



HAL
open science

Fusion of Polarimetric Features and Structural Gradient Tensors for VHR PolSAR Image Classification

Minh-Tan Pham

► **To cite this version:**

Minh-Tan Pham. Fusion of Polarimetric Features and Structural Gradient Tensors for VHR PolSAR Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, In press, 11 (10), pp.3732-3742. 10.1109/JSTARS.2018.2868545 . hal-02539261

HAL Id: hal-02539261

<https://hal.science/hal-02539261>

Submitted on 9 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fusion of Polarimetric Features and Structural Gradient Tensors for VHR PolSAR Image Classification

Minh-Tan Pham, *Member, IEEE*

Abstract—This article proposes a fast texture-based supervised classification framework for fully polarimetric synthetic aperture radar (PolSAR) images with very high spatial resolution (VHR). With the development of recent polarimetric radar remote sensing technologies, the acquired images contain not only rich polarimetric characteristics but also high spatial content. Thus, the notion of geometrical structures and heterogeneous textures within VHR PolSAR data becomes more and more significant. Moreover, when the spatial resolution is increased, we need to deal with large-size image data. In this work, our motivation is to characterize textures by incorporating (fusing) both polarimetric and structural features, and then use them for classification purpose. First, polarimetric features from the weighted coherency matrix and local geometric information based on the Di Zenzo structural tensors are extracted and fused using the covariance approach. Then, supervised classification task is performed by using Riemannian distance measure relevant for covariance-based descriptors. In order to accelerate the computational time, we propose to perform texture description and classification only on characteristic points, not all pixels from the image. Experiments conducted on the VHR F-SAR data as well as the AIRSAR Flevoland image using the proposed framework provide very promising and competitive results in terms of terrain classification and discrimination.

Index Terms—Polarimetric synthetic aperture radar, textures, gradient tensors, covariance descriptor, riemannian distance, supervised classification

I. INTRODUCTION

CLASSIFICATION of polarimetric synthetic aperture radar (PolSAR) images has become an active research topic for terrain interpretation and land-use understanding using remote sensing imagery in the past few years. Thanks to the ability to operate under any weather conditions and to capture different backscattering characteristics with polarimetric diversity, PolSAR systems offer great advantage compared to optical remote sensing imagery [1]. However, PolSAR image classification task still remains challenging due to the image perturbation caused by speckle noise and the pixel aggregation of heterogeneous terrain types with high intensity variation [2].

Classical methods have been so far using the polarimetric covariance or coherency matrices and their decomposition for classification purpose. Some popular methods were proposed such as the complex Wishart classifier [3], the entropy-based approach using the eigen-decomposition of coherency matrix [4], the improved Wishart classifier with H/α initialization [5] or various multi-component physical-based decomposition methods including the Freeman-Durden three-component model [6], the four-component model by Yamaguchi et al.

[7], the five-component model by Zhang et al. [8], etc. Based on these decomposition approaches, a great number of recent studies have been proposed to alternatively extract and model the coherency matrix's information such as the three-component Fisher-based feature weighting approach [9], the four-component with asymmetric scattering information [10], the generalized polar decomposition based on Mueller matrix [11], the manifold model based on the infinite-dimensional reproducing kernel Hilbert space [12], etc. However, all of these approaches are limited to the use of polarimetric features (derived from the estimated covariance or coherency matrices) and do not take into account the spatial or structural information from the image content.

The development of very high resolution (VHR) sensors offers PolSAR images including not only the fully polarimetric characteristics but also the significant spatial information. Heterogeneous textural and structural content becomes essential and should be taken into account to tackle classification task. To perform texture-based classification, literature studies have proposed to incorporate polarimetric features with several textural measures such as the local binary patterns (LBP), the edge histogram descriptors (EHD), the gray level cooccurrence matrix (GLCM), the Gabor filter banks (GFB) or the wavelet coefficients [13]–[16]. However, due to the influence of speckle noise, such intensity-based LBP, GLCM, Gabor or wavelet textural features do not seem to be robust for SAR images. They may considerably depend on the pre-processing stage of speckle filtering which remains another research issue in PolSAR image analysis. Beside that, inspired by the successful application of spectral-spatial frameworks using morphological profiles (MPs) and attribute profiles (APs) applied to optical images, some studies also proposed to incorporate polarimetric information with multilevel spatial features obtained by MPs and APs to perform classification task [17], [18]. Nevertheless, most of these works independently extract polarimetric, spatial and textural features and then concatenate them to form the combined feature descriptors exploited for classification step. Thus, their performance still depends on every single method to extract each type of features.

In this study, our main motivation is also to incorporate structural features and polarimetric characteristics to tackle the problem of PolSAR image classification. To do that, we propose to first estimate the PolSAR coherency matrix using a weighted multilooking operation based on patch similarity. Polarimetric features are derived from these estimated matrices. Next, the Di Zenzo structural gradient tensors, which were proposed for edge and contour analysis in multi-channel images [19], are extracted to capture the image local geometric properties. Both polarimetric and geometric features are then

M.T. Pham is with the IRISA laboratory - Université de Bretagne Sud, UMR 6074, Vannes 56000, France. e-mail: minh-tan.pham@irisa.fr.

This work was supported by the Région Bretagne grant.

fused together using a region covariance approach [20]. The obtained covariance-based descriptors are used to perform classification task using the support vector machine (SVM) classifier based on geometrical affine-invariant Riemannian distance metric [21]. In order to accelerate the classification process for a fast terrain interpretation, texture description and classification are performed only on characteristic points, not all pixels of the image. Here, the local extrema (i.e. local maximum and local minimum pixels in terms of intensity) will be exploited thanks to their capacity to represent and describe textures within both optical and SAR images [22]–[26].

The remainder of this paper is organized as follows. The next section introduces the proposed framework and describes each processing stage in details. In Section III, we present the experimental study conducted on two real PolSAR image data. Classification results provided by the proposed algorithm are compared to those yielded by several reference methods. Finally, Section IV concludes the paper and discusses some perspective works.

II. METHODOLOGY

An outline of the proposed method can be found in Figure 1. There are four main processing stages including the estimation of weighted coherency matrix, the extraction of structure gradient tensors, their fusion to generate covariance descriptors on keypoints and the final Riemannian kernel-based supervised classification. We now describe each of them in details.

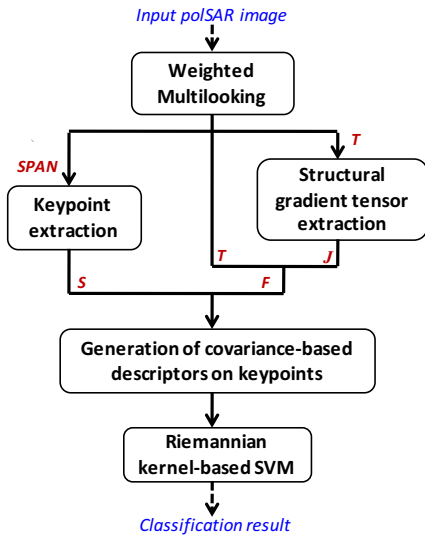


Fig. 1. Proposed texture-based classification framework for VHR PolSAR images. Notations: *SPAN*: total back-scattering power; *T*: coherency matrix; *J*: Di Zenzo structural tensors; *S*: keypoint set and *F*: feature image.

