

Analysis and Classification of SAR Textures Using Information Theory

Eduarda T. C. Chagas^{ID}, Alejandro C. Frery^{ID}, *Senior Member, IEEE*, Osvaldo A. Rosso^{ID},
and Heitor S. Ramos^{ID}, *Senior Member, IEEE*

Abstract—The use of Bandt–Pompe probability distributions and descriptors of information theory has been presenting satisfactory results with low computational cost in the time series analysis literature [1]–[3]. However, these tools have limitations when applied to data without time dependency. Given this context, we present a newly proposed technique for texture analysis and classification based on the Bandt–Pompe symbolization for SAR data. It consists of linearizing a 2-D patch of the image using the Hilbert–Peano curve, build an ordinal pattern transition graph that considers the data amplitude encoded into the weight of the edges, obtain a probability distribution function derived from this graph, and compute information theory descriptors (permutation entropy and statistical complexity) from this distribution and use them as features to feed a classifier. The ordinal pattern graph we propose considers that the edges’ weight is related to the absolute difference of observations, which encodes the information about the data amplitude. This modification considers the unfavorable signal-to-noise ratio of SAR images and leads to the characterization of several types of textures. Experiments with data from Munich urban areas, Guatemala forest regions, and Cape Canaveral ocean samples show the effectiveness of our technique in homogeneous areas, achieving satisfactory separability levels. The two descriptors chosen in this work are easy and quick to calculate and are used as input for a k-nearest neighbor classifier. Experiments show that this technique presents results similar to state-of-the-art techniques that employ a much larger number of features and, consequently, impose a higher computational cost.

Index Terms—Ordinal patterns transition graphs, permutation entropy, synthetic aperture radar (SAR), texture, terrain classification.

I. INTRODUCTION

TEXTURE is an elusive trait. When dealing with remotely sensed images, the texture of different patches carries relevant information that is hard to quantify and transform into

useful and parsimonious features. This may be since textures, in this context, is a synesthesia phenomenon that triggers tactile responses from visual inputs. This article presents a new way of extracting features from textures, both natural and resulting from anthropic processes, in synthetic aperture radar (SAR) imagery.

SAR systems are a vital source of data because they provide high-resolution images in almost all weather and day-night conditions. They provide basilar information, complementary to that offered by sensors that operate in other regions of the electromagnetic spectrum, for a variety of earth observation applications. Although they present rich information, such data have challenging characteristics. Most notably, they do not follow the usual Gaussian additive model, and the signal-to-noise ratio is usually low.

Yue *et al.* [4] provided a comprehensive account of how the physical properties of the target are translated into first- and second-order statistical properties of SAR intensity data. There is general agreement that nondeterministic textures are encoded in the second-order features, i.e., in the spatial correlation structure. Therefore, is frequent the use of covariance matrix and other measures that assume that a linear dependence, namely the Pearson correlation coefficient, suffices to characterize natural textures. However, in SAR imagery, texture is often visible only over large areas, and the multiplicative and non-Gaussian nature of speckle antagonizes with the additive assumption that underlies classical approaches, making complex the process of characterizing such data.

Surface classification and land use are among the most critical applications of the SAR image [5]. In recent years, handcrafted features and representation learning (supervised and unsupervised) algorithms have been proposed [6]–[8]. Algorithms of the unsupervised generative adversarial network have revolutionized the classification of SAR images, improving performance in small sample problems, and helping the interpretability of such data [9]. Among the supervised algorithms, support vector machine [10], random forest [11], and neural network (NN) [12] have been frequently used in remote sensing. The principle component analysis [13], autoencoder [14] and the Boltzmann machine [15] can to extract nonlocal resources and classify nonlabeled PolSAR pixels using an unsupervised approach. However, methods such as graph-based semisupervised deep learning algorithms [16] can improve classification accuracy in problems with few labeled samples.

Handcrafted features in SAR textures can be studied following two complementary approaches, namely analyzing the marginal

Manuscript received May 14, 2020; revised August 26, 2020 and September 24, 2020; accepted October 9, 2020. Date of publication October 19, 2020; date of current version January 6, 2021. This work was supported in part by the Coordination for the Improvement of Higher Education Personnel and in part by the National Council for Scientific and Technological Development. (Corresponding author: Heitor S. Ramos.)

Eduarda T. C. Chagas and Heitor S. Ramos are with the Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil (e-mail: eduarda.chagas@dcc.ufmg.br; ramosh@dcc.ufmg.br).

Alejandro C. Frery is with the School of Mathematics and Statistics, Victoria University of Wellington, Wellington 6140, New Zealand (e-mail: alejandro.frery@vuw.ac.nz).

Osvaldo A. Rosso is with the Instituto de Física, Universidade Federal de Alagoas, Maceió 57072-900, Brasil (e-mail: oarosso@if.ufal.br).

Digital Object Identifier 10.1109/JSTARS.2020.3031918

properties of the data (first-order statistics), and observing their spatial structure [4], [17]. In this work, we focus on the second approach, which shows relevant results using techniques from the image processing literature, such as co-occurrence matrices and Haralick's descriptors [18]. Through the gray-level co-occurrence matrices (GLCM), we can extract features that reflect statistical relationships of the pixel intensity values. On the other hand, Haralick's descriptors can capture information on intensity and amplitude based on global statistics of SAR images. Radford *et al.* [19] used textural information derived from GLCM, along with random forests, for geological mapping of remote and inaccessible localities; the authors obtained a classification accuracy of $\approx 90\%$, even when using limited training data ($\approx 0.15\%$ of the total data). Hagensieker and Waske [20] evaluated the synergistic contribution of multitemporal L-, C-, and X-band data to tropical land cover mapping, comparing classification outcomes of ALOS-2 [21], RADARSAT-2 [22], and TerraSAR-X [23] datasets for a study site in the Brazilian Amazon using a wrapper approach. The wrapper utilizes the GLCM texture information and a random forest classifier to estimate scene importance. Storie [24] proposed an open-source workflow for detecting and delineating the urban-rural boundary using Sentinel-1A SAR data. The author used a combination of GLCM information and a k-means classifier to produce a three-category map that distinguishes urban from rural areas. In higher resolution image classification activities, it is necessary to obtain more granular information from the data by extracting local characteristics such as scale and orientation. In this scenario, techniques such as Fourier power spectrum [25], random fields [26], Gabor filter [27], and wavelet transform [28] are usually applied.

In our approach, we opt to analyze the 1-D signals resulting from the linearization of the image samples, using nonparametric time series analysis techniques. With this approach, we reduce the dimensionality of the data while preserving the spatial correlation structure. Observations are then transformed into ordinal patterns with the Bandt–Pompe symbolization. We use information theory descriptors to analyze the distributions these patterns induce, both directly and by building transition graphs among subsequent patterns. Those descriptors are the entropy and the statistical complexity, which are easy to obtain and are interpretable. They reveal important features of the underlying process.

The following question guides us.

What is the best representation of a texture patch that allows extracting expressive information theory descriptors to characterize textures in the presence of speckle?

We verified that both the histogram of Bandt–Pompe ordinal patterns and classical transition graphs do not convey enough information for suitable applications as, for instance, classification.

Henceforth, we propose the weighted amplitude transition graph (WATG). This graph incorporates the absolute difference among observations as weights of the edges between nodes transitions. Such weights take part in the computation of the

probabilities and, thus, influence both entropy and statistical complexity.

This work's main contribution is the proposal of a new representation of SAR textures, which allows a low-dimensional characterization useful for, among other applications, their classification. We compare its performance with the classical histograms of Bandt–Pompe ordinal patterns and the regular transition graph. Since the proposed approach has a low computational cost, the results obtained suggest that this technique has good potential in other applications, such as texture segmentation tools of SAR images.

The rest of this article is structured as follows. Section II describes our proposed methodology. Section II-A presents the patch linearization process of the images. In Section II-B, we report the Bandt–Pompe symbolization process. Section II-C describes the ordinal patterns of transition graphs. Section II-E shows our technique of ordinal amplitude transition graph weighting by amplitudes. In Section II-F, we report the information theory descriptors used throughout this work. In Section II-D, we summarize the weighted ordinal patterns methods used to compare with our proposal. Section III describes the SAR image datasets, the analysis of ordinal pattern methods, experiments of sliding window selection, and a quantitative assessment. Finally, Section IV concludes this article.

II. METHODOLOGY

As outlined in Section I, we are interested in methods that employ Information Theory descriptors obtained from ordinal patterns. Such techniques are defined for time series, so the first step of our proposal consists of turning 2-D image patches into a 1-D signal; this is discussed in Section II-A. The second step, presented in Section II-B is obtaining the ordinal patterns. Such patterns can be used directly, or be the basis for building transition graphs; these graphs and their variants are described in Sections II-C and II-D. Our proposal is detailed in Section II-E, and its properties are assessed in Section II-G.

All these transformations produce different empirical probability distributions. The final features computed on these distributions are the entropy and statistical complexity, which are discussed in Section II-F.

A. Linearization of Image Patches

We perform a data dimensionality reduction by turning the 2-D patch into a 1-D signal. This could be accomplished by reading the data by lines, columns, or any transformation of 2-D indexes into a sequence of integers. In this work, we chose to use the Hilbert–Peano [29] curve, due to its low computational cost and its ability to preserve relevant properties of pixel spatial correlation.

