Visual textures as realizations of multivariate log-Gaussian Cox processes

Huu-Giao Nguyen, Ronan Fablet, Jean-Marc Boucher
Institut Telecom / Telecom Bretagne / LabSTICC
Université européenne de Bretagne
{huu.nguyen;ronan.fablet;jm.boucher}@telecom-bretagne.eu

Abstract

In this paper, we address invariant keypoint-based texture characterization and recognition. Viewing keypoint sets associated with visual textures as realizations of point processes, we investigate probabilistic texture models from multivariate log-Gaussian Cox processes. These models are parameterized by the covariance structure of the spatial patterns. Their implementation initially rely on the construction of a codebook of the visual signatures of keypoints. We discuss invariance properties of the proposed models for texture recognition applications and report a quantitative evaluation for three texture datasets, namely: UIUC, KTH-TIPs and Brodatz. These experiments include a comparison of the performance reached using different methods for keypoint detection and characterization and demonstrate the relevance of the proposed models w.r.t. state-of-the-art methods. We further discuss the main contribution of proposed approach, including the key features of a statistical model and complexity aspects.

1. Introduction

Texture information is among the key features of interest for the robust characterization and recognition of visual scenes. A variety of methods can be found in the literature for texture recognition applications, from the early Haralick cooccurrence features [8], statistics of the response to scale-space filters such as Gabor and wavelet analysis [22] or more recent methods embedding invariance properties such as keypoint-based settings [4, 5, 15], multifractal schemes [27], topographic map [26] or local binary pattern [10] descriptors.

The renewed interest in texture analysis emerged from the application of visual keypoints [1, 3, 16, 24] to reach texture description invariant to geometric and photometric image transforms, e.g. affine transforms, contrast changes. The classical keypoint-based setting consists in stating texture recognition as a voting-based output of the set of keypoints attached to a given visual texture [4, 13]. Such an approach guarantees to inherit the invariance properties of the local keypoints and advanced statistical learning strategy can efficiently be implemented including random forests [2] and SVM [12]. Such an approach only rely on the visual signatures of the keypoints and discard any spatial information in terms of spatial patterns formed by the keypoint sets.

Our previous work has shown that descriptive statistics of spatial point processes provide a relevant basis for jointly characterizing the spatial and visual signatures of the key-point set attached to a visual texture [21]. While no explicit references to spatial point processes and associated descriptive statistics, it can be noted that second-order descriptive statistics of spatial keypoint patterns were previously considered for application to scene categorization [15] and robot navigation [5]. Here we further investigate to which extent visual textures can be viewed as realizations of multivariate spatial point processes, but rather than descriptive statistics, we aim at delivering a specific formal model and explore in this context the relevance of a class of spatial processes, namely log-Gaussian Cox processes. Log Gaussian Cox processes introduced by Møller et al. [20] provide models for the spatial distribution of multivariate point processes. As they relate to a model of the covariance of count variables, they were shown to be easy to analyse and flexible for experiments in spatial statistical analysis, especially in environment-related sciences [20] or disease surveillance [6]. It might be noted that spatial point processes were previously investigated for texture analysis, e.g. Lafarge et al. [7] applied a spatial Gibbs point model and a Jump-Diffusion process for the extraction of geometric features in texture images. Such Gibbs models are however not suited for recognition issues. Overall the specification of the log-Gaussian Cox processes resort to the estimation of the covariance structure of count variables of the multivariate point processes. Simple estimation procedures can be derived for different types of the covariance structures and an invariant texture characterization follows from model parameters. Overall, the main contributions of this paper is three-fold:

• Addressing invariant texture characterization and mod-
Among the most popular descriptors, local keypoints were and photometric transformation of the images [3, 19, 27].

...has been given to visual signatures invariant to geometric feature space defined by local visual descriptors. The focus...generally covariance models for different types of keypoint detection and descriptors.

...Demonstrating the relevance of the proposed models for texture recognition with respect to previous work.

This paper is organized as follows. In Section 2, a brief overview of the proposed approach and related work is given. We present in Section 3 the proposed probabilistic keypoint-based texture model based on multivariate log-Gaussian Cox processes. Comparative evaluations of texture recognition performance are reported for several databases in Section 4. We further discuss the main contributions of proposed approach in Section 5.