A. Estimation of Weighted Coherency Matrix

Coherency matrix of PolSAR images has been proved to be able to provide significant polarimetric backscattering characteristics used for information extraction and surface interpretation of observed scenes. Generally, the spatial multilooking, i.e. averaging operator, is exploited to reduce the effect of speckle noise. However, such average filters may

induce the over-smoothing phenomenon and hence, cause the problem of losing fine details and local structures from the image content. That is the reason why in this work, we propose to compute the coherency matrix using a weighted multilooking operator which allows us to take into more consideration similar structures from the image during the averaging procedure.

The weighted coherency matrix can be estimated for each image pixel p as follows:

$$T_p = \frac{1}{\sum_{q \in \mathcal{N}_p} w_p(q)} \sum_{q \in \mathcal{N}_p} w_p(q) k_q k_q^H, \quad (1)$$

in which:

- \mathcal{N}_p stands for the neighborhood around pixel p ,
- k is the complex polarimetric scattering vector in Pauli representation [1]:

$$k = \frac{1}{\sqrt{2}} [S_{hh} - S_{vv} \quad 2S_{hv} \quad S_{hh} + S_{vv}]^T,$$

- $w_p(q)$ represents the weight function generated for each neighboring pixel q inside the neighborhood \mathcal{N}_p :

$$w_p(q) = \exp \left\{ -\frac{d^2(p, q)}{\sigma_p^2} \right\}, \quad (2)$$

where $d(p, q)$ involves a dissimilarity measure between two pixels p and q .

Several approaches can be considered to calculate this distance. For the sake of simplicity, we propose to compute $d(p, q)$ as the Euclidean distance measure between the total back-scattering power (SPAN) [1] of the small pixel patches (i.e. patch size from 3×3 to 7×7 pixels) around p and q . The scale parameter σ_p is defined as $\sigma_p = \hat{\sigma}_p = \sqrt{\frac{\pi}{2}} \text{MAD}$ where $\hat{\sigma}_p$ is an estimator based on the mean absolute deviation (MAD) of all the distance values $d(p, q), \forall q \in \mathcal{N}_p$, which is often adopted to estimate the scale parameter of normally distributed data [27].

At this point, it is observed that this weighted averaging estimator is analogous to the non-local SAR (NL-SAR) approach, presented in [28], which is quite effective in the scope of speckle filtering for SAR/PolSAR images. The reason why we do not directly apply the NL-SAR framework is its high complexity. As our main purpose is not the best denoising framework but an efficient coherency matrix estimation for polarimetric feature extraction, we would like to simplify this task to speed up the entire algorithm. For details, in order to compute the similarity weights between patches, the authors in [28] proposed the patch comparison using the generalized likelihood ratio test of pre-estimated coherency matrices which requires huge computation time. In our work, we propose to simplify this patch comparison by using the Euclidean distance between SPAN levels of pixel patches. Our estimated weights might not be as good as those in NL-SAR framework [28] for speckle filtering task but the computation time is much faster for such a feature extraction purpose.

B. Extraction of Di Zenzo Structural Gradient Tensors

Image gradients are considered as important parameters for human perception and analysis of visual features in computer vision and image processing domain. They have so far become key features in several local descriptors such as the scale-invariant feature transform (SIFT), the histograms of oriented gradients (HOG) or the gradient location and oriented histogram (GLOH) (see a survey in [29]). While the gradient computation in mono-channel images is straightforward, multi-channel gradient extraction needs to be considered when dealing with PolSAR images. In this work, we propose to exploit and adapt the principle of Di Zenzo structure gradient tensors which were proposed to tackle the problem of multi-channel image edge/contour analysis in [19]. In remote sensing imagery, this technique has been applied to perform edge detection and enhancement from multispectral and hyperspectral data [30], or tensor-based fusion of panchromatic with multispectral or hyperspectral images [31], [32], etc. Such an approach enables us to take into account the local properties of multi-component image pixels rather than their gray levels. It additionally considers the correlation between channels during the computation of local gradient features.

We now adapt the extraction of Di Zenzo gradient tensors to characterize local geometric structures of PolSAR data. As observed in Figure 1, they are generated from the previous 3×3 weighted coherency matrix T , computed in (1), by considering each matrix component as an image channel. Let us denote I the multi-channel image derived from the matrix T . Since T is symmetric in terms of intensity, the image I consists of 6 channels including 3 diagonal components and 3 off-diagonal components of T as follows:

$$I = [|T_{11}|, |T_{22}|, |T_{33}|, |T_{12}|, |T_{23}|, |T_{13}|]. \quad (3)$$

The three structural gradient tensors [19] from I are then extracted:

$$J_{xx} = \sum_{k=1}^6 \left(\frac{\partial I_k}{\partial x} \right)^2, \quad (4a)$$

$$J_{xy} = \sum_{k=1}^6 \left(\frac{\partial I_k}{\partial x} \right) \left(\frac{\partial I_k}{\partial y} \right), \quad (4b)$$

$$J_{yy} = \sum_{k=1}^6 \left(\frac{\partial I_k}{\partial y} \right)^2, \quad (4c)$$

where the computation of horizontal and vertical derivatives is adapted for SAR images using mean ratio operator [33]:

$$\frac{\partial I_k}{\partial x} = 1 - \min \left\{ \frac{I_k(x+1, y)}{I_k(x-1, y)}, \frac{I_k(x-1, y)}{I_k(x+1, y)} \right\}, \quad (5a)$$

$$\frac{\partial I_k}{\partial y} = 1 - \min \left\{ \frac{I_k(x, y+1)}{I_k(x, y-1)}, \frac{I_k(x, y-1)}{I_k(x, y+1)} \right\}. \quad (5b)$$

Figure 2 shows an example of the structural gradients extracted using the Di Zenzo approach. In Figure 2-a, the image is shown in Pauli color-coded representation (i.e. Red = $|\text{HH-VV}|$, Green = $|\text{HV}|$ and Blue = $|\text{HH+VV}|$). The gradient images in 2-b and 2-c are calculated using Eqs.

(4a) and (4c), respectively, to extract horizontal and vertical structures after accumulating local geometrical and structural information from different polarimetric components present in the coherency matrix T .