Nguyen *et al.* [30] first employed space-filling curves, to map texture into a 1-D signal. Carincotte *et al.* [31] used the Hilbert–Peano curve in the problem of change detection in pairs of SAR images. The authors noted that this transformation exploits the spatial locality and that its pseudorandomness of direction changes work well for a large family of images, especially natural ones.

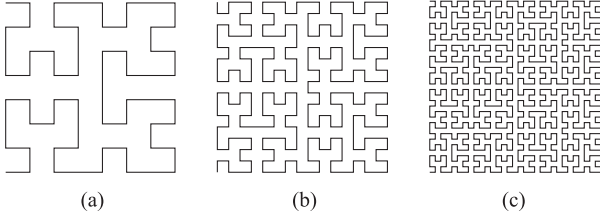


Fig. 1. Hilbert-Peano curves in areas of: (a) 8×8 , (b) 16×16 , and (c) 32×32 pixels.

Assuming an image patch is supported by an $M \times N$ grid, we have the following definition.

Definition 1: An image scan is a bijective function $f: \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ in the ordered pair set $\{(i, j) : 1 \leq i \leq M, 1 \leq j \leq N\}$, which denotes the points in the domain, for the closed range of integers $\{1, \dots, MN\}$. A scan rule is $\{f^{-1}(1), \dots, f^{-1}(MN)\}$.

This definition imposes that each pixel is visited only once and that all pixels are visited.

Space-filling curves, such as raster-1, raster-2, and Hilbert-Peano scanning techniques, stipulate a proper function f . Hilbert-Peano curves scan an array of pixels of dimension $2^k \times 2^k$, $k \in \mathbb{N}$, never keeping the same direction for more than three consecutive points, as shown in Fig. 1. Using the Hilbert-Peano curve, we reduce the data dimensionality while maintaining the patch's spatial dependence information. In this work, we use Hilbert-Peano patches of size 128×128 .

Fig. 6(a)–(e) shows five image patches with different textures. Fig. 6(f)–(j) presents their 1-D representation as signals.

B. Bandt-Pompe Symbolization

Bandt and Pompe [32] introduced the representation of time series by ordinal patterns as a transformation resistant to noise, and invariant to nonlinear monotonic transformations. The first step of the WATG subroutine is to calculate the ordinal patterns of the 1-D signal by Bandt-Pompe symbolization.

Consider $\mathcal{X} \equiv \{x_t\}_{t=1}^T$ a real valued time series of length T . Let \mathfrak{A}_D (with $D \geq 2$ and $D \in \mathbb{N}$) be the symmetric group of order $D!$ formed by all possible permutation of order D , and the symbol component vector $\boldsymbol{\pi}^{(D)} = (\pi_1, \pi_2, \dots, \pi_D)$ so every element $\boldsymbol{\pi}^{(D)}$ is unique ($\pi_j \neq \pi_k$ for every $j \neq k$). Consider for the time series $\mathcal{X} \equiv \{x_t\}_{t=1}^T$ its time delay embedding representation, with embedding dimension $D \geq 2$ and time delay $\tau \geq 1$ ($\tau \in \mathbb{N}$, also called “embedding time,” “time delay,” or “delay”)

$$\mathbf{X}_t^{(D, \tau)} = (x_t, x_{t+\tau}, \dots, x_{t+(D-1)\tau}) \quad (1)$$

for $t = 1, 2, \dots, N$ with $N = T - (D - 1)\tau$. Then, the vector $\mathbf{X}_t^{(D, \tau)}$ can be mapped to a symbol vector $\boldsymbol{\pi}_t^D \in \mathfrak{A}_D$. This mapping is such that preserves the desired relation between the elements $x_t \in \mathbf{X}_t^{(D, \tau)}$, and all $t \in \{1, \dots, T - (D - 1)\tau\}$ that share this pattern (also called “motif”) are mapped to the same $\boldsymbol{\pi}_t^D$.

We define the mapping $\mathbf{X}_t^{(D, \tau)} \mapsto \boldsymbol{\pi}_t^D$ by ordering the observations $x_t \in \mathbf{X}_t^{(D, \tau)}$ in increasing order. Consider the

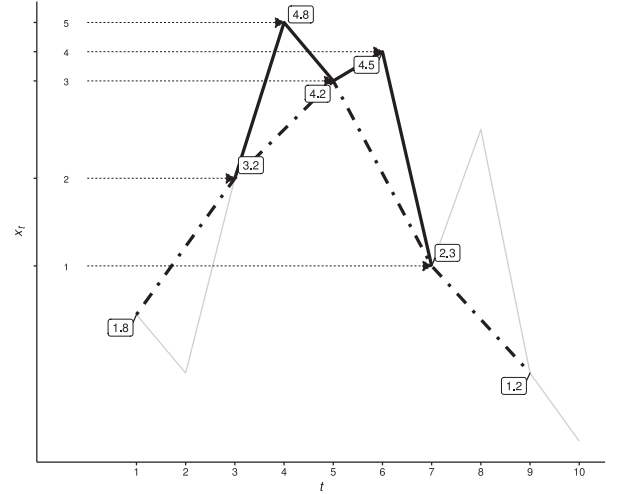


Fig. 2. Illustration of the Bandt and Pompe coding

time series $\mathcal{X} = (1.8, 1.2, 3.2, 4.8, 4.2, 4.5, 2.3, 1.2, .5)$ depicted in Fig. 2. Assume we are using patterns of length $D = 5$ with unitary time lag $\tau = 1$. The code associated to $\mathbf{X}_3^{(5, 1)} = (x_3, \dots, x_7) = (3.2, 4.8, 4.2, 4.5, 2.3)$, shown in black, is formed by the indexes in $\pi_3^5 = (1, 2, 3, 4, 5)$, which sort the elements of $\mathbf{X}_3^{(5, 1)}$ in increasing order: 51342. With this, $\tilde{\pi}_3^5 = 51342$, and we increase the counting related to this motif in the histogram of all possible patterns of size $D = 5$.

The dash-dot line in Fig. 2 illustrates $\mathbf{X}_1^{(5, 2)}$, i.e., the sequence of length $D = 5$ starting at x_1 with lag $\tau = 2$. In this case, $\mathbf{X}_1^{(5, 2)} = (1.8, 3.2, 4.2, 2.3, 1.2)$, and the corresponding motif is $\tilde{\pi}_1^5 = 51423$.

The classic approach to calculating the probability distribution of ordinal patterns is through the frequency histogram. Denote Π the sequence of symbols obtained by a given series $\mathbf{X}_t^{(D, \tau)}$. The Bandt-Pompe probability distribution is the relative frequency of symbols in the series against the $D!$ possible patterns $\{\tilde{\pi}_t^D\}_{t=1}^{D!}$

$$p(\tilde{\pi}_t^D) = \frac{\#\{\mathbf{X}_t^{(D, \tau)} \text{ is of type } \tilde{\pi}_t^D\}}{T - (D - 1)\tau} \quad (2)$$

where $t \in \{1, \dots, T - (D - 1)\tau\}$. These probabilities meet the conditions $p(\tilde{\pi}_t^D) \geq 0$ and $\sum_{i=1}^{D!} p(\tilde{\pi}_t^D) = 1$, and are invariant before monotonic transformations of the time series values. For example, the presence of α multiplicative noise in \mathcal{X} does not change the results of the patterns produced.

C. Graph of Transitions Between Ordinal Patterns

Alternatively, one may form an oriented graph with the transitions from $\tilde{\pi}_t^D$ to $\tilde{\pi}_{t+1}^D$. The ordinal pattern transition graph $G = (V, E)$ represents the transitions between two consecutive ordinal patterns over time t . The vertices are the patterns, and the edges the transitions between them: $V = \{v_{\tilde{\pi}_t^D}\}$, and $E = \{(v_{\tilde{\pi}_t^D}, v_{\tilde{\pi}_{t+1}^D}) : v_{\tilde{\pi}_t^D}, v_{\tilde{\pi}_{t+1}^D} \in V\}$ [33].

