUReT texture datasets and yields comparable performance with DLBP. Table 3 shows the performance of proposed texture classification method and other bag-of-words models. Lazebnik et al. [52] proposed to detect regions of interest using Harris-affine corner and Laplacian-affine blobs and then extract regions normalizing SPIN and RIFT, as texture pattern. Finally, texture classification is performed using a nearest neighbor classifier (NNC). Caputo et al. [55] used SVM kernel instead of nearest neighbor classifier and revealed that the SVM classifier could achieve reasonably better performance compared to NNC. Zhang et al. [53] proposed object and texture classification by analyzing different texture features and kernels. Recently, feature extraction

using global scale invariant methods highlights the researchers due to the pixel by pixel operations of local scale normalization which is usually slow. Xu et al. [57] and Quan et al. [58] tabulates image pixels into multiple regions by their gray intensities or local distribution of the feature descriptors and finally classify them. At varying scales multi-scale BIF [56]  $\sigma$ ,  $2\sigma$ ,  $4\sigma$ , and  $8\sigma$  give comparable performance than the other bag-of-words methods where NNC is the used as classifier. Because of the pyramid histogram along with time consuming shift matching process makes BIF as a discriminative feature. Also, the feature distribution length of BIF descriptor [56], CLBP [10], SSLBP [13] and PRIOLBP [48] are much larger ( $6^4 = 1296$ ,  $2200$ ,  $480 \times 5 = 1000$  and  $590 \times 2 = 1180$ ) compared to the proposed method ( $61 \times 19 = 1159$ ). Due to the absence of scale shifting scheme the performance of BIF, a bit [56] is reduced. Apart from the bag-of-words model, the performance of the proposed method is also compared with two conventional texton based methods, i.e VZ\_MR8 and VZ\_Patch. The performance of VZ\_MR8 and VZ\_Patch

**Table 2**

Comparative results of the classification performance of the proposed and variants of LBP methods, where '-' indicate the missing result.

Methods	Classifier	Classification accuracy (%)		
		KTH-TIPS [43]	Brodatz [44]	CUReT [45]
LBPV [9]	NNC	95.50	93.80	94.00
BRINT [12]	NNC	97.75	99.22	97.06
LBP <sub>1,8</sub> <sup>riu2</sup> [7]	NNC	82.67	82.16	80.63
DLBP <sub>3,24</sub> [11]	SVM	86.99	99.16	84.93
LBP <sub>1,8</sub> <sup>sri_su2</sup> [47]	NNC	89.73	69.50	85.00
LBP <sub>(1,8+2,16+3,24)</sub> [7]	NNC	95.17	91.60	95.84
CLBP_SMC [10]	NNC	97.19	94.80	97.40
SSLBP [13]	NNC	97.80	-	98.55
PRICoLBP <sub>g</sub> [48]	SVM	98.40	96.90	98.40
LBP <sub>PHF_S</sub> [50]	NNC	97.00	94.60	95.90
LBP <sub>PHF_S_M</sub> [51]	NNC	97.00	94.60	95.90
CoALBP [49]	NNC	97.00	94.20	98.00
CDCP [14]	NNC	97.90	97.20	-
LDZP [16]	NNC	97.82	97.20	97.76
LWP [3]	NNC	93.00	95.48	90.91
Proposed method	NRS	99.00	98.63	98.75

**Table 3**

Texture classification results of the proposed and other bag-of-words methods.