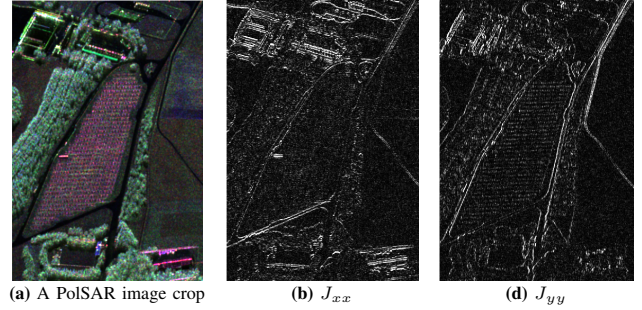


Fig. 2. Example of the Di Zenzo structure gradient tensors extracted from a PolSAR image crop: (a) Pauli color-coded image crop; (b) Gradient tensor J_{xx} ; (c) Gradient tensor J_{yy} .

C. Covariance-based Texture Description on Keypoints

Region covariance has been proposed as an efficient local descriptor for detection and recognition problems [20]. Considered as a natural way of fusing multiple features, covariance-based descriptor is able to take into account their correlated properties by providing both the variance of each feature (i.e. matrix's diagonal entries) as well as the joint distribution between them (i.e. matrix's off-diagonal entries). In remote sensing imagery, it was effectively applied to perform classification of VHR panchromatic optical images [24], [34]. In this work, such a covariance-based fusion technique is exploited to incorporate polarimetric characteristics and local geometric structures for texture characterization in PolSAR images. In fact, the proposed strategy attempts to incorporate geometric features within the well-known coherency matrix in PolSAR imaging. To this end, we construct the feature image F consisting of polarimetric and structural information achieved from the two previous steps as follows:

$$F = [T_{11}, T_{22}, T_{33}, \sqrt{2}T_{12}, \sqrt{2}T_{13}, \sqrt{2}T_{23}, J_{xx}, J_{xy}, J_{yy}]. \quad (6)$$

Due to the huge amount of data achieved by the VHR PolSAR technologies, we also propose to represent and characterize an input image using characteristic points (i.e. keypoints). For this purpose, the local extrema (i.e. local maximum and local minimum pixels in terms of intensity) are exploited since they have been proved to be able to capture significant contextual information of the image in both optical as well as SAR imaging [22]–[26]. Within these works, local extrema pixels were successfully adopted to perform non-dense approach for VHR optical image classification [23], [24], [35] and SAR image change detection [25], [36]. In case of mono-channel images (e.g. panchromatic optical or single-pol SAR), these keypoints can be directly extracted from the image. Here, to deal with multi-channel PolSAR data, we propose to extract them from the SPAN image, which represents the total back-scattering power captured by each pixel. Let us denote S the local extrema set extracted using a

search window $\omega \times \omega$ pixels. Here is its definition:

$$p \in S \Leftrightarrow \left\{ p = \arg \max_{q \in \mathcal{N}_\omega(p)} \text{Span}(q) \mid p = \arg \min_{q \in \mathcal{N}_\omega(p)} \text{Span}(q) \right\}, \quad (7)$$

where $\mathcal{N}_\omega(p)$ represents the set of neighboring pixels of p within the $\omega \times \omega$ search window.

Now, let us consider a neighborhood of size $W \times W$ around a studied keypoint $p \in S$. The feature covariance matrix dedicated to p is estimated as follows:

$$C_p = \frac{1}{W^2} \sum_{q \in \mathcal{N}_W(p)} (f_q - \mu)(f_q - \mu)^T, \quad (8)$$

where $\mathcal{N}_W(p)$ is the studied neighborhood; $f_q \in \mathbb{R}^9$ denotes the feature vector at pixel q ; and $\mu = \frac{1}{W^2} \sum_{q \in \mathcal{N}_W(p)} f_q$ represents the estimated mean feature vector $\mathcal{N}_W(p)$.

The matrix C_p provides a combination of polarimetric characteristics and local geometric structures of the input image located at keypoint p . Hence, it can be considered as a local polarimetric-geometric descriptor dedicated to p . It should be noted that the proposed strategy is not limited to the use of only six polarimetric features and three structural tensors as described above. Other polarimetric, statistical or geometrical, spatial signatures can be also integrated into the generation of F in (6) to perform more complex feature description purpose.

D. Supervised Classification based on AIR Kernel-based SVM

The final stage of the proposed algorithm is to perform supervised classification based on the extracted covariance descriptors. Since covariance matrices possess a semi-positive definite (SPD) structure and do not lie on the Euclidean space, non-Euclidean metrics should be applied to measure the distances between them. In this paper, we propose to exploit the affine-invariant Riemannian (AIR) metric proposed in [21]. This metric has been employed in several works related to covariance descriptors [20], [21], [24], [37], [38]. Another alternative could be to exploit the log-Euclidean (LE) metric which has been proved to be also effective for SPD matrices, especially in terms of computational time [39], [40]. Let us first remind the definition of the LE and AIR distances. We note that some dedicated experimental results will be shown in Section III to compare these two metric and justify our choice of AIR distance. The LE distance d_{LE} and the AIR distance d_{AIR} between two covariance matrices C_1 and C_2 are defined:

$$d_{LE}(C_1, C_2) = \|\log(C_1) - \log(C_2)\|, \quad (9)$$

$$d_{AIR}(C_1, C_2) = \sqrt{\sum_{\ell=1}^d \ln^2 \lambda_\ell}, \quad (10)$$

where λ_ℓ is the ℓ^{th} generalized eigenvalue which satisfies $\lambda_\ell C_1 \chi_\ell - C_2 \chi_\ell = 0$, $\ell = 1 \dots d$ (in this work, $d = 9$) and χ_ℓ is the corresponding eigenvector to λ_ℓ .

To this end, the SVM classifier with pre-computed kernels (available in LIBSVM [41]) is adopted to perform our classification stage. To define such a Riemannian kernel-based SVM

exploited by [41], the selected AIR distance (10) is embedded into a Gaussian kernel of standard deviation σ as follows:

$$K(C_1, C_2) = \exp \left\{ -\frac{d_{AIR}^2(C_1, C_2)}{\sigma^2} \right\}. \quad (11)$$

In the next section, we also investigate classification experiments using the popular radial basis function (RBF) kernel (based on the classical Euclidean metric) as well as the log-Euclidean kernel (by replacing d_{AIR} in (11) with d_{LE}) to compare with the adopted AIR kernel-based SVM.

