

Spatial statistics of visual keypoints for texture recognition

Huu-Giao Nguyen, Ronan Fablet, Jean-Marc Boucher

⁽¹⁾Institut Telecom / Telecom Bretagne / LabSTICC
CS 83818 - 29238 Brest Cedex 3 - France

⁽²⁾Université européenne de Bretagne
{huu.nguyen, ronan.fablet, jm.boucher}@telecom-bretagne.eu

Abstract. In this paper, we propose a new descriptor of texture images based on the characterization of the spatial patterns of image keypoints. Regarding the set of visual keypoints of a given texture sample as the realization of marked point process, we define texture features from multivariate spatial statistics. Our approach initially relies on the construction of a codebook of the visual signatures of the keypoints. Here these visual signatures are given by SIFT feature vectors and the codebooks are issued from a hierarchical clustering algorithm suitable for processing large high-dimensional dataset. The texture descriptor is formed by cooccurrence statistics of neighboring keypoint pairs for different neighborhood radii. The proposed descriptor inherits the invariance properties of the SIFT w.r.t. contrast change and geometric image transformation (rotation, scaling). An application to texture recognition using the discriminative classifiers, namely: k-NN, SVM and random forest, is considered and a quantitative evaluation is reported for two case-studies: UIUC texture database and real sonar textures. The proposed approach favourably compares to previous work. We further discuss the properties of the proposed descriptor, including dimensionality aspects.

1 Introduction

The analysis of the texture content of images is among the critical issues for numerous application domains ranging from multimedia applications including content-based image and video indexing [1], automated scene analysis, archaeology [2] to more environment-oriented domains such as geosciences and remote sensing [3]. From the early 1970s, numerous advances have been reported in the definition of efficient while compact descriptors of visual textures. In the literature, the analysis of image textures initially mostly relied on a statistical analysis. Visual textures were regarded as the realization of random fields which could be characterized from relevant statistics such as covariance statistics [4, 5], co-occurrence statistics [6–9] or statistics of the response to scale-space filters such as Gabor and wavelet analysis [10].

More recently, a renewed interest in texture analysis emerged from the development of texture descriptors invariant to geometric and photometric transformations of the images. Visual keypoints initially proposed for object recognition

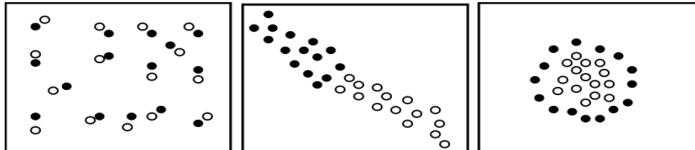


Fig. 1. Examples of difference spatial distributions of two marked-points.

were shown to be particularly efficient for texture analysis [11, 12]. The resulting texture characterization inherits the robustness of visual keypoints in terms of invariance to geometric image distortions and contrast change. Among the most popular descriptors is the SIFT descriptor based on the characterization of gradient orientations in scale invariant regions [13]. The application to texture recognition generally consists in learning classification models from these signatures of the visual keypoints [11, 12].

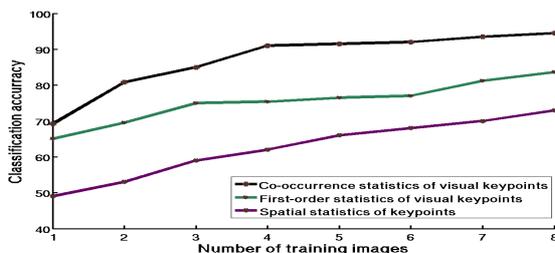


Fig. 2. Relevance of the combination of visual signatures of the keypoints to descriptors of their spatial organization: correct classification rate as a function of the number of training images for the UIUC texture database using only spatial statistics of the keypoint set (magenta), only statistics of the visual signatures of the keypoints (green), second-order co-occurrence statistics of the visual keypoint set (black) jointly describing the visual content and the spatial organization.

This typical approach however generally ignores the spatial organization of the visual signatures. As sketched by Fig.1, for similar relative occurrences of visual signatures, different spatial patterns revealing differences in visual content of the textures may be observed. In this work, we aim at jointly characterizing the local visual signatures of the textures and their spatial layout. Formally, we exploit spatial statistics to propose novel texture descriptors. Here, a texture is regarded as the realization of 2D marked point process referring to the set of the visual keypoints along with the associated visual descriptors, e.g., SIFT feature vectors. Texture descriptors are extracted as second-order co-occurrence statistics of the multivariate 2D point process. Applied to supervised texture

recognition for two different texture databases, UIUC textures and real sonar textures, these descriptors favorably compare to previous work [7–11]. As illustrated by Fig.2, we show the relevance of the combination of statistics of the visual signatures to statistics of the spatial layout of the keypoints.

