COMPUTER METHODS IN MATERIALS SCIENCE

2020, vol. 20, no. 3, 95–106
https://doi.org/10.7494/cmms.2020.3.0724

A hybrid statistical approach for texture images classification based on scale invariant features and mixture gamma distribution

Said Benlakhdar*[1], Mohammed Rziza[1], Rachid Oulad Haj Thami[2]

[1] LRIT URAC 29, Faculty of Sciences, Mohammed V University in Rabat, Morocco.
[2] RIITM, ENSIAS, Mohammed V University in Rabat, Morocco.

Abstract

Image classification refers to an important process in computer vision. The purpose of this paper is to propose a novel approach named GGD-GMM and based on statistical modeling in the wavelet domain to describe textured images and rely on a number of principles that give its internal coherence and originality. Firstly, we propose a robust algorithm based on the combination of the wavelet transform and Scale Invariant Feature Transform. Secondly, we implement the aforementioned algorithm and fit the result using the finite mixture gamma distribution (GMM). The results, obtained for two benchmark datasets show that the proposed algorithm has a good relevance as it provides higher classification accuracy than some other well-known models (Kohavi, 1995). Moreover, it shows other advantages relied upon Noise-resistant and rotation invariant.

Keywords: statistical image modeling, SIFT, mixture gamma distribution, uniform discrete curvelet transform, classification.

1. Introduction

Statistical modeling provides a flexible and rich methodological framework which allows to solve various imaging problems using a variety of that allows solving various imaging problems using various tools. The techniques of classification allow capturing morphological properties, properties related to color, the texture of images, etc. For example, Sutton and Hall (1972) lead texture analysis of X-ray images by using the classification technique of pulmonary diseases. Chen et al. (1989) employ fractal texture analysis to classify ultrasound images of the liver. As for the diagnosis of bone diseases, particularly osteoporosis, some authors lead texture analysis on bone radiographs to discriminate between osteoporotic patients and controls.

Most of the earliest image processing analyses focus only on the magnitude of the wavelet describing the image (the real part of the complex representation). Nevertheless, several recent studies analyze, in addition to the magnitude of the wavelet, the phase that contains more information about the features of the image. Oppenheim and Lim (1981) is considered as one of the earliest work that begins to include the phase in their analysis.

Various approaches are recently developed for image processing analysis, particularly for analyzing the phase in the wavelet decomposition of the image, such as the generalized Gaussian density (GGD) (Oppenheim & Lim, 1981). The phase's estimation and its fitting involve the use of the standard circular distributions, where wrapped Cauchy (WC) and Vonn are considered the two most popular ones (see Mallat, 1999). Regarding the

*Corresponding author: said-benlakhdar@um5s.net.ma
ORCID ID's: 0000-0003-0558-2428 (S. Benlakhdar), 0000-0003-4974-5903 (M. Rziza), 0000-0001-9736-7260 (R.O.H. Thami)

wrapped Cauchy distribution, it is more accurate while it is not good for relative phase pdfs with Gaussian shapes (e.g., Moulin & Liu, 1999). In addition, the Vonn distribution fits well with behaviors of relative phases from various real images, including texture images.

Most current research is based on the assumption that certain invariant characteristics are common to an entire class of objects. Most classification methods characterize objects by their global appearance, usually of the entire image. These methods are not robust to occlusion or variations such as rotation or scale.

Moreover, these methods are only applicable to rigid objects. Local invariant features have become very popular to give a solution to the limitations of these methods in object detection, recognition, and classification.

Scale invariant feature transform is an algorithm that allows corner points automatically with subpixel resolution. When a set of images seems to be similar, for example with regard to scale, orientation, etc. The simple corner detectors are found to be useful (see Vo et al., 2011). However, the later techniques become less effective when the images look different regarding scales and/or orientation. In this situation, SIFT algorithm appears to more effective for the image processing analysis. In fact, this algorithm well locates the points of the image in the spatial and frequency domains, and preserves relative stability of the abstracted point's features concerning the visual angle, noise, affine transformation, and some other distinctive characteristics (see Manickam et al., 2019).

In this work, we demonstrate how SIFT algorithm provides better accuracy, what is fitted by a mixture gamma distribution, at the image description level. We also show how this algorithm can describe the characteristics of a typical image through a small number of parameters. This allows to speed up the processing analysis of the studied image by the algorithm. Moreover, we examine the accuracy of the classification related to our approach and compare it to the accuracy of GGD-Vonn and GGD-WC (see Vo & Oriaintara, 2010; Vo et al., 2011).

The remainder of the paper is organized as follows. Section 2 presents Materials and Methods. Section 3 discusses classification methods. Section 4 presents and discusses the experimental results. Finally, section 5 provides the conclusion and implications of our paper.

2. Materials and methods

2.1. Materials

The steps feature extraction and classification were implemented using the MATLAB® language. The experiments were performed using a notebook (Acer E5-574-592S) Intel Core i5-6200U, and 6 GB RAM.

We conduct the experiment on two sets of texture images, the Brodatz and VisTex databases, and we select 40 image textures from the VisTex databases used by e.g., Do & Vetterli, (2002a, 2002b) for our experiments.

Brodatz is one of the most popular and all-purpose datasets, which includes natural textures provides by Brodatz through photographs scanned after print. It established a texture catalog for the purpose of synthesizing textured images (Brodatz, 1966). This catalog now constitutes a benchmark test base for texture researchers. These are 112 textures (sand, grass, bubbles, water, wood, etc.), each having sixteen different variations with a resolution of 640 × 640 pixel and in 8 bit (256 gray values). To expand the Brodatz database, each image was divided into sixteen 128 × 128 non-overlapping sub-images, thereby forming 1248 texture samples. For each image in the database, the uniform discrete curvelet transform was applied with four scales and six orientations angles per scale (0°, 30°, 60°, 90°, 120° and 150°). These images are conventionally used to validate the different methods of texture analysis (see Fig. 1).

The Vision Texture Database was produced by MIT University; it contains color texture images. Texture images representative of real-world conditions are the primary focus of VisTex. All of the images are in 128 × 128 size and belong to one of 40 classes. Each class has just one sample. To provide enough samples for each class, each texture image should be divided into some non-overlap windows of the same size.