III. EXPERIMENTAL STUDY

A. Data Description

In order to validate the performance of the proposed framework, experiments were conducted on two PolSAR image data. The first one is an S-band image acquired in Kaufbeuren, Germany in June 2010 by the VHR airborne F-SAR system operated by the German Aerospace Center (DLR). The entire image consists of 8500×17152 pixels with the pixel spacing in azimuth and range of 0.34 m and 0.3 m, respectively. In this work, a ROI (region of interest) of 1500×1500 pixels was extracted to perform the experiments. The studied ROI is shown in Figure 3-a using RGB Pauli color-coded representation (i.e. Red = $|\text{HH-VV}|$, Green = $|\text{HV}|$ and Blue = $|\text{HH+VV}|$). From Figures 3-b to 3-d, we show a reference optical image (taken from Google Maps), the thematic ground truth including five classes (i.e. solar panel, tree, grass pasture, bare soil and building) and the training set. We note that the thematic ground truth (Fig. 3-c) were manually generated with the help of reference maps from OpenStreetMap and Google Maps to provide the most relevant land-cover texture interpretation in our point of view.

For the second dataset, we exploited the popular L-band PolSAR image of Flevoland, the Netherlands, acquired by the NASA/JPL AIRSAR sensor. The image size is 750×1204 pixels with the pixel spacing of 6.6 m in range and 12.1 m in azimuth. Although this is not a very high resolution image, it was chosen to perform our experimental study thanks to its popularity (ground truth is available) in the field. Moreover, despite the relevance of the proposed technique to VHR PolSAR data (the main target of this work), it can be also applied to medium or high resolution images. Back to the dataset, the image scene consists of 11 thematic classes of land cover including different vegetation types, forest, bare soil and water. The image's RGB Pauli representation and its ground truth are shown in Figure 4 for visualization.

B. Experiment Setup

To evaluate the effectiveness of the proposed framework, supervised classification was conducted on our two datasets following the processing description in the previous section. Four experiments were carried out as follows:

a) Experiment 1 - Comparison to reference methods:

Classification results yielded by the proposed strategy will be compared to those from reference approaches including the SPAN (i.e. only use total backscattering power information); the H/ α /A decomposition [4] and their combination with

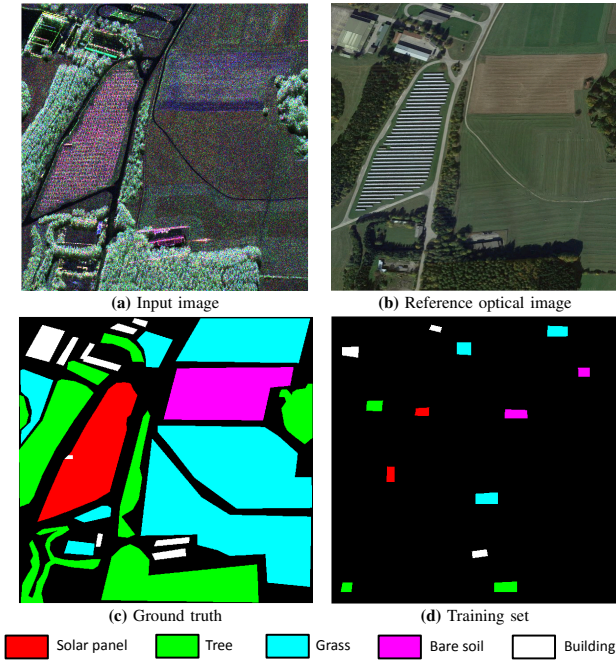


Fig. 3. Dataset 1: the 1500×1500 Kaufbeuren image collected by the VHR F-SAR sensor. Five thematic classes are found in the scene including solar panel, tree, grass field, bare soil and man-made building.

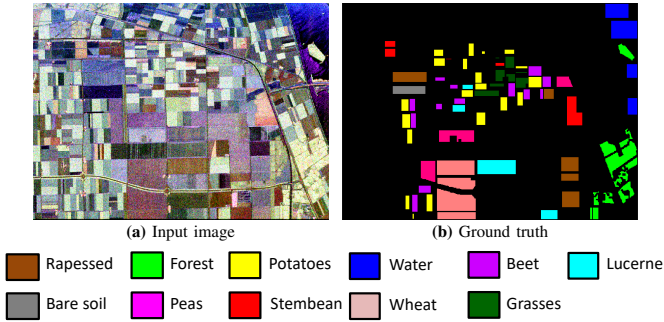


Fig. 4. Dataset 2: the 750×1024 Flevoland data acquired by the NASA/JPL AIRSAR system. The image consists of 11 thematic classes of different vegetation types, forest, bare soil and water.

SPAN and Pauli color information; three popular texture-based methods: the GLCM [42], the Gabor filter banks (GFB) [43], the Weber local descriptor (WLD) [44] and their combination with polarimetric features. We also investigate the spectral-spatial approach using the morphological attribute profiles (AP) [45], [46] combined with polarimetric information for comparison. For these methods, we both extracted their features from the SPAN image (namely $GLCM_{span}$, GFB_{span} , WLD_{span} and AP_{span}) and also on three channels of the Pauli color image, i.e. marginal approaches ($GLCM_{pauli}$, GFB_{pauli} , WLD_{pauli} and AP_{pauli}). For GLCM, 5 common Haralick measures including *contrast*, *correlation*, *energy*, *homogeneity* and *entropy* were extracted by setting a distance offset $(dx, dy) = (0, 1)$, 8 graylevels and 4 directions (0° , 45° , 90° and 135°) to create 20-D GLCM feature vector from each single channel. Similarly, a GBF description vector of 20 Gabor features was extracted with 5 scales and 4 orientations. The WLD parameters were set following the related paper

[44] with $M = 4$, $T = 3$ and $S = 2$ to create the 24-D feature vector. Finally, the commonly-used *area* and *standard deviation* attributes with automatic threshold selection [47] (5 thresholds for each attribute) were adopted to generate 21-D multilevel AP feature vector.

For our method and all above techniques, a sliding window of 7×7 pixels was first applied to estimate the weighted coherency matrix T using (1) (i.e. size of \mathcal{N}_p). Local extrema keypoints were extracted using a 3×3 search window, i.e. $\omega = 3$ in (7). Then, a neighborhood window of 15×15 pixels was applied to the generation of all related local descriptors (i.e. $W = 15$ in (8) for our algorithm). For classification stage, the SVM classifier [41] with RBF kernel was applied to reference methods while the AIR kernel was adopted for our technique.

b) Experiment 2 - Role of structural information: The second experiment aimed at confirming the significant role of structural gradient tensors within the proposed fusion method. The method performance was investigated with and without the use of the tensors J (cf. Section II-B) to fuse with polarimetric characteristics within the feature image F in (6).

c) Experiment 3 - Comparison of different distance metrics: As mentioned in Section II-D, some experiments were also conducted to compare and justify our choice of affine-invariant Riemannian distance metric with the classical Euclidean as well as the log-Euclidean measures.

d) Experiment 4 - Sensitivity to the neighborhood window size: We finally investigated the sensitivity of the proposed scheme to its parameter (W) in terms of both computation time as well as classification performance by varying this window size from 9×9 pixels to 35×35 pixels.