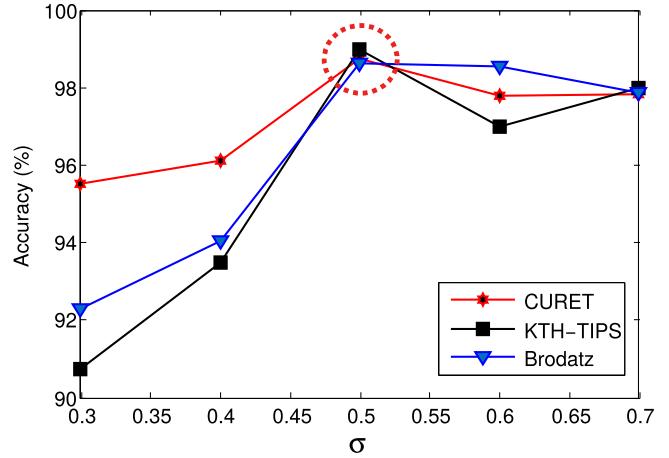
Methods	Classifier	Classification accuracy (%)		
		KTH-TIPS [43]	Brodatz [44]	CUReT [45]
VZ-MR8 [21]	NNC	94.50	94.62	97.43
VZ-Patch [22]	NNC	92.40	87.10	98.03
Lazebnik et al. [52]	NNC	91.30	88.20	72.50
Zhang et al. [53]	SVM	96.10	95.90	95.30
Liu et al. [54]	SVM	-	94.20	98.50
Capato et al. [55]	SVM	94.80	95.00	98.50
BIF [56]	Shift NNC	98.50	98.47	98.60
S [20]	NNC	-	-	95.25
LM [19]	NNC	-	-	94.65
Proposed method	NNC	91.00	94.30	94.50
Proposed method	SVM	94.00	98.20	97.50
Proposed method	NRS	99.00	98.63	98.75

is quite lower than the proposed method. The feature extraction and matching complexity is also higher than the proposed method, because for each texture image both methods initially find the dense 38 filter responses, whereas the proposed method extract only six DtGs response, so called 6-jet and achieves invariance to scales, rotations or reflection. The overall observations from **Table 3** show that the proposed method exceeds the classification performance of several bag-of-words methods. **Table 3** also indicates that the proposed method with NNC and SVM classifiers provides a comparable performance with some other methods. The proposed method achieves reasonably high performance due to the following characteristic. In the proposed approach, first a set of derivative of Gaussian (DtGs) filter responds upto 2nd order ( $\mathbb{R}^6$ ) which stacks to form a jet space representation of a texture image called local jet vector (Ljv), satisfying the statistical in variance characteristics of scale space, where the combinations of local jet are preserved and then abstract representation of local image structure is done in a hierarchical way and invariant to image scaling, translation, and rotation or reflection. Then, the jet textons dictionary is learned from DtGs responses, followed by a Weber law contrast normalization pre-processing step using K-means clustering algorithm. Finally, the feature distribution of jet texton is considered as a texture model which is utilized to classify texture using a non-parametric nearest regularized subspace (NRS) classifier and enjoys the following properties of NRS: the testing sample is approximated from available training samples within the class via a linear combination. The best approximation of the test sample, derived from class-wise training samples is

**Table 4**

The classification accuracy (%) of few state-of-the-art texton based methods on CUReT database with different training conditions along with the number of textons.

Method	Number of textons	Training samples/class ( $L$ )			
		46	23	12	6
VZ_MR8	2440	97.79	95.03	90.48	82.90
VZ_Patch	2440	97.66	94.58	89.40	81.06
Varma and Garg	2048	97.50	94.69	89.74	81.67
CLBP	2248	97.39	94.19	80.72	79.88
Proposed method	1159	98.75	97.75	91.10	83.50