The literature reports two approaches to compute the weight of edges. Some authors employ unweighted edges [34], [35],

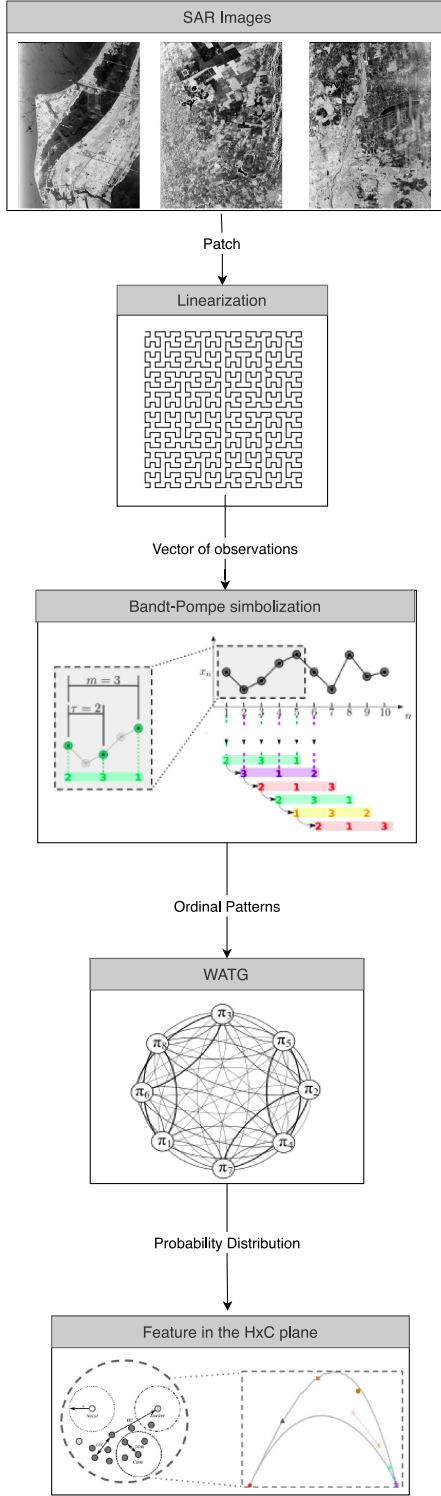


Fig. 3. Outline of the methodology used for the classification of textures.

which represent only the existence of transitions, while others apply the frequency of transitions [36], [37]. The weights $\mathbb{W} = \{w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} : v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D} \in V\}$ assigned to each edge describe the chance of transitions between the patterns $(v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D})$. The weights are calculated as the relative frequency of each

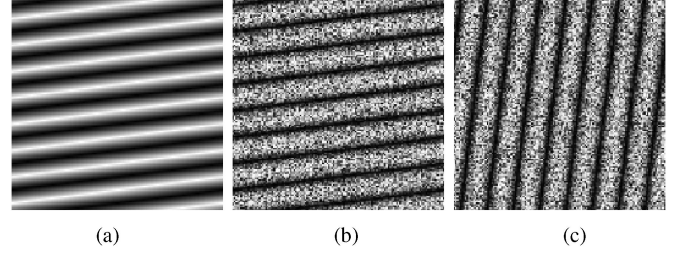


Fig. 4. Ground truth, speckled, and speckled transposed versions. (a) z. (b) I(1). (c) IT(1).

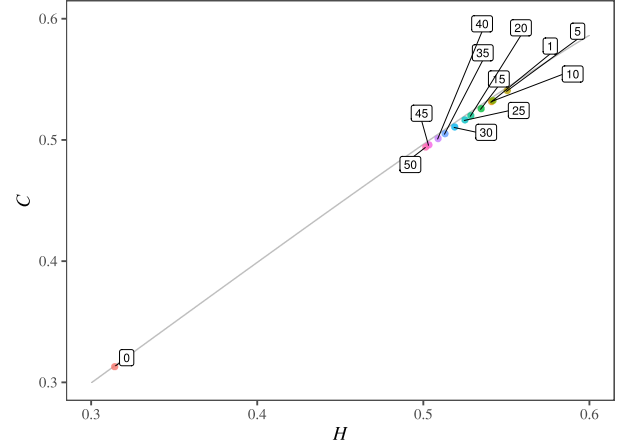


Fig. 5. Modifications to the $H \times C$ plane features by adding different multiplicative noises.

transition, i.e.,

$$w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = \frac{|\Pi_{\tilde{\pi}_i^D, \tilde{\pi}_j^D}|}{T - (D - 1)\tau - 1} \quad (3)$$

where $|\Pi_{\tilde{\pi}_i^D, \tilde{\pi}_j^D}|$ is the number of transitions from pattern $\tilde{\pi}_i^D$ to pattern $\tilde{\pi}_j^D$, $\sum_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = 1$, and the denominator is the number of transitions between sequential patterns in the series of motifs of length $T - (D - 1)\tau$.

D. Weighted Ordinal Patterns Methods

Recent works proposed using weights in the calculation of relative frequencies for ordinal patterns. They all aim at incorporating the information coded in the amplitude of the observations back into the permutation entropy. We summarize in the following those that we used for comparison with our proposal.

1) *Weighted Permutation Entropy*: The weighted permutation entropy (WPE) was proposed by Fadlallah *et al.* [38].

Denote $\bar{X}_t^{(D, \tau)}$ the arithmetic mean

$$\bar{X}_t^{(D, \tau)} = \frac{1}{D} \sum_{k=1}^D x_{t+(k-1)}. \quad (4)$$

The weight w_t is the sample variance of each vector $X_t^{(D, \tau)}$

$$w_t = \frac{1}{D} \sum_{k=1}^D \left[x_{t+(k-1)} - \bar{X}_t^{(D, \tau)} \right]^2. \quad (5)$$

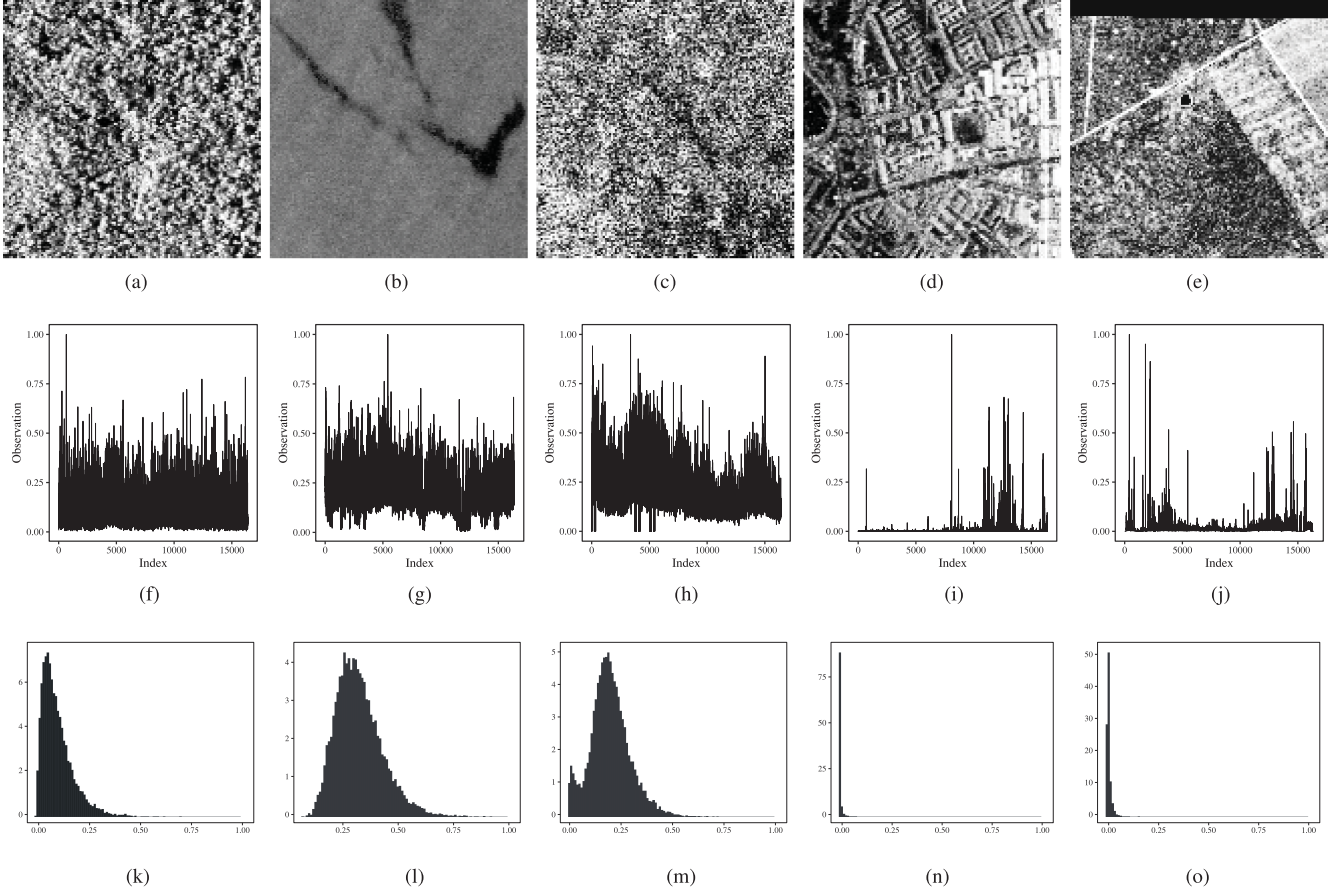


Fig. 6. Types of regions (Guatemala forest, Canaveral ocean types 1 and 2, Munich urban area, and pasture area), their signal representation, and histograms. (a) Forest. (b) Sea – lower contrast. (c) Sea – higher contrast. (d) Urban. (e) Pasture. (f) Signal – forest. (g) Signal – sea. (h) Signal – sea. (i) Signal – urban. (j) Signal – pasture. (k) Histogram – forest. (l) Histogram – sea. (m) Histogram – sea. (n) Histogram – urban. (o) Histogram – pasture.

Then, the probability distribution is given from the weighted relative frequencies

$$p(\tilde{\pi}_t^D) = \frac{\sum_{i:\{\mathbf{X}_i^{(D,\tau)} \mapsto \tilde{\pi}_t^D\}} w_i}{\sum_{i=1}^{T-(D-1)\tau} w_i}. \quad (6)$$

Fig. 7(b) shows the weighted graph produced by the urban area shown in Figs. 6(d) (as image) and (i) (as signal).