This paper is organised as follows. In Section 2, an overview of the proposed approach is described. We present in Section 3 the proposed textural features based on spatial statistics of visual keypoints. Section 4 discusses the application to texture recognition. In Section 5, texture recognition performances are reported for two texture databases. The main contributions of proposed approach with respect to previous work are further discussed in Section 6.

2 Overview of proposed method

The sketch of the proposed approach is reported in Fig.3. A texture is regarded as the realization of a marked point process of visual keypoints. Descriptive statistics of this marked point process will form a set of texture descriptors.

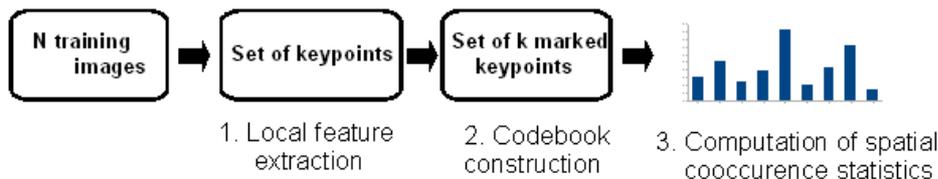


Fig. 3. Principal steps of our method for the extraction of a texture descriptor jointly characterizing the spatial layout of visual keypoints and their visual signatures.

The initial step consists in detecting and characterizing visual keypoints. Various local descriptors have been proposed for object and texture recognition. The SIFT descriptor achieves a good and stable performance in the domain of texture classification [14]. SIFT keypoints [13] correspond to local extrema of difference-of-Gaussian (DoG) filters at different scales. Each pixel in the DoG images is compared to its 26 neighbors in a $3 \times 3 \times 3$ neighborhood that spans adjacent DoG images. Each keypoint is characterized by the distribution of the orientations of the gradient of the intensity in the sixteen 4×4 windows around the considered point. This description ensures contrast invariance and partial invariance to affine transform. Orientation being quantized over eight values, the resulting SIFT feature vector is 128-dimensional, denoted by $X(s_i) = \{x_{i1}, \dots, x_{i128}\}$. As an output of this first step, a texture sample can be regarded as the realization of a marked point process, corresponding to the extracted keypoints with marks given by the SIFT feature vectors.

Our goal is to characterize the spatial patterns formed by the visual keypoints for a given texture image. To this end, we consider the joint density that pairs of neighboring keypoints s_1 and s_2 occur with given visual signatures x and y

$$p(X(s_1) = x, X(s_2) = y, s_2 \in V(s_1)), \quad (1)$$

where $V(s_1)$ specifies the neighborhood of point s_1 . This density is parameterized by the visual signatures and the neighborhood structure. Considering different neighborhood sizes, this density conveys second-order information on the relations between the visual signatures and their spatial layout.

The dimensionality of the visual signatures, i.e. the 128-dimensional SIFT vectors, makes untractable the non-parametric computation of the above second-order co-occurrence density. Rather than investigating parametric models such as Poisson point processes [15], a non-parametric approach based on an initial adaptive quantization of the SIFT feature space is proposed. We first build a codebook of visual keypoints from their feature vectors using a k-means-like method. Any visual keypoint is then associated with a discrete mark corresponding to the assigned category of visual signatures, denoted by $m_{s_i} = p$ where $p = \{1, \dots, k\}$. We then resort to a discrete approximation of the continuous density (Eq.1) for cooccurrences of visual words. The resulting second-order cooccurrence statistics are exploited for texture recognition.

The different steps of this approach are detailed in the subsequent.

3 Co-occurrence statistics of visual keypoints

Modeling and characterizing spatial point patterns is an active area of research, especially in environment-related sciences [16]. The general goal is to reveal underlying phenomena from the statistical analysis of some spatially-referenced observations or measures. In our case, a set of spatial points with some associated discrete signatures is regarded as the realization of a multivariate 2D point process, for which relevant descriptive statistics should be defined.

3.1 Spatial point process

A *spatial point process* \mathbb{S} is defined as a locally finite random subset of a given bounded region $B \subset \mathbb{R}^2$, and a realization of such a process is a spatial point pattern $s = \{s_1, \dots, s_n\}$ of n points contained in B . Considering a realization of a point process, the moments of the random variable are meaningful descriptive quantities, such as the expected number $\mu(B)$ of points falling in region B [17]. This first-order moment is associated with the intensity measure ρ of \mathbb{S} :

$$\mu(B) = E\{\#B\} = \int_B \rho(s) ds \quad (2)$$

where $E\{\cdot\}$ denotes the number of expected points falling in B , $\rho(s)ds$ is the probability that one point falls in an infinitesimally small area ds around s . The

normalized first-order moment $K() = \mu(B)/|B|$, where $|B|$ is the area of region B , is a popular descriptive spatial statistics, known as Ripley's K-function [16].