2. Proposed approach and related work

The general goals of this paper are the characterization and modeling of visual textures from the spatial patterns formed by visual keypoints using point process models. The initial step then consists in detecting local keypoints in texture images(Fig.1a). As a way to making easier statistical estimation issues, we build a codebook of visual keypoints from their visual signatures using adapted clustering techniques such that any visual keypoint is assigned to a category(Fig.1b). Regarding visual keypoint sets as finite spatial random sets, log-Gaussian Cox models are investigated to reach an invariant texture characterization from the covariance structure of spatial patterns. The state-of-the-art approaches of visual keypoints are detailed below.

When addressing matching and recognition issues in images, the typical approach relies on learning models in the feature space defined by local visual descriptors. The focus has been given to visual signatures invariant to geometric and photometric transformation of the images [3, 19, 27]. Among the most popular descriptors, local keypoints were shown to be particularly efficient [13, 18] compared to the early feature developed for texture analysis such as Gabor features [22] and cooccurrence matrices.[8]

Numerous approaches have been proposed to detect regions or points of interest in images. Among the most popular the Harris detector [9] detects corners, i.e. the points at which significant intensity changes in two directions oc-
cur. It relies on the eigen-decomposition of the structure tensor of the intensity function. Scale-space approaches based on the analysis of the Hessian matrix were also proposed to address scale adaption [14]. Scale-spaces of Difference of Gaussians (DoG) are also widely considered as an approximation of the Laplacian [16]. More recently, Mikolajczyk et al.[19] combined Harris or Hessian detector and the Laplacian operator (for scale adaption) to propose two scale-invariant feature detectors, namely: Harris-Laplace (Har-Lap) and Hessian-Laplace (Hes-Lap). Bay et al.[1] presented the Fast-Hessian (FH) detector based on the Hessian matrix in the integral images. Other categories of keypoint detectors may be cited, for instance the maximally stable extremal region (MSER) detector [17], the edge-based region (EBR) detector, the intensity extrema-based region(IBR) detector [25] or entropy-based region (such as salient regions) detector [11]. Comparisons between the different detectors are given in [1, 13, 19].

Given the pixel coordinates of the extracted keypoints, many different schemes have been proposed to extract a feature vector of each keypoint $s_i$ and invariances to contrast change and geometric transforms, typically affine transforms, are embedded [10, 18, 28]. The SIFT descriptor is certainly among the most popular and relevant ones. It is formed by the distribution of the orientations of the gradient of the intensity in 4x4 windows around the considered point [16]. This description ensures contrast invariance and partial invariance to affine transforms. Orientations are typically quantized over eight values such that the SIFT feature vector is 128-dimensional. Several extensions of the original SIFT descriptor have been proposed including GLOH, PCA-SIFT, RIFT (see [13] for a review).

For instance, an intensity-domain spin image [13] is a 2D histogram encoding the distribution of the intensity value and the distance from the reference point. Rather than considering gradient orientations, the SURF descriptor [1] rely

<table>
<thead>
<tr>
<th>Descriptor size</th>
<th>Keypoint density</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoG+Sift</td>
<td>1382</td>
</tr>
<tr>
<td>FH+Surf</td>
<td>758</td>
</tr>
<tr>
<td>(Har-Lap)+(Sift-Spin)</td>
<td>538</td>
</tr>
<tr>
<td>(Hes-Lap)+Daisy</td>
<td>1216</td>
</tr>
<tr>
<td>FH+Brief</td>
<td>758</td>
</tr>
</tbody>
</table>

Table 1. Number of detected keypoints and size of signature vector of the different detector-descriptor types. The processed texture pattern is the image displayed in Fig.1a.
on the distribution of Haar-wavelet responses whereas the Daisy descriptor [24] exploits responses to oriented Gaussian filters; the Brief descriptor [3] was issued from a relatively small number of intensity of different image patch using binary string.

From these reviews, we investigate five robust detector/descriptor types reported with the best score performance in [1, 3, 16, 24, 28], respectively: FH+Surf, FH+Brief, DoG+Sift, (Hes-Lap)+Daisy and (Har-Lap)+(Sift-Spin). As illustrated in Table 1 for the texture sample displayed in Fig.1a, these different combinations lead to different complexity levels as well as large differences in the number of detected keypoints, a critical aspect when considering keypoint statistics.

