Key-point Based Analysis of Sonar Images: Application to Seabed Recognition

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

Keypoint-Based Analysis of Sonar Images: Application to Seabed Recognition

Huu-Giao Nguyen, Ronan Fablet, Axel Ehrhold, and Jean-Marc Boucher

***Abstract*—In this paper, we address seabed characterization and recognition in sonar images using keypoint-based approaches. Keypoint-based texture recognition has recently emerged as a powerful framework to address invariances to contrast change and geometric distortions. We investigate here to which extent keypoint-based techniques are relevant for sonar texture analysis which also involves such invariance issues. We deal with both the characterization of the visual signatures of the keypoints and the spatial patterns they form. In this respect, spatial statistics are considered. We report a quantitative evaluation for sonar seabed texture data sets comprising six texture classes such as mud, rock, and gravely sand. We clearly demonstrate the improvement brought by keypoint-based techniques compared to classical features used for sonar texture analysis such as cooccurrence and Gabor features. In this respect, we demonstrate that the joint characterization of the visual signatures of the visual keypoints and their spatial organization reaches the best recognition performances (about 97% of correct classification w.r.t. 70% and 81% using cooccurrence and Gabor features). Furthermore, the combination of difference of Gaussian keypoints and scale-invariant feature transform descriptors is recommended as the most discriminating keypoint-based framework for the analysis of sonar seabed textures.**

***Index Terms*—Acoustic remote sensing, log-Gaussian Cox process, maerly sand, megaripples, sonar texture, visual keypoint.**

I. Introduction

IN RECENT years, the analysis of the physical properties of the seafloor and the mapping of seabed habitats has become increasingly important for both marine resource management and scientific research [1]–[4]. Many projects in coastal marine environment survey have been initiated to provide a baseline knowledge of seafloor and marine habitats [5]–[8]. The innovations in underwater acoustic signal processing have played a major role in the emergence of sidescan sonar as the privileged remote sensing device for *in situ* seabed observation. As an example, a sidescan sonar survey of the REBENT project has been carried out by IFREMER since 2000 [7]. An EdgeTech DF1000 sidescan sonar is deployed to survey a 200-km^2 area in the Bay of Concarneau on the south Brittany coast at different times, and it resulted in a high-resolution mapping of the marine habitats (Fig. 1). Sidescan sonar data refer to the backscatter energy from an insonified region of the seabed. Each line of a sonar image is built from one beam of the multibeam sidescan sonar system. The physical and biological properties of the seabed typically lead to the formation of textured images, different seabed types such as rock, mud, sand corresponding to different texture types (Fig. 4). Hence, sidescan sonar imaging is increasingly used for a variety of applications such as environmental monitoring, marine geosciences, and biology, as well as the oil industry or defense [9], [10], and the automated recognition and classification of sonar images in terms of seabed types are among the key issues [11]–[13].

Following the early development of texture descriptors in the field of image processing, sonar texture analysis [14], [15] usually relies on the extraction of statistics of the response to scale-space filters, as Gabor and wavelet analysis [15]–[17] or cooccurrence matrix [10]. However, as shown in Fig. 3, sonar texture characteristics strongly depend on the incidence angle. Variations of the incidence angle between the sonar beam and the insonified surface result both in local contrast changes as well as geometric distortions. These issues are not intrinsically embedded in the aforementioned techniques and are usually dealt with using texture models learned for different angular sectors. Recently, meaningful advances have been reported for visual texture recognition invariance to contrast and viewpoint change [18]–[21]. They mainly involved keypoint-based techniques, which primarily relied on the detection of interest points in visual textures [22]–[25]. Visual signatures of such keypoints also embedding such invariance properties were then proposed for image classification and categorization issues. The resulting texture characterization was regarded as the statistical analysis from the descriptor vector and local position of visual keypoints [26]–[28]. To our knowledge, the relevance of keypoint-based schemes in sonar seabed imaging has not been investigated. Here, we aim at evaluating to which extent keypoint-based techniques can improve the characterization and recognition of seabed textures. Regarding keypoint sets as realizations of multivariate point processes, our methodological contributions lie in the development of multivariate statistical descriptors and models of spatial keypoint patterns for sonar texture recognition. We focus on the analysis of the spatial covariance of keypoint patterns, namely, using second-order descriptive statistics and log-Gaussian Cox processes. These results open the door for other application of keypoint-based approaches in sonar imaging, e.g., registration-related issues.

This paper is organized as follows. In Section II, a sonar texture data set for seabed classification is introduced.

H.-G. Nguyen, R. Fablet, and J.-M. Boucher are with the Institut Telecom/ Telecom Bretagne/LabSTICC and Université Européenne de Bretagne, 29238 Brest CEDEX 3, France (e-mail: huu.nguyen@telecom-bretagne.eu; ronan.fablet@telecom-bretagne.eu; jm.boucher@telecom-bretagne.eu).

A. Ehrhold is with the Department of Coastal Environment Dynamics—IFREMER, 29280 Plouzané, France (e-mail: axel.ehrhold@ifremer.fr).

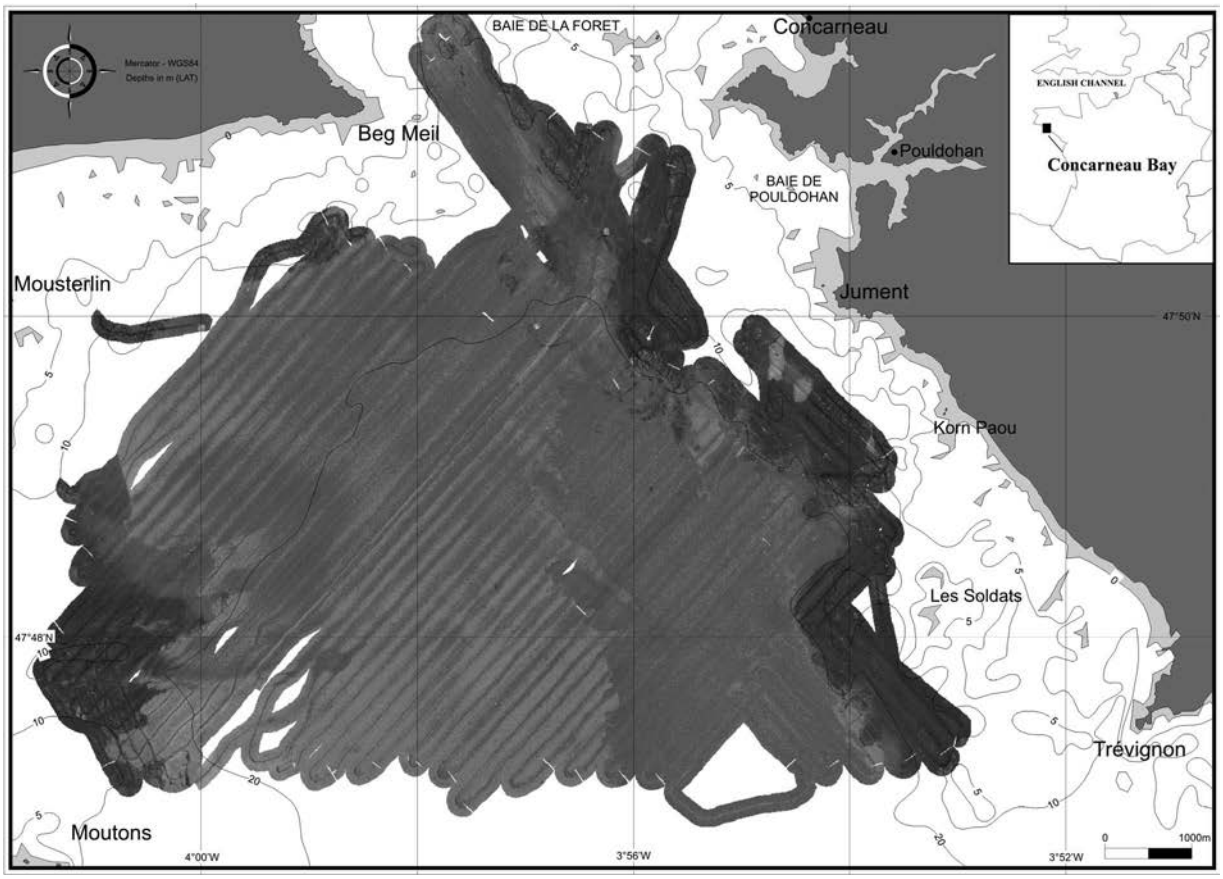

Fig. 1. Illustration of the surveyed area for marine habitat mapping as part of the REBENT project. The surveyed zone is located in the Bay of Concarneau (Britanny, France).

Keypoint detection and characterization in sonar seabed images are reviewed in Section III. In Section IV, we introduce keypoint statistics and associated probabilistic models, including spatial statistics of visual keypoint patterns. The application to sonar texture recognition and performance evaluations is reported in Section V. The main contributions of the proposed approach with respect to the previous work are further outlined in Section VI.

II. Sonar Texture Database

The sonar images considered in this paper were obtained from an EdgeTech DF1000 sidescan sonar as part of project REBENT, IFREMER. This database was used to survey coastal benthic habitats and to evaluate biodiversity changes in a 200-km^2 area in the Bay of Concarneau on the South Brittany, France (Fig. 1). The sidescan acquisition settings were chosen for the period from February 26, 2003, to March 21, 2003, as follows: 100-kHz backscatter signal, 110 m for swath width, and vertical beam titled down 20° from the horizontal. Moreover, 25 underwater video profile and diver observations and 93 grab samples were collected to provide a groundtruth knowledge of the structure of the seafloor.