**Fig. 9.** The sigma ( $\sigma$ ) of the DtGs kernel vs. classification accuracies for the proposed method.

**Table 5**

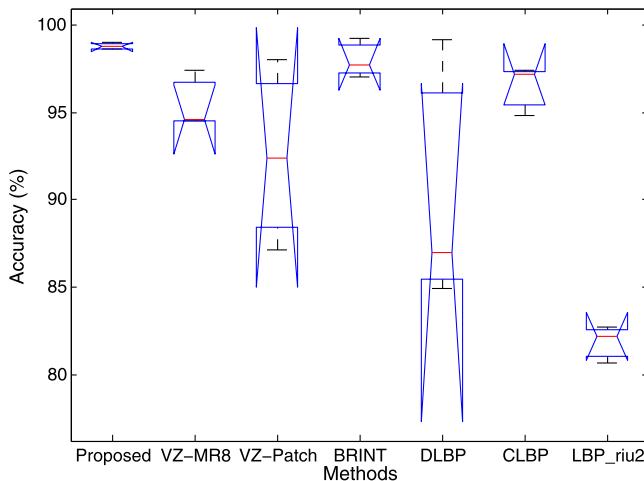
The result of one way statistical ANOVA test with significance level  $\alpha = 0.05$  for KTH-TIPS, Brodatz, and CUReT texture databases.

Source	SS	df	MS	F	Prob (p) > F
Groups	624.958	06	104.160	7.58	<b>0.0009</b>
Error	192.463	14	013.747		
Total	817.420	20			

used to assign its class label. The linear combination of hypotheses is calculated in a stable manner using a distance-weighted Tikhonov regularization.

The performance of the statistical texton based learning method varies with the number of samples present in the training set. The classification accuracy (%) of the proposed and a few state-of-the-art texton based methods for CUReT database under different training conditions are shown in **Table 4**. It is observed that the classification performance degrades due to reduction of the number of training samples. The performances of proposed Jet Texton Learning approach for different sigma ( $\sigma$ ) values of DtGs kernel are evaluated on KTH-TIPS, Brodatz and CUReT databases and shown in **Fig. 9**. It indicates that the best result is achieved for  $\sigma = 0.5$  in all three databases.

Though the results of **Tables 2** and **3** clearly indicate the improved results, we conducted another experiment using one way statistical analysis of variance (ANOVA) test [59] to cross check the significant improvement of the classification performance. ANOVA test is a way to find whether experimental results are significant or not, and used to evaluate the group means differences in a sample set. The null hypothesis  $H_0$  of the test points out that, *the group means are very close to each other*. In case any group average is totally out of the box as compared to other group averages, it signifies that the selected significant level is more than the  $p$ -value and the  $H_0$  can be rejected. In order to perceive the difference between the performance of



**Fig. 10.** The box plot (classification methods vs. accuracy) of proposed and other state-of-the-art methods corresponding to ANOVA test on KTH-TIPS, Brodatz and CUReT texture databases.

proposed method and other popular methods like VZ-MR8, VZ-Patch, BRINT, DLBP, CLBP\_S/M/C, and LBP<sup>riu2</sup>, an one way ANOVA test is experimented with selecting the significance level  $\alpha = 0.05$ . The test results are tabulated in Table 5. The result in Table 5 indicate that the selected significant level  $\alpha = 0.05$  is more than the  $p$ -value (0.0009), which implies the proposed method's performance significantly differs from other methods and hence cancels the hypothesis  $H_0$ . Moreover, the ANOVA test results are also displayed using a box plot as depicted in Fig. 10. The box plot results convey the improvement in the mean performance of the proposed method over other state-of-the-art methods such as VZ-MR8 [21], VZ-Patch [22], BRINT [12], DLBP [11], CLBP\_S/M/C [10], and LBP<sup>riu2</sup> [7].

To show the class-wise performance of the proposed texture classification method, the confusion matrix for KTH-TIPS, Brodatz, and CUReT databases are shown in Fig. 11. To make it more visible, we adopt a color bar which represents the class-wise classification or misclassification rate. In confusion matrix the horizontal and vertical axes represent the true and predicted class labels, and the principle diagonal points represent the class-wise performance rate whereas outside the diagonal points shows miss classification rate in percentage. In confusion matrix, the best classification along the diagonal is as dark red as possible and away from the diagonal the points should be as dark blue as possible.

The texture classification performance of the proposed method is also compared with the state-of-the-art deep CNN based methods, such as PCANet [30], PCANet<sup>riu2</sup> [30], FV-AlexNet [60], FV-VGGM [27,61], FV-VGGVD [27,61] in Table 6. This result comparison indicates that the proposed method provides significant improvement of the classification performance as compared to the state-of-the-art deep CNN based methods. Moreover, the dimension of final feature of proposed method is less as compared to others, except PCANet<sup>riu2</sup>.