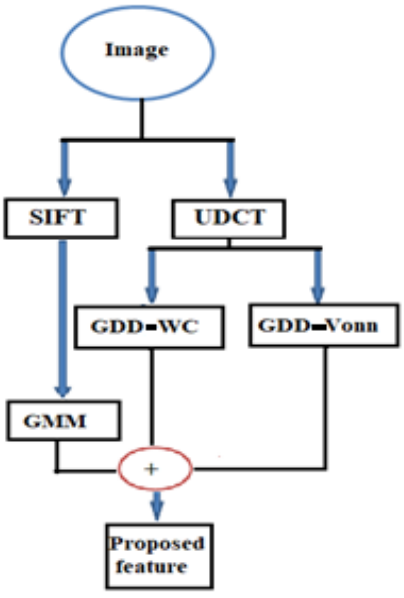

Fig. 1. Scheme of the proposed feature extraction approach (extracted characteristic)

Figure 1 presents the methodological sequence employed in this study. This figure illustrates the steps of texture feature extraction used in the proposed algorithm. The texture feature extraction was applied based on the application of wavelet and ranklet transforms, obtaining the vertical, horizontal, and diagonal subbands. Next, these vectors were classified through the evaluation of different classifiers.

The GGD parameters of the real coefficients in each subband will be estimated as presented by Do and Vetterli (2002a). A feature based on a real part model using the GGD as well as an imaginary part using the (WC) for fitted the relative phase model is named GGD-WC. For GGD-Vonn, the image is analyzed using the same decomposition, except that the finest scale is fitted to the Vonn distribution. In this new approach, the standard vector is based on GGD, and the scale invariant feature transform is fitted with gamma mixture distribution that we will call GGD-GMM.

2.2. Methods

The basic theory of SIFT algorithm, calculation of points of interest and descriptors developed by Lowe in 2004, the implement feature transform is a method to transform an image into a set of feature vectors that are invariant by usual geometrical transformations (rotation, homothety) (see Lowe, 2004). It is used for extracting a distinctive invariant feature from images to serve reliable matching between different views of a scene or an object.

Two main steps are required to implement the Lowe method. Firstly, it is necessary to extract the characteristics of an object and to calculate its descriptors. In other words, it detects the characteristics that are most likely to represent this object, to define and discriminate it by comparing it with others. Secondly, it is necessary to set up a matching procedure. This is the eventual goal of the method (see Fig. 2).

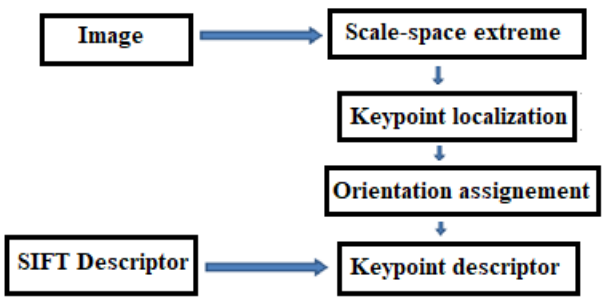

Fig. 2. Proposed scheme for feature extraction approach (see Lowe, 2004)

We will elaborate in the following main steps on how to transform an image into a set of descriptor vectors.

Step 1: scale-space extrema detection
Using a Gaussian difference function, we start with a search on all scales and image locations to identify the potential points of interest that are invariant to scale and orientation. In other words, we can obtain the candidate keypoints by locating the extrema from the Difference of Gaussian (DOG) pyramid.

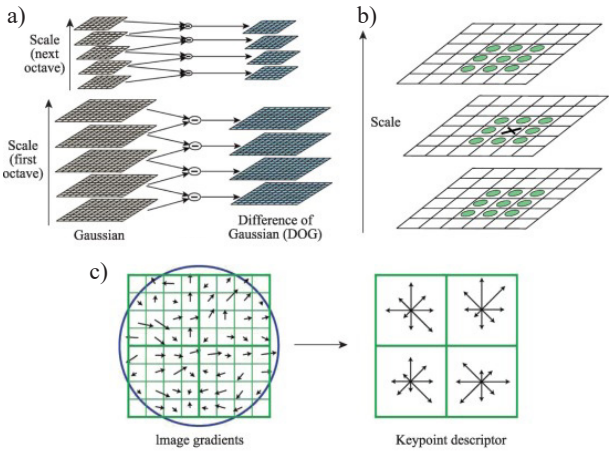

Fig. 3. Construction of SIFT descriptor (see Lowe, 2004): a) image pyramid, b) extrema detection for DOG pyramid, c) creation of keypoint descriptor

Step 2: Keypoint localization
In the interest of obtaining stable keypoints, three processes are applied in this step: By using the 3rd order Taylor polynomial, the first process is done to find the accurate location of keypoints. The second process is focused on eliminating the keypoints with low contrast. In the last process, the keypoints which are in the edge will be eliminated by using the principal curvature.

Step 3: orientation assignment
To each keypoint location, based on local image gradient directions, one or more orientations are assigned (see Figs. 3 and 4).

Step 4: orientation assignment to points of interest
The calculation of the orientation histograms according to the neighborhood is used to justify the invariance of the descriptors with respect to the rotation.

Step 5: calculation of the descriptors
The generation of the descriptor vectors associated with each point of interest requires the calculation of the keypoint descriptor at each point in the window, orientation and gradient magnitude. For each sub-region based on gradient magnitude, an orientation histogram that represents eight cardinal directions are calculated.