3. Multivariate Log Gaussian Cox process

3.1. Multivariate point process and associated descriptive statistics

A spatial point process \( S \) is defined as a locally finite random subset of a given bounded region \( B \subset \mathbb{R}^2 \). A realization of such a process is a spatial point pattern \( s = \{s_1, ..., s_n\} \) of \( n \) points contained in \( B \). Considering a realization of the point process, the moments of random variable are relevant descriptive statistics. In the general case, the \( p^{th} \)-order moment of \( S \) is defined as:

\[
m^{(p)}(B_1 \times ... \times B_p) = E\{N(B_1)...N(B_p)\}
\]

where \( E\{\cdot\} \) denotes the expectation. \( N(B_i) \) is the number of random points contained in a given Borel set \( B_i \). Focusing on intensity measure of \( S \), the first-order moment is evaluated with \( p = 1 \):

\[
m(B) = E\sum_{s \in S} 1_B(s) = \int_B \rho(s) ds
\]

where \( 1_B(s) \) is an indicator function that takes the value 1 when \( s \) falls in region \( B \), \( \rho(s) ds \) is the probability that one point falls in an infinitesimally small area \( ds \) of the neighborhood of point \( s \). The normalized first-order moment \( \lambda = \mu(B)/|B| \) is the mean density of expected points per surface unit, \( |B| \) is the surface of region \( B \). This quantity fully characterizes Poisson point processes. For a homogeneous process, this density is spatially constant.

Beyond the first-order moment, the covariance structure of points of the finite random set, can be characterized by the second-order moment \( \mu^{(2)} \) of \( S \) parameterized as:

\[
\mu^{(2)}(B_1 \times B_2) = E\sum_{s_1 \in B_1} \sum_{s_2 \in B_2} 1_{B_1}(s_1)1_{B_2}(s_2)
\]

\[
= \int_{B_1 \times B_2} \rho^{(2)}(s_1, s_2) ds_1 ds_2
\]

where second-order density \( \rho^{(2)}(s_1, s_2) \) is interpreted as the density, per surface unit, of the pair of points \( s_1 \) and \( s_2 \) in infinitesimally small areas \( ds_1 \) and \( ds_2 \). For a stationary and isotropy point process, this density function \( \rho^{(2)}(s_1, s_2) \) states the correlation of pairs of points and only depends on distance \( ||s_1 - s_2|| \) [23]. In the spatial point fields, the second-order measure \( \mu^{(2)} \) is frequently replaced by the factorial moment measure \( \alpha^{(2)} \) as:

\[
\alpha^{(2)}(B_1 \times B_2) = E\sum_{s_1 \in B_1} \sum_{s_2 \in B_2} 1_{B_1}(s_1)1_{B_2}(s_2)
\]

where the relation between the second-order measure \( \mu^{(2)} \) and the factorial moment measure \( \alpha^{(2)} \) is given by:

\[
\alpha^{(2)}(B_1 \times B_2) = \mu^{(2)}(B_1 \times B_2) - \mu(B_1 \cap B_2)
\]

A multivariate point process \( \Psi \) is defined as a spatial point process for which a discrete mark \( m_i \) is associated to each point \( s_i \) in \( B \). Second-order moment in Eq.4 can be extended to multivariate point patterns. Considering circular study region \( D(\cdot, r) \) with radius \( r \) (Fig.1b), the second-order cooccurrence statistics of \( \Psi \) are characterized by the factorial moment measure as follows.

\[
\alpha^{(2)}_{i,j}(r) = \frac{1}{2} \sum_{h \neq h} \delta(m_h) \delta(m_i m_i) I(||s_h - s_i|| \leq r)
\]

where \( \delta(m_h) \) equals 1 if the mark \( m_h \) of point \( s_h \) is \( i \) and 0 otherwise. For statistical interpretation of second-order moment \( \mu^{(2)} \) [23], Ripley’s K function that is usually used to analyse the mean number of points of type \( j \) located in a study region of radius \( r \) centered at the points of type \( i \) (which itself is excluded) is measured as:

\[
K_{ij}(r) = (\lambda_i \lambda_j)^{-1} \alpha^{(2)}_{i,j}(r)
\]