Actually, the organization of spatially stationary point process is only provided by the higher-order moments. The covariance structure of the count variables is measured by the second-order factorial moment $\mu^{(2)}$ of \mathbb{S} :

$$\mu^{(2)}(B_1 \times B_2) = E\left\{ \sum_{s_1, s_2 \in S}^{\#} \mathbb{I}_{B_1}(s_1) \mathbb{I}_{B_2}(s_2) \right\} = \int_{B_1 \times B_2} \rho^{(2)}(s_1, s_2) ds_1 ds_2 \quad (3)$$

where $\sum^{\#}$ is the sum runs only over pairs of points, $\mathbb{I}[\cdot]$ is an indicator function that takes the value 1 when s_i falls in the region B_i . Second-order density $\rho^{(2)}(s_1, s_2) ds_1 ds_2$ is interpreted as the density of the pair of points s_1 and s_2 in infinitesimally small areas ds_1 and ds_2 . Density function $\rho^{(2)}$ states the correlation of pairs of points [5]. Considering a stationary and translation-invariant point process, density $\rho^{(2)}(s_1, s_2)$ only depends on distance $\|s_1 - s_2\|$.

3.2 Descriptive statistics of the spatial patterns of visual keypoints

The above second-order moment of spatial point process can be extended to multivariate spatial point patterns. In our case, each visual keypoint is associated with a discrete mark m_{s_i} . Let $\Psi = \{s_i; m_i\}$ be a multivariate marked point process. Extending (Eq.3), the second-order characteristics of Ψ are characterized by the factorial moment measure $\mu_f^{(2)}(B_1 \times B_2)$ for regions B_1 and B_2 :

$$\mu_f^{(2)}(B_1 \times B_2) = E\left\{ \sum_{[s_1; m_1], [s_2; m_2] \in \Psi}^{\#} f(m_1, m_2) \mathbb{I}_{B_1}(s_1) \mathbb{I}_{B_2}(s_2) \right\} \quad (4)$$

where f is an arbitrary measurable non-negative function. Following [17], considering a translation-invariant and stationary point process, spatial sets $B_1 \times B_2$ are parameterized as $\{s_1 \in B\} \times \{s_2 \in D(s_1, r)\}$ where $D(s_1, r)$ is the disk of center s_1 and radius r . Focusing on cooccurrence statistics, i.e. function f set as $\delta_i(\cdot) \cdot \delta_j(\cdot)$ with i, j mark indices and δ the Kronecker function, we resort to the following second-order cooccurrence moment:

$$\mu_{ij}^{(2)}(\cdot, r) = E\left\{ \sum_{[s_1; m_1], [s_2; m_2] \in \Psi}^{\#} \delta_i(m_1) \cdot \delta_j(m_2) \mathbb{I}(\|s_1 - s_2\| \leq r) \right\} \quad (5)$$

Function $\mu_{ij}^{(2)}(\cdot, r)$ states the number of points of type j in a disk of radius r centered at a point of the spatial pattern of type i . These statistics express the covariance structure of the spatial pattern for points of type i and j (Fig.4a).

In practice, to exploit the above second-order descriptive statistics of keypoints in image, edge effects have to be dealt with. The area of the analysis region affects the accurate counting-based estimation of the spatial statistics.

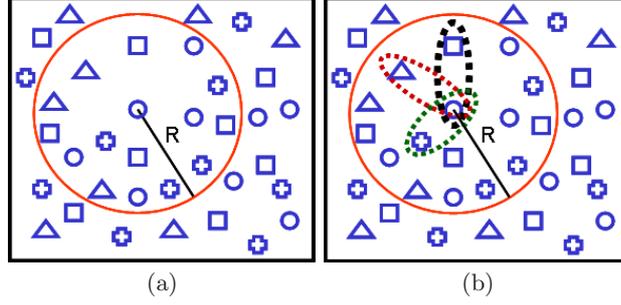


Fig. 4. Second-order spatial statistics for multivariate point process: (a) spatial statistics for any pair of keypoints within a circular region centered at a point of the pattern, (b) cooccurrence statistics for pairs of keypoints associated with given visual categories within a circular region centered at a point of the pattern.

Several corrections for edge effects for points located near the boundary of the image have been proposed in the literature[16, 18].

We then consider a normalized version of the second-order cooccurrence statistics (Eq.5) denoted by $\Gamma_{ij}^I(r)$ whose estimate is given by:

$$\Gamma_{ij}^I(r) = \frac{1}{N} \sum_{p=1}^N \frac{1}{A(s_p)} \sum_{j=1, j \neq i}^N \delta_i(m_p) \delta_j(m_q) \mathbb{I}[\|s_p - s_q\| \leq r] \quad (6)$$

where $A(s_p)$ is the actual area of the circular study region of radius r for keypoint s_p with p, q mark indices, N is the total number of keypoints in the image.

3.3 Scaling effects

The radius value in the computation of the proposed second-order cooccurrence statistics can be viewed as a scale-space parameter for the analysis of the spatial patterns of visual keypoints. The effects of image scaling should then be further analyzed to reach invariance to such image transformations.