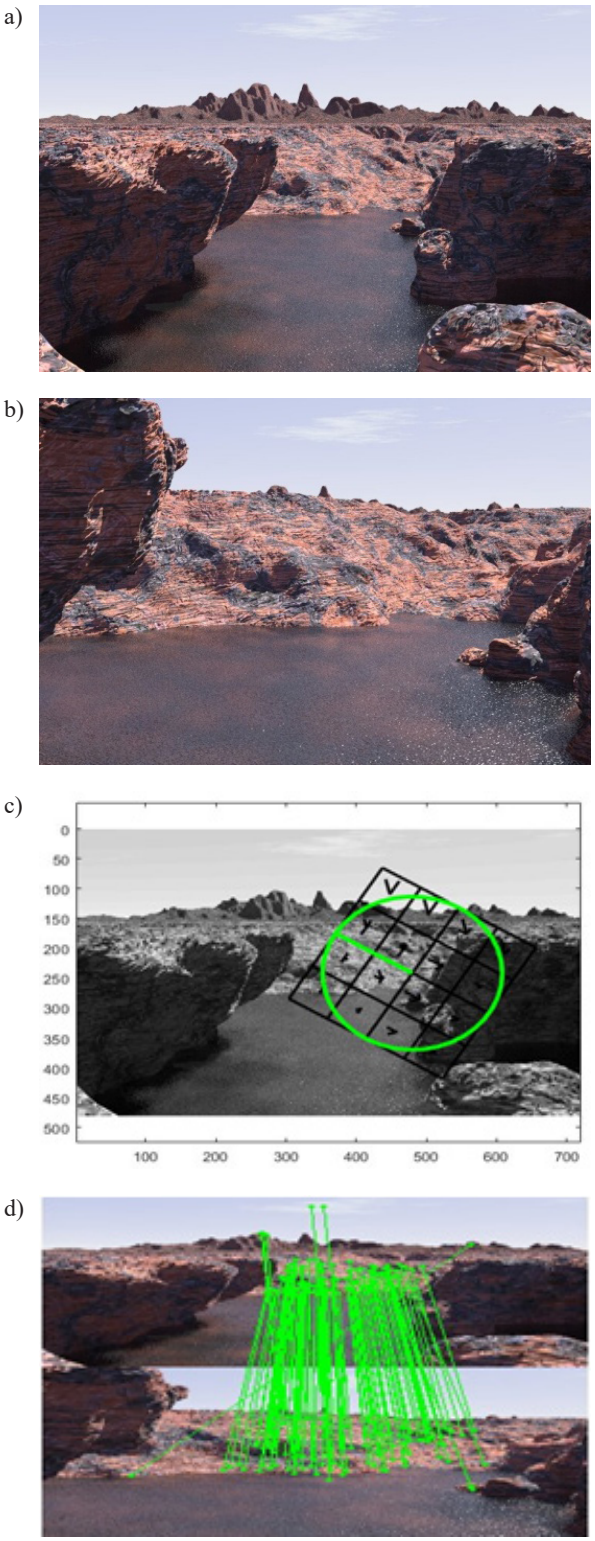

Fig. 4. Landscape image (a), landscape after zoom (b), the representation of dominant direction assignment (c), and correspondences linked with green lines (Matlab implementation of feature extraction and marked matching results) (d)

Four sub-images (bark, bubbles, wood and leather) from Brodatz database are presented in Figure 5.

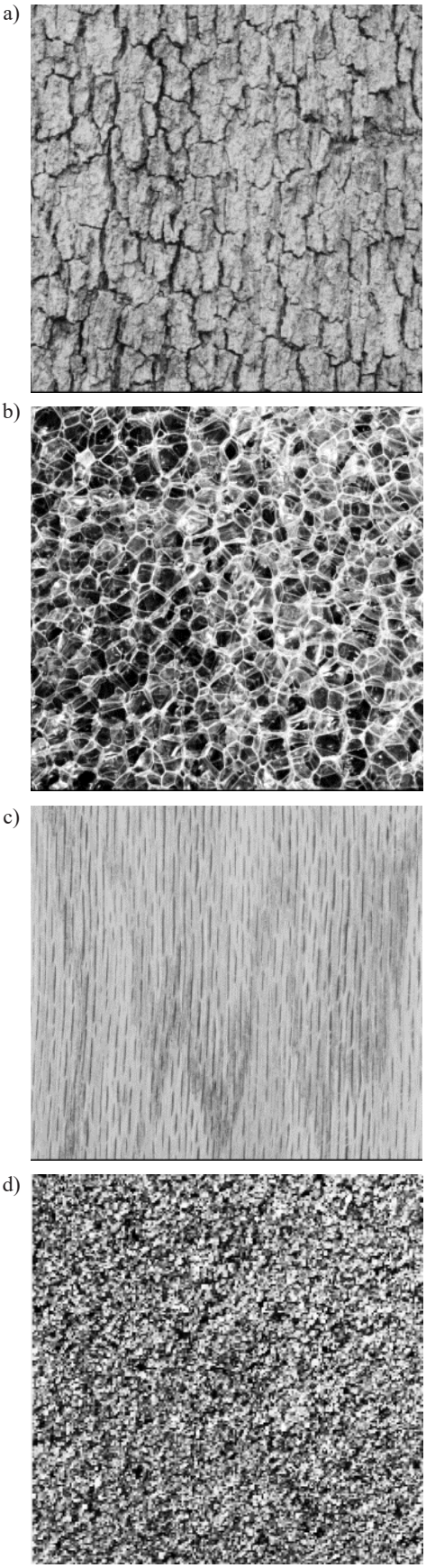

Fig. 5. Four sub-images: a) bark, b) bubbles, c) wood, d) leather with the size of 128 × 128 from the Brodatz database

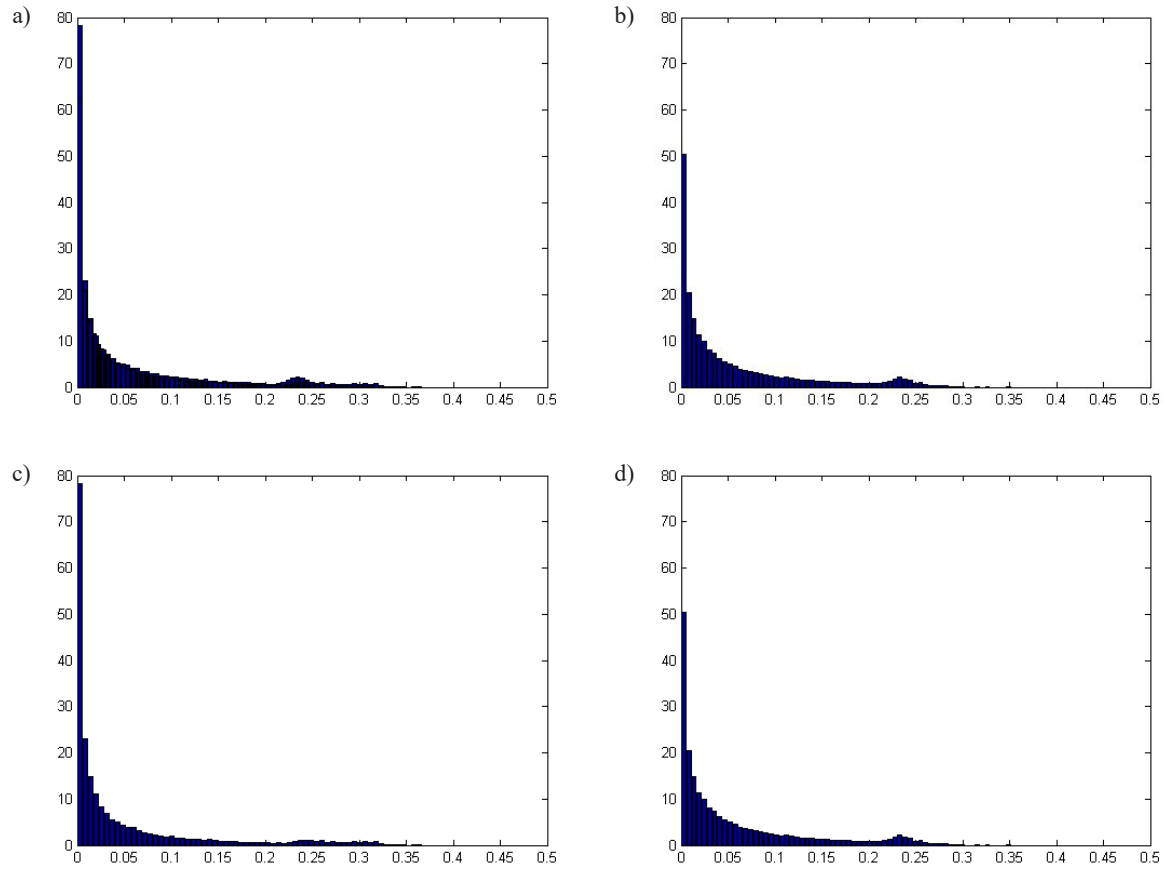

Fig. 6. SIFT descriptors histogram for the textured image: a) bark, b) bubbles, c) wood, d) leather extract from the Brodatz database