3.2. Log-Gaussian Cox model

A Cox process \( X \) with random intensity function \( Z \) is a point process such that \( X|Z \) is a Poisson process with intensity function \( Z \) [20, 23]. For an univariate log Gaussian Cox process \( X \) on a locally finite subset \( S \subset \mathbb{R}^2 \), the random intensity function is given by \( Z = \exp(Y) \), where \( Y \) is a Gaussian field on \( S \) characterized by its mean \( \mu = EY(s) \) and covariance function \( c(r) = \text{Cov}(Y(s_1), Y(s_2)) \), where \( r = ||s_1 - s_2|| \) are defined and finite for all bounded \( B \subset S \). An important property of log-Gaussian Cox process is that the characteristics of the Gaussian field \( Y \) relate to the first and second-order moments of the point process. More precisely, the following relations hold [20]:

\[
\rho(s) = \lambda = \exp(\mu + \sigma^2/2)
\]

\[
\rho^{(2)}(s_1, s_2)/(\rho(s_1)\rho(s_2)) = g(r) = \exp(c(r))
\]
are respectively the intensity and the pair correlation function, where \( \sigma^2 = Var(Y(s)) \) is the variance of the Gaussian process. We report an example of the intensity estimation issued from log Gaussian Cox processes in Fig.2.

Extending to a multivariate log-Gaussian Cox process, Cox processes \( \{X_i\} \) are conditionally independent w.r.t. a multivariate intensity field \( Z = \{Z_i\} \) and that \( X_i | Z_i \) is a Poisson process with intensity measure \( \lambda Z_i \). \( Z \) relates to a multivariate Gaussian field \( Y \) as \( Z_i = \exp(Y_i) \). The multivariate Gaussian random field is characterized by its mean \( \mu_i(s) \) and covariance functions \( c_{ij}(r) = \text{Cov}(Y_i(s), Y_j(s)) \). And the intensity and pair correlation function become:

\[
\lambda_i = \exp(\mu_i + \sigma^2/2) \quad g_{ij}(r) = \exp(c_{ij}(r))
\]

Fitting a stationary parametric log-Gaussian Cox process comes to estimating mean and covariance parameters for the associated Gaussian field. Following [20, 23], an estimation procedure relies on the relation between the pair correlation function \( g_{ij} \) and the K-function of Gaussian processes as:

\[
K_{ij}(R) = 2\pi \int_0^R r g_{ij}(r) dr
\]

where \( R \) is a pre-defined value of radius. Combining Eq.8 and Eq.11, the pair correlation function can be estimated as:

\[
g_{ij}(r) = \frac{1}{2\pi \lambda_i \lambda_j \sum_h \sum_{l \neq h} \delta_i(m_h) \delta_j(m_l) \xi(||s_h - s_l||, r) b_{s_h}}
\]

where \( \xi(\cdot) \) is a kernel (here a Gaussian kernel is considered), \( \lambda_i \) is the intensity for class \( i \) estimated from (Eq.2), \( b_{s_h} \) is the proportion of the circumference of the study circle lying within the image. In practice, the computation of the above second-order descriptive statistics take account edge effects [21]. Considering the edge effect correction, \( g_{ij} \) is not symmetric in \( i \) and \( j \). Hence, the non-parametric estimation of the covariance function is defined as:

\[
c_{ij}(r) = \log \left( \frac{\lambda_i g_{ij}(r) + \lambda_j g_{ij}(r)}{\lambda_i + \lambda_j} \right)
\]

Table 2. Different correlation functions of \( \mathbb{L}(\beta, r) \).

<table>
<thead>
<tr>
<th>Correlation function</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>( \exp(-\frac{r}{\beta}) )</td>
</tr>
<tr>
<td>Cardinal sine</td>
<td>( \sin(r/\beta) / (r/\beta) )</td>
</tr>
<tr>
<td>Hyperbolic</td>
<td>( (1 + r/\beta)^{-1} )</td>
</tr>
</tbody>
</table>

To resort to a compact probabilistic model for the representations of visual textures, we investigate parametric forms of the covariance function \( c \). Given a chosen parameterization \( \mathbb{L}(\beta, r) \) in Tab.2, model parameters are estimated from the minimization of the following criterion:

\[
\int_0^R \{ \sigma^2 \mathbb{L}(\beta, r) - c_{ij}(r) \}^2 dr
\]

A gradient-based optimization procedure is applied to solve for this minimization. The proposed probabilistic keypoint-based texture model is eventually given by intensity parameters \( \lambda_i \), variances \( \sigma_{ij} \) and scale parameters \( \beta_{ij} \).