Considering two images of a given texture sample at two different scales α_1 and α_2 , the first-order moments of the point process in these two images are related as follows:

$$\mu_1(r_1) \approx \mu_2(r_2) \quad (7)$$

$$\pi(\alpha_1 r)^2 \cdot K_1(\alpha_1 r) \approx \pi(\alpha_2 r)^2 \cdot K_2(\alpha_2 r) \quad (8)$$

with $r_1 = \alpha_1 r$ and $r_2 = \alpha_2 r$ and r the reference radius at a reference scale set to 1. As the detection of the visual keypoints in the images is scale-invariant, this property should be held in practice in our case. Interestingly, for an infinite circular region, the first-order moments of the two texture samples refer to the average point densities per surface unit and they only differ up to a factor depending on the rate α_1/α_2 . This property then provides the mean for scale

adaption. We proceed as follows. We first set a reference scale, i.e. $\alpha_1 = 1$ corresponding to expected mean point density per surface unit $K_1(\infty)$. For any texture sample, we estimate the associated scale factor α_2 with respect to the reference scaling from relation (Eq.8). The texture descriptor is then formed by the normalized second-order cooccurrence statistics for radius values $\{\alpha_2 r_i\}$, where the r_i 's are predefined radius values at the reference scale. In practice, the r_i 's are set according to a logarithmic sampling. Fig.5 clearly illustrates the benefit for the scale adaption in the computation of the cooccurrence statistics when comparing two texture samples differing by their scale factor.

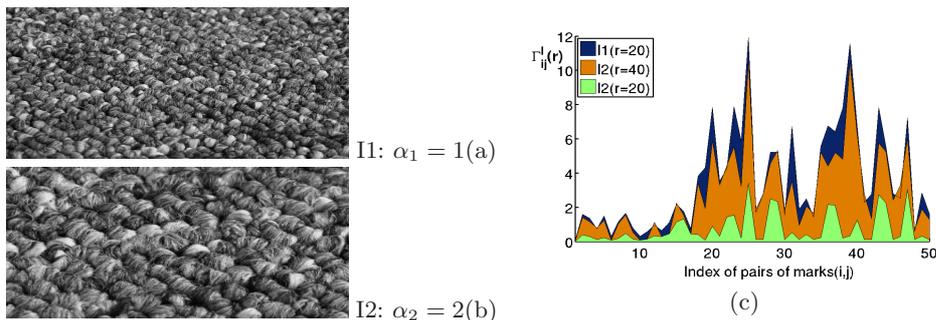


Fig. 5. Scaling effect on second-order cooccurrence statistics: images of the same texture at two scale factors (left), comparison of cooccurrence statistics without (blue vs. green) and with (orange vs. blue) scale adaption for the computation of the second-order cooccurrence statistics (right). The plot depicts the value of $\Gamma_{ij}^I(r)$ as a function of an index assigned to the pair of discrete marks (i, j) .

3.4 Dimensionality of the feature space

Given a codebook of keypoints with k visual words, the feature vector issued from the cooccurrence statistics in Eq.6 is $N_r k^2$ -dimensional, where N_r is the number of radius r of circular study region. For instance, considering a codebook of 40 visual words and $N_r = 20$, the feature space is 32000-dimensional. Such a high-dimensional description may affect recognition performance.

To address this issue, we suggest determining categories of pairs of visual keypoints to reduce the dimension of the second-order cooccurrence statistics. The codebook of keypoint pairs is issued from the application of a classical k-means method to the set of pairs of categorized keypoints within the training images. The size k^* of this codebook of pairs of visual keypoints is chosen to be typically in the order of k , the size of the codebook of visual words. Let us denote by $M(s_i, s_j) = \{1, \dots, k^*\}$ the category assigned to the pair of keypoints s_i and s_j . The second-order cooccurrence statistics (Eq.6) are then computed as:

$$\Gamma_u^I(r) = \frac{1}{N} \sum_{p=1}^N \frac{1}{A(s_p)} \sum_{j=1, j \neq i}^N \delta_u(M(s_i, s_j)) \mathbb{I}[\|s_p - s_q\| \leq r] \quad (9)$$

for a given category u of the codebook of pairs of visual keypoints. This procedure can be regarded as an adapted quantization of the normalized cooccurrence statistics defined by Eq.(1)

4 Texture classification

Given the textural features defined in the previous section, supervised texture classification is addressed using k nearest-neighbor (k-NN), support vector machines (SVM) and random forest (RF) classifiers:

- Nonparametric k-NN classifier is considered in this work because of its simplicity and its computational efficiency.
- Regarding SVM classifiers[19], a one-versus-all strategy is exploited to train a multi-class SVM with a Gaussian kernel.
- RF classifier relies on the construction of an ensemble of classification trees using some form of randomization. A texture sample is classified by sending it down every tree and aggregating the reached leaf distributions. RF classifier uses a voting rule to assign a class to an unknown texture sample[20].