2) *Fine-Grained Permutation Entropy*: The fine-grained permutation entropy (FGPE) was introduced in [39].

Let β_t be the difference series

$$\beta_t = \{|x_{t+1} - x_t|, \dots, |x_{t+(D-1)} - x_{t+(D-2)}|\}. \quad (7)$$

The weight w_t quantifies such differences

$$w_t = \left\lfloor \frac{\max\{\beta_t\}}{\alpha s(\beta_t)} \right\rfloor \quad (8)$$

where s is the sample standard deviation, α is a user-defined parameter, and $\lfloor \cdot \rfloor$ is the floor function. Then, w_t is added as a symbol at the end of the corresponding pattern, leading to an update of Π

$$\pi_t^D = \{\tilde{\pi}_t^D \cup w_t\}. \quad (9)$$

Finally, the probability distribution is calculated as

$$p(\pi_t^D) = \frac{\#\{\mathbf{X}_t^{(D,\tau)} \text{ is of type } \pi_t^D\}}{T - (D-1)\tau}. \quad (10)$$

3) *Amplitude-Aware Permutation Entropy*: The amplitude-aware permutation entropy (AAPE) was proposed in [40]. It consists of weighting the amplitude of ordinal patterns by both the mean and the differences of the elements. For this, only an additional parameter $A \in [0, 1]$ is required

$$w_t = \frac{A|x_t|}{D} + \sum_{k=1}^{D-1} \left(\frac{A|x_{t+k}|}{D} + \frac{(1-A)|x_{t+k} - x_{t+k-1}|}{D-1} \right). \quad (11)$$

The probability distribution is given from the weighted relative frequencies

$$p(\tilde{\pi}_t^D) = \frac{\sum_{i:\{\mathbf{X}_i^{(D,\tau)} \mapsto \tilde{\pi}_t^D\}} w_i}{\sum_{i=1}^{T-(D-1)\tau} w_i}. \quad (12)$$

E. Weighted Graph of Transitions Between Ordinal Patterns – WATG

Most man-made targets are anisotropic scatterers because their particular shape determines the relationship between their scatterings and vision directions. Natural targets, such as lawns and forests, are generally isotropic in flat areas because they produce volume scatterings with random phases. In this way, we can associate nonstationary signals with man-made targets and stationary signals with natural targets. Water surfaces cause the mirror-like reflection of the incident electromagnetic wave. It results in an unusually low backscatter, which, thanks to multiplicative noise, translates into a homogeneous region, almost without characteristics and without textures [41]. In any case, speckle produces low signal-to-noise ratio images.

Our proposal hereinafter referred to WATG, differs from the traditional ordinal pattern transition graph by incorporating the absolute difference between successive patterns. Since WATG encodes the amplitude of the signal into the edges' weights, the information on the intensity of dispersion can be captured and used to distinguish classes of regions. Our hypothesis is that such encoding is effective to discriminate the regions captured by SAR imagery.

First, each \mathcal{X} time series is scaled to $[0, 1]$, since we are interested in a metric able to compare datasets

$$\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \mapsto x_i \quad (13)$$

where x_{\min} and x_{\max} are, respectively, the minimum and maximum values of the series. This transformation is relatively stable before contamination, e.g., if instead of x_{\max} we observe kx_{\max} with $k \geq 1$, the relative values are not altered. Nevertheless, other more resistant transformations as, for instance, z scores, might be considered.

Each $\mathbf{X}_t^{(D, \tau)}$ vector is associated with a weight β_t that measures the largest difference between its elements

$$\beta_t = \max\{|x_i - x_j|\} \quad (14)$$

where $x_i, x_j \in \mathbf{X}_t^{(D, \tau)}$.

We propose that the weight assigned to each edge is proportional to the amplitude difference observed in the transition

$$w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = \sum_{i: \{\mathbf{X}_t^{(D, \tau)} \mapsto \tilde{\pi}_i^D\}} \sum_{j: \{\mathbf{X}_t^{(D, \tau)} \mapsto \tilde{\pi}_j^D\}} |\beta_i - \beta_j|. \quad (15)$$

Thus, the probability distribution taken from the weighted amplitude transition graph is given as follows:

$$\begin{cases} \lambda_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = 1, & \text{if } (v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}) \in E, \\ \lambda_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = 0, & \text{otherwise} \end{cases}, \text{ and} \quad (16)$$

$$p(\tilde{\pi}_i^D, \tilde{\pi}_j^D) = \frac{\lambda_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} \cdot w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}}}{\sum_{v_{\tilde{\pi}_a^D}, v_{\tilde{\pi}_b^D}} w_{v_{\tilde{\pi}_a^D}, v_{\tilde{\pi}_b^D}}}. \quad (17)$$

Note that $p(\tilde{\pi}_i^D, \tilde{\pi}_j^D) \geq 0$ and $\sum_{\tilde{\pi}_i^D, \tilde{\pi}_j^D} p(\tilde{\pi}_i^D, \tilde{\pi}_j^D) = 1$, so p is a probability function.

Thus, series with uniform amplitudes have edges with the probability of occurrence well distributed along with the graph, while those with large peaks have edges with the probability of occurrence much higher than the others.

Fig. 7(b) shows the weighted graph produced by the urban area shown in Fig. 6(d) (as image) and (i) (as signal) using this approach. Notice the difference in weights when compared with Fig. 7(b), which was obtained with WPE.

F. Information-Theoretic Descriptors

We chose two information theory descriptors: Shannon entropy and statistical complexity. Computing these quantities is the last step of the algorithm, i.e., obtaining the point in the $H \times C$ plane.

Entropy measures the disorder or unpredictability of a system characterized by a probability measure \mathbb{P} . Let $\mathbb{P} = \{p(\tilde{\pi}_1^D, \tilde{\pi}_1^D), p(\tilde{\pi}_1^D, \tilde{\pi}_2^D), \dots, p(\tilde{\pi}_{D!}^D, \tilde{\pi}_{D!}^D)\} = \{p_1, \dots, p_{D!^2}\}$ be the probability function obtained from the 1-D signal weighted amplitude transition graph \mathbb{X} . Its normalized Shannon entropy is given by

$$H(\mathbb{P}) = -\frac{1}{2 \log D!} \sum_{\ell=1}^{D!^2} p_\ell \log p_\ell. \quad (18)$$

The entropy's ability to capture system properties is limited, so it is necessary to use it in conjunction with other descriptors to obtain a complete analysis. Other interesting measures are the distances between \mathbb{P} and a probability measure that describes a noninformative process, typically the uniform distribution.

The Jensen–Shannon distance to the uniform distribution $\mathbb{U} = (\frac{1}{D!^2}, \dots, \frac{1}{D!^2})$ is a measure of how similar the underlying dynamics is to a noninformative process. It is calculated as

$$Q'(\mathbb{P}, \mathbb{U}) = \sum_{\ell=1}^{D!^2} \left(p_\ell \log \frac{p_\ell}{u_\ell} + u_\ell \log \frac{u_\ell}{p_\ell} \right). \quad (19)$$

This quantity is also called “disequilibrium.” The normalized disequilibrium is $Q = Q' / \max\{Q'\}$.

Conversely to entropy, the statistical complexity seeks to find interaction and dependence structures among the elements of a given series, being an extremely important factor in the study of dynamic systems. The statistical complexity is defined as [42]

$$C(\mathbb{P}, \mathbb{U}) = H(\mathbb{P})Q(\mathbb{P}, \mathbb{U}). \quad (20)$$

In our analysis, each time series can, then, be described by a point $(H(\mathbb{P}), C(\mathbb{P}, \mathbb{U}))$. The set of all pairs $(H(\mathbb{P}), C(\mathbb{P}, \mathbb{U}))$ for any time series described by patterns of length D lies in a compact subset of \mathbb{R}^2 : the entropy-complexity plane. Martín *et al.* [43] obtained explicit expressions for the boundaries of this closed manifold, which depend only on the dimension of the probability space considered, i.e., $D!$ for the traditional Bandt–Pompe method, $D! \times D!$ in our case. Through such a tool it is possible to discover the nature of the series, determining if it corresponds to a chaotic (or other deterministic dynamics) or stochastic sequences.

Algorithm 1 outlines our methodology. Line 1 transforms the input texture patch P in a 1-D signal with a Hilbert–Peano

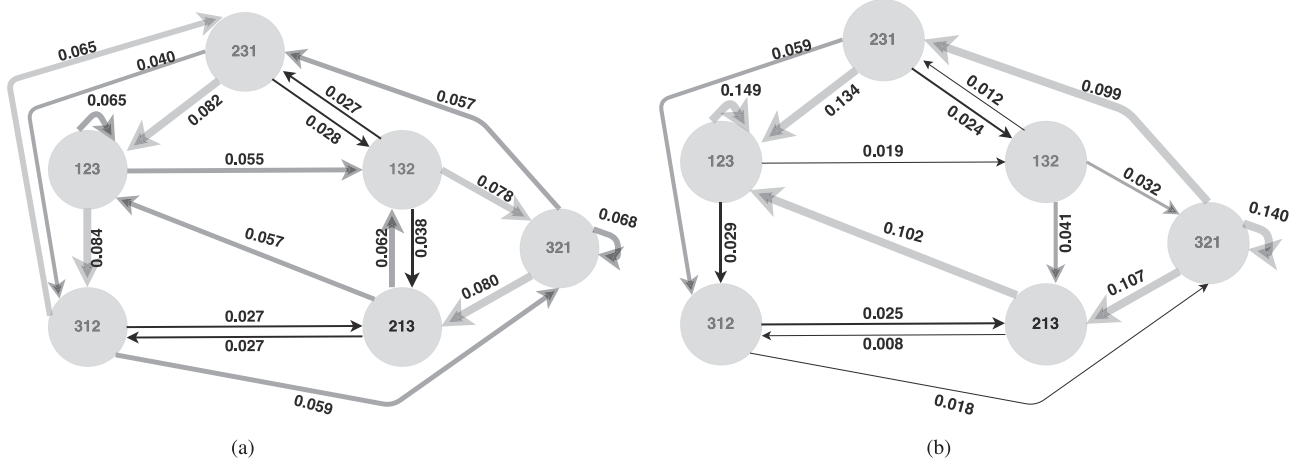


Fig. 7. Difference of edges weights between the transition graph and the weighted graph of ordinal patterns transitions; urban area, with dimension 3 and delay 1. (a) Transition Graph. (b) WATG.