Considering sidescan sonar data interpretation, acoustic backscatter can be regarded as a function of the incidence angle (i.e., the angle of incidence of the acoustic wavefront to the seafloor), surface roughness, impedance contrast across the sediment water interface, topography, and volume reverberation [29]. Sonar images are issued from the measurements of the backscattered echo of the seabed for successive sonar swaths. An example of sidescan sonar images with incident angles from $-85°$ to $+85°$ is shown in Fig. 2(a). The different seabeds correspond to different textural features. Fig. 2(a) contains two different seabed types, namely, maerly and gravelly sand and cleanly sand. Maerl biotopes [Fig. 2(c) and (d)] can build up over millennia to create carbonate-rich gravel deposits that often have high benthic biodiversity and productivity [30]. Maerl beds are protected by OSPAR convention. At the north of Mouton Island (Fig. 1), maerl beds are sculpted into a large-scale ripple pattern [Fig. 2(a) and (b)], indicating differential distributions of live and dead thalli between the tops of the ridges and the bottoms of the gullies. A better discrimination of megaripples sonar texture is important to recognize this sensitive habitat. Each column of this sonar image corresponds to a 25-m-wide band of the seafloor. For a given seabed type, the mean backscatter clearly depends on various incidence angles [9], [10]. Fig. 3(a) shows the mean backscatter evolution of two different seabed types in Fig. 2(a). In particular, for vertical incidences, poor discrimination among seabed types can be expected. Moreover, textural patterns may also vary depending on incidence angles, as shown in Fig. 3(b) and (c), where, in the specular domain $[5°, 40°]$, a loss in contrast is observed for maerly sand megaripples compared with the sector $[80°, 85°]$.

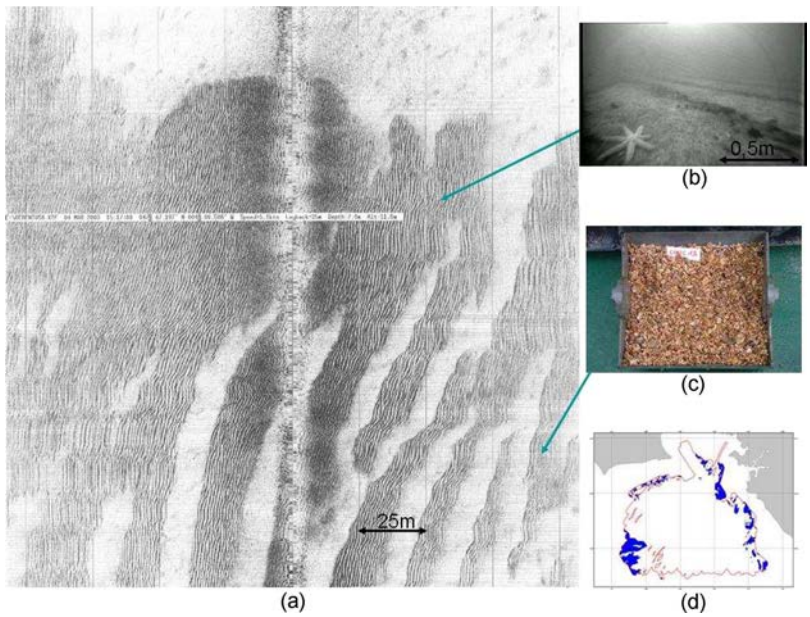

Fig. 2. (a) Example of sidescan sonar image (REBENT, IFREMER). Image extracted from (b) a video frame and (c) a grab observation of maerly and gravelly sand collected at the survey areas (in blue) in the (d) Bay of Concarneau.

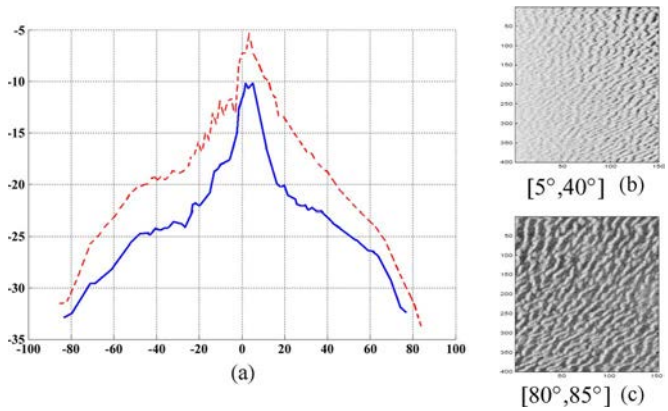

Fig. 3. (a) Backscatter evolution as a function of the incidence angle for two different seabed types: maerly-gravelly (M-G) sand (dotted line) and cleanly sand (solid line), from -85° to $+85^\circ$ with the incidence angle in Fig. 2. (b) and (c) Sonar image samples of the M-G sand for two angular sectors (REBENT, IFREMER).

Given the different physical and biological characteristics of the surveyed area, 50 sonar images were collected (Fig. 4). Six different seabed classes are identified for this paper, namely, *mud, sandy mud, maerly and gravelly (M-G) sand, clearly sand, rock, and mixed sediment*. Here, we used a database of 240 sonar textures images. Each class comprises 40 256 $\times$ 256 images with strong variations of incidence angles and scaling. This image database is made available.[1]

III. Keypoint Detection and Characterization in Sonar Images

We aim here at evaluating to which extent visual keypoints introduced for computer vision applications may be applied to sonar seabed imaging. Numerous approaches have been proposed to detect regions or points of interest in images. Among the most popular, the Harris detector detects corners [31], i.e., the points at which significant intensity changes in two directions occur. 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 [32]. Scale-spaces of difference of Gaussians (DoG) are also widely considered as an approximation of the Laplacian [24]. More recently, Mikolajczyk *et al.* [33] has 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.* [22] presented the fast-Hessian (FH) detector based on the Hessian matrix in the integral images. Other categories of keypoint detectors may be cited, e.g., the maximally stable extremal region detector [34], the edge-based region detector, the intensity extrema-based region detector [35], or the entropy-based region (such as salient regions) detector [36]. Comparisons between the different detectors for computer vision applications are given in [22], [33], and [37].

Given the pixel coordinates of the extracted keypoints, denoted by $\{s_1, \ldots, s_N\}$, many different schemes have been proposed to extract an invariant feature vector of each keypoint s_i [18], [26], [38]. The scale-invariant feature transform (SIFT) descriptor is certainly among the most popular. It is formed by local distributions of the orientations of the gradient of the intensity [24]. Intensity-domain features such as spin feature [37] may also be cited. The later relies on 2-D histogram encoding the distribution of the intensity value and the distance from the reference point. Rather than considering gradient orientations, the SURF descriptor [22] relies on the distribution of Haar-wavelet responses, whereas the Daisy descriptor [25] exploits responses to oriented Gaussian filters. The Brief descriptor

[1]This database is available for download at http://perso.telecom-bretagne.eu/ronanfablet.

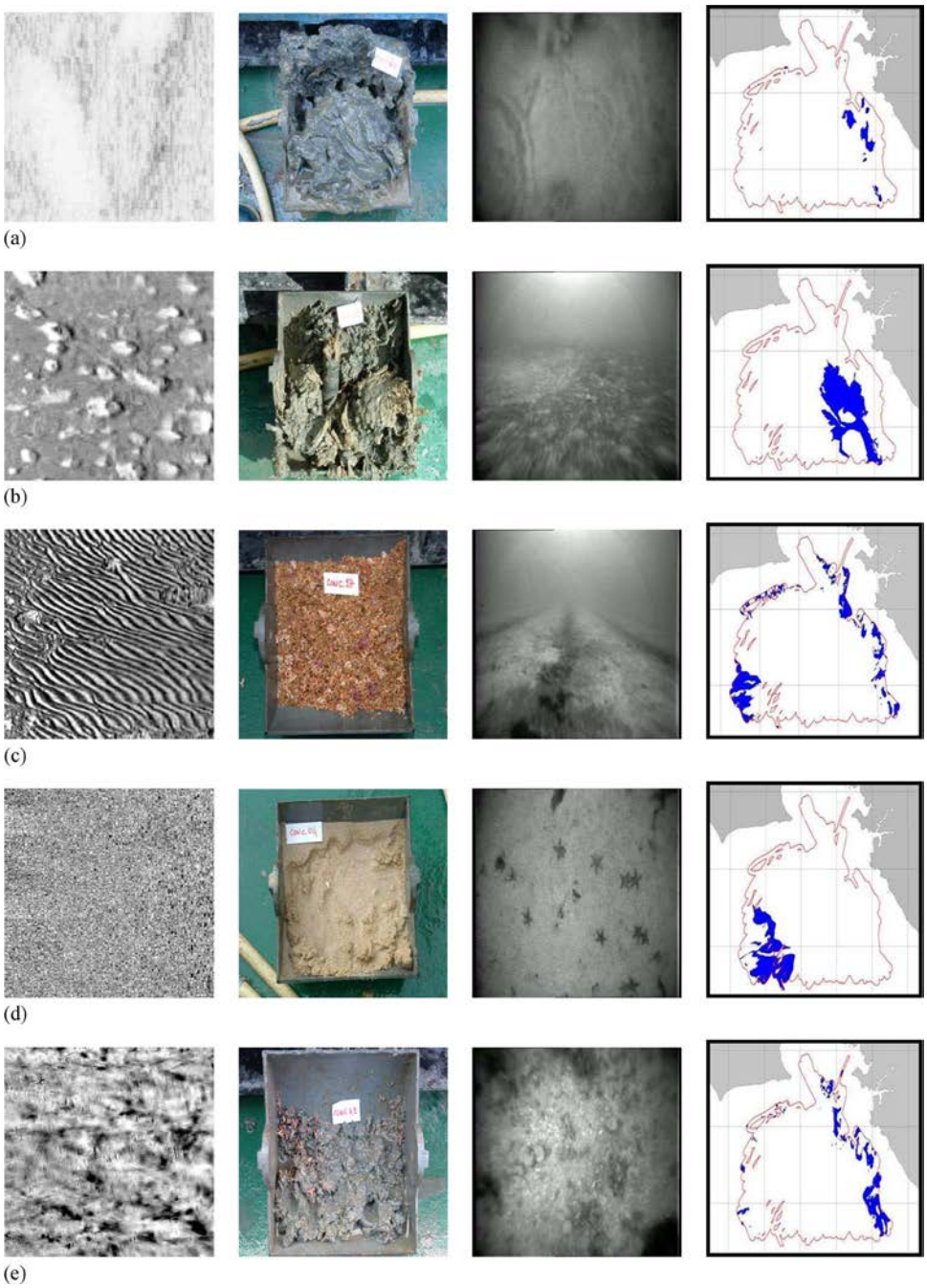

Fig. 4. Examples of sonar texture images with the corresponding video, grab observations, and survey location for the different seabed types. (a) Mud. (b) Sandy mud. (c) M-G sand. (d) Clearly sand. (e) Mixed sediment.