SVM and k-NN classifiers here require the definition of a similarity measure in the considered feature space. Three different distances reported in Tab.1 accounting for the characteristics of the cooccurrence statistics are investigated.

Euclidean distance	χ^2 distance	Jeffrey divergence
$d(H, K) = \sum_i h_i - k_i ^2$	$d(H, K) = \sum_i \frac{(h_i - m_i)^2}{m_i}$	$d(H, K) = \sum_i \left(h_i \log \frac{h_i}{m_i} + k_i \log \frac{k_i}{m_i} \right)$

Table 1. Three different similarity measures, where $m_i = \frac{h_i + k_i}{2}$.

Randomization-based learning strategy: Randomization-based machine learning procedures are particularly appealing for classification from very large training dataset and high-dimensional feature space [19, 20]. In our case, a huge dataset of keypoints is used as input for the construction of the codebook of keypoint signatures from their SIFT feature vectors. For instance, a typical texture image involves 4.000 keypoints and 100 640x480 UIUC texture images leads to over 4.10^5 samples points in the 128-dimensional feature space defined by the SIFT descriptors. Adapted clustering techniques are required to perform the initial determination of the categories of visual keypoints. We suggest using a hierarchical clustering algorithm [21]. It relies on an agglomerative algorithm to generate a clustering solution. A hierarchical subtree is first built for each cluster. It then re-agglomerates these clusters to build a final hierarchical tree.

The initial clustering of visual keypoints might be regarded as a critical step in the proposed procedure. We can turn it into an advantage to build a randomized set of classifiers and improve the robustness to this initial clustering step. We proceed as follows. First, we carry out a random subsampling of visual keypoints

within the training dataset to select a subset tractable for the clustering scheme. Typically, 10^5 sample keypoints and up to 150 clusters can be considered. Given extracted categories of visual keypoints, second-order cooccurrence statistics are computed for a predefined set of radius values for each texture sample in the training set and a classifier is trained in the resulting feature space. From repeated initial random subsampling steps, a randomized ensemble of classifiers is built. Regarding the recognition step with a trained classifier ensemble, the classification of an unknown texture results from a simple voting rule.

5 Experiments

5.1 UIUC texture classification

The first experiment relies on 1000 640x480 texture images of UIUC database. This database involves 25 texture classes and each class contains 40 images with strongly varying viewpoint, scale and illumination conditions. Examples are reported in Fig.5a. The evaluation involves the computation of classification performances for model learning with N_t training texture samples per class. Training images are randomly selected among the 40 samples per class. The remaining $40 - N_t$ images per class are used as test images. The random selection of training samples is repeated 200 times to evaluate the mean and the standard deviation of the correct classification rate.

For comparison purposes, a set of texture descriptors such as Gabor filter [10], co-occurrence matrix [9], spatial statistics of the keypoints [15], bag-of-keypoints(BoK) [22], Zhang’s method [11], Xu’s method [7] and Varma’s method [8] were selected to evaluate the relevance of our contribution compared to the state-of-the-art techniques. The results on UIUC database of Varma’s method was reported in [8]. For the other methods, we report the performance with the best parameter setting. BoK was tested with $k = \{60, 120, 150\}$ classes. For cooccurrence matrices, the following neighborhood types were considered: $[0,1]$, $[1,1]$. Gabor features were computed for the frequencies $f = \{0, 4, 8\}$ and the orientation $\theta = \{0, \frac{\pi}{3}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. We also tested different parameter settings for Xu’s method: density level $ind = \{1, 8\}$, dimension of MFS $f = \{26, 64\}$ and iteration level $ite = \{8, 10\}$. SVM classifiers and Jeffrey divergence are used.

The parameter setting for our approach is as follows. A set of 10^5 random sampling keypoints is exploited for each hierarchical clustering step. The numbers of categories of visual keypoints k and of visual keypoint pairs k^* are respectively set to $k = 150$ and $k^* = 60$. $N_{r_{ref}} = 10$ circular study regions with reference radius values $r_{ref} = \{10, 20, 40, 60, 80, 120, 150, 190, 220, 240\}$ are considered for the computation of the second-order cooccurrence statistics. In Tab.2, we report the following results for our approach: cooccurrence statistics $\Gamma_{ij}^I(r)$ in Eq.6, cooccurrence statistics with dimensional feature reduction $\Gamma_u^I(r)$ in Eq.9.