3.3. Feature dimension reduction

Considering the parameters of the log-Gaussian Cox model as descriptors of the spatial patterns of visual keypoints, each texture image is associated with a \( k(k+2) \)-dimensional feature vector, where \( k \) is the size of the codebook of visual words. In practice, the high-dimensional feature may affect recognition performance. State-of-the-art methods based on visual keypoints typically involve dimensionality in the range of \( k \), e.g. bag-of-keypoint [4], (Har+Lap)(Sift+Spin) scheme [28]. Hence dimension reduction issues should be further analyzed.

A dimensional reduction procedure of second-order statistics was introduced in [21] from the determination of categories of keypoint pairs. The codebook of keypoint pairs, denoted by \( u = M(s_i, s_j) \), are issued from an adapted clustering technique applied for each set of two categorized keypoints \( s_i \) and \( s_j \). Therefore, the non-parametric estimation of the covariance function is given by:

\[
c_u(r) = \log \left( \frac{1}{2\pi \lambda_u \sum_h \sum_{l \neq h} \delta_u(M(s_h, s_l) \xi(||s_h - s_l||, r) b_{s_h}}
\]

The estimation of intensity parameter \( \lambda_u \), variances \( \sigma_u \) and scale parameters \( \beta_u \) for each category of keypoint pairs follows as previously from minimization (Eq.14). Overall this procedure results in downsizing the proposed texture descriptor to \( 3k^* \)-dimensional vectors, where \( k^* \) is the size of codebook of keypoint pairs.

3.4. Invariance properties

Invariance properties of the resulting texture characterization are inherited from the characteristics of the chosen
4. Experimental evaluation

Given the textural features defined in the previous section, an application to texture recognition is considered, i.e. an unknown texture sample is assigned to one of a set of known texture classes using a discriminative classifier. The evaluation of the proposed descriptor involves the computation of classification performances for model learning with $N_t$ training texture samples per class. Training images are randomly selected among the $N$ samples available in each class. The remaining $N - N_t$ images are used as test images. The random selection of training samples is repeated 50 times to evaluate the mean and the standard deviation of the correct classification rate. These experimentations are carried out with three texture datasets, namely UIUC, Brodatz, KTH-TIPS databases.

We exploit random forest classifiers [2]. They rely on the construction of an ensemble of classification trees using some form of randomization. A sample is classified by sending it down every tree and aggregating the distribution of classes for the reached leaves. Random forest uses a voting rule to assign a class to an unknown sample.

4.1. Parameter setting

A set of texture features such as Gabor filter [22], co-occurrence matrix [8], local multifractal features [27], s-of-keypoints(BoK) [4], combination scheme of local keypoints [28] and descriptive statistics of visual keypoints [15, 21] were selected to evaluate the relevance of our contribution compared to the state-of-the-art techniques. Here, we report the best performance result of each approach obtained from the different parameter settings detailed as follows.

Given a texture sample, texture features were characterized as the statistics of the response to scale-space filters such as Gabor wavelets at the orientation $\theta = \{0, \pm \frac{\pi}{4}, \pi\}$ and the frequencies $f = \{0, 4, 8\}$. In contrast, co-occurrence matrices measure the intensity or grayscale values of texture image at a neighborhood distance $d = \{1, 2, 4\}$ with a set of orientation $\theta = \{0, \pm \frac{\pi}{4}, \pm \frac{\pi}{2}, \pm \frac{3\pi}{4}, \pi\}$. In addition to these classical texture descriptors, Xu’s approach [27] was also tested. It relies on a multifractal description of textures with invariance to viewpoint changes, non-rigid deformations and local affine contrast change. We tested different parameter settings for Xu’s method: density level $ind = \{1, 8\}$, dimension of MFS $f = \{16, 64\}$ and iteration level $ite = \{8, 10\}$.