C. Results and Discussions

a) Experiment 1: Figures 5 and 6 show the keypoint-based classification maps of the two image data obtained by the proposed method compared to different reference approaches. For a quantitative evaluation, Tables I and II provide the comparative study on feature dimension and classification performance of all methods. Here, we report the classification accuracy for each thematic class, the *overall accuracy* (OA), the *average accuracy* (AA) as well as the *kappa coefficient* (κ). The feature dimensions of reference methods have been discussed in the experimental setup. For our approach, since covariance matrices are symmetrical, it is sufficient to store only the entries from their upper (or lower) triangular part in the description vector. In our algorithm, the matrix size is 9×9 so that the feature dimension is 45.

For the Kaufbeuren F-SAR data (Dataset 1), the number of keypoints extracted from the whole image (using the 3×3 window) was 48229 from which 1165 were used for training and 33481 were used for testing. In terms of classification performance, we can observe that the proposed framework (Fig. 5-p) provided the best result regarding to the ground truth (Fig. 5-a). The best OA of 98.42% ($\kappa = 0.9743$) was achieved, with a gain of 4.29% compared to the second best AP_{pauli} (Fig. 5-o). Our method also yielded the best classification accuracy for all 5 classes. Most reference techniques produced a lot of misclassified points, in particular from the solar panel



Fig. 5. Classification maps of Dataset 1 obtained by the proposed fusion framework (p) compared to several reference approaches (from b to o). The figure is better visualized with colors.

class (red) and the tree class (green), which contain rich textural information; as well as from the building class (white) from which structural information is very important. Only marginal approaches which extract spatial/textural features from the Pauli-color representation, i.e. $GLCM_{pauli}$ (Fig. 5-g) with $OA = 92.2\%$, GFB_{pauli} (Fig. 5-j) with $OA = 94.06\%$ and AP_{pauli} (Fig. 5-o) with $OA = 94.13\%$, could yield quite clear classification maps with sufficient thematic coherence of the scene. However, they still produced several misclassified points especially between grass and bare soil classes, or between the tree and building classes (i.e. many cyan points within the homogeneous magenta field, a mix of green and white points on the right). For quantitative example, $GLCM_{pauli}$ and GFB_{pauli} provided poor performance

for the tree class (85.04% and 89%) and particularly for the building class (74.32% and 84.01%) compared to the proposed algorithm (98.88% for tree and 94.95% for building).

Next, qualitative and quantitative classification performance of the Flevoland AIRSAR data are showed in Figure 6 and Table II. For this data, 23992 keypoints were detected to perform the algorithm. Due to the lack of split training/test sets, we randomly selected 20% of each class for training and the rest for testing. Training and test sets were ensured to be similar for all methods at each run. All experiments were run 10 times and the average performance was reported. Again, we observe the similar behavior to the first dataset in terms of classification performance of different methods. Most marginal methods ($GLCM_{pauli}$, GFB_{pauli} , WLD_{pauli} and AP_{pauli})

TABLE I
CLASSIFICATION PERFORMANCE OF DATASET 1 YIELDED BY THE PROPOSED APPROACH COMPARED TO REFERENCE METHODS.

Method	Dimension	Per-class accuracy (%)					Overall performance		
		Solar panel	Tree	Grass	Soil	Building	OA (%)	AA (%)	κ
SPAN	1	66.69	63.53	98.26	40.65	47.12	79.26	63.25	0.6550
H/ α /A	3	81.72	80.44	84.00	83.13	70.91	82.86	80.04	0.7343
H/ α /A+SPAN	3 + 1 = 4	82.00	80.71	84.70	85.43	73.68	83.81	81.31	0.7479
H/ α /A+Pauli	3 + 3 = 6	93.33	86.89	83.22	84.94	81.01	84.90	85.88	0.7688
GLCM _{span}	20	35.78	66.20	98.09	62.94	68.75	81.11	66.35	0.6857
GLCM _{span} +Pauli	20 + 3 = 23	75.39	78.08	96.40	75.56	77.88	87.90	80.66	0.8027
GLCM _{pauli}	20 \times 3 = 60	91.30	85.04	96.53	85.94	74.32	92.20	87.03	0.8726
GFB _{span}	20	49.56	66.35	98.71	67.42	74.04	83.75	71.22	0.7287
GFB _{span} +Pauli	20 + 3 = 23	78.45	79.59	96.38	78.35	80.29	88.93	82.61	0.8196
GFB _{pauli}	20 \times 3 = 60	93.86	89.00	97.31	88.65	84.01	94.06	90.57	0.9031
WLD _{span}	24	72.61	75.40	86.22	80.64	69.95	82.17	76.97	0.7179
WLD _{span} +Pauli	24 + 3 = 27	85.46	81.92	92.84	82.53	65.99	88.26	81.75	0.8105
WLD _{pauli}	24 \times 3 = 72	93.70	83.55	94.81	81.22	75.00	90.37	85.66	0.8447
AP _{span}	21	84.03	87.37	87.57	48.34	69.71	79.65	75.41	0.6864
AP _{span} +Pauli	21 + 3 = 24	92.98	93.50	90.82	82.42	75.48	89.45	87.04	0.8349
AP _{pauli}	21 \times 3 = 63	95.04	98.73	95.66	87.13	83.65	94.13	92.04	0.9059
Proposed method	45	99.13	98.88	98.94	96.57	94.95	98.42	97.69	0.9743

provided good classification results (over 94% in general). More importantly, the proposed approach still produced the best classification map where different thematic classes were well discriminated from the observed scene (Fig. 6-l). With an OA of 98.31% ($\kappa = 0.9809$), our scheme achieved a gain of 2.14% compared to the second best method (GFB_{pauli} with $OA = 96.17\%$ in Fig.6-g). In terms of per-class performance, our framework also yielded the best rate for 8/11 classes including rapeseed, forest, water, lucerne, bare soil, peas, wheat and grass. To this end, the comparative results from the two figures (5 and 6) and two tables (I and II) have validated the effectiveness of the proposed framework compared against reference approaches for both datasets.

b) Experiment 2: Table III shows the classification performance of the proposed framework with and without the incorporation of the Di Zenzo structural gradient tensors J with polarimetric features to form the feature image F in (6). As observed, by adding J , the κ score was increased by 4.4% and 3.11% for the Dataset 1 and Dataset 2, respectively. This confirms the important role of structural tensor information within the proposed fusion approach. Our experiments of the Kaufbeuren dataset also showed that the accuracy of the solar panel class and the building class was significantly increased where J was integrated. This remark is encouraging since the objective of exploiting the structural gradients J is to enhance classification performance on highly-structured zones within VHR PolSAR images.