Mean classification rates and standard deviations over 200 random selections are reported in Tab.2 as a function of the number of training images. The proposed descriptor favourably compares to the other approaches. Observing

N_t	1	5	10	15	20
Gabor filter	31.22 ± 3.14	45.14 ± 2.54	57.37 ± 1.93	61.25 ± 1.52	67.78 ± 1.28
Cooccurrence matrix	44.17 ± 2.93	62.25 ± 2.34	70.33 ± 1.75	73.67 ± 1.53	79.17 ± 1.37
Spatial statistic	48.69 ± 2.85	69.25 ± 2.45	75.42 ± 1.66	82.66 ± 1.28	87.34 ± 1.34
BoK	67.25 ± 2.75	76.38 ± 2.15	81.12 ± 1.45	86.35 ± 1.20	91.28 ± 1.15
Xu[7]	61.14 ± 2.90	83.33 ± 2.07	89.68 ± 1.65	91.34 ± 1.45	93.85 ± 1.31
Varma[8]	–	85.35 ± 1.69	91.64 ± 1.18	94.09 ± 0.98	95.40 ± 0.92
Zhang[11]	72.53 ± 2.45	88.62 ± 1.33	93.17 ± 1.15	95.33 ± 0.98	96.67 ± 0.93
$\Gamma_{ij}^I(r)$	75.43 ± 2.65	89.22 ± 1.47	93.48 ± 0.98	96.21 ± 0.66	97.17 ± 0.42
$\Gamma_u^I(r)$	75.66 ± 1.65	91.67 ± 0.93	94.33 ± 0.78	96.54 ± 0.53	97.34 ± 0.25

Table 2. Classification rates and standard deviations over 200 random selections on UIUC texture database.

the result in the case of 20 training samples, our proposed method reaches up to 97.34% of correct classification. Our descriptor gets a gain greater than 6% compared with BoK and than 3% compared with local fractal feature extraction methods. Our descriptor is shown to be slightly more robust and stable than Zhang’s method , 97.34 ± 0.25 w.r.t. 96.67 ± 0.93 . Greater gain in performances are observed when only few training images are used. From 5 training images per class, the proposed approach gets a correct classification greater than 91.67% whereas all the other methods are below 88.62%. Besides, while reducing the computational complexity, the dimension reduction technique also leads to a more robust texture recognition when few training samples are available.

5.2 Real sonar textures

Sonar imaging provides a remote sensing tool to observe and characterize the physical properties of the seafloor and is increasingly used for a variety of applications such as environmental monitoring, marine geosciences and ecology, as well as oil industry or defense [3, 23]. Sonar images are issued from the measurements of the backscattered echo of the seabed for successive sonar swaths corresponding to various incidence angles. In Fig.6a, an example of sidescan sonar images with incident angles from -85° to $+85^\circ$ was obtained from a DF1000 sonar. The different seabeds here correspond to different textural features. For a given seabed type, the mean backscatter clearly depends on the incidence angle [3, 23]. Especially, for vertical incidences, poor discrimination among seabed types can be expected. Besides, textural patterns may also vary depending on incidence angles as shown in Fig.6(b,c), where in the specular domain $[5^\circ, 40^\circ]$ a loss in contrast is observed for sand ripples compared with the sector $[80^\circ, 85^\circ]$.

We used a database of 180 sonar texture images involving 6 different seabed classes which are extracted from sidescan sonar images, e.g. Fig.6. Each class comprises 30 texture images. Sonar texture samples are 256×256 images with strong variations of incidence angles, scale and illumination conditions. We randomly choose $N = \{1, 5\}$ samples of each class to build a training dataset, the

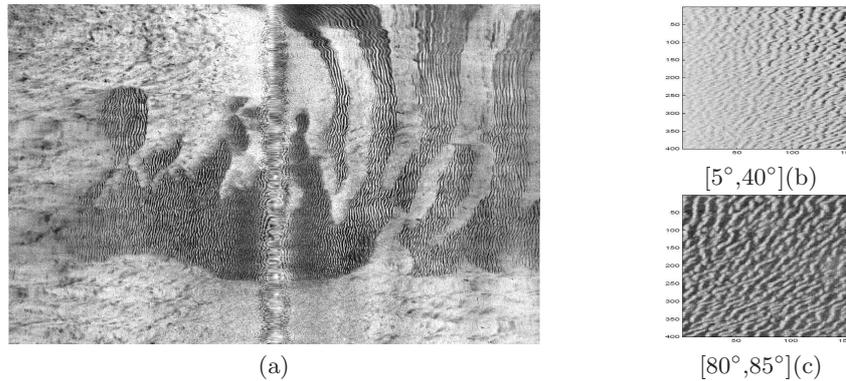


Fig. 6. A sidescan sonar image(a) and textures of a seabed type sample of sand ripples for two angular sectors(b,c) (Rebent, IFREMER).

other images being used as the query images. For these experiments, the following parameter setting is used: $k=50$, $r = \{10, 20, 40, 80, 120\}$ and $k^* = 35$. For comparison purposes, except for Varma’s method, we used all methods similarly to the experiment conducted with the UIUC texture database.

	k-NN			SVM			RF
	Euclidean	χ^2	Jeffrey	Euclidean	χ^2	Jeffrey	default
3 samples	89.2%	88.9%	91.3%	88.7%	89.4%	91.6%	91.8%
5 samples	94.2%	93.6%	95.1%	93.8%	94.7%	96.4%	96.4%

Table 3. Correct classification rate for the sonar texture database using second-order cooccurrence statistics (Eq.10) with different classifiers (k-NN, SVM, RF) and similarity measures (Euclidean, χ^2 , Jeffrey divergence).