Regarding schemes based on visual keypoints, we implemented bags-of-keypoints, i.e. the relative occurrence statistics of the different visual words based on SIFT descriptor [4]. We considered also the most popular keypoint combination scheme (Har+Lap)(Sift+Spin) introduced in [28]. These methods were computed with the different size of categories, here, $k = \{60, 120, 150\}$. These approaches were selected to examine the contribution of spatial information of keypoints to texture recognition.

The quantitative evaluation also included co-occurrence statistics of visual keypoints, here Ling’s method [15] and
Nguyen’s method [21]. They were implemented with logarithmically increased neighborhood sizes, $N_r = 128 \log(x)$ where $x$ varies between 1 and $\exp(1)$ according to a 0.05 linear step. The computation of the second-order descriptive statistics reported in [21] involved a correction of edge effects and scaling factor as well as feature dimension reduction with $k^* = 60$ categories of visual keypoints pairs. The similar parameter setting was used for the proposed approach based on multivariate log-Gaussian Cox model.

### 4.2. Performance results

#### 4.2.1 UIUC dataset

The UIUC dataset involves 25 texture classes and each class contains 40 640x480 images with strongly varying viewpoint, scale and illumination conditions (Fig.4a). Regarding the comparison of the proposed descriptor to previous work in Tab.3, the mean correct classification rate and standard deviations over 50 random selections are reported for each approach as a function of the number of training samples $N_t$. Observing the result of the classical approaches such as Gabor filters, cooccurrence matrix and multifractal (Xu’s method), these experiments clearly demonstrate the interest of inheriting the robustness of the visual keypoints for texture recognition in terms of invariance to geometric image distortions and contrast change: respectively, $67.78\% \pm 1.28$, $80.12\% \pm 1.30$, $93.85\% \pm 1.31$ vs. $97.84\% \pm 0.32$ of our method with $N_t = 20$.

Considering state-of-the-art keypoint-based approaches, reported performances stress the relevance of the proposed probabilistic texture models. The gain is greater than 6.5% compared with the BoK and 1.5% compared with the most popular local keypoint (Har+Lap)/(Sift+Spin) scheme of Zhang’s method when 20 training images are considered. These results emphasize the efficiency of the spatial statistical analysis of the visual keypoints. On the other hand, the proposed descriptor gets a more robust texture recognition than the techniques of cooccurrence statistics of visual keypoints of Ling’s method and Nguyen’s method, respectively $91.87\% \pm 1.38$, $97.34\% \pm 0.25$ vs. $97.84\% \pm 0.32$ when 20 training images are considered.

#### 4.2.2 Brodatz dataset

We also evaluated recognition performances for 111 different texture classes from Brodatz album. Each class of this dataset comprises 9 170x170 sub-images (Fig.4b) extracted from images. It might be noted that this dataset does not include scale and illumination changes. The proposed descriptor favorably compares to the other approaches in Tab.4 with a random selection of 1 or 3 training images per class. Our approach reaches up to $96.14\% \pm 0.41$ of correct classification with 3 training images. Greater improvements of than 12% are reported to compare with Gabor and cooccurrence features. All other methods reached classification performances below 95.67%. The proposed descriptor is shown to be slightly more robust and stable than Nguyen’s method and Zhang’s method, here $88.81\% \pm 0.92$ vs. $87.67\% \pm 0.81$ and $86.63\% \pm 1.05$ when $N_t = 1$.

#### 4.2.3 KTH-TIPs dataset

Similar conclusion can be drawn from the third experiment for the KTH-TIPs texture dataset. This dataset involves 10 material classes. Each class contains 81 images. Texture samples are 200x200 images (except some samples of two classes: brown-bread and cracker) for different scales, illumination directions and object poses (Fig.4c). The classification performance comparisons for this dataset are showed in Tab.7. The proposed approach has a gain of about 1% compared to the best score of all other approaches, namely Nguyen’s method, $95.74\% \pm 0.45$ vs. $95.09\% \pm 0.41$ when 40 training images are used.