c) Experiment 3: Table IV shows the performance of our method when exploiting the SVM classifier with the selected AIR kernel compared to the RBF kernel (based on the classical Euclidean distance) as well as the LE kernel. We first observe that both AIR and LE metrics are more relevant than the Euclidean distance since they take into account the geometric structure of SPD matrices within their Riemannian space rather than from the classical Euclidean space [21], [39]. Then, when choosing between the AIR and LE metrics, there is a trade-

off between the classification accuracy and the computational time. From the table, we observe the LE metric provided lower classification performance (for both datasets) compared to the AIR metric but in terms of computational time, it is much faster (i.e. complexity comparable to the classical Euclidean distance while the AIR distance requires the matrix's eigen-decomposition). To illustrate, the training (resp. testing) time of AIR kernel applied to the Kaufbeuren dataset is 10.9 times (resp. 78.4 times) slower than that of LE kernel (All implementations were carried out using Matlab 2017a on a PC of 16GB RAM/3.5GHz). However, we argue that in our case of measuring low-dimensional covariance matrix (i.e. 9×9) and only performing classification at keypoint positions (not on dense pixel map), we propose to exploit the AIR metric in order to reach the best classification accuracy. Nevertheless, if one would like to deal with higher dimensional matrices or higher amount of data, the log-Euclidean metric may be a better option to reduce the calculation time. Some further discussion about the AIR and LE metrics applied to covariance descriptors can be also consulted in [48].

d) Experiment 4: The proposed algorithm involves two important parameters: the window size (ω) for extrema keypoint extraction and the neighborhood size (W) for covariance descriptor generation. In fact, the influence of ω on the classification results is not so significant since it controls directly the number of extracted keypoints but does not get involved in the descriptor generation. A smaller or higher ω just provides a denser or coarser keypoint-based image representation and classification. In our experiments, ω was fixed to 3 to provide a high number of keypoints for dense keypoint-based classification maps. When increasing ω , fewer keypoints are extracted and coarser maps are produced, but the classification accuracy does not considerably varied.

Being more important, the parameter W involves directly in the generation of covariance descriptors. It represents the level of exploiting information from neighboring pixels to support

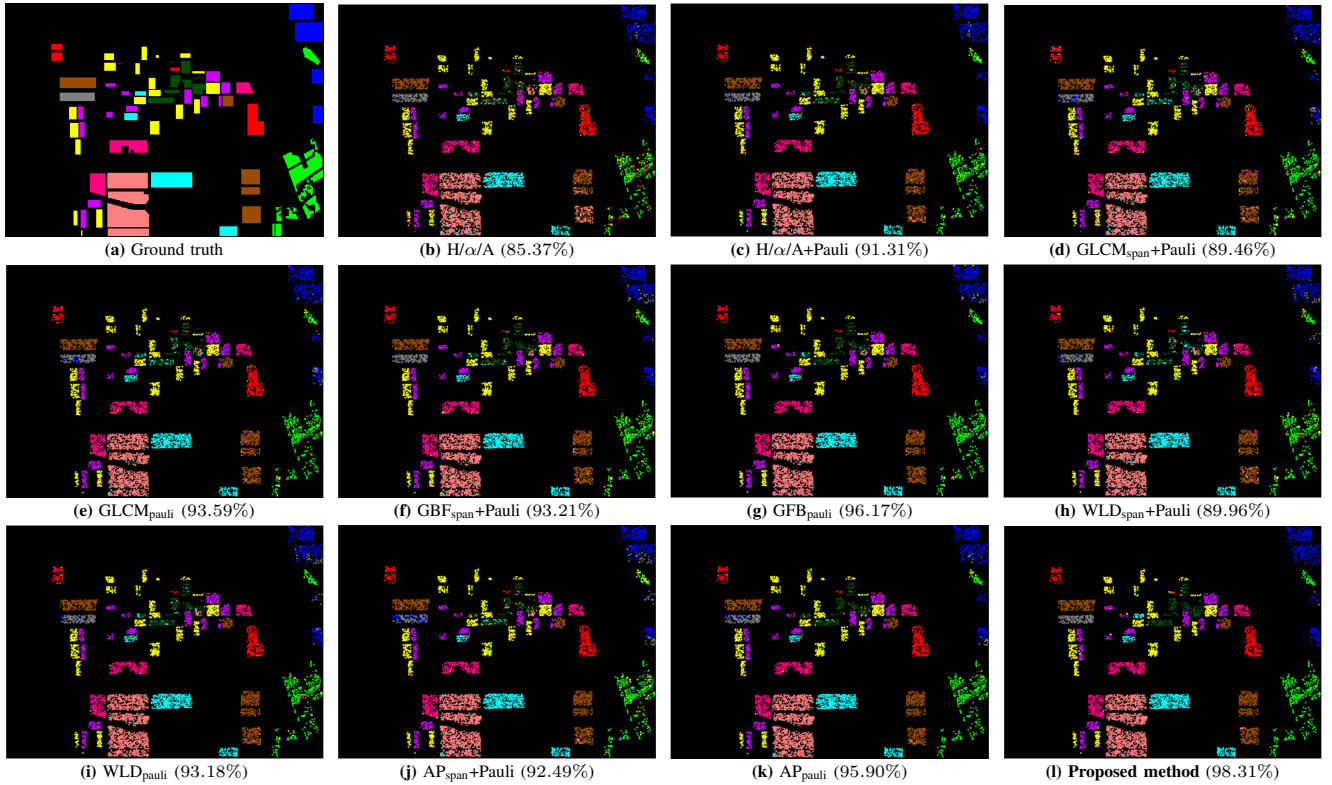


Fig. 6. Classification maps of Dataset 2 obtained by the proposed fusion framework (I) compared to reference approaches (from b to k). The figure is better visualized with colors.

TABLE II
CLASSIFICATION PERFORMANCE OF DATASET 2 YIELDED BY THE PROPOSED APPROACH COMPARED TO REFERENCE METHODS.