We first compare the performances of the different classifiers and similarity measures. Jeffrey divergence improves the classification performance with an approximate gain greater than 2% compared to χ^2 and Euclidean distance when 3 (or 5) training images are considered in Tab.3. Differences in the performance among the classifiers, k-NN, SVM and RF, are not obvious for this dataset. However, RF may be slightly more robust and stable than k-NN or SVM.

Regarding the comparison of the proposed descriptor to previous work, the mean correct classification rate is reported for each approach as a function of the number of training samples (Fig.7). The proposed method reaches up to 96.4% of correct classification when 5 training images are considered. It favourably compares to previous work for which the best score is 93.5%(Zhang’s method). More than 6% of correct classification is gained when one image per class is used for training. These results demonstrate the robustness of the proposed descriptor to geometric and photometric image distortions.

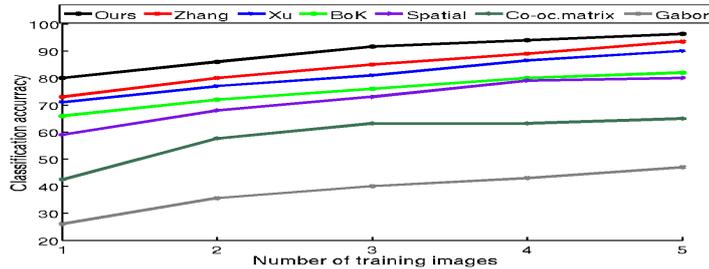


Fig. 7. Average classification rate on sonar images database

6 Discussion and future developments

We have proposed a novel texture descriptor based on the statistical characterization of the spatial patterns of visual keypoints in texture images. The key feature of our approach is its joint description of the visual signatures of the texture and of their spatial organization. We further discuss these contributions with respect to previous work.

- *Spatial information for texture recognition:* Numerous previous studies have investigated the use of spatial information for texture description [7–9]. The focus is generally given to the characterization of local texture patterns in a neighbourhood of a point. As an example Varma et al.[8] and Xu et al. [7] proposed a local fractal feature based on the analysis of the distribution of intensities as a function of the distance to a center point. In contrast, our approach encodes spatial information not at a local scale but at some object-related scale. The proposed descriptors aim at describing the spatial patterns formed by the set of visual keypoints, not the spatial variations of the intensity in a neighborhood of each keypoint.

These two aspects should be regarded as complementary. In the reported results, key points extracted as local DOG extrema and visual signatures provided by the SIFT descriptor were considered. It should be noted that any other local visual signature, including for instance the local fractal feature, could be considered in a similar manner. In future work, the evaluation of the robustness and the distinctiveness of other texture descriptors may be investigated such as GLOH[14], SURF[24], DAISY[25], CS-LBP (see [14, 26] for a review).

- *Image modelling from spatial point processes:* It might be noted that spatial point process were previously investigated for texture analysis. For instance, Linnett et al.[15] modeled the spatial distribution of the grey-levels by a 2D Poisson process. More recently, cooccurrence statistics of visual keypoints were investigated for different applications, namely scene categorization [27, 28], robot navigation [29]. However, those methods are again aimed at characterizing the variations of image intensities in local neighbourhoods. It should be noted that the link to spatial statistics also provide an unbiased estimation of the cooccurrence statistics based on a correction of edge effects.

Recently Gibbs point process models have also been applied to extract geometric objects in texture images by Lafarge et al. [30]. The Gibbs model provide an a priori model for the spatial organization of elementary geometric objects. Such a model cannot be applied to texture recognition. In contrast, our approach aims at deriving a feature vector for texture recognition based on spatial statistics. It should be noted that considered spatial statistics can be regarded as the sufficient statistics defining a multivariate log-Gaussian Cox process [5].

- *Invariance and dimensionality issues:* Local descriptors have emerged as a powerful tool for invariant texture characterization and classification compared to the early feature developed for texture analysis such as Gabor features and cooccurrence matrices. Approaches based on descriptive statistics of a set of visual keypoints may benefit from the robustness of their visual signatures in terms of invariance to photometric and geometric image transformations while providing a more compact representation of the information. In that case, each texture image is associated with a feature vector such that the size of the training database equals the number of training images. The BoK method, i.e. the distribution of the occurrences of the visual words in each texture sample, are a first example. As reported here, the associated recognition performances are however degraded compared to the classical keypoint classifier. The method described here also falls in this category. In contrast to BoK, both the visual signatures and the spatial patterns are characterized from second-order cooccurrence statistics. These statistics could convey scale-invariance. Improved texture recognition performances were initially obtained at the expense of the dimensionality of the feature space compared to BoK, i.e. $N_r k^2$ vs. k where N_r is the number of circular analysis region and k and the number of visual words. We have shown that dimensionality could be downsized up to the range of $N_r k$ simultaneously to a more robust recognition of texture sample.