#### 4.2.4 Performance comparison among different keypoint types and different covariance functions

We also report a detailed analysis of the classification performances on UIUC texture dataset reached by the pro-
Table 3. Classification rates and standard deviations of proposed model compared with the-state-of-the-art approaches on UIUC dataset.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31.22±3.14</td>
<td>45.33±3.03</td>
<td>67.25±2.75</td>
<td>67.62±2.93</td>
<td>61.14±2.90</td>
<td>72.83±2.45</td>
<td>75.66±1.65</td>
</tr>
<tr>
<td>5</td>
<td>45.14±2.54</td>
<td>61.58±2.14</td>
<td>76.38±2.15</td>
<td>78.42±2.33</td>
<td>83.33±2.07</td>
<td>86.62±1.33</td>
<td>91.67±0.93</td>
</tr>
<tr>
<td>10</td>
<td>57.37±1.93</td>
<td>70.67±1.72</td>
<td>81.12±1.45</td>
<td>84.14±1.72</td>
<td>89.68±1.65</td>
<td>93.17±1.15</td>
<td>94.33±0.78</td>
</tr>
<tr>
<td>15</td>
<td>61.25±1.52</td>
<td>73.85±1.34</td>
<td>86.35±1.20</td>
<td>86.38±1.25</td>
<td>81.34±1.45</td>
<td>85.33±0.98</td>
<td>96.54±0.53</td>
</tr>
<tr>
<td>20</td>
<td>67.78±1.28</td>
<td>80.12±1.30</td>
<td>91.28±1.13</td>
<td>91.87±1.38</td>
<td>93.85±1.31</td>
<td>96.67±0.93</td>
<td>97.34±0.25</td>
</tr>
</tbody>
</table>

Table 4. Classification rates and standard deviations over 50 random selections on Brodatz texture database.

| $N_t$ | Gaussian Cardinal sine Hyperbolic |
|-------|--------------------------|--------------------------|--------------------------|
| 1     | 78.52±1.72               | 75.42±1.73               | 83.16±1.50               |
|       | 83.43±1.63               | 85.95±0.91               | 93.41±0.73               |
| 3     | 85.14±1.41               | 83.22±1.04               | 92.78±0.91               |
|       | 93.17±0.87               | 94.34±0.43               | 95.67±0.33               |
| 5     | 91.96±1.13               | 91.63±1.17               | 91.32±1.19               |
|       | 94.35±0.75               | 94.72±0.85               | 95.43±0.71               |
| 10    | 95.42±0.71               | 95.35±0.75               | 94.72±0.85               |
|       | 96.17±0.63               | 95.43±0.71               | 96.85±0.38               |
| 15    | 96.87±0.65               | 97.15±0.42               | 98.85±0.38               |
| 20    | 97.84±0.32               | 97.15±0.42               | 98.85±0.38               |

Table 5. Comparison performance of proposed model with the different covariance functions on UIUC dataset.

Table 6. Comparison performance of proposed model with and without scaling effect and dimensional reduction.

<table>
<thead>
<tr>
<th>$N_t$</th>
<th>Without scaling effect</th>
<th>Without dimensional reduction</th>
<th>Model complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>73.27±2.08</td>
<td>75.65±1.82</td>
<td>75.21±1.75</td>
</tr>
<tr>
<td>5</td>
<td>89.12±1.27</td>
<td>91.67±1.15</td>
<td>91.96±1.13</td>
</tr>
<tr>
<td>10</td>
<td>94.35±0.75</td>
<td>95.15±0.73</td>
<td>95.42±0.71</td>
</tr>
<tr>
<td>15</td>
<td>95.12±0.81</td>
<td>96.27±0.71</td>
<td>96.87±0.65</td>
</tr>
<tr>
<td>20</td>
<td>95.89±0.54</td>
<td>97.12±0.35</td>
<td>97.84±0.32</td>
</tr>
</tbody>
</table>

5. Discussion

In this paper, we have further explored texture description and recognition from the joint characterization of the spatial and visual patterns of keypoints in texture images. Viewing keypoint sets as realizations of finite spatial random sets we have shown that beyond descriptive statistics proposed in [15, 21] probabilistic keypoint-based models can be developed for visual textures. The proposed models embed invariance properties with respect to contrast change and geometric image transforms to reach robust texture recognition performance as proven by the quantitative evaluation to state-of-the-art schemes.