The algorithm is implemented in MATLAB 2016 environment and executed on Intel® Core™2 Duo CPU T6400 @ 2.00 GHz × 2 processor and 4 GB RAM with UBUNTU 16.04 Lts operating system. According to three benchmark texture database, Table 7 lists out the average training and matching (or classification) time cost per-image is carried out using the proposed method. It has been observed from the experimental result that the proposed method can be performed in reasonable time.

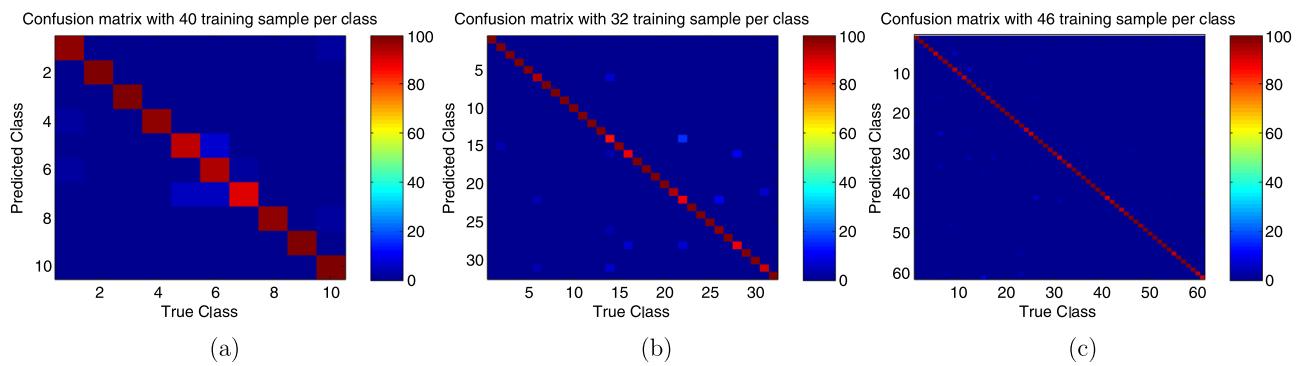
The computational complexity of the proposed jet texton learning method is described as follows. The jet texton learning

**Table 6**  
Texture classification results of the proposed and deep CNN based methods.

Method	Classifier	Feature dimension	Results on texture datasets		
			Brodatz	CUReT	KTH-TIPS
PCANet [30]	NNC	2 048	90.89	92.03	–
PCANet <sup>riu2</sup> [30]	NNC	80	85.70	81.48	–
FV-AlexNet [60]	SVM	32 768	98.20	98.50	98.60
FV-VGGM [27,61]	SVM	65 536	98.50	98.40	98.80
FV-VGGVD [27,61]	SVM	65 536	98.60	98.60	98.80
Proposed method	NRS	1 159	98.63	98.75	99.00