Algorithm 1: $H \times C$ Point from a Patch Using WATG.

Input: Patch of texture P , dimension D and time delay τ

Output: $H \times C$ feature

```

1: signal.1-D  $\leftarrow$  hilbertCurve( $P$ )
2: Probs  $\leftarrow$  WATG(signal.1-D,  $D$ ,  $\tau$ )
3:  $H \leftarrow$  ShannonEntropy(Probs)
4:  $C \leftarrow$  StatisticalComplexity(Probs)
5: return  $H \times C$ 
6: function WATG(signal.1-D,  $D$ ,  $\tau$ )
7:   patterns  $\leftarrow$  BPSymbolization(signal.1-D,  $D$ ,  $\tau$ )
8:   transitions  $\leftarrow$  transitions(patterns)
9:   graph  $\leftarrow$  weighthGraph(signal.1-D, transitions)
10:  Probs  $\leftarrow$  as.vector(graph)
11:
12:
13:  return Probs
14: end Function

```

sequence. With this, the spatial information is encoded into a 1-D signal. Line 2 computes the probability distribution of the weighted transition graph induced by the 1-D signal. The WATG function is detailed in Lines 7–10. Lines 3 and 4 compute the two descriptors of the patch.

The WATG function consists of three steps as follows.

- 1) Each subsequence of size (dimension) D of observations at delay τ is transformed into an ordinal pattern using the Bandt–Pompe symbolization (function `BPSymbolization`, Line 7).
- 2) Function `transitions` (Line 8) calculates the sequence of alternations of the ordinal patterns.
- 3) Function `weighthGraph` (Line 9) generates the incidence matrix of the graph using the as weights the amplitude differences between the time series elements.

Finally, the probability distribution is obtained by turning the transition matrix into a vector (Line 10). These steps are also depicted in Fig. 3.

G. Properties

We conducted two experiments to analyze the response of WATG to different noise levels and image rotations. Our truth is the deterministic image generated by the function

$$z(x, y) = \sin(4x + 0.5y)$$

where $x, y \in [-2\pi, 2\pi]$. Fig. 4(a) shows this function as a 128×128 -pixel patch.

The speckle noise was modeled as outcomes of independent, identically distributed unitary-mean Gamma random variables with shape parameter L (the number of looks, which controls the signal-to-noise ratio) $W(L) \sim \Gamma(L, L)$, with $L \in \{1, 5, 10, 15, \dots, 45, 50\}$. The observed images $I(L)$ are the pixelwise product of z and $w(L)$. Fig. 4(b) shows the product $I(1)$.

Fig. 5 shows how the point in the $H \times C$ plane varies according to the level of noise introduced. The ground truth (identified as “0”) has relatively low entropy and is close to the maximum complexity (the continuous line is the upper bound). This behavior is typical of deterministic sequences. Observe that when we inject single-look speckle ($L = 1$), the entropy increases along with the complexity. Thus, the technique is able to identify the deterministic component even when it is embedded in the strongest possible speckle noise. The point “1” shifts towards “0” when the signal-to-noise progressively increases.

We also verified that the points in the $H \times C$ plane are almost insensitive to rotations. Fig. 4(c) shows the transpose of $I(1)$. In all cases, the coordinates (h, c) of the transposed noisy images were equal, up to the fourth decimal place, to those of the original version.

These experiments provide evidence that the WATG mapping is little sensitive to rotations (thanks to the use of Hilbert–Peano curves), and that it can identify the presence of underlying structural information in the presence of varying levels of speckle.

H. Image Datasets

We used the HH backscatter magnitudes of three quad-polarimetric L-band SAR images from the NASA Jet Propulsion Laboratory's uninhabited aerial vehicle synthetic aperture radar (UAVSAR) sensor with $L = 36$ nominal looks as follows.

- 1) Forest and pasture region of Sierra del Lacandn National Park, Guatemala, (acquired on April 10, 2015)¹. The image has 8917×3300 pixels with 10×2 m resolution.
- 2) Ocean regions from Cape Canaveral Ocean (acquired on September 22, 2016). The image has 7038×3300 pixels with 10×2 m resolution.
- 3) Urban area of the city of Munich, Germany (acquired on June 5, 2015)². The image has 5773×3300 pixels with 10×3 m resolution.

We manually selected 200 samples of size 128×128 to compose the dataset used in the experiments. It is organized as follows: 40 samples from Guatemalan forest regions; 40 samples from Guatemalan pasture regions; 80 samples from the oceanic regions of Cape Canaveral, divided into two types with different contrast; and 40 samples of urban regions of the city of Munich. Fig. 6 shows examples of each. We applied equalization to the patches to improve their visualization. On the one hand, the Bandt–Pompe representation does not change with it because it is a monotonic transformation. On the other hand, WATG does change as the result of altering the observed values. Notice that the subsequent texture analysis employs the unprocessed data. In our analysis, both types of ocean images are grouped.

We randomly split the samples in training (85%) and test (15%) sets. We used the first set to train a k -nearest neighbor classifier algorithm with tenfold cross validation.

III. RESULTS

In this section, we describe the classification process, and the results of applying WATG. To assess the performance of the technique here proposed, we analyze the impact of its parameters and compare its results in the classification with other methods.

A. Analysis of Ordinal Patterns Methods

Fig. 6 shows examples of the ocean, forest, urban, and pasture both as image patches and as 1-D sequences after the linearization process: the first row shows the equalized patches, the second row presents the original data as a signal, and the third row shows the histograms of the original data after scaling with (13).

The variation in the magnitude of the targets' backscatter and, consequently, in the intensity of the image pixels, depends on the intrinsic properties of the region under analysis. Urban targets usually exhibit the strongest variation, followed by forest, pasture, forests, and finally, water bodies. By adding such information related to the amplitude, the proposed method is able

to increase, compared to traditional methods, the granularity of information captured by ordinal patterns.

As already described in Section II-E, our proposal weights the edges in terms of the difference of amplitudes. As expected, the greatest impact is observed on the transition graphs obtained from urban areas. The urban area 1-D signal shown in Fig. 6 has the largest dynamic range. Fig. 7 shows how this information alters the weights of the transition graph. Notice, in particular, that $(v_{\pi_{123}^3}, v_{\pi_{123}^3})$ almost doubled, while $(v_{\pi_{312}^3}, v_{\pi_{231}^3})$ and $(v_{\pi_{213}^3}, v_{\pi_{132}^3})$ became negligible. We highlight the impact of the weighting on the probability distribution in the two extreme cases observed as follows.

- 1) If the 1-D signal presents a low amplitude variation and intensity peaks between, then the transitions of ordinal patterns that represent the latter have larger weights. This contributes so that the probability distribution becomes less uniform among the symbols since it will be more concentrated in these edges. This will also cause a drop in entropy when compared to the traditional method.
- 2) In 1-D signal that shows a uniform amplitude variation, the weights are well distributed between their edges, giving rise to a more random probability distribution, thus obtaining larger entropy.

Fig. 8 shows the impact of using the data amplitude information on the weights of the transition graphs.

The Bandt–Pompe symbolization was the first method based on ordinal patterns proposed in the literature. As shown in Fig. 8 left, it provides a limited separation of the textures. Transition graphs (see Fig. 8 center) improve the spread of the features, but with some amount of confusion. Our proposal, shown in Fig. 8 right, produces well-separated features. In this way, we were able to obtain, for this experiment, a perfect characterization and, consequently, the high descriptive power of the regions.

B. Experiments on Sliding Window Selection

In this section, we analyze the parameters of the proposed method and its impact on textures classification. McCullough *et al.* [34] reported that inadequate values may hinder important characteristics of the phenomenon under analysis. The two parameters of the transition graph are the dimension D , and the delay τ . In the experiments below, we present the results in the classification using different values of these parameters.

The classification method's performance based on ordinal patterns is sensitive to window size, the embedding dimension, and the delay. In techniques based in Bandt–Pompe symbolization, for a fixed signal, as the size of the embedding dimension decreases, more ordinal patterns are produced. Therefore, we acquire a higher granularity of information about the dynamics of the system and, consequently, we capture more spatial dependencies between the elements.