[23] was issued from a relatively small number of intensity of different image patches using binary string.

From computer vision reviews, we investigate five robust detector/descriptor types which were associated with the best performance for computer vision applications in [22]–[25] and [38], respectively: DoG+SIFT, (Har-Lap)+(SIFT–Spin), (Hes-Lap)+Daisy, FH+SURF, and FH+Brief. We briefly review the details of the implementation of each combination.

1) **DoG+SIFT**: This combination initially relies on the DoG detector to detect image keypoints. For each keypoint, a set of orientation histograms computed in 4×4 pixel neighborhoods with eight bins is evaluated. We therefore obtain a 128-dimensional feature vector. The implementation of DoG+SIFT is available at David Lowe's page.[2]
2) **(Har-Lap)+(SIFT–Spin)**: Following [33], we first apply the Har-Lap approach to detect the image keypoints. A 178-dimensional feature vector is formed from the 128-dimensional SIFT descriptor and the 50-dimensional Spin descriptor. The spin image is computed from the ten-bin normalized histogram of intensity in each of the five rings centered on the region. The implementation of

[2]http://www.cs.ubc.ca/~lowe/keypoints/

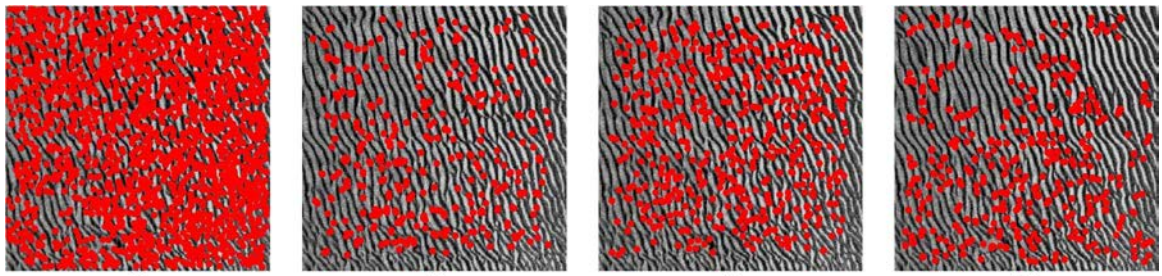

Fig. 5. Illustration of image keypoints detected in a sample of maerly and gravely sand using the different approaches. (a) DoG detector. (b) Har-Lap detector. (c) Hes-Lap detector. (d) FH detector. The mean number of keypoints per image for each method is presented in Table I.

this combination is available from the Visual Geometry group's page.[3]

3) **(Hes-Lap)+Daisy**: Following [25], we first apply the Hes-Lap approach to detect the image keypoints. The Daisy descriptor relies on histograms of oriented Gaussian filters. Here, we consider eight orientations, i.e., $(\pi/4)i$, $i = 1, 8$ at three levels of rings, where each level has eight rings with radius $R = 15$. The resulting feature vector is made of $8 + 8 \times 3 \times 8 = 200$-dimensional, extracted from 25 locations and 8 orientations. The implementation of this descriptor was taken from E. Tola's page.[4]
4) **FH+SURF**: A 64-dimensional orientation histogram of SURF descriptor is calculated from the distribution of four bins of Harr-wavelet responses in 4×4 windows around the FH keypoints. The code of this combination is available.[5]
5) **FH+Brief**: Following [23], we first apply the FH approach to detect the image keypoints, and the Brief descriptor is computed as the binary comparison between two values of N pairs of FH keypoints. The size of this vector descriptor is $N/8$. In our implementation, the dimension of Brief descriptor is fixed to 256. The code is available.[6]

As an example of the application of these keypoint detection schemes to sonar seabed images, we report detection results for a sonar texture corresponding to maerly and gravely sand (Fig. 5). Visually, detection results significantly differ among methods. While the DoG detector leads for this texture sample to a dense and homogeneous set of keypoints, Hes-Lap and Har-Lap detectors extract fewer points, and surprisingly, no keypoints are detected for some subregions. To further illustrate these aspects, we compare detection results using Har-Lap and DoG schemes for representative samples of each seabed texture class (Fig. 6) and report the mean number of detected points detected in the samples of each class (Table I).

Overall, visual keypoints in sonar seabed images are typically detected along the boundaries of shadow regions [e.g., for the mixed sediment class, Fig. 6(f)] or high-echo zones (e.g., for the sandy mud samples). As actual corners are not particularly characteristic of sonar images, the greater sensitivity of the DoG detector seems to make it more suited to the analysis of sonar images than Hessian-based detector. As detailed in the following section, keypoint-based recognition relies on keypoint statistics. In this respect, the DoG detector results in the greatest number of detected points, typically 1.8 (2.7 and 6.65) times more than Hes-Lap (respective Har-Lap and F.H) setting. This property is expected to be beneficial for characterization and classification issues.

[3]http://www.robots.ox.ac.uk/~vgg/research/affine/index.html
[4]http://cvlab.epfl.ch/~tola/daisy.html
[5]http://www.vision.ee.ethz.ch/~surf/
[6]http://cvlab.epfl.ch/software/brief/

IV. Keypoint-Based Statistics for Sonar Texture Characterization

Keypoint-based texture recognition typically relies on a statistical description of keypoint patterns. The first category of approaches relies on directly learning keypoint classification models in the feature space defined by the visual signatures of the keypoints [19], [20], [37]. Given a set of local keypoints in a texture image, the classification then relies on a simple voting procedure over all detected keypoints. A drawback of such approach is the requirement for learning classification models from very large training databases of keypoints.

By contrast, we investigate here actual statistical texture characterization and models for visual keypoint sets. As detailed in the subsequent section, such approaches 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 [39], [40]. The bag-of-keypoints (BoKs) method [39], i.e., the distribution of the occurrences of the visual words in each texture sample, is the first solution. Moreover, with a view to jointly characterize the visual signatures of the keypoints along with their spatial distribution, we consider spatial descriptive statistics [40] and models [41] of spatial keypoint patterns. As shown in Fig. 7, for similar relative occurrences of visual signatures, different spatial patterns revealing differences in visual content of the textures may be observed.

A. BoKs

BoKs were inspired by bag-of-words characteristics widely used for text characterization and retrieval [39]. Given a set of keypoints, the BoK method relies on the construction of a codebook of the visual signatures of the keypoints using a k-means-like method. A discrete value is then assigned for