As pointed out in [6, 7, 22], it is difficult to recommend a priori the best model among many models for spatial point pattern analysis in the literature, e.g. Neyman-Scott, shot-noise Cox or Gibbs processes. Here, multivariate log-Gaussian Cox models were selected. They have several appealing features:

- They are fully characterized from the underlying Gaussian fields, hence the associated mean and covariance features. This makes simple the interpretation of model parameters as well as their estimation.
- Through the various parametric and non-parametric forms that can be considered for the covariance structure, these models are highly flexible to cover a wide range of covariance structures.

Besides compared to simple descriptive statistics these models have several major advantages:

- They are independent on the a priori selection of the study regions (number of regions and radius sizes) which is critical setting for the computation of descriptive statistics and provide a description of keypoint patterns intrinsically free of edge effects and image size.
These aspects will be further explored in future work. While
sure can be defined from distances between Gaussian fields
similarity measures from model distances. This is a key fea-
ture spaces are also preferred.

\[ k \times k \] dimensional feature space vs. 

\[ (N_r + 1)k^* \] complexity required by descriptive statistics

[15, 21]. This compactness is of great benefit in the applica-
tion of learning techniques for which the lower-dimensional
feature spaces are also preferred.

The proposed probabilistic models also offer additional
generalization properties. They are associated with an ana-
lytical formulation of the likelihood functions of a keypoint-

patterns and simple simulation schemes. These features are
of great interest for various applications. For instance, they
could benefit to the definition of well-founded texture simi-
larities measures from model distances. This is a key fea-
ture of the log-Gaussian Cox model as the similarity mea-
ure can be defined from distances between Gaussian fields
for which analytical formulation may be derived. While
not investigated here as random forests where considered
for comparison to previous work, such similarity measures
would be of great interest for kernel-based learning as well as
other applications such as image indexing or segmentation.
These aspects will be further explored in future work.

References


and Mapping in the Space of Appearance. IJRR, 27(6):647–665,
2008.

for on-line spatio-temporal disease surveillance. Environmetrics,

traction by a multi-marked point process. PAMI, 32(9):1579–1609,
2010.

[8] R. Haralick. Statistical and structural approaches to textures. Pro-


regions with local binary patterns. Pattern Recognition, 42(3):425–


niew of classification and combining techniques. Artificial Intelli-


[14] T. Lindeberg. Feature detection with automatic scale selection. IJCV,


stereo from maximally stable extremal regions. BMVC, pages 384–
393, 2002.


F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine


[22] T. Randen and J. H. Husøy. Filtering for texture classification: A


[27] Y. Xu, H. Ji, and C. Fermuller. Viewpoint invariant texture descrip-

[28] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local fea-
tures and kernels for classification of texture and object categories: a

\[
\begin{array}{ccccccccc}
N_r & \text{DoG+Sift}[16] & \text{FH+Surf}[1] & \text{FH+Brief}[3] & \text{(Hes-Lap++Daisy)[24]} & \text{(Har-Lap++Sift-Spun)[28]} & \text{our method} \\
1 & 75.21 \pm 1.75 & 75.05 \pm 1.94 & 75.43 \pm 1.71 & 75.18 \pm 1.69 & 74.85 \pm 1.87 \\
5 & 91.96 \pm 1.13 & 90.73 \pm 1.11 & 91.42 \pm 1.23 & 92.13 \pm 1.19 & 91.15 \pm 1.41 \\
10 & 95.42 \pm 0.71 & 95.15 \pm 0.91 & 95.22 \pm 0.85 & 95.47 \pm 1.08 & 95.23 \pm 0.72 \\
15 & 96.87 \pm 0.65 & 96.14 \pm 0.63 & 96.43 \pm 0.51 & 96.75 \pm 0.58 & 96.37 \pm 0.61 \\
20 & 97.84 \pm 0.32 & 97.65 \pm 0.74 & 97.25 \pm 0.34 & 97.67 \pm 0.35 & 97.14 \pm 0.37 \\
\end{array}
\]

Table 7. Classification rates and standard deviations over 50 random selections on KTH-TIP's texture database.

Table 8. Comparison performance of proposed model with the different detector-descriptor types on UIUC dataset.

- They deliver a compact representation of keypoint-
based information, here \(3k^*\) –dimensional feature space vs.

\( (N_r + 1)k^* \) complexity required by descriptive statistics

[15, 21]. This compactness is of great benefit in the applica-
tion of learning techniques for which the lower-dimensional
feature spaces are also preferred.