Fig. 9 shows the ROC plane for different values of $D \in \{3, 4, 5, 6\}$ and $\tau \in \{1, 2, 3, 4, 5\}$ to select the best configuration. The configurations that extracted most information from the 1-D signal and, thus, that presented the best results in the experiments, are $(D = 3, \tau = 1)$ and $(D = 4, \tau = 1)$. The

¹Online. [Available]: https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=Lacandn_30202_15043_006_150410_L090_CX_01#dados

²Online. [Available]: https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=munich_19417_15088_002_150605_L090_CX_01#data

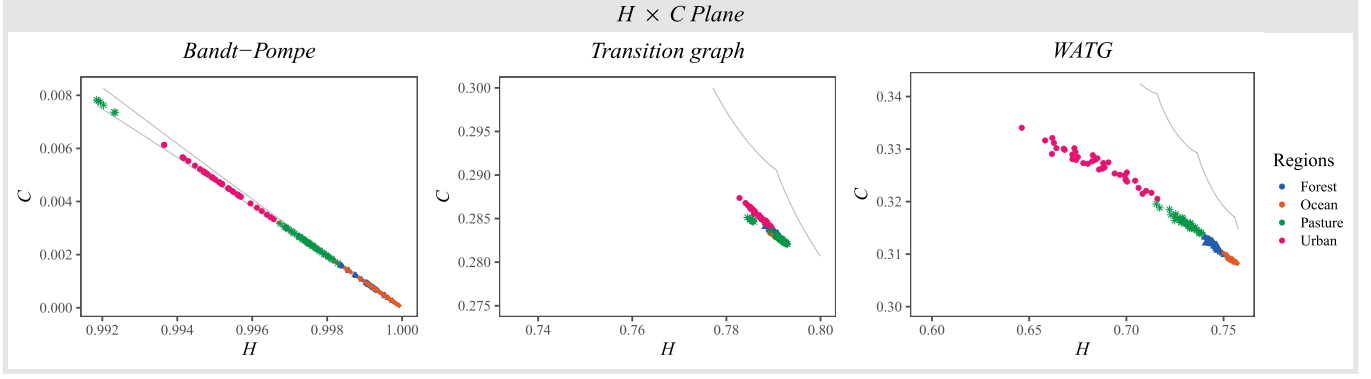


Fig. 8. Location of Guatemala (forest), Cape Canaveral (ocean), and Munich (urban) in the $H \times C$ plane for dimension 3 and delay 1. The continuous curves correspond to the maximum and minimum values of C as a function of H .

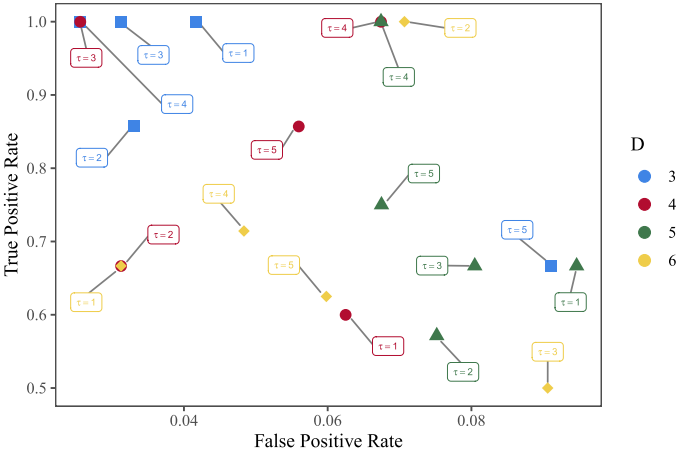


Fig. 9. Evaluation of the sliding window parameters using ROC curve

technique, thus, shows its best performance choosing the parameters with the lowest computational cost.

Fig. 10 shows the points in the $H \times C$ produced by the same samples with all the parameters mentioned previously. The spatial distribution of the points changes with the parameters and specific configurations promote better separation. This figure shows that the discrimination ability decreases with increasing τ . Larger values of delay dilute the spatial dependence, as neighboring points in the sample tend to be more distant in the image. For this reason, we use $\tau = 1$. Fig. 10 suggests that only one feature (H or C) is sufficient to discriminate the classes studied. Although this is true for the experiments herein conducted, we opt to preserve the most common ordinal pattern analysis, which uses both features. As we studied only homogeneous patches, we still do not know how this approach performs with heterogeneous patches. For this last situation, we may need both features.

Considering $\tau = 1$ (see first column of Fig. 10), we also notice that $D = 3$ produces the best separation among classes. Increasing D also increases the statistical complexity; this is noticeable for the forest class. The other effect of considering larger values of D is an increased entropy of ocean and an undesirable overlap with urban samples.

C. Quantitative Evaluation

We present a comparison between our proposal and other methods for texture characterization and classification. We use the following ten methods: Gabor filters [44], Histogram of oriented gradients (HOG) [45], GLCM [46], speeded-up robust features (SURF) [47], short time Fourier transform (STFT) [48] with SURF, Bandt-Pompe probability distribution [32], Ordinal patterns transition graphs [33], WPE [38], FGPE [39] with $\alpha = 0.5$, and AAPE [40] with $A = 0.5$. As in [49], we computed four statistics from co-occurrence matrices: contrast, correlation, energy, and homogeneity. Likewise, we implemented the Gabor filters in five scales and eight orientations; using the energy, we obtained an 80-D feature vector for each patch. For the HOG technique, we used image pixels divided into equal cells of 3×3 pixels, and for each cell, we computed 6-bin histograms ranging from 0° to 180° or 0° to 360° .

We classified the features using the k -nearest neighbor algorithm with Euclidean distance, selecting the value of k with the automatic grid search method of the Caret R package [50]. For validation, we used 10-fold cross validation. More details about the classifier and the sampling can be seen in [51].

Table I presents the number of features each method produces, as well as its performance at classifying the 200 samples. We assessed the effectiveness of each approach using the following metrics. We used the first two metrics (recall and precision) to evaluate classifiers' per class performance and the last three metrics (average accuracy, micro F1-score, and macro F1-score) to evaluate the overall performance of the multiclass classifiers. We denote TP_i , TN_i , FP_i , and FN_i as the true positives, true negatives, false positives, and false negatives counts of a given class i , among a set of K classes, respectively.

- 1) Recall or true positive rate of the class i (TPR_i)

$$TPR_i = \frac{TP_i}{TP_i + FN_i}.$$

- 2) Precision or positive predictive value of the class i (PPV_i)

$$PPV_i = \frac{TP_i}{TP_i + FP_i}.$$

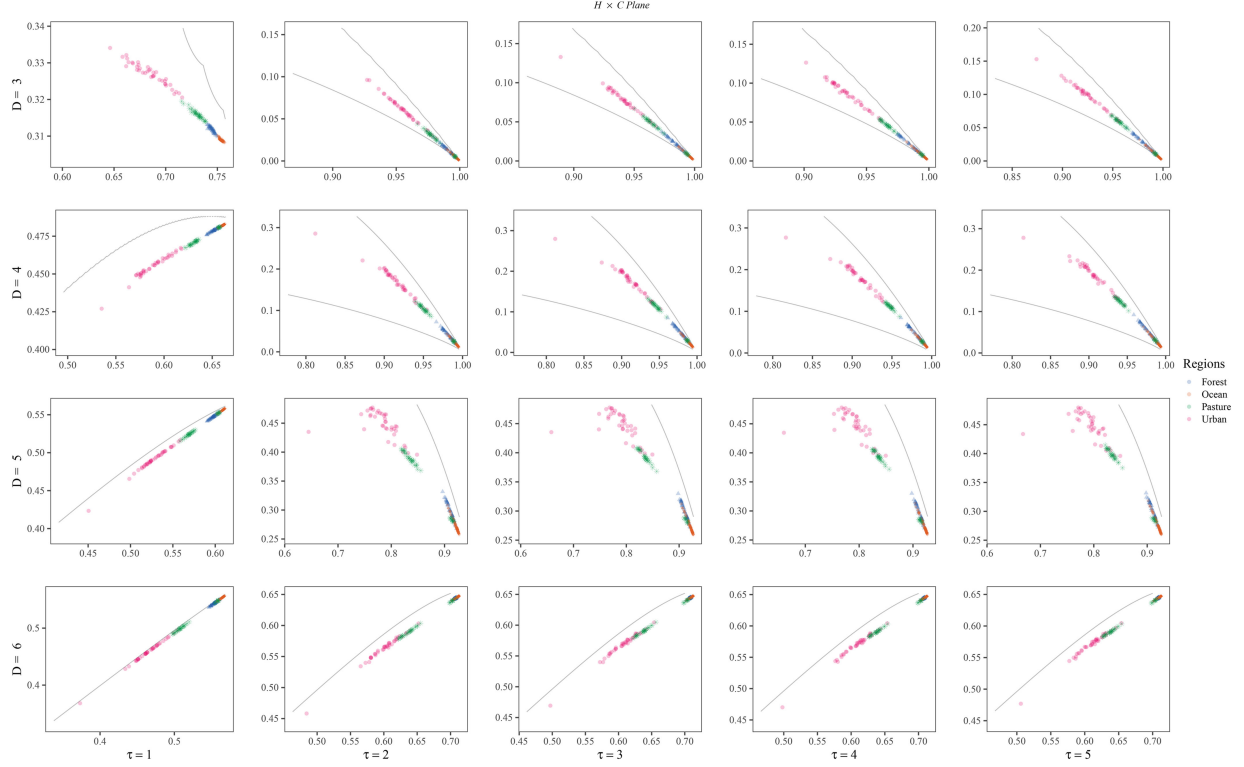


Fig. 10. Characterization resulting in $H \times C$ plane from the application of the Hilbert–Peano curve in WATG on textures of different regions: Guatemala (forest), Cape Canaveral (ocean) and Munich (urban). The continuous curves correspond to the maximum and minimum values of C as a function of H .