Method	Per-class accuracy (%)										Overall performance			
	Rapeseed	Forest	Potatoes	Water	Beet	Lucerne	Bare soil	Peas	Stembeans	Wheat	Grass	OA (%)	AA (%)	κ
H/α/A	84.88	77.30	90.13	95.24	84.37	86.73	92.18	86.17	84.32	87.88	55.26	85.37	84.04	0.8343
H/α/A+SPAN	84.97	80.13	89.87	95.08	86.29	88.35	92.73	87.37	87.89	88.44	56.65	86.36	85.25	0.8456
H/α/A+Pauli	94.73	88.42	94.66	96.85	88.02	91.29	93.45	89.73	96.34	92.72	64.51	91.31	90.07	0.9018
GLCM _{span} +Pauli	92.64	89.01	96.10	90.40	86.53	86.19	84.00	90.72	94.55	90.99	59.54	89.46	87.33	0.8808
GLCM _{pauli}	96.56	94.87	98.52	91.21	91.62	90.69	86.55	95.33	97.84	95.68	67.40	93.59	91.48	0.9275
GBF _{span} +Pauli	95.73	90.00	96.14	91.37	93.05	92.25	94.18	94.07	97.75	94.31	74.91	93.21	92.16	0.9232
GBF _{pauli}	97.91	97.11	99.37	93.15	96.83	93.93	94.73	97.84	99.06	96.45	80.12	96.17	95.13	0.9567
WLD _{span} +Pauli	91.63	91.25	96.73	87.02	90.60	87.21	75.27	91.62	93.80	93.93	53.99	89.96	86.64	0.8863
WLD _{pauli}	95.42	94.67	97.55	88.47	93.95	91.65	81.82	94.67	96.81	95.60	68.21	93.18	90.80	0.9229
AP _{span} +Pauli	96.13	87.43	95.65	83.08	93.51	94.01	89.03	94.83	98.16	96.27	64.23	92.49	90.21	0.9151
AP _{pauli}	98.07	95.26	98.22	92.26	94.41	95.99	95.48	95.31	99.83	96.98	83.71	95.90	95.05	0.9537
Proposed method	98.39	99.87	97.77	99.76	94.91	97.48	98.55	99.40	97.84	99.61	95.61	98.31	98.11	0.9809

TABLE III
IMPACT OF THE STRUCTURAL TENSORS J .

Dataset	Without J		With J	
	OA(%)	κ	OA(%)	κ
Kaufbeuren	95.65	0.9303	98.42	0.9743
Flevoland	95.56	0.9498	98.31	0.9809

the construction of each descriptor. It was set to 15 in all above experiments. We now study the sensitivity of the method by varying W from 9 to 35 pixels. Table V shows the feature extraction time as well as the classification performance in function of W . Experiments were conducted on the Dataset 1. First, it is understandable that the computational time increases when W increases, but this extraction stage is still very fast

(i.e. by using a window of 15×15 pixels, only 1.53 seconds was required). In terms of classification accuracy, when W increases, OA (and also κ) was considerably enhanced at the beginning (from 95.82% to 98.42% when W varies from 9 to 15), then slightly decreased (W from 15 to 25). If we continue to increase W , the accuracy was rapidly decreased (from 98.23% down to 96.58% when W increases from 25 to 35). Here are our explanations. At the first stage, when W increases, more information of the local neighborhood around each keypoint is taken into account. Thus, the performance of the covariance descriptor is improved. When we continue to increase the size of this support neighborhood, although more information is exploited, we may lose the notion of local feature and signal stationarity. This reduces the capacity of our

descriptor to discriminate local structures and textures. Hence, the descriptor performance is decreased. This behavior occurs in fact in most of local descriptors. To this end, according to Table V, the proposed method is not very sensitive to W . A wide range of neighborhood size, i.e. from 10×10 to 30×30 pixels for example, could provide effective classification results for the studied PolSAR data.

TABLE IV
PERFORMANCE OF THE AIR KERNEL COMPARED TO THE RBF KERNEL AND THE LE KERNEL EXPLOITED FOR THE FINAL CLASSIFICATION STEP.

SVM kernel	Accuracy		Time (s)	
	OA (%)	κ	t_{train}	t_{test}
Dataset 1				
RBF	94.05	0.9318	1.21	3.61
LE	97.35	0.9569	1.38	4.86
AIR	98.42	0.9743	13.25	380.94
Dataset 2				
RBF	93.68	0.9312	0.36	1.03
LE	96.12	0.9540	0.47	1.29
AIR	98.31	0.9809	3.15	82.14

TABLE V
SENSITIVITY OF THE PROPOSED METHOD TO THE NEIGHBORHOOD WINDOW SIZE W FOR GENERATION OF COVARIANCE DESCRIPTORS. EXPERIMENTS WERE CONDUCTED ON THE DATASET 1.

Window ($W \times W$)	Time (s)	OA (%)	AA (%)	κ
9×9	0.95	95.82	94.38	0.9321
13×13	1.31	98.19	96.80	0.9705
15×15	1.53	98.42	97.69	0.9743
19×19	2.05	98.19	97.94	0.9705
23×23	2.69	98.23	98.12	0.9712
27×27	3.47	97.90	97.66	0.9659
31×31	3.92	96.98	96.57	0.9509
35×35	4.74	96.58	96.28	0.9445

e) *Further discussions:* The results from Tables IV and V show that the proposed algorithm is very fast in feature extraction stage (only a few seconds for an image containing 1500×1500 pixels like Dataset 1) while classification stage involves the trade-off between the use of AIR and LE metrics as discussed within *Experiment 3*. In general, for an $N \times M$ image, the complexity of our approach has an order of $\mathcal{O}(NM\omega^2 + nW^2)$ in which the first term relates to keypoint detection stage and the latter corresponds to the covariance feature extraction. Here, n is the number of detected keypoints while ω and W were defined in Section II-C. In case of performing a dense feature extraction for all pixels, the complexity becomes $\mathcal{O}(NMW^2)$ (ω is usually much smaller than W). Since $n \ll NM$, such a keypoint-based approach has reduced a lot of computational complexity. Nevertheless, one could remark that this keypoint strategy significantly depends on how much detected keypoints could represent the information of the image. As observed, the use of local extrema is very effective to tackle classification task applied to land-use and land-cover zones as in our experimental data. On the contrary, in case of classification/detection of linear and thin objects (e.g. as narrow streets, roads, power lines),

local extrema keypoints may be not relevant since they usually appear on the object borders. In this case, a more relevant keypoint detection technique should be exploited.

IV. CONCLUSION

We have presented a texture-based classification framework for fully PolSAR data with very high spatial resolution by fusing polarimetric characteristics and structural gradient features. Compared to classical approaches, the proposed method involves three key features: the keypoint-based image representation and classification; the exploitation of structure tensors within covariance-based fusion technique; and the use of geometric affine-invariant Riemannian metric for classification. Throughout this paper, the effectiveness of the proposed method has been confirmed and validated by different experiments conducted on the VHR F-SAR Kaufbeuren data and the AIRSAR Flevoland image. The strategy is quite promising to provide a fast and accurate classification map for terrain interpretation when dealing with large-size and rich-texture PolSAR data. Useful discussions about the trade-off between the use of affine-invariant Riemannian metric or log-Euclidean metric for covariance descriptors have been also addressed.

Future work can continue to improve the performance of the proposed method by incorporating alternative polarimetric, geometric or textural information. One may prefer to generate dense classification maps from non-dense maps provided by our framework by performing interpolation or major voting strategies. Also, covariance-based approach may be exploited for fusing multimodal features from multi-sensor remote sensing data (i.e. SAR/PolSAR, optical and Lidar) to tackle more complex classification tasks.

V. ACKNOWLEDGEMENT

The author would like to thank Dr. Grégoire Mercier, CTO at eXo maKina (ex-professor at Télécom Bretagne, France for his fruitful discussions and Dr. Andreas Reigber from DLR, Germany for providing the VHR airborne F-SAR data.