Figure 6 shows the histogram of SIFT, which offers four different images extract from the Brodatz database. As can be seen, this distribution has a particular shape, and can be interpreted through a statistical model.

2.3. Gamma model

The gamma distribution is a continuous distribution whose support is the set of strictly positive reals. With the classical parameterization, this distribution denoted $f(\alpha, \beta)$ admits for measurement:

$$f(\alpha,\beta)d(x) = \frac{\exp^{\left(-\frac{x}{\beta}\right)}}{\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1} \tag{1}$$

where the gamma function is defined as $\Gamma(\alpha)$ and two positive parameters, α and β for shape and scale, respectively.

Note that this law is sometimes parameterized not by the parameter but according to its inverse. In this paper, the classical parameterization will always correspond to the parameterization by the pair (α, β) introduced in equation (1). When the parameter β is a parametric set, the families are a natural exponential family.

2.4. Mixture gamma distribution

In this section, a probabilistic formalization is proposed to resolve the problem based on a special case of a mixed model.

Let $\chi = \{x_j\}$, $j = 1, \ldots, N$ be a set of samples, the density function for the finite mixture gamma distributions takes the following form:

$$f(x;\alpha,\beta) = \sum_{i=1}^{k} \pi_i f_i(x;\alpha_i,\beta_i) \tag{2}$$

$$f_i(x;\alpha,\beta) = \frac{x^{(\alpha_i-1)}\exp-\left(\frac{x}{\beta_i}\right)}{\beta_i^{\alpha_i}\Gamma(\alpha_i)} \tag{3}$$

where: $x > 0$, $\alpha_i > 0$, $\beta_i > 0$, $i = 1, 2, \ldots, k$.

Here, k denotes the number of components in the mixture. $\pi_1, \pi_2, \ldots, \pi_k$ are the proportions that satisfy the conditions $0 < \pi_i < 1$, $\sum_{i=1}^{k} \pi_i = 1, \forall i = 1, 2, \ldots, k$. α_i denote the shape of the i-th component of the mixture distribution and β_i their scale parameters. Where $\Gamma(\alpha_i)$ is the gamma function defined as $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \exp(-x)dx$, for $x > 0$.

The EM (Expectation-Maximization) algorithm will allow us to find the parameters of this mixture gamma distribution, starting from random values and adjusting them progressively until the likelihood of this model is maximum.

2.5. EM algorithm

The choice of standard algorithms comes from the fact that we are looking for performance and relevant algorithms, thus the most popular, which helps us to estimate the parameters of proposed models. Expectation-Maximization algorithm is a general method for finding the estimated maximum likelihood of a given set of parameters of a distribution from a sample. Using the general representation of log-likelihood function given in McLachlan and Peel, the finite mixture gamma model is given as follows:

$$L_n(X,\theta) = \sum_{t=1}^{n} \log\left(\sum_{i=1}^{M} \pi_i f(x_t, \alpha_i, \lambda_i)\right) \tag{4}$$

Where $X = \{x_1, x_2, \ldots, x_t\}$ denotes the set of N observation vectors available to estimate the parameters of the model and θ regroups the parameters that be estimated. The quantities $\pi_1, \ldots, \pi_m$ are called the mixing proportions or weights. The $f(x_t, \alpha_i, \lambda_i)$ is the density component of the mixture. It is assumed that the number of components M is known so that only the vector $\theta = (\pi_1, \ldots, \pi_m, (\alpha_1, \lambda_1), \ldots, (\alpha_n, \lambda_n))$ must be estimated.

$$L_n(X,\theta) = \sum_{t=1}^{n} \log\left(\sum_{i=1}^{M} \pi_i \left(\frac{x_j^{(\alpha_i - 1)} \exp - \left(\frac{x_j}{\beta_i}\right)}{\beta_i^{\alpha_i} \Gamma(\alpha_i)}\right)\right) \tag{5}$$

The space parameter is denoted by Θ, i.e.:

$$\Theta = \{(\theta_1, \theta_2, \ldots, \theta_n)\} : \theta_i = (\pi_i, \alpha_i, \lambda_i),\ 1 \le i \le M\}$$

The EM algorithm is used to estimate the gamma mixture parameters in the following manner.

Let s_t, $t \ge 1$ be a sequence of i.i.d. random variables with distribution $P(s_t = i) = \pi_i$. We can associate $(x_1, x_2, \ldots, x_n)$ with $(s_1, s_2, \ldots, s_n)$ as follows: assuming $s_t = i$, x_t has a gamma distribution with parameters (α_i, λ_i). We call $(x_1, r_1, x_2, \ldots, x_n, s_n)$ the augmented data, its likelihood is given by:

$$L_n^c(X,r) = \prod_{t=1}^{n} \pi_{s_t} f(x_t, \alpha_{s_t}, \lambda_{s_t}) \tag{6}$$

Instead of finding the optimal likelihood estimation, the EM algorithm optimizes the conditional logarithmic likelihood:

$$\max_{\theta \in \Theta} Q(X, \theta, \theta^{(k)}) := \mathrm{E}[\log(\log L(x; \alpha, \beta) \mid \theta^{(k)}, X) \tag{7}$$

That is:

$$\theta^{(k+1)} = \mathrm{argmax}_{\theta}\, Q(X, \theta, \theta^{(k)}) \tag{8}$$

The computation gives:

$$Q(X,\theta,\theta^{(k)}) = \sum_{t=1}^{n}\sum_{i=1}^{M}[\log(\pi_i) + \log(f(x_t, \alpha_t, \lambda_t))]p(i \mid x_t, \theta^{(k)}) \tag{9}$$

where:

$$p(i \mid x_t, \theta^{(k)}) = \frac{\pi_i^{(k)} f(x_t, \alpha_i^{(k)}, \lambda_i^{(k)})}{\sum_{j=1}^{M} \pi_j^{(k)} f(x_t, \alpha_j^{(k)}, \lambda_j^{(k)})} \tag{10}$$