framework comprises three steps which include (i) formation of jet representation for a given image, (ii) building the texton dictionary and (iii) building the histogram for each texton. The computation complexity to compute the filter response (jet representation) using separable convolutions is  $\mathcal{O}(2 \times MNk)$ , where  $M$  and  $N$  are the height and width of the image and  $k$  is the kernel size. The complexity to compute each filter response with normalization process becomes  $\mathcal{O}(2 \times MNk) + \mathcal{O}(MN)$ , where  $\mathcal{O}(MN)$  represents the complexity of normalization process for each filter response. Now, each image having 6 filter responses and  $L$  number of sample images is taken from each class for training. So, the complexity to compute filter responses for  $CL$  samples of  $C$  class becomes  $\mathcal{O}(CL \times 6 \times 2 \times MNk) + \mathcal{O}(CL \times 6 \times MN)$ . The complexity to find the dictionaries of  $K$  jet texton for each class using K-means clustering (Lloyd's algorithm and its most variants) is  $\mathcal{O}(nKdP)$ , where  $n = 6L$  is the number of  $d$ -dimensional ( $d = MN$ ) vectors to be clustered,  $K$  is the number of clusters and  $P$  is the number of iterations needed until convergence. So, the complexity to find the dictionaries of  $K$  jet texton for each class is  $\mathcal{O}(6 \times LKMNP)$  and for  $C$  number of classes it becomes  $\mathcal{O}(6 \times CKLMNP)$ . The complexity to build the histogram of each jet texton is  $\mathcal{O}(MN)$ . So, the complexity to build the histogram for  $C \times K$  textons becomes  $\mathcal{O}(CK \times MN)$ . Therefore, the computational complexity of the proposed jet texton learning model is  $(\mathcal{O}(CL \times 6 \times 2 \times MNk) + \mathcal{O}(CL \times 6 \times MN)) + \mathcal{O}(6 \times CKLMNP) + \mathcal{O}(CK \times MN)$  and the final computational complexity without the constant term can be written as  $\mathcal{O}(CLMNk) + \mathcal{O}(CLMN) + \mathcal{O}(CKMN) + \mathcal{O}(CKMN)$ .

Finally we highlight the differences and common characteristic of VZ\_MR8 and the proposed method. The proposed method and VZ\_MR8 both have classified an image in two steps: texton dictionary learning and classification. In texton learning, stage the proposed method finds six DtGs responses ( $\mathbb{R}^6$ -Jet space which satisfies the statistical invariance properties of scale space [40]) to extract a 6-dimensional feature vector corresponding to each pixel and a texton dictionary where a total  $61 \times 19 = 1159$  textons have been learned using K-means clustering algorithm. Whereas VZ\_MR8 requires 38 filters which extract an 8-dimensional feature vector corresponding to each pixel and a texton dictionary containing a total of  $61 \times 40 = 2440$  learned textons. For both techniques the histogram is build based on texton frequency. In classification stage VZ\_MR8 utilizes NNC classifier for dissimilarity measurement whereas the proposed method uses NRS classifier and enjoys the following properties of NRS: the testing sample is approximated from available training samples within the class via a linear combination. The best approximation of the test sample, derived from class-wise training samples is used to assign its class label. The linear combination of hypotheses is calculated in a stable manner using a distance-weighted Tikhonov regularization.



**Fig. 11.** The confusion matrix of texture classification using the proposed method for (a) KTH-TIPS, (b) Brodatz, and (c) CUReT databases.

**Table 7**

The average training and matching (testing) time cost per-image on three benchmark texture database.

Method	Dataset	Training time (s)		Test time (s)
		Build texton	Build histogram	
Proposed model	Brodatz	0.02	0.06	0.07
	CUReT	1.83	2.63	0.82
	KTH-TIPS	2.37	2.74	0.92
VZ-MR8 [21]	CUReT	2.23	–	–
	KTH-TIPS	1.7	–	–
VZ-Patch [22]	CUReT	26.1	–	–
	KTH-TIPS	18.6	–	–

## 5. Conclusion

This work presents an efficient jet texton learning based texture classification method. The proposed method includes jet texton dictionary learning and classification phase. In jet texton dictionary learning, the proposed method extracts six DtGs responses ( $\mathbb{R}^6$ -jet space) and finds 6 dimensional feature vector for each pixel of training images. Then a jet texton dictionary is learned using  $K$ -means clustering algorithm. Finally, the model of training images is formed by extracting the histogram of learned jet textons. In the classification phase, the class of test image is recognized by comparing its model histogram with trained models using NRS classifier. To analyze the texture classification performances of the proposed method and to compare the outcomes with other state-of-the-art methods, the experiments are conducted on three well-known KTH-TIPS, Brodatz and CUReT databases. The experimental results show that the proposed method provides better performance compared to the other state-of-the-art methods and maintains a good classification performance even in large databases such as CUReT with less number of textons. The proposed method can be used in real applications like industrial inspection, document image recognition, biomedical image analysis etc.

## Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.asoc.2019.105910>.

## Acknowledgments

The authors would like to gratefully thank the Associate Editor and the anonymous reviewers for their outstanding comments and suggestions, which greatly helped to improve the technical quality and presentation of the work.

## References

- [1] X. Xie, M. Mirmehdi, A galaxy of texture features, in: *Handbook of Texture Analysis*, World Scientific, 2008, pp. 375–406.
- [2] S. Umer, B.C. Dhara, B. Chanda, A novel cancelable iris recognition system based on feature learning techniques, *Inform. Sci.* 406 (2017) 102–118.
- [3] S.R. Dubey, S.K. Singh, R.K. Singh, Local wavelet pattern: a new feature descriptor for image retrieval in medical ct databases, *IEEE Trans. Image Process.* 24 (12) (2015) 5892–5903.
- [4] G. JayaBrindha, E.S.G. Subbu, Ant colony technique for optimizing the order of cascaded svm classifier for sunflower seed classification, *IEEE Trans. Emerg. Top. Comput. Intell.* 2 (1) (2018) 78–88, <http://dx.doi.org/10.1109/TETCI.2017.2772918>.
- [5] S. Brahma, L.C. Jain, L. Nanni, A. Lumini, et al., *Local Binary Patterns: New Variants and Applications*, Springer, 2014.
- [6] Y. Akbulut, Y. Guo, A. Şengür, M. Aslan, An effective color texture image segmentation algorithm based on hermite transform, *Appl. Soft Comput.* 67 (2018) 494–504.
- [7] T. Ojala, M. Pietikäinen, T. Maenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (7) (2002) 971–987.
- [8] A.R. Yadav, R. Anand, M. Dewal, S. Gupta, Multiresolution local binary pattern variants based texture feature extraction techniques for efficient classification of microscopic images of hardwood species, *Appl. Soft Comput.* 32 (2015) 101–112.
- [9] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using LBP variance (LBPV) with global matching, *Pattern Recognit.* 43 (3) (2010) 706–719.
- [10] Z. Guo, L. Zhang, D. Zhang, A completed modeling of local binary pattern operator for texture classification, *IEEE Trans. Image Process.* 19 (6) (2010) 1657–1663.
- [11] S. Liao, M.W. Law, A.C. Chung, Dominant local binary patterns for texture classification, *IEEE Trans. Image Process.* 18 (5) (2009) 1107–1118.
- [12] L. Liu, Y. Long, P.W. Fieguth, S. Lao, G. Zhao, Brint: binary rotation invariant and noise tolerant texture classification, *IEEE Trans. Image Process.* 23 (7) (2014) 3071–3084.
- [13] Z. Guo, X. Wang, J. Zhou, J. You, Robust texture image representation by scale selective local binary patterns, *IEEE Trans. Image Process.* 25 (2) (2016) 687–699.
- [14] S.K. Roy, B. Chanda, B.B. Chaudhuri, D.K. Ghosh, S.R. Dubey, A complete dual-cross pattern for unconstrained texture classification, in: 4th Asian Conference on Pattern Recognition (ACPR 2017), Nanjing, China, 2017, pp. 741–746.
- [15] S.K. Roy, B. Chanda, B.B. Chaudhuri, S. Banerjee, D.K. Ghosh, S.R. Dubey, Local directional zigzag pattern: A rotation invariant descriptor for texture classification, *Pattern Recognit. Lett.* 108 (2018) 23–30.
- [16] S.K. Roy, B. Chanda, B.B. Chaudhuri, D.K. Ghosh, S.R. Dubey, Local morphological pattern: A scale space shape descriptor for texture classification, *Digit. Signal Process.* 82 (2018) 152–165.
- [17] S.K. Roy, D.K. Ghosh, R.K. Pal, B.B. Chaudhuri, Affine differential local mean zigzag pattern for texture classification, in: TENCON 2018–2018 IEEE Region 10 Conference, IEEE, 2018, pp. 0488–0493.
- [18] S.K. Roy, S.R. Dubey, B.B. Chaudhuri, Local zigzag max histograms of pooling pattern for texture classification, *Electron. Lett.* 55 (7) (2019) 382–384.
- [19] T. Leung, J. Malik, Representing and recognizing the visual appearance of materials using three-dimensional textons, *Int. J. Comput. Vis.* 43 (1) (2001) 29–44.
- [20] C. Schmid, Constructing models for content-based image retrieval, in: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, Vol. 2, IEEE, 2001, pp. II–II.