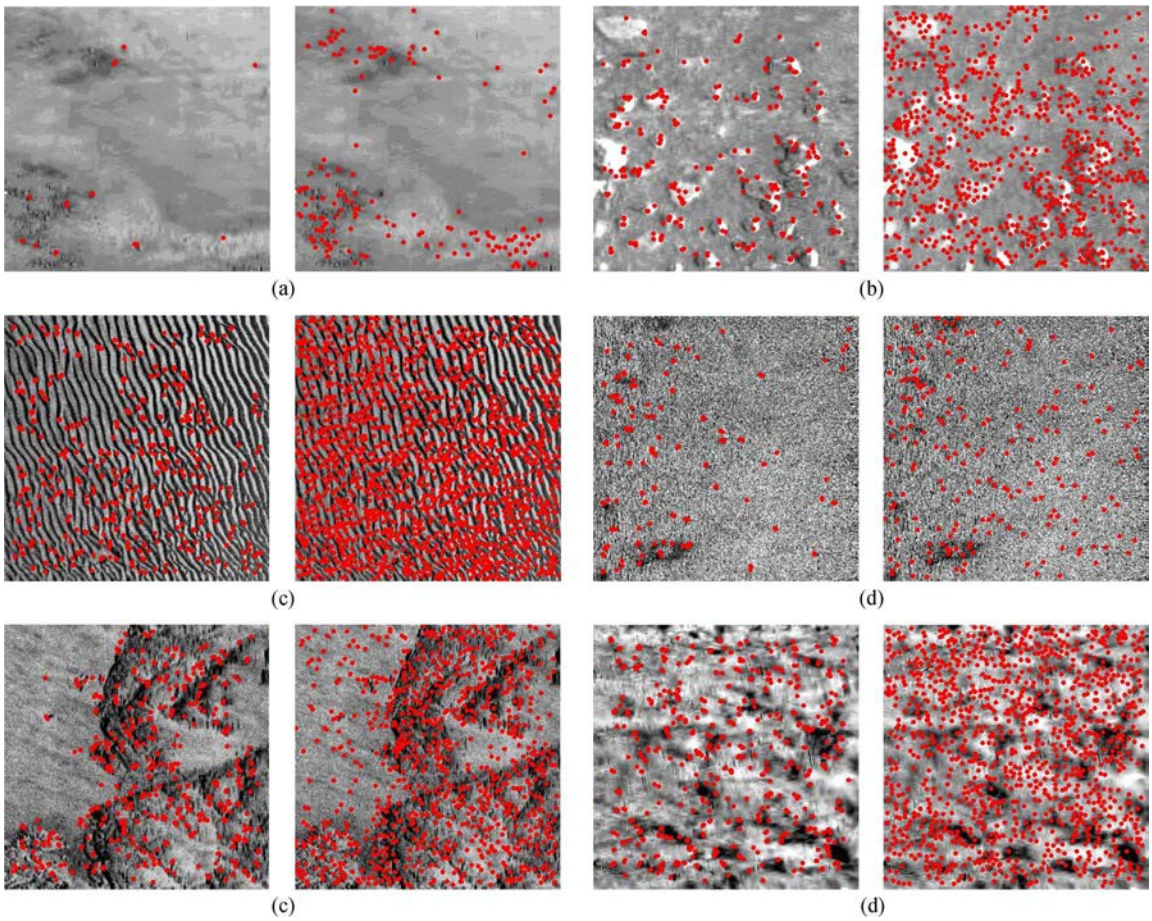

Fig. 6. Example of different spatial distributions of keypoint sets in sonar texture samples of the different seabed classes using the (left) Har-Lap detector and (right) DoG detector. (a) Mud: (Har-Lap) 19 points; (DoG) 190 points. (b) Sandy mud: (Har-Lap) 309 points; (DoG) 921 points. (c) M-G sand: (Har-Lap) 529 points; (DoG) 2287 points. (d) Clearly sand: (Har-Lap) 160 points; (DoG) 269 points. (e) Rock: (Har-Lap) 527 points; (DoG) 1002 points. (f) Mixed sediment: (Har-Lap) 569 points; (DoG) 1286 points.

TABLE I
MEAN NUMBER OF KEYPOINTS DETECTED FROM THE CONSIDERED SONAR TEXTURE DATABASE

	DoG	FH	Har-Lap	Hes-Lap
Mud	**334.85**	35.125	20.925	182.325
Sandy mud	**670.5**	166.65	230.95	534.65
M-G sand	**1777.275**	297.975	476.425	400.3
Clearly sand	**377.525**	143.825	205.55	368.5
Rock	**1070.7**	307.65	573.75	775.65
Mixed Sediment	**1350.175**	353.525	554.85	834.525
Overall	**930**	217.45	344	516

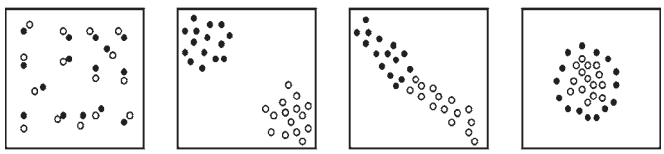

Fig. 7. Examples of different spatial distributions of marked point patterns.

each keypoint group, and the image is characterized by the occurrence statistics of each keypoint category (often referred to as visual words). Compared to approaches directly learning classification models in the feature space of the keypoint descriptors, each texture image is associated here with a feature vector such that the size of the training database is equal to the number of training images. However, BoK also ignores the spatial organization of the visual keypoints, as shown in Fig. 7.

B. Spatial Keypoint Statistics

With a view to jointly characterize the visual signatures of the keypoints along with their spatial distribution, we propose descriptors formed by spatial statistics of keypoint patterns [40]. Our approach consists in regarding the set of visual keypoint attached to a given sonar texture sample as the realization of a spatial point process.

A *spatial point process* $\mathbb{S}$ is defined as a locally finite random subset of a given bounded region $B \subset \mathbb{R}^2$. A realization of such a process is a spatial point pattern $s = \{s_1, \ldots, s_n\}$ 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 pth-order moment of $\mathbb{S}$ is defined as

$$\mu^{(p)}(B_1 \times \ldots \times B_p) = E\left\{\mathbb{N}(B_1) \ldots \mathbb{N}(B_p)\right\} \qquad (1)$$

where $E\{.\}$ denotes the expectation and $\mathbb{N}(B_i)$ is the number of random points contained in a given Borel set B_i. The first-order moment is evaluated with $p = 1$

$$\mu(B) = E \sum_{s \in \mathbb{S}} \mathbb{I}_B(s) = \int_B \rho(s) ds \qquad (2)$$

where $\mathbb{I}_B(s)$ is an indicator function that takes the value of 1 when s falls in region B and $\rho(s)ds$ is the probability that one point falls in an infinitely 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, and $|B|$ is the surface of region B. This quantity fully characterizes Poisson point processes. For a homogeneous process, this density is spatially constant. For a Poisson process, the individual points of a realization are fully independent.

Beyond the first-order moment, with a view to encode spatial dependences, the covariance structure of the count variable, i.e., the descriptive statistics of the pairs of points of the finite random set, can be characterized by the second-order moment $\mu^{(2)}$ of $\mathbb{S}$ given by

$$\mu^{(2)}(B_1 \times B_2) = E \sum_{s_1 \in \mathbb{S}} \sum_{s_2 \in \mathbb{S}} \mathbb{I}_{B_1}(s_1)\mathbb{I}_{B_2}(s_2) \tag{3}$$

$$= \int_{B_1 \times B_2} \rho^{(2)}(s_1, s_2) ds_1 ds_2 \tag{4}$$

where the 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 infinitely small areas ds_1 and ds_2. For a stationary and isotropic point process, this density function $\rho^{(2)}(s_1, s_2)$ states the correlation of pairs of points and only depends on distance $\|s_1 - s_2\|$ [42]. In the theory of spatial point processes [42]–[44], the second-order measure $\mu^{(2)}$ is frequently replaced by the factorial moment measure $\alpha^{(2)}$ as

$$\alpha^{(2)}(B_1 \times B_2) = E \sum_{s_1 \in \mathbb{S}} \sum_{\substack{s_2 \in \mathbb{S} \\ (s_2 \neq s_1)}} \mathbb{I}_{B_1}(s_1)\mathbb{I}_{B_2}(s_2) \tag{5}$$

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

$$\alpha^{(2)}(B_1 \times B_2) = \mu^{(2)}(B_1 \times B_2) - \mu(B_1 \cap B_2). \tag{6}$$

In our application of spatial point process to image keypoint sets, each point of the realization is associated with a visual signature. Such spatial patterns can be regarded as realizations of a *marked point process*. A *marked point process*Ψ is defined as a spatial point process for which a mark m_i is associated to each point s_i in B. Following the BoK setting, we resort here to discrete marks and encode the signature associated with a given keypoint as a keypoint category. Such point processes associated with discrete marks can be referred to as multivariate point process [42], [44].

Similar to previous discussion, second-order moments [in (4)] can be derived for multivariate point patterns. Considering circular study region $D(.,r)$ with radius r [Fig. 5(f)], the second-order spatial cooccurrence statistics (SCS) of Ψ are characterized by the factorial moment measure as follows:

$$\alpha^{(2)}_{i,j}(r) = E\left\{\sum_h \sum_{l \neq h} \delta_i(m_h)\delta_j(m_l)\mathbb{I}\left(\|s_h - s_l\| \leq r\right)\right\} \tag{7}$$

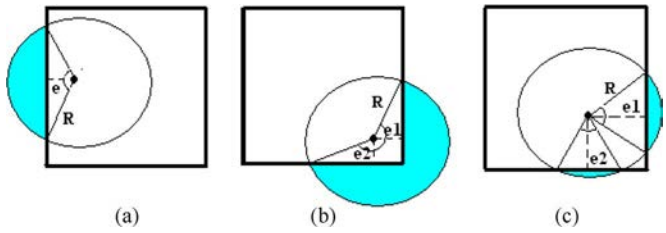

Fig. 8. Intersection cases of circular study window with the boundary of the image are presented. A very large neighborhood compared with the size of the image is not necessary to analyze the spatial texture pattern, i.e., we do not consider radius values greater than the width and the height of the image. Therefore, all cases of three-edge or four-edge effects and some particular cases of two-edge effect are not addressed. Considering the points located near the image's boundary with conditions of radius r in Table II, we obtain a total of 13 configurations of edge effect. The outside region is colored in green.

TABLE II
EXPLICIT FORMULAS OF ACTUAL CIRCUMFERENCE $b = 2\pi r - \tilde{b}$ OF THE CIRCULAR STUDY

Condition of radius	Outside area of the circular study
no-edge $r \leq e_1, e_2, e_3, e_4$	$\tilde{b} = 0$
one-edge $\begin{cases} r > e_1 \\ r \leq e_2, e_3, e_4 \end{cases}$	Fig. 8(a) $\tilde{b} = 2\arccos(\frac{e_1}{r})r$
two-edge1 $\begin{cases} e_1, e_2 < r \\ e_3, e_4 \geq r \\ r^2 > e_1^2 + e_2^2 \end{cases}$	Fig. 8(b) $\tilde{b} = \left(\frac{\pi}{2} + \arccos(\frac{e_1}{r}) + \arccos(\frac{e_2}{r})\right) r$
two-edge2 $\begin{cases} e_1, e_2 < r \\ e_3, e_4 \geq r \\ r^2 \leq e_1^2 + e_2^2 \end{cases}$	Fig. 8(c) $\tilde{b} = \left(\arccos(\frac{e_1}{r}) + \arccos(\frac{e_2}{r})\right) r$

where $\delta_i(m_h)$ is equal to one if the mark m_h of point s_h is i and zero otherwise. For statistical interpretation of second-order moment $\mu^{(2)}$ [42], Ripley's K function that is usually used to analyze 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