Solving $\dfrac{\partial Q(X,\theta,\theta^{(k)})}{\partial \theta} = 0$, yield:

$$\pi_i^{(k+1)} = \frac{1}{n}\sum_{t=1}^{n} p(i \mid x_t, \theta^{(k)}) \tag{11}$$

$$\lambda_i^{(k+1)} = \frac{\alpha_i^{(k)} \sum_{t=1}^{n} x_t p(i \mid x_t, \theta^{(k)})}{\sum_{t=1}^{n} x_t p(i \mid x_t, \theta^{(k)})} \tag{12}$$

Next, we proceed to update α. Note that:

$$\frac{\partial Q(X,\theta,\theta^{(k)})}{\partial \alpha_i} = \sum_{t=1}^{n} \frac{\partial f(x_t, \alpha_i, \lambda_i)}{\partial \alpha_i} p(i \mid x_t, \theta^{(k)}) = \sum_{t=1}^{n} [\log(x_t) + \log(\lambda_i) - \varphi(\alpha_i)] p(i \mid x_t, \theta^{(k)}) \tag{13}$$

where $\varphi(\alpha_i) = \dfrac{\partial \log(\Gamma(x))}{\partial(x)}$.

Since $\partial Q(X,\theta,\theta^{(k)})/\partial\alpha_i = 0$ has no closed-form expression, we do not have the optimal updating scheme for α aviable. So, we can update α_i in its gradient direction:

$$\alpha_i^{(k+1)} = \alpha_i^{(k)} + \alpha_k G_{\alpha_i}(X, \theta^{(k)}) \tag{14}$$

where is a step size that will be specified later and:

$$G_{\alpha_i}(X,\theta^{(k)}) = \left.\frac{\partial Q(X,\theta,\theta^{(k)})}{n\partial\alpha_i}\right|_{\theta=\theta^{(k)}} \tag{15}$$

$$\frac{1}{n}\sum_{t=1}^{n}[\log(x_t) + \log(\lambda_i^{(k)}) - \varphi(\alpha_i^{(k)})]p(i \mid x_t, \theta^{(k)}) \tag{16}$$

where x_t has a gamma distribution with parameters (α_i, λ_i), α_i is the shape of the i-th component of the mixture distribution, with λ is the rate parameter.

2.6. Uniform discrete curvelet transform

The complex wavelet transform has main advantages compared with the discrete wavelet transform (DWT), like good directional selectivity and the shift invariant property (e.g., Meeker et al., 1998; Selesnick et al., 2005).

The transform is named the uniform discrete curvelet transform (UDCT) (see Nguyen & Orraintara, 2008), this is due to the positioning on a uniform lattice at each resolution the centers of the curvelet functions. At each resolution, the UDCT basis functions are located on a uniform integer grid. In general, the UDCT can have $3 \times 2n$ directional subbands, where $n \geq 0$. Compared with the existing transforms, the new discrete transform has several advantages, such as ease implementation, hierarchical data structure, and lower redundancy ratio. For more information on the detailed construction of the UDCT, the reader is referred to (Candès et al., 2006; Nguyen & Chauris, 2008).

2.7. Wrapped Cauchy distribution

The wrapped Cauchy is a unimodal and symmetric distribution obtained from a wrapping of the Cauchy distribution with density around the unit circle. The distribution (WC) closely resembles a von Mises distribution for many values of ρ (see Mardia & Jupp, 2009; Jammalamadaka & SenGupta, 2001), and it has the probability density function defined by:

$$p(\theta) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)},\ -\pi \leq \theta \leq \pi \tag{17}$$

where $\rho = e^{-\sigma}$, σ is the scale parameter, $-\pi \leq \mu \leq \pi$ is the location parameter and $0 \leq \rho < 1$ is the scale parameter.

When, $\rho \to 0$, the wrapped Cauchy distribution tends towards the uniform distribution.

2.8. Vonn distribution

Vonn distribution of relative phases at a spatial location (i, j) is defined as the difference between the phase of two adjacent complex wavelet coefficients (see Vo & Oraintara, 2010):

$$\theta(i,j) = Lz(i,j) - Lz(i,j+1) \tag{18}$$

where $z(i,j)$ is the coefficient at position (i, j).

It is noted that to treat the circularity of the phases for complex coefficient z, Lz it is necessary to returns the angle of phase in radians. The angles lie between $\pm\pi$.

The Vonn density distribution of relative phase θ is defined by:

$$p(\theta) = \frac{1-\lambda^2}{2\pi(1-c^2)}\left[1 - \frac{c\cos^{-1}(c)}{\sqrt{1-c^2}}\right] \tag{19}$$

where $c = \lambda\cos(\theta - \mu + \pi)$, $-\pi \leq \theta, \mu \leq \pi$, and $0 \leq \lambda \leq 1$.

The Vonn distribution is unimodal and symmetrical about $\theta = \mu$ (is the mean direction and λ is the correlation parameter). The Vonn distribution parameters can be estimated by using the maximum-likelihood estimator (ML).

Let $\theta_1, \theta_2, \ldots, \theta_n$ be a set of observations from a Vonn distribution and (μ, λ) are the two parameters, with θ_1, θ_2, …, θ_n are i.i.d. This likelihood function is given by:

$$L(\mu,\lambda \mid \theta_1,\theta_2,\ldots,\theta_n) = \prod_{i=1}^{n} p(\theta_i;\ \mu,\lambda) \tag{20}$$

where μ and λ are parameters to be estimated as follows:

$$[\hat{\mu}\hat{\lambda}] = \arg\max_{[\mu\lambda]}\sum_{i=1}^{n}\log p(\theta_i;\mu,\lambda) \tag{21}$$

Differentiating log (L) and equating to zero, we obtain the likelihood equations:

$$f(\mu) = \frac{\partial \log L(\mu,\lambda \mid \theta_1,\theta_2,\ldots,\theta_n)}{\partial\mu} = 0 \tag{22}$$

$$g(\lambda) = \frac{\partial \log L(\mu,\lambda \mid \theta_1,\theta_2,\ldots,\theta_n)}{\partial\lambda} = 0 \tag{23}$$

These equations can be solved numerically to find the parameters μ and λ. Nevertheless, μ can also be estimated by the mean direction:

$$\hat{\mu} = \arctan\left(\frac{\sum_{i=1}^{n}\sin(\theta_i)}{\sum_{i=1}^{n}\cos(\theta_i)}\right) \tag{24}$$

To oversimplify the estimation problem, we propound to estimate μ mean using direction and the Newton Raphson iterative method to find a solution of the equation $g(\lambda) = 0$ with $\mu = \hat{\mu}$.