- [21] M. Varma, A. Zisserman, A statistical approach to texture classification from single images, *Int. J. Comput. Vis.* 62 (1–2) (2005) 61–81.
- [22] M. Varma, A. Zisserman, A statistical approach to material classification using image patch exemplars, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (11) (2009) 2032–2047.
- [23] M. Varma, R. Garg, Locally invariant fractal features for statistical texture classification, in: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, IEEE, 2007, pp. 1–8.
- [24] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, M. Pietikäinen, From bow to cnn: Two decades of texture representation for texture classification, *Int. J. Comput. Vis.* 127 (1) (2019) 74–109.
- [25] S.K. Roy, G. Krishna, S.R. Dubey, B.B. Chaudhuri, Hybridsn: Exploring 3-d-2-d cnn feature hierarchy for hyperspectral image classification, *IEEE Geosci. Remote Sens. Lett.* (2019).
- [26] S.R. Dubey, S. Chakraborty, S.K. Roy, S. Mukherjee, S.K. Singh, B.B. Chaudhuri, dffgrad: An optimization method for convolutional neural networks, *CoRR abs/1909.11015*.
- [27] M. Cimpoi, S. Maji, I. Kokkinos, A. Vedaldi, Deep filter banks for texture recognition, description, and segmentation, *Int. J. Comput. Vis.* 118 (1) (2016) 65–94.
- [28] J. Bruna, S. Mallat, Invariant scattering convolution networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8) (2013) 1872–1886.
- [29] L. Sifre, S. Mallat, Rotation, scaling and deformation invariant scattering for texture discrimination, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013*, pp. 1233–1240.
- [30] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, Y. Ma, Pcanet: A simple deep learning baseline for image classification?, *IEEE Trans. Image Process.* 24 (12) (2015) 5017–5032.
- [31] L. Liu, J. Chen, G. Zhao, P. Fieguth, X. Chen, M. Pietikäinen, Texture classification in extreme scale variations using ganet, *IEEE Trans. Image Process.* (2019).
- [32] W. Li, E.W. Tramel, S. Prasad, J.E. Fowler, Nearest regularized subspace for hyperspectral classification, *IEEE Trans. Geosci. Remote Sens.* 52 (1) (2014) 477–489.
- [33] T. Lindeberg, Scale-space theory: A basic tool for analyzing structures at different scales, *J. Appl. Stat.* 21 (1–2) (1994) 225–270.
- [34] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* 60 (2) (2004) 91–110.
- [35] J.D. Victor, B.W. Knight, Simultaneously band and space limited functions in two dimensions, and receptive fields of visual neurons, in: *Perspectives and Problems in Nonlinear Science*, Springer, 1994, pp. 375–419.
- [36] R.A. Young, R.M. Lesperance, W.W. Meyer, The gaussian derivative model for spatial-temporal vision: i. cortical model, *Spatial Vis.* 14 (3) (2001) 261–319.
- [37] J.-B. Martens, The hermite transform-theory, *IEEE Trans. Acoust. Speech Signal Process.* 38 (9) (1990) 1595–1606.
- [38] L. Florack, B.T.H. Romeny, M. Viergever, J. Koenderink, The Gaussian scale-space paradigm and the multiscale local jet, *Int. J. Comput. Vis.* 18 (1) (1996) 61–75.
- [39] L.D. Griffin, The second order local-image-structure solid, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (8) (2007) 1355–1366.
- [40] B. Markussen, K.S. Pedersen, M. Loog, Second order structure of scale-space measurements, *J. Math. Imaging Vision* 31 (2–3) (2008) 207–220.