TABLE I
EXPERIMENTAL RESULTS USING k -NN

Method	# features	TPR				PPV				AA	F1-Score $_{\mu}$	F1-Score $_M$
		Forest	Pasture	Ocean	Urban	Forest	Pasture	Ocean	Urban			
Gabor	80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
HOG	54	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GLCM	32	0.833	1.000	1.000	0.833	1.000	0.857	0.923	1.000	0.967	0.980	0.970
SURF	1856	0.500	0.000	1.000	0.000	1.000	0.000	0.444	0.000	0.467	0.666	0.572
STFT + SURF	1856	0.166	0.000	0.833	0.166	0.250	0.000	0.416	0.500	0.300	0.462	0.292
Bandt-Pompe	2	0.333	1.000	0.750	1.000	0.500	0.857	0.750	0.857	0.600	0.776	0.633
Transition Graph	2	0.833	0.666	0.833	1.000	0.833	0.800	0.769	1.000	0.767	0.929	0.875
WPE	2	1.000	0.833	1.000	0.833	0.857	0.833	1.000	1.000	0.933	0.868	0.779
AAPE	2	0.666	1.000	1.000	1.000	1.000	0.857	0.923	1.000	0.833	0.947	0.896
FGPE	2	0.666	0.666	1.000	1.000	0.800	0.666	0.923	1.000	0.767	0.868	0.711
WATG	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

3) Average accuracy (AA)

$$AA = \sum_{i=1}^K \left\{ \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \right\}.$$

4) Micro F1-score

$$PPV_{\mu} = \frac{\sum_{i=1}^K TP_i}{\sum_{i=1}^K TP_i + \sum_{i=1}^K FP_i}$$

$$TPR_{\mu} = \frac{\sum_{i=1}^K TP_i}{\sum_{i=1}^K TP_i + \sum_{i=1}^K FN_i}$$

$$F1\text{-score}_{\mu} = 2 \frac{PPV_{\mu} \times TPR_{\mu}}{PPV_{\mu} + TPR_{\mu}}.$$

5) Macro F1-score

$$PPV_M = \frac{1}{K} \sum_{i=1}^K \frac{TP_i}{TP_i + FP_i}$$

$$TPR_M = \frac{1}{K} \sum_{i=1}^K \frac{TP_i}{TP_i + FN_i}$$

$$F1\text{-score}_M = 2 \frac{PPV_M \times TPR_M}{PPV_M + TPR_M}.$$

Table I shows that, among the methods of weighting ordinal patterns, FGPE produced the worst results: AA = 76.7%,

$F1\text{-score}_\mu = 86.8\%$, and $F1\text{-score}_M = 71.1\%$. WPE also produced a low F1-score, but it produced consistently better results in the other metrics, presenting $AA = 93.3\%$. AAPE achieved one of the best F1-score results: $AA = 83.3\%$, $F1\text{-score}_\mu = 94.7\%$, and $F1\text{-score}_M = 89.6\%$. WATG, considering the transition graph of ordinal patterns, can better describe the textures presented, as it achieves the best performance achievable in all metrics 100%.

STFT + SURF produced the worst results: $AA = 30.0\%$, $F1\text{-score}_\mu = 46.2\%$, and $F1\text{-score}_M = 29.2\%$. SURF alone provided a better performance: $AA = 46.7\%$, $F1\text{-score}_\mu = 66.6\%$, and $F1\text{-score}_M = 57.2\%$. HOG and Gabor filters achieved the highest success rates among all the handcrafted methods here considered: $AA = 100\%$, $F1\text{-score}_\mu = 100\%$, and $F1\text{-score}_M = 100\%$. However, WATG achieves that same performance using only 2 features. This reduction implies less computational power requirement and avoids the curse of dimensionality [52]. Moreover, the features it is based upon are fully interpretable.

IV. CONCLUSION

We presented and assessed a new method of analysis and classification of SAR image textures. This method consists of three steps: 1) linearization, 2) computing the weighted ordinal pattern transition graph, and 3) obtaining information theory descriptors. A simple k -NN algorithm applied to the pairs entropy-statistical complexity classifies the data with 100% performance. In addition to such perfect separation among urban, pasture, ocean, and forest areas, the proposed descriptors are interpretable in terms of the degree and structure of the spatial dependence among observations.

Experiments using patches from UAVSAR images showed that the proposal performs better than GLCM, Bandt–Pompe, transition graphs, SURF, STFT + SURF, and other techniques, which also employ amplitude information in the analysis of ordinal patterns. Our approach provides the same quality of results obtained with Gabor filters and HOG. However, while Gabor filters employ 80 features and HOG uses 54 features, our proposal requires only two. This such reduced dimensionality consists of a huge advantage over the other techniques, with added values: First, by reducing the dimension of the features to 2-D, we can visualize the differences between the classes of regions analyzed. Second, for machine learning algorithms, the smaller the number of dimensions, the faster the training process is, and the less storage space is required. Third, we managed to avoid overfitting, a recurring problem in data with high dimensionality.

We also observed that only one feature (H or C) is enough to discriminate the classes with the same reported performance. We opted to preserve both features because we consider that this study shed light on a novel way of SAR image analysis. Thus, we preferred to stick to the most common ordinal pattern analysis using the $H \times C$ plane. As future work, we consider investigating the discriminative power of these features in more complex situations.

Our approach is robust to rotations and the presence of speckle noise. The behavior showed in Fig. 5 shows that our approach can capture the speckle contamination adequately.

Since the application of this work is limited to texture patches from homogeneous regions, we aim to study the possible impacts of heterogeneous areas, such as mixed culture and urban regions.

V. REPRODUCIBILITY AND REPLICABILITY

Following the guidelines presented in [53], the text, source code, and data used in this study are available at the SAR-WATG repository <https://github.com/EduardaChagas/SAR-WATG>. The information includes a link to download the 200 labeled samples we employed in the analysis.

REFERENCES

- [1] A. L. L. Aquino, H. S. Ramos, A. C. Frery, L. P. Viana, T. S. G. Cavalcante, and O. A. Rosso, "Characterization of electric load with information theory quantifiers," *Physica A*, vol. 465, pp. 277–284, 2017.
- [2] O. A. Rosso, R. Ospina, and A. C. Frery, "Classification and verification of handwritten signatures with time causal information theory quantifiers," *PLoS One*, vol. 11, no. 12, 2016, Art. no. e0166868.
- [3] T. A. Schieber, L. Carpi, A. C. Frery, O. A. Rosso, P. M. Pardalos, and M. G. Ravetti, "Information theory perspective on network robustness," *Phys. Lett. A*, vol. 380, pp. 359–364, 2016.
- [4] D.-X. Yue, F. Xu, A. C. Frery, and Y.-Q. Jin, "A generalized Gaussian coherent scatterer model for correlated SAR texture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2947–2964, Apr. 2020.
- [5] J.-S. Lee, M. R. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 722–731, Apr. 2004.
- [6] P. Han, B. Han, X. Lu, R. Cong, and D. Sun, "Unsupervised classification for PolSAR images based on multi-level feature extraction," *Int. J. Remote Sens.*, vol. 41, no. 2, pp. 534–548, 2020.
- [7] Z. Huang, C. O. Dumitru, Z. Pan, B. Lei, and M. Datcu, "Classification of large-scale high-resolution SAR images with deep transfer learning," in *Proc. IEEE Geosci. Remote Sens. Letters*, 2020, doi: 10.1109/LGRS.2020.2965558.
- [8] W. Xie, G. Ma, F. Zhao, H. Liu, and L. Zhang, "PolSAR image classification via a novel semi-supervised recurrent complex-valued convolution neural network," *Neurocomput.*, vol. 388, pp. 255–268, 2020.
- [9] F. Liu, L. Jiao, and X. Tang, "Task-oriented GAN for PolSAR image classification and clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2707–2719, Sep. 2019.
- [10] C. Sukawattananavijit, J. Chen, and H. Zhang, "GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 284–288, Mar. 2017.
- [11] H. McNairn, A. Kross, D. Lapen, R. Caves, and J. Shang, "Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 28, pp. 252–259, 2014.
- [12] Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, "Deep convolutional highway unit network for SAR target classification with limited labeled training data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1091–1095, Jul. 2017.
- [13] R. Ressel, A. Frost, and S. Lehner, "A neural network-based classification for sea ice types on X-band SAR images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3672–3680, Jul. 2015.
- [14] R. Wang and Y. Wang, "Classification of PolSAR image using neural non-local stacked sparse autoencoders with virtual adversarial regularization," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1038.
- [15] F. Qin, J. Guo, and W. Sun, "Object-oriented ensemble classification for polarimetric SAR imagery using restricted boltzmann machines," *Remote Sens. Lett.*, vol. 8, no. 3, pp. 204–213, 2017.
- [16] H. Bi, J. Sun, and Z. Xu, "A graph-based semisupervised deep learning model for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2116–2132, Apr. 2019.