The Newton iteration can be stated as:

$$\lambda^{[k+1]} = \lambda^{[k]} - \frac{g(\lambda^{[k]})}{g'(\lambda^{[k]})} \tag{25}$$

We derive $g(\lambda)$ and $g'(\lambda)$ (see Vo et al., 2011). They are given by:

$$g(\lambda) = \sum_{i=1}^{n} h(\lambda,-1) + h(\lambda,+1) - 1.5[h(\lambda,-x_i) + h(\lambda,x_i)] + \frac{b'(\lambda,x_i)}{b(\lambda,x_i)} \tag{26}$$

$$g'(\lambda) = \sum_{i=1}^{n} -[h^2(\lambda,-1) + h^2(\lambda,+1) + 1.5[h^2(\lambda,-x_i) + h^2(\lambda,x_i)] + \frac{b'\ddot{b} - b'^2}{b^2(\lambda,x_i)} \tag{27}$$

where:

$$x_i = \cos(\theta_i - \hat{\mu} + \pi), \quad h(\lambda,x_i) = \frac{x_i}{1+\lambda x_i}$$

$$b(\lambda,x_i) = \sqrt{1-\lambda^2 x_i^2} - \lambda x_i \cos^{-1}(\lambda x_i)$$

$$b'(\lambda,x_i) = -x_i \cos^{-1}(\lambda x_i), \quad \ddot{b}(\lambda,x_i) = \frac{x_i^2}{\sqrt{1-\lambda^2 x_i^2}}$$

We propose using the correlation coefficient as a good initial value for the root of $g(\lambda)$ as follows:

$$\lambda^{[0]} = \frac{|\Psi_{12}|}{\sqrt{\Psi_{11}\Psi_{12}}} \tag{28}$$

where the covariance of complex wavelet coefficients in a subband is:

$$C_z = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix} \tag{29}$$

where Ψ_{11} is the covariance.

With the initial value as in (7), our ML estimator converges with a few iterations.

3. Statistical classifiers

3.1. Support vector machines

In this section, we present the method of support vector machines (SVM), developed by Vapnik (1998). It is a method of classification by supervised learning that seeks to find the optimal function called hyperplane at the maximum margin, which allows optimal separation between data belonging to two different classes (Vapnik, 1998; Zhang et al., 2004).

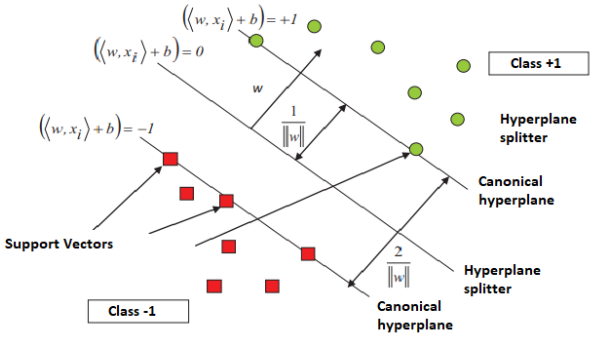

Fig. 7. The separator hyperplane for two-dimensional data (see Huang et al., 2002)

Let $X = V \times Y$ be the space of labeled instances and $D = \{x_1, x_2, \ldots, x_n\}$ be a dataset consisting of n labeled instances, where $x_i = \langle v_i \in V, y_i \in Y \rangle$.

The accuracy of a classier C is the probability of correctly classifying a randomly selected instance, i.e., acc = $Pr(C(v) = y)$ for a randomly selected instance $\langle v, y \rangle \in \chi$, where the probability distribution over the instance space is the same as the distribution that was used to select instances for the inducer's training set.

Once we obtain the support vectors, we try to classify the training samples and also test samples where we know the class labels. From the know class labels, we can compute the true positive, false positive, and false negatives. That will give the accuracy of the classification from the obtained Support vectors.

3.2. *k*-nearest neighbors (KNN)

The KNN algorithm is one of the simplest machine learning algorithms. In the context of classifying a new observation x, the simple founding idea is to have the

closest neighbors of this observation vote. The class of x is determined based on the majority class among the k-nearest neighbors of observation x.

The KNN method is therefore a neighborhood-based, non-parametric method by this means that the algorithm makes it possible to make a classification without making an assumption on the function $y = f(x_1, x_2, \ldots, x_p)$, which relates the dependent variable to the independent variables. The flexibility and efficiency of this method have motivated several authors to introduce this approach into functional statistics. Among the pioneering works in this theme, we cite Burba et al. (2009), Karaa et al. (2017), Kudraszow and Vieu (2013), and Lian (2011)

$$\sum_{r=1}^{c} k_r = k \tag{30}$$

Thus a new observation is predicted in class l with:

$$l = \max_r(k_r) \tag{31}$$

This prevents the predicted class from being determined only from a single observation. The degree of locality of this technique is determined by the parameter k: for $k = 1$, we use the method of the only nearest neighbor as maximum local technique, for $l = k/n$ we use the majority class on the integral set of observations (this implying a constant prediction for each new observation to be classified) (see Deng et al., 2016).

3.3. Cross validation

To minimize the influence of the choice of partitioning of all the examples, cross-validation subdivides the initial training set into k disjoint subsets, $D_1, D_2, D_3, \ldots, D_k$ of the same size. The training and the test are done k times. At iteration i, the subset D_i is reserved for the test, and the rest of the examples are used to train the model. The final model accuracy is equal to the average of the k test accuracies.

The Leave-One-Out method (see Hamel, 2011) is a special case of cross-validation where $k = N$. At each iteration, the model is trained on $N - 1$ examples and tested on the example excluded from the training. In the end, N precisions are obtained, the precision of the model is equal to their average.

3.4. Convolutional neural networks

Convolutional neural networks are directly inspired by the visual cortex of vertebrates. A convolutional neural network also called ConvNet (for "Convolutional Network").

CNN can be divided into two parts. The first part, which we call the convolutional part of the model, and the second part, which we will call the classification part of the model, which corresponds to an MLP (Multi Layers Perceptron) model. It is a multi-layered neural network, and more precisely, it is a deep network composed of multiple layers which are generally organized in blocks (CONV → RELU → POOL).There are four types of layers for a convolutional neural network: the convolutional layer, the pooling layer, the ReLU correction layer, and the fully-connected layer.