$$K_{ij}(r) = \frac{1}{\lambda_i \lambda_j} \alpha^{(2)}_{ij}(r). \tag{8}$$

1) ***Correction of edge effects***: In practice, the computation of the aforementioned second-order descriptive statistics takes into account the edge effects. Several corrections for edge effects for points located near the boundary of the image have been proposed in the literature [45]. In Fig. 8, we present the visualization of one-edge effect [Fig. 8(a)] and two-edge effects [Fig. 8(b) and (c)] which are analyzed in this paper. Let us denote with e_1, e_2, e_3, e_4 the distances between a given point and the four sides of an image. In Table II, we present the explicit formulas of the actual circumference of the study circle for the intersection cases with the boundary of image.
2) ***Feature dimension reduction***: The aforementioned descriptive statistics refer to the mean occurrence of a keypoint of a given category in a ball of radius r centered at a keypoint of the pattern associated with another keypoint category. The feature vector size is $N_r k^2$-dimensional, where N_r is the number of balls of radius r and k is the number of keypoint categories. In practice, such

high-dimensional feature may affect recognition performance, where BoK [39] leads to a k-dimensional feature space. The spatial statistics combine cooccurrence statistics, i.e., occurrence statistics of pairs of keypoint categories in an image, and both spatial dependences through varying ball radii. A dimensional reduction procedure of second-order statistics was introduced in [40] from the determination of categories of keypoint pairs. The codebook of keypoint pairs, denoted by $u = M(s_h, s_l)$, is issued from an adapted clustering technique applied for each set of two categorized keypoints s_h and s_l. The second-order SCS in (7) are calculated as

$$\alpha_u^{(2)}(r) = E\left\{\sum_h \sum_{l\neq h} \delta_u\left(M(s_h, s_l)\right) \mathbb{I}\left(\|s_h - s_l\| \leq r\right)\right\}. \tag{9}$$

C. LGCM

The proposed second-order spatial statistics are the sufficient statistics describing the log-Gaussian Cox model (LGCM). Beyond their theoretical interest, parametric forms of these point process models provide a more compact representation of the spatial keypoint patterns.

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 [42], [46]. For a *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 functions $c(r) = \mathrm{Cov}(Y(s_1), Y(s_2))$, where $r = \|s_1 - s_2\|$ is defined and finite for all bounded $B \subset S$. The mean and covariance structure of Gaussian field Y relate to the first- and second-order moments of the point process [46]. More precisely, the following relations hold for intensity function ρ and pair correlation function g (see the Appendix):

$$\begin{cases} \rho(s) = \lambda = \exp(\mu + \sigma^2/2) \\ \rho^{(2)}(s_1, s_2)/\left(\rho(s_1)\rho(s_2)\right) = g(r) = \exp\left(c(r)\right) \end{cases} \tag{10}$$

where $\sigma^2 = \mathrm{Var}(Y(s))$ is the variance of the Gaussian process.

The extension to a *multivariate log-Gaussian Cox process* is derived as follows. Cox processes $\{X_i\}$ are conditionally independent w.r.t. a multivariate intensity field $Z = \{Z_i\}$, and $X_i|Z_i$ is a Poisson process with intensity measure $\{Z_i\}$. 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 $c_{ij}(r) = \mathrm{Cov}(Y_i(s_1), Y_j(s_2))$. Moreover, the intensity and pair correlation function become

$$\begin{cases} \lambda_i = \exp\left(\mu_i + \sigma_i^2/2\right) \\ g_{ij}(r) = \exp\left(c_{ij}(r)\right). \end{cases} \tag{11}$$

Fitting a stationary parametric log-Gaussian Cox process comes to the estimation of the mean and covariance parameters of the associated Gaussian field. Following [42] and [46], the proposed estimation procedure relies on the relation between the pair correlation function g_{ij} and the K function of Gaussian processes as

$$K_{ij}(R) = 2\pi \int_0^R r g_{ij}(r) dr \tag{12}$$

where R is a predefined value of the radius. Combining (8) and (12), the pair correlation function can be estimated as

$$g_{ij}(r) = \frac{1}{2\pi r \lambda_i \lambda_j} \sum_h \sum_{l\neq h} \delta_i(m_h)\delta_j(m_l)\xi\left(\|s_h - s_l\|, r\right) b_{s_h} \tag{13}$$

where $\xi(.)$ is a kernel (here, a Gaussian kernel is considered), λ_i is the intensity for class i estimated from (2), and b_{s_h} is the proportion of the circumference of the study circle lying within the image. Considering the edge effect correction for the computation of the descriptive statistics as detailed in the previous section, g_{ij} is not symmetric in i and j. The nonparametric estimation of the covariance function is then defined as

$$\begin{cases} c_{ii}(r) = \log\left(g_{ii}(r)\right) \\ c_{ij}(r) = \log\left(\frac{\lambda_i g_{ij}(r) + \lambda_j g_{ji}(r)}{\lambda_i + \lambda_j}\right). \end{cases} \tag{14}$$

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

$$\int_0^R \left\{\sigma_{ij}^2 \mathbb{L}(\beta_{ij}, r) - c_{ij}(r)\right\}^2 dr. \tag{15}$$

A gradient-based optimization procedure is applied to solve this minimization. The proposed probabilistic keypoint-based texture model is eventually given by intensity parameters λ_i, variances σ_{ij}, and scale parameters β_{ij}. The feature vector size is reduced from $k(k+2)$ to $3k^*$, where k and k^* are the number of categories of keypoints or pairs of keypoints.

TABLE III
COVARIANCE FUNCTIONS OF $\mathbb{L}(\beta, r)$

Exponential	Cardinal sine	Hyperbolic
$\exp(-(r/\beta)^\alpha)$	$\sin(r/\beta)/(r/\beta)$	$(1 + r/\beta)^{-1}$

1) **Feature dimension reduction**: The log-Gaussian Cox covariance model estimation gives a feature vector of dimensionality $k(k+2)$. The same procedure of complexity reduction as SCS is applied, i.e., the nonparametric estimation of the covariance function is calculated for the pairs of keypoint categories as

$$c_u(r) = \log\left(\frac{1}{2\pi r \lambda_u} \sum_h \sum_{l\neq h} \delta_u\left(M(s_h, s_l)\right) \times \xi\left(\|s_h - s_l\|, r\right) b_{s_h}\right) \tag{16}$$

and the estimation of intensity parameter λ_u, variances σ_u, and scale parameters β_u for each category of keypoint pairs follows as that previously from minimization (15).

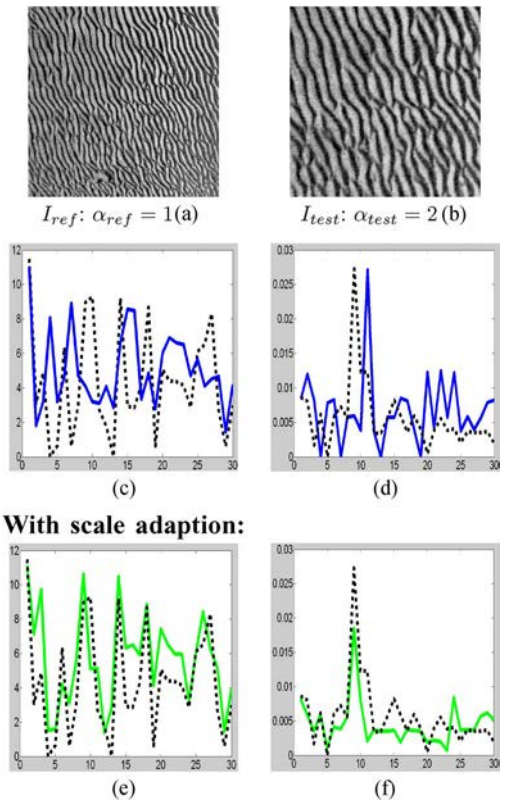

Fig. 9. Scaling effect on the estimation of the parameters of the LGCM of visual keypoint sets. (a) Reference image I_{ref} and (b) test image I_{test} at two scale factors. The dotted lines are the values of (c) and (d) variances σ_u and (e) and (f) scale parameters β_u of the reference image I_{ref} in all plots. The results of parameter estimation without (or with) scale adaption of the test image I_{test} are, respectively, showed on the second (or third) lines. These experiments were carried out with feature dimensionality reduction.

D. Invariance Properties

For the proposed approach based on the statistics of visual keypoint sets, invariance issues should be further analyzed. In all cases, invariance to contrast change is fully inherited from the definition of the keypoints and their visual signatures.

This is also the case for the invariance to geometric distortions of the BoK representation. For spatial statistics and associated models, image scaling clearly affects the second-order moments of the spatial patterns, where the radius value of the circular study region can be viewed as a scale-space parameter. Assuming that the detection and characterization of visual keypoints are scale invariant, scale adaption can be addressed. It should be stressed that this assumption is linked to the robustness of keypoint and characterization which is widely acknowledged in computer vision applications. Scale adaption proceeds here as follows. A scale factor is estimated from the rate of average point densities per surface unit compared to this rate for a reference image (corresponding to the reference (typically one) scale factor). The actual radius values R_i of the proposed estimation scheme were chosen depending on its scale factor. Fig. 9 shows the stability of the proposed scale-adapted features for different image scalings in the case of the log-Gaussian model.

V. Experimental Evaluation

Given the textural features defined in the previous section, an application to sonar texture classification is addressed, 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 models learned from 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 50 times to evaluate the mean and the standard deviation of the correct classification rate. The reported experiments comprise both an evaluation of different parameter settings, particularly in terms of considered type of keypoint and discriminative classifiers. Moreover, a comparison to state-of-the-art techniques is also carried out. The later involves both standard sonar texture descriptors based on cooccurrence statistics and Gabor features [17], [47] as well as state-of-the-art techniques for invariant visual texture recognition [21], [38].