- [41] H. Roy, D. Bhattacharjee, Heterogeneous face matching using geometric edge-texture feature (getf) and multiple fuzzy-classifier system, *Appl. Soft Comput.* 46 (2016) 967–979.
- [42] Y. Song, S. Zhang, B. He, Q. Sha, Y. Shen, T. Yan, R. Nian, A. Lendasse, Gaussian derivative models and ensemble extreme learning machine for texture image classification, *Neurocomputing* 277 (2018) 53–64.
- [43] E. Hayman, B. Caputo, M. Fritz, J.-O. Eklundh, On the significance of real-world conditions for material classification, in: *European Conference on Computer Vision*, Springer, 2004, pp. 253–266.
- [44] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover Pubns, New York, NY, USA, 1966.
- [45] K.J. Dana, B. Van Ginneken, S.K. Nayar, J.J. Koenderink, Reflectance and texture of real-world surfaces, *ACM Trans. Graph. (TOG)* 18 (1) (1999) 1–34.
- [46] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, in: *International Workshop on Analysis and Modeling of Faces and Gestures*, Springer, 2007, pp. 168–182.
- [47] Z. Li, G. Liu, Y. Yang, J. You, Scale-and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift, *IEEE Trans. Image Process.* 21 (4) (2012) 2130–2140.
- [48] X. Qi, R. Xiao, C.-G. Li, Y. Qiao, J. Guo, X. Tang, Pairwise rotation invariant co-occurrence local binary pattern, *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (11) (2014) 2199–2213.
- [49] R. Nosaka, Y. Ohkawa, K. Fukui, Feature extraction based on co-occurrence of adjacent local binary patterns, in: *Pacific-Rim Symposium on Image and Video Technology*, Springer, 2011, pp. 82–91.
- [50] T. Ahonen, J. Matas, C. He, M. Pietikäinen, Rotation invariant image description with local binary pattern histogram fourier features, *Image Anal.* (2009) 61–70.
- [51] G. Zhao, T. Ahonen, J. Matas, M. Pietikainen, Rotation-invariant image and video description with local binary pattern features, *IEEE Trans. Image Process.* 21 (4) (2012) 1465–1477.
- [52] S. Lazebnik, C. Schmid, J. Ponce, A sparse texture representation using local affine regions, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (8) (2005) 1265–1278.
- [53] J. Zhang, M. Marszałek, S. Lazebnik, C. Schmid, Local features and kernels for classification of texture and object categories: A comprehensive study, *Int. J. Comput. Vis.* 73 (2) (2007) 213–238.
- [54] L. Liu, P. Fieguth, Texture classification from random features, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (3) (2012) 574–586.
- [55] B. Caputo, E. Hayman, M. Fritz, J.-O. Eklundh, Classifying materials in the real world, *Image Vis. Comput.* 28 (1) (2010) 150–163.
- [56] M. Crosier, L.D. Griffin, Using basic image features for texture classification, *Int. J. Comput. Vis.* 88 (3) (2010) 447–460.
- [57] Y. Xu, H. Ji, C. Fermüller, Viewpoint invariant texture description using fractal analysis, *Int. J. Comput. Vis.* 83 (1) (2009) 85–100.
- [58] Y. Quan, Y. Xu, Y. Sun, A distinct and compact texture descriptor, *Image Vis. Comput.* 32 (4) (2014) 250–259.
- [59] F. Anscombe, The validity of comparative experiments, *J. R. Stat. Soc. A (Gen.)* 111 (3) (1948) 181–211.
- [60] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Advances in Neural Information Processing Systems, 2012*, pp. 1097–1105.
- [61] M. Cimpoi, S. Maji, A. Vedaldi, Deep filter banks for texture recognition and segmentation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015*, pp. 3828–3836.