To create such compact representations, deep networks typically adjust several hundred thousand artificial synapses to achieve their learning goal, a training process dominated by trial and error in the design phase and many hours, even days of actual training (see Khoshnood et al., 2017).

3.5. ROC curve and AUC

The ROC curve (receiver operating characteristic) was initially developed in the 1950s for military purposes (use of Radar data). Its interest in the medical field was underlined in 1960 by Lee Lusted (1960). Since then, this statistical tool has been used in particular in the pharmaceutical field (e.g., Lusted, 1970; Au, 2004), in radiology (Landais et al., 1994), and in biology. The ROC curve is a graphic representation of the relationship between the sensitivity and the specificity of a test, calculated for all possible threshold values. The closer the ROC curve is to the upper left corner of the graph, the more efficient is the test since the true positive rate approaches 1 ($Se = 1$), and the false positive rate tends towards 0 ($Sp = 1$). The upper left point which is furthest from the bisecting diagonal and usually corresponds to the best threshold value (Tipples, 2002).

The area under the ROC curve (AUC) is one of the most widely used overall test performance measures. It varies between 0.5 in the case of a non-informative test to 1 in the case of perfect performance. So, an AUC of 0.50 means the test is bad. The higher the area under the curve, the better is the test. When several tests are studied simultaneously, their areas under the curve can also be used to compare their performance (DeLong et al., 1988). However, these comparisons are made tricky when the ROC curves intersect. Local comparisons of the sensitivity and specificity of the different tests make it possible to refine the comparative study (Delacour et al., 2005). Other indices, such as likelihood ratios (positive or negative) or expected information capacity (CIA), are also used in addition to the overall measurement.

4. Experimental results and discussion

The experimental process began with testing how well the mixture gamma model fit phase features derived from the SIFT transform. The dataset available through the Brodatz and VisTex Databases was used for the evaluation of the accuracy and precision of the proposed approach. We compare the proposed method with GGD-Vonn and GGD-WC features using the UDCT curvelet transform.

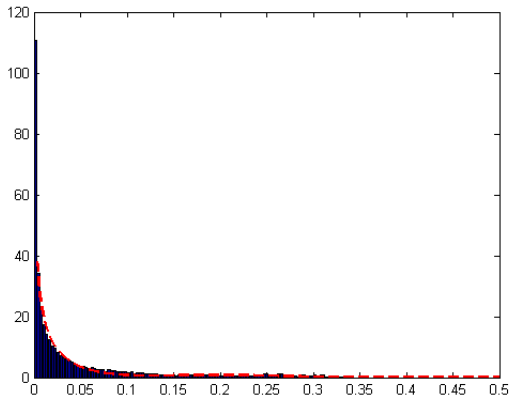

Fig. 8. Wood (128 × 128), mixture gamma distributions fitted the histogram (using 256 bins) $(\pi_1, \pi_2) = (0.87, 0.13)$, $(\alpha_1, \beta_1) = (0.58, 0.04)$, $(\alpha_2, \beta_2) = (11.87, 0.02)$

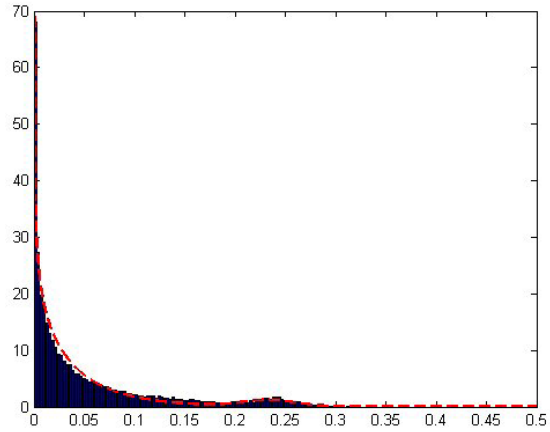

Fig. 9. Leather (128 × 128), mixture gamma distributions fitted the histogram (using 256 bins) $(\pi_1, \pi_2) = (0.94, 0.06)$, $(\alpha_1, \beta_1) = (0.47, 0.07)$, $(\alpha_2, \beta_2) = (61.62, 0.004)$

Figures 8 and 9 show the histograms of scale invariant feature transform descriptors for two different images. The tested images are wood and lather. Clearly, the proposed mixture gamma model fits the data well. In addition, the estimated parameters are different for both images. This suggests the use of these parameters for discriminating the different database images.

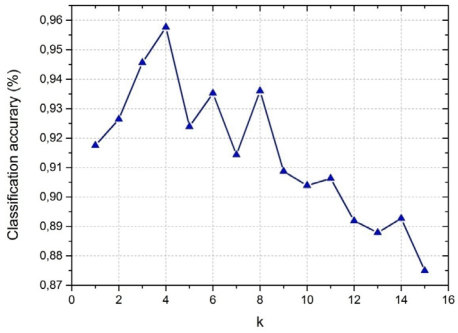

Fig. 10. KNN classification results, according to the values of k (number of nearest neighbors), using the proposed GGD-GMM

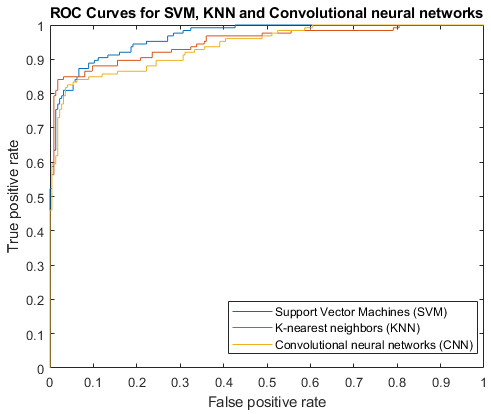

Fig. 11. ROC curves obtained using the different transforms for the classification

The Figure10 shows that the best classification accuracy for the k-nearest neighbors classification algorithm is obtained forthe value $k = 4$.

We compare our approach with three well-known classifiers, i.e., KNN, Neural networks, and SVM. One can see from Table 1 that the proposed approach GGD-

Table 1. Classification accuracy achieved with state of art feature and ours

Rate of accurately	Neural networks		SVM		KNN	
			SVM-CrossVal		KNN-KFold	
Database	Brodatz	VisTex	Brodatz	VisTex	Brodatz	VisTex
GGD-WC	90.54	88.82	88.07	85.64	87.12	85.03
GGD-Vonn	91.89	92.79	91.15	85.82	89.85	86.19
GGD-GMM	90.99	**94.59**	**96.03**	**93.16**	**95.83**	**92.47**

GMM outperforms the state-of-the-art methods in both databases, Brodatz and VisTex. Besides, in the experiments, the proposed features GGD-GMM increase significantly the overall accuracy rate up to 96.03%. On the other hand, GGD-WC and GGD-Vonn achieve an accuracy rate of 88.07% and 91.15% for the SVM classification algorithm, respectively.