A. Discriminative Classifiers

These experiments were carried out using different discriminative classifiers. We selected the K nearest-neighbor (k-NN) classifier for its nonparametric nature and its simplicity, support vector machine (SVM) [48], and random forest (RF) [49], which are among the most accurate classification schemes [50].

The following parameter settings were considered.

1) The nonparametric k-NN classifier was implemented with a varied k parameter depending on the number of training samples N_t

$$\begin{cases} k = N_t & \text{if } N_t \leq 5 \\ k = 5 & \text{if } N_t > 5. \end{cases} \tag{17}$$

2) Regarding SVM classifiers [48], a one-versus-all strategy is exploited to train a multiclass SVM, and a Gaussian kernel is chosen

$$G(H, K) = \exp\left(-\frac{d(H, K)}{2\sigma^2}\right) \tag{18}$$

where H and K are the feature vectors of the texture sample and σ is the standard deviation of the Gaussian distribution.

3) The 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 by aggregating the reached leaf distributions. The RF classifier uses a voting rule to assign a class to an unknown texture sample [49]. We used the RF with 500 trees (default value), and 100 features were randomly selected for the optimization of each split of the tree $(\mathrm{mtry} = 100)$.

SVM and k-NN classifiers require the definition of a distance in the considered feature space. We investigate here different distances accounting for the characteristics of the considered

TABLE IV
RECOGNITION PERFORMANCE OF THE LGCM TEXTURE FEATURE COMBINED WITH DIFFERENT CLASSIFIERS (K-NN, SVM, AND RF) AND SIMILARITY MEASURES (EUCLIDEAN, χ^2, AND JEFFREY DIVERGENCE)

N_t	k-NN			SVM			RF
	Euclidean	χ^2	Jeffrey	Euclidean	χ^2	Jeffrey	
5 samples	87.45%	88.17%	89.56%	88.11%	88.95%	91.34%	91.34%
10 samples	93.12%	94.83%	95.37%	94.18%	95.07%	96.95%	97.14%

TABLE V
RECOGNITION PERFORMANCE OF THE LGCM TEXTURE FEATURE USING DIFFERENT DETECTOR/DESCRIPTOR TYPES, NAMELY, DOG+SIFT, FH+SURF, (HAR-LAP)+(SIFT–SPIN), (HES-LAP)+DAISY, AND FH+BRIEF. AN RF CLASSIFIER WAS TRAINED IN EACH CASE

N_t	DoG+Sift	FH+Surf	(Har-Lap)+(Sift-Spin)	(Hes-Lap)+Daisy	FH+Brief
1 sample	73.85±1.75	71.15±1.91	72.67±1.95	73.67±1.52	**74.85±1.65**
3 samples	**82.95±1.13**	79.95±1.41	82.31±1.55	82.35±1.23	81.62±1.28
5 samples	**91.34±0.72**	90.17±1.37	91.05±1.31	91.24±0.81	90.84±1.13
8 samples	**93.87±0.61**	92.08±0.82	92.12±0.95	92.75±0.73	92.07±1.11
10 samples	**97.14±0.37**	95.67±0.51	96.12±0.62	96.63±0.47	95.92±0.59

descriptive statistics and models. Namely, three different dissimilarity measures are evaluated:

1) the Euclidean distance: $d_E(H,K) = \sum_i |h_i - k_i|^2$;
2) the χ^2 distance: $d_{\chi^2}(H,K) = \sum_i ((h_i - m_i)^2/m_i)$;
3) the Jeffrey divergence

$$d_J(H,K) = \sum_i \left(h_i \log \frac{h_i}{m_i} + k_i \log \frac{k_i}{m_i} \right) \qquad (19)$$

where $m_i = (h_i + k_i)/2$.

B. Parameter Setting of the Evaluated Texture Descriptors

Overall, we evaluated eight texture feature sets, namely, cooccurrence matrix [47] and Gabor [17] features which are classically used for sonar texture classification and state-of-the-art techniques for visual texture recognition such as Xu's method [21], BoK [39], Zhang's method [38], and Ling's method [51]. The most discriminative features from the different settings of each method were reported. The following parameter settings were considered.

1) **Gabor features** [17]: Gabor features are extracted as the statistics of the response to Gabor scale-space filters. The Gabor texture features were considered in this paper at orientations $\theta = \{0, \pm(\pi/2), \pi\}$ and frequencies $f = \{0, 4, 8\}$.
2) **Cooccurrence features** [47]: Cooccurrence matrix evaluates the occurrences of pairs of intensity values for neighboring pixels. Considered neighborhoods were parameterized by a distance $d = \{1, 2, 4\}$ of an orientation $\theta = \{0, \pm(\pi/4), \pm(\pi/2), \pm(3\pi/4), \pi\}$.
3) **BoK** [39]: The BoK method exploits relative occurrence statistics of the different visual words based on the SIFT descriptor. The number of classes of visual keypoints was set to $k = \{35, 50, 60\}$. We used the same k-means-like technique as that for spatial statistics.
4) **Ling's method** [51]: Ling's feature is a histogram of cooccurrence statistics of pair visual keypoint categories. We extend Ling' feature to a set of logarithmically increased neighborhood sizes $N_r = 128\log(x)$, where x varies between 1 and exp(1) according to a 0.05 linear step.

TABLE VI
RECOGNITION PERFORMANCE OF THE LGCM TEXTURE FEATURE WITH AND WITHOUT SCALE ADAPTION AND DIMENSION REDUCTION SCHEMES. HERE, A COMPLETE MODEL IS OBTAINED WHEN TWO OF THE AFOREMENTIONED SCHEMES ARE APPLIED

N_t	without scaling effect	without dimensional reduction	complete model
1 sample	72.63±1.75	**74.41±1.62**	73.85±1.75
3 samples	82.27±1.53	**83.11±1.25**	82.95±1.13
5 samples	89.85±1.11	91.05±0.95	**91.34±0.72**
8 samples	91.72±0.85	92.37±0.62	**93.87±0.61**
10samples	95.15±0.55	96.28±0.45	**97.14±0.37**

TABLE VII
RECOGNITION PERFORMANCE OF THE LGCM TEXTURE FEATURE FOR DIFFERENT COVARIANCE MODELS

N_t	Gaussian	Cardinal sine	Hyperbolic
1 sample	**73.85±1.75**	73.60±1.79	73.42±1.91
3 samples	**82.95±1.13**	82.47±1.21	82.15±1.23
5 samples	**91.34±0.72**	90.84±0.67	90.52±0.81
8 sample	**93.87±0.61**	93.11±0.65	92.57±0.75
10samples	**97.14±0.37**	96.81±0.48	96.12±0.58

5) **Xu's method** [21]: Xu's approach relies on a multifractal description of textures invariant to viewpoint changes, nonrigid 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 $ite = \{8, 10\}$.
6) **Zhang's method** [38]: Zhang *et al.* proposed a technique of texture classification using SVM classifier and earth move distance for the combination of two keypoint detectors (Har+Lap) and two local descriptors (SIFT+Spin) with the same number of clusters of visual keypoints $k = \{35, 50, 60\}$.
7) **SCS** [40]: The same parameter setting as the Ling's method was used for the SCS approach. However, SCS involves a correction of edge effects and scale adaption as well as a feature dimension reduction with $k^* = 30$ categories of visual keypoint pairs.
8) **LGCM** [41]: LGCM was implemented with $k = \{35, 50, 60\}$ and $k^* = 30$, which are the categories of visual keypoints and visual keypoint pairs.

TABLE VIII
CLASSIFICATION RATES AND STANDARD DEVIATIONS OF THE EVALUATED TEXTURE RECOGNITION METHOD FOR THE CONSIDERED SONAR TEXTURE DATABASE, NAMELY, GABOR FEATURE, COOCCURRENCE FEATURE, BOKS, KEYPOINT COOCCURRENCE STATISTICS [51], MULTIFRACTAL FEATURE [21], KEYPOINT CLASSIFIER [38], SCS, AND LGCM

N_t	1	3	5	8	10
Gabor filter[17]	51.71±3.24	55.34±2.84	59.27±1.97	62.53±1.72	69.81±1.48
Cooccurrence matrix[47]	62.13±3.17	69.72±2.44	72.15±1.53	76.22±1.41	81.21±1.27
BoK[39]	67.15±2.55	76.27±2.08	85.42±1.56	90.35±1.42	92.03±1.21
Ling[51]	66.83±2.33	78.12±2.21	85.27±1.83	90.68±1.35	91.92±1.24
Xu[21]	67.54±2.49	79.33±2.11.	87.12±1.91	90.14±1.63	91.85±1.12
Zhang[38]	73.33±2.17	82.62±1.83	90.67±1.15	93.43±0.91	94.25±0.73
SCS[40]	**74.57±1.69**	**83.23±1.19**	91.17±1.08	93.25±0.58	96.67±0.35
LGCM	73.85±1.75	82.95±1.13	**91.34±0.72**	**93.87±0.61**	**97.14±0.37**

C. Performance Result

We first evaluated the importance of the parameter setting (keypoint types, dimension reduction, scaling invariance, classifier type, and associated parameterization) for the proposed spatial keypoint statistics and models. We detail here these experiments for the descriptors issued from the LGCMs. Similarly, results were obtained for the SCS. Considering the performances of the different classifiers and similarity measures in Table IV, Jeffrey divergence improves the classification performance with an approximate gain greater than 1.5% (3%) compared to χ^2 distance (respectively, Euclidean distance) when five (or ten) training images were used. SVM and RF classifiers lead to similar recognition performances for our sonar data set, but the gain compared to k-NN classifier was greater than 2%. We further evaluated the relevance of the different types of visual keypoints (Table V). It should be pointed out that recognition performances appeared relevant whatever the considered keypoint type is. DoG+SIFT descriptors however outperform the other keypoint types, with a gain slightly greater than 1% over (Har-Lap)+(SIFT–Spin) and (Hes-Lap)+Daisy and 2% over FH+SURF and FH+Brief. These results might be explained by the greater number of keypoints using DoG+SIFT schemes compared to the other combinations such that a finer characterization of the textures can be reached as well as a more robust estimation of the considered statistics.