- [17] F. N. Numbisi, F. Van Coillie, and R. De Wulf, "Multi-date Sentinel SAR image textures discriminate perennial agroforests in a tropical forest-savannah transition landscape," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 1, 2018.
- [18] Q. Yu, M. Xing, X. Liu, L. Wang, K. Luo, and X. Quan, "Detection of land use type using multitemporal SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1534–1537.
- [19] D. D. Radford, M. J. Cracknell, M. J. Roach, and G. V. Cumming, "Geological mapping in Western Tasmania using radar and random forests," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3075–3087, Sep. 2018.
- [20] R. Hagensieker and B. Waske, "Evaluation of multi-frequency SAR images for tropical land cover mapping," *Remote Sens.*, vol. 10, no. 2, pp. 257–272, 2018.
- [21] Y. Kankaku, S. Suzuki, and Y. Osawa, "ALOS-2 mission and development status," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2013, pp. 2396–2399.
- [22] L. Morena, K. James, and J. Beck, "An introduction to the RADARSAT-2 mission," *Can. J. Remote Sens.*, vol. 30, no. 3, pp. 221–234, 2004.
- [23] H. Breit, T. Fritz, U. Balss, M. Lachaise, A. Niedermeier, and M. Vonavka, "TerraSAR-X SAR processing and products," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 727–740, Feb. 2009.
- [24] C. D. Storie, "Urban boundary mapping using Sentinel-1A SAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 2960–2963.
- [25] J. B. Florindo and O. M. Bruno, "Fractal descriptors based on Fourier spectrum applied to texture analysis," *Physica A*, vol. 391, no. 20, pp. 4909–4922, 2012.
- [26] T. Zhu, F. Li, G. Heygster, and S. Zhang, "Antarctic sea-ice classification based on conditional random fields from RADARSAT-2 dual-polarization satellite images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2451–2467, Jun. 2016.
- [27] C. O. Dumitru, S. Cui, G. Schwarz, and M. Datcu, "Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 4, pp. 1635–1650, Apr. 2014.
- [28] G. Akbarizadeh, "A new statistical-based kurtosis wavelet energy feature for texture recognition of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4358–4368, Nov. 2012.
- [29] J.-H. Lee and Y.-C. Hsueh, "Texture classification method using multiple space filling curves," *Pattern Recognit. Lett.*, vol. 15, no. 12, pp. 1241–1244, 1994.
- [30] P. Nguyen and J. Quinqueton, "Space filling curves and texture analysis," in *Proc. IEEE Int. Conf. Pattern Recognit.*, 1982, pp. 282–285.
- [31] C. Carincotte, S. Derrode, and S. Bourennane, "Unsupervised change detection on SAR images using fuzzy hidden Markov chains," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 432–441, Feb. 2006.
- [32] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, 2002, Art. no. 174102.
- [33] J. Borges, H. Ramos, R. Mini, O. A. Rosso, A. C. Frery, and A. A. F. Loureiro, "Learning and distinguishing time series dynamics via ordinal patterns transition graphs," *Appl. Math. Comput.*, vol. 362, 2019, Art. no. 124554.
- [34] M. McCullough, M. Small, T. Stemler, and H. H.-C. Iu, "Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems," *Chaos*, vol. 25, no. 5, 2015, Art. no. 053101.
- [35] C. W. Kulp, J. M. Chobot, H. R. Freitas, and G. D. Sprechini, "Using ordinal partition transition networks to analyze ECG data," *Chaos*, vol. 26, no. 7, 2016, Art. no. 073114.
- [36] T. Sorrentino, C. Quintero-Quiroz, A. Aragonese, M. Torrent, and C. Masoller, "Effects of periodic forcing on the temporally correlated spikes of a semiconductor laser with feedback," *Opt. Exp.*, vol. 23, pp. 5571–5581, 2015.
- [37] J. Zhang, J. Zhou, M. Tang, H. Guo, M. Small, and Y. Zou, "Constructing ordinal partition transition networks from multivariate time series," *Sci. Rep.*, vol. 7, 2017, Art. no. 7795.
- [38] B. H. Fadlallah, B. Chen, A. Keil, and J. C. Príncipe, "Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information," *Phys. Rev. E, Stat., Nonlinear, Soft Matter Phys.*, vol. 87 2, 2013, Art. no. 022911.
- [39] L. Xiao-Feng and W. Yue, "Fine-grained permutation entropy as a measure of natural complexity for time series," *Chin. Phys. B*, vol. 18, no. 7, 2009, Art. no. 2690.
- [40] H. Azami and J. Escudero, "Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation," *Comput. Methods Programs Biomed.*, vol. 128, pp. 40–51, 2016.
- [41] W. Wu, H. Guo, and X. Li, "Man-made target detection in urban areas based on a new azimuth stationarity extraction method," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1138–1146, Jun. 2013.
- [42] P. W. Lamberti, M. T. Martín, A. Plastino, and O. A. Rosso, "Intensive entropic non-triviality measure," *Physica A*, vol. 334, no. 1/2, pp. 119–131, 2004. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378437103010963>
- [43] M. T. Martín, A. Plastino, and O. A. Rosso, "Generalized statistical complexity measures: Geometrical and analytical properties," *Physica A*, vol. 369, no. 2, pp. 439–462, 2006.
- [44] T. P. Weldon, W. E. Higgins, and D. F. Dunn, "Efficient Gabor filter design for texture segmentation," *Pattern Recognit.*, vol. 29, no. 12, pp. 2005–2015, 1996.
- [45] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 886–893.
- [46] A. Kourgli, M. Ouarzeddine, Y. Oukil, and A. Belhadj-Aissa, "Texture modelling for land cover classification of fully polarimetric SAR images," *Int. J. Image Data Fusion*, vol. 3, no. 2, pp. 129–148, 2012.
- [47] H. Bay, T. Tuytelaars, and L. Van Gool, *SURF: Speeded Up Robust Features*, vol. 3951. A. Leonardis, H. Bischof, A. Pinz Eds., Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2006, doi: 10.1007/11744023_32.
- [48] M. Portnoff, "Time-frequency representation of digital signals and systems based on short-time fourier analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 1, pp. 55–69, Feb. 1980.
- [49] D. Guan, D. Xiang, X. Tang, L. Wang, and G. Kuang, "Covariance of textural features: A new feature descriptor for SAR image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3932–3942, Oct. 2019.
- [50] M. Kuhn, "Building predictive models in R using the caret package," *J. Stat. Softw.*, vol. 28, no. 5, pp. 1–26, 2008.
- [51] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.
- [52] N. Altman and M. Krzywinski, "The curse(s) of dimensionality," *Nat. Methods*, vol. 15, no. 6, pp. 399–400, May 2018.
- [53] A. C. Frery, L. Gomez, and A. C. Medeiros, "A badging system for reproducibility and replicability in remote sensing research," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4988–4995, Aug. 2020.



Eduarda T. C. Chagas received the graduate degree in computer science from the Federal University of Alagoas, Maceió, Brazil, in 2018. She has been working toward the master's degree in computer science with Federal University of Minas Gerais, Horizonte, Brazil, since 2019.

Her research interests include time series analysis, information theory, urban computing, and statistical computing.



Alejandro C. Frery (Senior Member, IEEE) received the B.Sc. degree in electronic and electrical engineering from the Universidad de Mendoza, Mendoza, Argentina, in 1985, the M.Sc. degree in applied mathematics (statistics) from the Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil, in 1990, and the Ph.D. degree in applied computing from the Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil, in 1993.

He is currently a Professor of Statistics and Data Science with the School of Mathematics and Statistics, Victoria University of Wellington, Wellington, New Zealand, and holds a Huashan Scholar position (2019–2021) with the Key Lab of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, Xian, China. His research interests include statistical computing and stochastic modeling.



Osvaldo A. Rosso was born in Rojas, Buenos Aires, Argentina, in October 15, 1954. He received the M.Sc. and Ph.D. degrees in physics from the Universidad Nacional de La Plata, La Plata, Buenos Aires, in 1978 and 1984, respectively.

He is an Adjoint Professor with Instituto de Física, Universidade Federal de Alagoas, Maceió, Alagoas, Brazil. He is a Huashan Scholar, School of Artificial Intelligence, Xidian University, Xian, China, (2019–2021). From March 1985 to December 2019, he was at Permanent Research Position with the Argentinean

Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina, being Principal Researcher. He has authored or coauthored more than 290 research publications including about 164 papers published in international journals (Hirsch index $H = 41$, Scopus Total of citations = 6278). His current research interests include time series analysis, nonlinear dynamics, information theory, complex networks and their applications to physics, biological, and medical sciences.



Heitor S. Ramos (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Universidade Federal da Paraíba, João Pessoa, Brazil, in 1992, the M.Sc. degree in computing modeling from the Universidade Federal de Alagoas, Maceió, Brazil, in 2004, and the Ph.D. degree in computer science from the Universidade Federal de Minas Gerais, Horizonte, Brazil, in 2012.

He is currently an Associate Professor with the Department of Computer Science, Federal University of Minas Gerais, DCC/UFGM, Brazil. His research interests include wireless networks, sensors networks, mobile and ad hoc networks, and urban computing.