In terms of AUC rate, according to the ROC curves shown in Figure 11, the proposed approach achieves the highest rate of AUC using the SVM classifier, which is consistent with the results presented in Table 1.

5. Conclusion and implications

In this study, we introduced a SIFT algorithm fitted by the GMM (GGD-GMM) to describe the characteristics of texture images by only a small number of parameters. This allows to speed up the processing image analysis. By using Brodatz and VisTex dataset, the experiments showed that our algorithm provides more accurate classification results than GGD-Vonn and GGD-WC methods. We, above all, sought to set up a generic approach so as not to be dependent on the content of the image. This system is based on local extraction and rich characterization of image information and a classification approach by the statistical classifiers.

Ethical statements

I declare Said Benlakhdar that I have no conflict of interest.

I declare Mohammed Rziza that I have no conflict of interest.

I declare Rachid Oulad Haj Thami that I have no conflict of interest.

References

Au, Y.H.J., Eissa, S., & Jones, B.E. (2004). Receiver operating characteristic analysis for the selection of threshold values for detection of capping in powder compression. *Ultrasonics*, 42(1–9), 149–153.

Brodatz, P. (1966). *Textures: a photographic album for artists and designers*. Dover Pub.

Burba, F., Ferraty, F., & Vieu, P. (2009). k-Nearest Neighbour method in functional nonparametric regression. *Journal of Nonparametric Statistics*, 21(4), 453–469.

Candès, E., Demanet, L., Donoho, D., & Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3), 861–899.

Chen, C.C., DaPonte, J.S., & Fox, M.D. (1989). Fractal feature analysis and classification in medical imaging. *IEEE Transactions on Medical Imaging*, 8(2), 133–142.

Delacour, H., Servonnet, A., Perrot, A., Vigezzi, J.F., & Ramirez, J.M. (2005). La courbe ROC (receiver operating characteristic): principes et principales applications en biologie Clinique. *Annales de Biologie Clinique*, 63(2), 145–154.

DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 837–845.

Deng, Z., Zhu, X., Cheng, D., Zong, M., & Zhang, S. (2016). Efficient kNN classification algorithm for big data. *Neurocomputing*, 195, 143–148.

Do, M.N., & Vetterli, M. (2002a). Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models. *IEEE transactions on multimedia*, 4(4), 517–527.

Do, M.N., & Vetterli, M. (2002b). Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, 11(2), 146–158.

Erkel, A.R., van, & Pattynama, P. (1998). Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology, *European Journal of Radiology*, 27(2), 88–94.

Greiner, M., Pfeiffer, D., & Smith, R.D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine*, 45(1–2), 23–41.

Hamel, L.H. (2011). *Knowledge discovery with support machines* (Vol. 3). John Wiley & Sons.

Huang, C., Davis, L.S., & Townshend, J.R.G. (2002). An assessment of support vector machines for land cover classification. *International Journal of remote sensing*, 23(4), 725–749.

Jammalamadaka, S.R., & SenGupta, A. (2001). *Topics in Circular Statistics* (Vol. 5). World Scientific Press.

Karaa, L.Z., Laksaci, A., Rachdi, M., & Vieu, P. (2017). Data-driven kNN estimation in nonparametric functional data analysis. *Journal of Multivariate Analysis*, 153, 176–188.

Khoshnood, B., Lelong, N., Houyel, L., Bonnet, D., Ballon, M., Jouannic, J.M., & Goffinet, F. (2017). Impact of prenatal diagnosis on survival of newborns with four congenital heart defects: a prospective, population-based cohort study in France (the EPICARD Study). *BMJ Open*, 7(11), e018285.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (Vol. 2, pp. 1137–1145).

Kudraszow, N.L., & Vieu, P. (2013). Uniform consistency of *k*NN regressors for functional variables. *Statistics & Probability Letters*, 83(8), 1863–1870.

Landais, P., Besson, C., & Jais, J.P. (1994). Evaluation of the diagnostic contribution of a test. Main information indices, *Journal de Radiologie*, 75(2), 141–50.

Lian, H. (2011). Convergence of functional *k*-nearest neighbor regression estimate with functional responses. *Electronic Journal of Statistics*, 5, 31–40.

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.

Lusted, L.B. (1960). Logical analysis in roentgen diagnosis: memorial fund lecture. *Radiology*, 74(2), 178–193.

Lusted, L.B. (1971). Signal detectability and medical decision-making. *Science*, 171(3977), 1217–1219.

Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.

Manickam, A., Devarasan, E., Manogaran, G., Priyan, M.K., Varatharajan, R., Hsu, C.H., & Krishnamoorthi, R. (2019). Score level based latent fingerprint enhancement and matching using SIFT feature. *Multimedia Tools and Applications*, 78(3), 3065–3085.

Mardia, K.V., & Jupp, P.E. (2009). *Directional statistics* (Vol. 494). John Wiley & Sons.

Meeker, W.Q., Escobar, L.A., & Lu, C.J. (1998). Accelerated degradation tests: modeling and analysis. *Technometrics*, 40(2), 89–99.

Moulin, P., & Liu, J. (1999). Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. *IEEE transactions on Information Theory*, 45(3), 909–919.

Nguyen, T.T., & Oraintara, S. (2008). The shiftable complex directional pyramid-Part II: Implementation and applications. *IEEE Transactions on Signal Processing*, 56(10), 4661–4672.

Oppenheim, A.V., & Lim, J.S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69(5), 529–541.

Peel, D., & McLachlan, G.J. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing*, 10(4), 339–348.

Selesnick, I.W., Baraniuk, R.G., & Kingsbury, N.C. (2005). The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6), 123–151.

Sutton, R.N., & Hall, E.L. (1972). Texture measures for automatic classification of pulmonary disease. *IEEE Transactions on Computers*, 100(7), 667–676.

Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin & Review*, 9(2), 314–318.

Vapnik, V.N. (1998). *Statistical learning theory* (1st ed.), Wiley.

Vo, A., & Oraintara, S. (2010). A study of relative phase in complex wavelet domain: Property, statistics and applications in texture image retrieval and segmentation. *Signal Processing: Image Communication*, 25(1), 28–46.

Vo, A., Oraintara, S., & Nguyen, N. (2011). Vonn distribution of relative phase for statistical image modeling in complex wavelet domain. *Signal Processing*, 91(1), 114–125.

Zhang, L., Zhou, W., & Jiao, L. (2004). Wavelet support vector machine. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1), 34–39.