Selecting DoG+SIFT descriptors as the reference keypoint setting, we emphasize the relevance of the proposed scale adaption and dimension reduction schemes in Table VI. While scale adaption leads to a gain of about 2%, dimension reduction also improves the correct classification rate by about 1.5% when 5–10 training images are available. The choice of the covariance model also affects recognition performance (Table VII). The best performances were obtained with a Gaussian covariance model with 97.14 ± 0.37 versus 96.81 ± 0.48 for a cardinal sine model and 96.12 ± 0.58 for a hyperbolic model when ten training images are used.

The comparison of the different categories of texture descriptors is reported in Table VIII. These results clearly stressed the relevance of the proposed spatial keypoint statistics for sonar texture recognition. Compared to classical cooccurrence and Gabor texture features, the gains were, respectively, greater than 16% and 27%. These improvements were direct benefits from the robustness and invariance of visual keypoints for the analysis of sonar textures which depicted local contrast variations as well as geometric distortions. It should be pointed out that, even

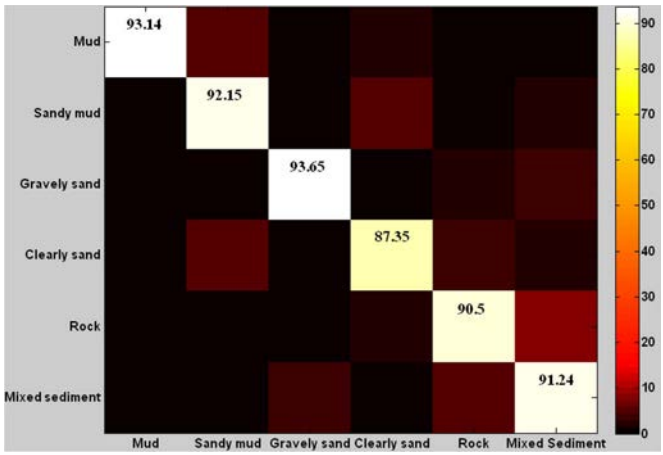

Fig. 10. Confusion matrix of sonar texture classification using the LGCM texture feature. These experiments were carried out using five training images per class.

with the increase of the number of training samples available for each class, cooccurrence and Gabor features would not lead to such high recognition performances greater than 95%. In this respect, all methods exploiting visual keypoints [38], [39], [51] or invariant local features [21] reached correct classification rates greater than 90% when five training images or more are available.

This quantitative evaluation also demonstrated the meaningfulness of the considered spatial keypoint statistics. They outperformed by, respectively, 5% and 2% the BoK feature and the most popular local keypoint (H+L)(SIFT+Spin) method of Zhang's method. Where these two approaches only rely on the discrimination of visual keypoint signatures, spatial keypoint statistics jointly characterize the visual information conveyed by these visual keypoint signatures as well as the spatial distributions of the keypoints. As expected from the observation of keypoint distributions in sonar textures (Fig. 6), the spatial organization of keypoint sets is also a discriminative information that is used to distinguish seabed types. It might be noted that straighforward keypoint cooccurrence statistics proposed by Ling *et al.* [51] were significantly outperformed (97.14% versus 91.92% when ten training images were considered). The combination of scale adaption, dimension reduction, and edge effect corrections explained these improvements.

The proposed spatial statistics, namely, SCS and estimated covariance models, led to very similar recognition

performances, although covariance models were slightly more robust, e.g., 97.14% $\pm$ 0.37 versus 96.67% $\pm$ 0.35 when ten training images were considered. Covariance models should however be preferred as they provide a more compact representation (typically, covariance models lead to $3k^*$-dimensional feature space when the dimensionality of the SCS reach $(N_r + 1)k^*$, where k^* is the size of the codebook of keypoint pairs and N_r is the number of bounded regions; here, $N_r = 20$, and $k^* = 30$) and benefit from a solid theoretical background.

We further detail the class-by-class recognition performances issued from the parametric LGCM when five training images were used (Fig. 10). The reported confusion matrix was nicely balanced over seabed classes, with the mean correct classification rate ranging from 87.35% to 93.65%. The greatest confusions retrieved between clearly sand and sandy mud (5.75%) were consistent in terms of visual similarities of the sonar images.

VI. Conclusion

This paper has addressed the development of invariant descriptors of sonar textures for seabed classification using keypoint-based approaches. We have showed that visual keypoints, developed for computer vision applications and visual texture analysis, also provide meaningful signatures of sonar images to deal with local contrast change and geometric distortions observed in the sonar observations of the seabed. In this respect, statistical descriptors and models of the covariance structure of the multivariate spatial patterns formed by keypoint sets in sonar textures were highly relevant to improve sonar texture characterization and recognition.

The analysis of keypoint detection statistics for various seabed types indicated that DoG keypoints may be more appropriate to deal with the structures observed in sonar images. Whereas visual keypoints were initially developed as corner detectors in natural images, sonar images do not intrinsically involve such corner-like structures. They are rather characterized by the presence of highly reflective areas along with shadow zones. Whereas some keypoint detectors seem too conservative to spatially cover all the structures of interest (e.g., sand ripples in Fig. 2), the detection of the DoG keypoints appears highly correlated to the geometric structures of interest in sonar images. This qualitative evaluation was confirmed by the reported results for seabed recognition. These results open the door for other applications of keypoint-based sonar imaging, including autonomous underwater vehicle navigation [52] and mine detection [53]. Future work might also pursue the development of sonar-specific local signatures.

Regarding the application to seabed recognition, we have reported a quantitative evaluation for a sonar data set involving six different types of seabed, namely, mud, sandy mud, maerly and gravelly sand, clearly sand, rock, and mixed sediment. We have demonstrated that keypoint-based sonar texture characterization can actually achieve very high recognition performances compared to classical cooccurrence and Gabor texture features, which are typically exploited in sonar imaging, while only considering a few training images to learn seabed texture models (typically, 5–10 training samples per class). This aspect is of critical importance for operational applications, in which only a limited subset of the acquired sonar data set is generally groundtruthed by an expert. We believe that this paper could provide a reference benchmarked database for future work on sonar texture analysis. In this respect, the sonar data set is made freely available to the community through a Web access.[7]

From a methodological point of view, we have shown that spatial point processes provide a mathematically sound framework to analyze the spatial organization and the visual signatures of keypoint sets in sonar images. These descriptors that are invariant to local contrast change as well as geometric distortions, especially scaling, provided significant improvements over state-of-the-art techniques. The advantages of parametric point process models were further pointed out in their ability to provide more compact representation of the sonar texture patterns. In future work, the potential of such models will be further explored. Compared to simple descriptive statistics, they can be associated with powerful statistical tools, including goodness-of-fit or hypothesis test. Such tools are of key interest for marine seabed applications to statistically evaluate the significance of the changes of seabed textures over time [54]. Other classes of statistical models may also be investigated, such as Neyman–Scott, shot-noise Cox, or Gibbs processes [42], [55], keeping in mind that we should balance model flexibility and complexity. Combining spatial keypoint process and advanced segmentation schemes is also among the key issues that should be addressed in future work.

Appendix
Intensity and Pair Correlation Function Computation

Considering a univariate log-Gaussian Cox process X on a locally finite subset $S \subset \mathbb{R}^2$, the random intensity function is given by $Z = \exp(Y)$ and covariance functions $c(r) = Cov(Y(s_1), Y(s_2))$, where Y is a Gaussian field on S and $r = \|s_1 - s_2\|$.

1) **The intensity function**

$$\begin{aligned}\rho(s) = E[Z] &= \int_B \exp(y) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_B \exp\left(-\frac{y^2 - 2y(\mu+\sigma^2) + \mu^2}{2\sigma^2}\right) dy \\ &= \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\frac{(\mu+\sigma^2)^2}{2\sigma^2}\right) A_1 \\ &= \exp(\mu + \sigma^2/2) \end{aligned} \quad (20)$$

where $A_1 = (1/\sqrt{2\pi\sigma^2}) \int_B \exp(-((y - (\mu+\sigma^2))^2/2\sigma^2))dy = 1$.

2) **The pair correlation function**:

Regarding the case of $\mu = 0$ with $\hat{r}$ as the correlation coefficient of the Gaussian field, the second-order

[7]The sonar image database is made available from the authors' Web page.

moment is given by

$$\begin{aligned}
&\rho^{(2)}(s_1,s_2)\\
&\quad = E[Z_1,Z_2]\\
&\quad = \int\limits_{B_1\times B_2} \frac{\exp(y_1+y_2)\exp\left(-\frac{\left(y_1^2-\hat{r}y_1y_2+y_2^2\right)}{(1-\hat{r}^2)\sigma^2}\right)}{2\pi\sigma^2\sqrt{1-\hat{r}^2}}dy_1dy_2\\
&\quad = \int\limits_{B_1\times B_2} \frac{exp\left(-\frac{A_2}{(1-\hat{r}^2)\sigma^2}\right)}{2\pi\sigma^2\sqrt{1-\hat{r}^2}}dy_1dy_2 \qquad (21)
\end{aligned}$$

where $A_2 = y_1^2-(2\hat{r}y_1y_2+2(1-\hat{r}^2)\sigma^2(y_1+y_2))+y_2^2$.