In future work, other lower-dimensional representation could be investigated, especially from parametric and semi-parametric models of the covariance functions underlying the considered second-order cooccurrence statistics [5].

References

1. Ndjiki-Nya, P., Makai, B., Blattermann, G., Smolic, A., Schwarz, H., Wiegand, T.: A content-based video coding approach for rigid and non-rigid textures. *IEEE Conf. on Im. Proc., ICIP (2006)* 3169–3172
2. Zalesny, A., der Maur, D.A., Paget, R., Vergauwen, M., Gool, L.V.: Realistic textures for virtual anastylosis. *CVPR Workshop 1 (2003)* 14–20
3. Karoui, I., Fablet, R., Boucher, J.M., Augustin, J.M.: Seabed segmentation using optimized statistics of sonar textures. *IEEE Trans. on Geos. and Rem. Sens.* **47(6)** (2009) 1621–1631
4. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. *BMVC (2002)* 384–393
5. Møller, J., Syversveen, A., Waagepetersen, R.: Log gaussian cox processes. *Scandinavian Journal of Statistics* **25(3)** (1998) 451–482
6. Sivic, J., Russell, B., Efros, A., Zisserman, A., Freeman, B.: Discovering objects and their location in images. *ICCV 1 (2005)* 370–377

7. Xu, Y., Ji, H., Fermuller, C.: Viewpoint invariant texture description using fractal analysis. *IJCV* **83(1)** (2009) 85–100
8. Varma, M., Garg, R.: Locally invariant fractal features for statistical texture classification. *ICCV* **1** (2007) 1–8
9. Haralick, R.: Statistical and structural approaches to textures. *Proceedings of the IEEE* **67** (May,1979) 786–804
10. Randen, T., Husøy, J.H.: Filtering for texture classification: A comparative study. *IEEE Trans. on PAMI* **21** (1999) 291–310
11. Zhang, J., Marszałek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: a comprehensive study. *IJCV* **73** (2007) 213–238
12. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. *IEEE Trans. on PAMI* **27** (2005) 1265–1278
13. Lowe, D.: Distinctive image features from scale-invariant keypoints. *IJCV* **60(2)** (2004) 91–110
14. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Trans. on PAMI* **27(10)** (2005) 1615–1630
15. Linnett, L., Carmichael, D., Clarke, S.: Texture classification using a spatial-point process model. *IEE Vision, Image and Signal Processing* **142** (1995) 1–6
16. Goreaud, F., Péliissier, R.: On explicit formulas of edge effect correction for ripley’s k-function. *Journal of Vegetation Science* **10** (1999) 433–438
17. Stoyan, D., H.Stoyan: *Fractals, random shapes and point fields*. Wiley,Chichester (1994)
18. Nguyen, H.G., Fablet, R., Boucher, J.M.: Invariant descriptors of sonar textures from spatial statistics of local features. *ICASSP* (2010) 1674–1677
19. Kotsiantis, S., Zaharakis, I., Pintelas, P.: Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review* **26** (2006) 159–190
20. Breiman, L.: Random forests. *Machine learning* **45** (2001) 5–32
21. Zhao, Y., Karypis, G.: Hierarchical clustering algorithms for document datasets. *Data Mining and Knowledge Discovery* **10** (2005) 141–168
22. Csurka, G., Bray, C., Dance, C., Fan, L.: Visual categorization with bags of keypoints. *ECCV* (2004) 1–22
23. Chenadec, G.L., Boucher, J.M., Lurton, X.: Angular dependence of k-distributed sonar data. *IEEE Trans. on Geos. and Rem. Sens.* **45** (2007) 1124–1235
24. Bay, H., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. *ECCV* **1** (2006) 404–417
25. Tola, E., Lepetit, V., Fua, P.: Daisy: An efficient dense descriptor applied to wide baseline stereo. *IEEE Trans. on PAMI* (2009)
26. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with local binary patterns. *Pattern Recognition* **42(3)** (2009) 425–436
27. Ling, H., Soatto, S.: Proximity distribution kernels for geometric context in category recognition. *ICCV* (2007) 1–8
28. Savarese, S., Winn, J.M., Criminisi, A.: Discriminative object class models of appearance and shape by correlatons. *CVPR* **2** (2006) 2033–2040
29. Cummins, M., Newman, P.: FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. *IJRR* **27** (2008) 647–665
30. F.Lafarge, Gimel’farb, G., Descombes, X.: Geometric feature extraction by a multi-marked point process. *IEEE Trans. on PAMI* **99(1)** (2009)