Considering the identification

$$\begin{aligned}
&y_1^2-\left(2\hat{r}y_1y_2+2(1-\hat{r}^2)\sigma^2(y_1+y_2)\right)+y_2^2\\
&\quad = (y_1-\kappa)^2-2\hat{r}\left((y_1-\kappa)(y_2-\kappa)\right)+(y_2-\kappa)^2+A_3\\
&\quad = y_1^2-(2\hat{r}y_1y_2-(2\hat{r}\kappa-2\kappa)(y_1+y_2))\\
&\qquad +y_2^2+A_3+2\kappa^2-2\hat{r}\kappa^2. \qquad (22)
\end{aligned}$$

We get

$$\begin{cases} 2(1-\hat{r}^2)\sigma^2 = -2(\hat{r}-1)\kappa \\ A_3+2\kappa^2-2\hat{r}\kappa^2 = 0 \end{cases} \qquad (23)$$

$$\begin{cases} \kappa = \sigma^2(1+\hat{r}) \\ A_3 = 2\kappa^2(\hat{r}-1) = 2\hat{r}^4(\hat{r}^2-1)(\hat{r}+1). \end{cases} \qquad (24)$$

Replacing κ and A_3 into (21), we have

$$\rho^{(2)}(s_1,s_2) = \exp\left(-\frac{A_3}{2(1-\hat{r}^2)\sigma^2}\right) = \exp\left((1+\hat{r})\sigma^2\right). \qquad (25)$$

The correlation function is given by

$$\begin{aligned}
g(s_1,s_2) &= \frac{\rho^{(2)}(s_1,s_2)}{\rho^{(1)}(s_1)\rho^{(1)}(s_2)} = \frac{\exp\left((1+\hat{r})\sigma^2\right)}{\exp(\sigma^2/2)\exp(\sigma^2/2)}\\
&= \exp(\hat{r}\sigma^2) = \exp\left(c(r)\right) \qquad (26)
\end{aligned}$$

where $\hat{r}\sigma^2 = E[Y_1,Y_2] = c(r)$.

REFERENCES

[1] C. Brown, K. Cooper, W. Meadows, D. Limpenny, and H. Rees, "Small-scale mapping of sea-bed assemblages in the eastern English Channel using sidescan sonar and remote sampling techniques," *Estuarine, Coastal Shelf Sci.*, vol. 54, no. 2, pp. 263–278, 2002.

[2] G. Cochrane and K. Lafferty, "Use of acoustic classification of sidescan-sonar data for mapping benthic habitat in the Northern Channel Islands, California," *Continental Shelf Res.*, vol. 22, pp. 683–690, 2002.

[3] W. Dierking, "Mapping of different sea ice regimes using images from Sentinel-1 and ALOS synthetic aperture radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1045–1058, Mar. 2010.

[4] X. Li, C. Li, Q. Xu, and W. Pichel, "Sea surface manifestation of along-tidal-channel underwater ridges imaged by SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2467–2477, Aug. 2009.

[5] J. Davies, J. Baxter, M. Bradley, D. Connor, J. Khan, E. Muray, W. Sanderson, C. Turnbull, and M. Vincent, *Marine Monitoring Handbook (UK Marine SACs Project)*. Peterborough, U.K.: Joint Nature Conservation Committee, 2001.

[6] S. Degraer, V. van Lancker, G. Moerkerke, G. V. Hoey, M. V. P. Jacobs, and J. P. Henriet, "Intensive evaluation of the evolution of a protected benthic habitat: HABITAT. Final report of the Federal Office for Scientific, Technical and Cultural Affairs (OSTC) e ministry of the Flemish community, environment and infrastructure," Waterways Marine Affairs Admin., Coastal Waterways, Belgium, 2002.

[7] A. Ehrhold, D. Hamon, and B. Guillaumont, "The REBENT monitoring network, a spatially integrated, acoustic approach to surveying nearshore macrobenthic habitats: Application to the Bay of Concarneau (France)," *ICES J. Mar. Sci.*, vol. 63, no. 9, pp. 1604–1615, 2006.

[8] C. Hu, F. Muller-Karger, B. Murch, D. Myhre, J. Taylor, R. Luerssen, C. Moses, C. Zhang, L. Gramer, and J. Hendee, "Building an automated integrated observing system to detect sea surface temperature anomaly events in the Florida keys," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 6, pp. 1607–1620, Jun. 2009.

[9] G. L. Chenadec, J.-M. Boucher, and X. Lurton, "Angular dependence of k-distributed sonar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1124–1235, May 2007.

[10] I. Karoui, R. Fablet, J.-M. Boucher, and J.-M. Augustin, "Seabed segmentation using optimized statistics of sonar textures," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 6, pp. 1621–1631, Jun. 2009.

[11] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 1216–1231, Jul. 2000.

[12] J.-M. Bell, M.-J. Chantler, and T. Wittig, "Sidescan sonar: A directional filter or seabed texture?," *Proc. Inst. Elect. Eng.—Radar, Sonar Navig.*, vol. 146, no. 1, pp. 65–72, Feb. 1999.

[13] S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," *IEEE J. Ocean. Eng.*, vol. 28, no. 1, pp. 90–105, Jan. 2003.

[14] A. Martin, "Comparative study of information fusion methods for sonar images classification," in *Proc. Int. Conf. Inf. Fusion*, 2005, vol. 2, pp. 1420–1426.

[15] Y. Wang, Z. Liu, E. Sang, and H. Ma, "Sonar image classification based on directional wavelet and fuzzy fractal dimension," in *Proc. IEEE Conf. Ind. Electron. Appl.*, 2007, pp. 118–120.

[16] N. Sun and T. Shim, "Sonar images classification of seabed physiognomy based on the information fusion methods," in *Proc. Congr. Image Signal Process.*, 2008, pp. 813–819.

[17] T. Randen and J. H. Husøy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 291–310, Apr. 1999.

[18] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognit.*, vol. 42, no. 3, pp. 425–436, Mar. 2009.

[19] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution grayscale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[20] M. Varma and A. Zisserman, "Classifying images of materials: Achieving viewpoint and illumination independence," in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 255–271.

[21] Y. Xu, H. Ji, and C. Fermuller, "Viewpoint invariant texture description using fractal analysis," *Int. J. Comput. Vis.*, vol. 83, no. 1, pp. 85–100, Jun. 2009.

[22] H. Bay, T. Tuytelaars, and L. Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.

[23] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Proc. Eur. Conf. Comput. Vis.*, 2010, vol. IV, pp. 778–792.

[24] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[25] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.

[26] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[27] J. Liu and M. Shah, "Scene modeling using co-clustering," in *Proc. Int. Conf. Comput. Vis.*, 2007, pp. 1–7.

[28] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. Comput. Vis. Pattern Recog.*, 2006, pp. 2169–2178.

[29] R. J. Urick, *Principles of Underwater Sound*. New York: McGraw-Hill, 1983, p. 423.

[30] D. Birkett, C. Maggs, and M. Dring, Marl (Volume V). An Overview of Dynamic and Sensitivity Characteristics for Conservation Management of Marine SACs. Scottish Association for Marine Science (UK Marine SACs Project), p. 1161998.

[31] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, pp. 147–151.
[32] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79–116, Nov. 1998.
[33] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1/2, pp. 43–72, Nov. 2005.
[34] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. Brit. Mach. Vis. Conf.*, 2002, pp. 384–393.
[35] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vis.*, vol. 59, no. 1, pp. 61–85, Aug. 2004.
[36] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 345–457.
[37] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.
[38] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213–238, Jun. 2007.
[39] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 1–22.
[40] H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Spatial statistics of visual keypoints for texture recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2010, vol. IV, pp. 764–777.
[41] H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Visual textures as realizations of multivariate log-Gaussian Cox processes," in *Proc. CVPR*, 2011, pp. 2945–2952.
[42] D. Stoyan and H. Stoyan, *Fractals, Random Shapes and Point Fields*. Chichester, U.K.: Wiley, 1994.
[43] P. Diggle, *Statistical Analysis of Spatial Point Patterns*. London, U.K.: Academic, 1983.
[44] M. Schlather, "On the second-order characteristics of marked point process," *Bernoulli*, vol. 7, no. 1, pp. 99–117, Feb. 2001.
[45] F. Goreaud and R. Pélissier, "On explicit formulas of edge effect correction for Ripley's K-function," *J. Vegetation Sci.*, vol. 10, no. 3, pp. 433–438, Jun. 1999.
[46] J. Møller, A. Syversveen, and R. Waagepetersen, "Log Gaussian Cox processes," *Scand. J. Statist.*, vol. 25, no. 3, pp. 451–482, Sep. 1998.
[47] R. Haralick, "Statistical and structural approaches to textures," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.
[48] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, Jun. 1998.
[49] F. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[50] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *Proc. Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
[51] H. Ling and S. Soatto, "Proximity distribution kernels for geometric context in category recognition," in *Proc. Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
[52] S. Botelho, P. Junior, M. Figueiredo, C. Rocha, and G. Oliveira, "Appearance-based odometry and mapping with feature descriptors for underwater robots," *J. Brazilian Comput. Soc.*, vol. 15, no. 3, pp. 47–54, 2009.
[53] T. Wang, I. Gu, and T. Tjahjadi, "Enhanced landmine detection from low resolution IR image sequences," in *Proc. Comput. Anal. Images Patterns*, 2009, vol. 5702, pp. 1236–1244.
[54] T. Nelson, S. Gillanders, J. Harper, and M. Morris, "Nearshore aquatic habitat monitoring: A seabed imaging and mapping approach," *J. Coastal Res.*, vol. 27, no. 2, pp. 348–355, 2011.
[55] F. Lafarge, G. Gimel'farb, and X. Descombes, "Geometric feature extraction by a multi-marked point process," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1597–1609, Sep. 2010.