REFERENCES

- [1] J.-S. Lee and E. Pottier, *Polarimetric radar imaging: From basics to applications*. CRC press, 2009.
- [2] X. Deng, C. López-Martínez, J. Chen, and P. Han, "Statistical modeling of polarimetric SAR data: A survey and challenges," *Remote Sensing*, vol. 9, no. 4, p. 348, 2017.
- [3] J.-S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sens.*, vol. 15, no. 11, pp. 2299–2311, 1994.
- [4] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, 1997.
- [5] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2258, 1999.
- [6] A. Freeman and S. L. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 963–973, 1998.
- [7] Y. Yamaguchi, T. Moriyama, M. Ishido, and H. Yamada, "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1699–1706, 2005.
- [8] L. Zhang, B. Zou, H. Cai, and Y. Zhang, "Multiple-component scattering model for polarimetric sar image decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 603–607, 2008.

- [9] B. Chen, S. Wang, L. Jiao, R. Stolkin, and H. Liu, "A three-component Fisher-based feature weighting method for supervised PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 731–735, 2015.
- [10] B. Zou, Y. Zhang, N. Cao, and N. P. Minh, "A four-component decomposition model for PolSAR data using asymmetric scattering component," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 3, pp. 1051–1061, 2015.
- [11] H. Wang, Z. Zhou, J. Turnbull, Q. Song, and F. Qi, "Pol-SAR classification based on generalized polar decomposition of Mueller matrix," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 4, pp. 565–569, 2016.
- [12] X. Yang, W. Yang, H. Song, and P. Huang, "Polarimetric SAR image classification using geodesic distances and composite kernels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2018.
- [13] Y. Zhang, J. Zhang, G. Huang, and Z. Zhao, "Object-oriented classification of polarimetric SAR imagery based on texture features," in *Int. Symp. Image and Data Fusion (ISIDF)*. IEEE, 2011, pp. 1–4.
- [14] S. Uhlmann and S. Kiranyaz, "Integrating color features in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2197–2216, 2014.
- [15] G. Akbarzadeh and M. Rahmani, "Efficient combination of texture and color features in a new spectral clustering method for PolSAR image segmentation," *National Academy Science Letters*, vol. 40, no. 2, pp. 117–120, 2017.
- [16] C. He, S. Li, Z. Liao, and M. Liao, "Texture classification of PolSAR data based on sparse coding of wavelet polarization textons," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 8, pp. 4576–4590, 2013.
- [17] P. R. Marpu, K.-S. Chen, and J. A. Benediktsson, "Spectral-spatial classification of polarimetric SAR data using morphological attribute profiles," in *Image Sig. Proc. for Remote Sens. XVII*, vol. 8180. International Society for Optics and Photonics, 2011, p. 81800K.
- [18] P. Du, A. Samat, B. Waske, S. Liu, and Z. Li, "Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features," *ISPRS J. Photogram. Remote Sens.*, vol. 105, pp. 38–53, 2015.
- [19] S. Di Zenzo, "A note on the gradient of a multi-image," *Comp. vis., graphics, image proc.*, vol. 33, no. 1, pp. 116–125, 1986.
- [20] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *ECCV*. Springer, 2006, pp. 589–600.
- [21] W. Förstner and B. Moonen, "A metric for covariance matrices," in *Geodesy-The Challenge of the 3rd Millennium*. Springer, 2003, pp. 299–309.
- [22] M.-T. Pham, G. Mercier, and J. Michel, "Textural features from wavelets on graphs for very high resolution panchromatic Pléiades image classification," *RFPT*, vol. 208, pp. 131–136, 2014.
- [23] M.-T. Pham, G. Mercier, and J. Michel, "Pointwise graph-based local texture characterization for very high resolution multispectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 1962–1973, 2015.
- [24] M.-T. Pham, G. Mercier, and J. Michel, "PW-COG: an effective texture descriptor for VHR satellite imagery using a pointwise approach on covariance matrix of oriented gradients," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3345–3359, 2016.
- [25] M.-T. Pham, G. Mercier, and J. Michel, "Change detection between SAR images using a pointwise approach and graph theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2020–2032, 2016.
- [26] M.-T. Pham, G. Mercier, E. Trouvé, and S. Lefèvre, "SAR image texture tracking using a pointwise graph-based model for glacier displacement measurement," in *IGARSS*. IEEE, 2017, pp. 1083–1086.
- [27] L. Sachs, *Applied statistics: A handbook of techniques*. Springer Science & Business Media, 2012.
- [28] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, "NL-SAR: A unified nonlocal framework for resolution-preserving (pol)(in) sar denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2021–2038, 2015.
- [29] T. Tuytelaars, K. Mikolajczyk, et al., "Local invariant feature detectors: a survey," *Foundations and trends® in computer graphics and vision*, vol. 3, no. 3, pp. 177–280, 2008.
- [30] M. Marin-McGee and M. Velez-Reyes, "Coherence enhancement diffusion for hyperspectral imagery using a spectrally weighted structure tensor," in *WHISPERS*. IEEE, 2016, pp. 1–4.
- [31] P. Scheunders and S. De Backer, "Fusion and merging of multispectral images with use of multiscale fundamental forms," *JOSA A*, vol. 18, no. 10, pp. 2468–2477, 2001.
- [32] J. Qu, J. Lei, Y. Li, W. Dong, Z. Zeng, and D. Chen, "Structure tensor-based algorithm for hyperspectral and panchromatic images fusion," *Remote Sensing*, vol. 10, no. 3, p. 373, 2018.
- [33] E. J. Rignot and J. J. Van Zyl, "Change detection techniques for ERS-1 SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 4, pp. 896–906, 1993.
- [34] M.-T. Pham, G. Mercier, and J. Michel, "Pointwise approach on covariance matrix of oriented gradients for very high resolution image texture segmentation," in *IGARSS*. IEEE, 2015, pp. 1008–1011.
- [35] M.-T. Pham, G. Mercier, O. Regniers, and J. Michel, "Texture retrieval from VHR optical remote sensed images using the local extrema descriptor with application to vineyard parcel detection," *Remote Sensing*, vol. 8, no. 5, p. 368, 2016.
- [36] M.-T. Pham, G. Mercier, and J. Michel, "A keypoint approach for change detection between SAR images based on graph theory," in *Multi-Temp*. IEEE, 2015, pp. 1–4.
- [37] M.-T. Pham, G. Mercier, and J. Michel, "Covariance-based texture description from weighted coherency matrix and gradient tensors for polarimetric SAR image classification," in *IGARSS*. IEEE, 2015, pp. 2469–2472.
- [38] M.-T. Pham, G. Mercier, and L. Bombrun, "Color texture image retrieval based on local extrema features and Riemannian distance," *J. Imaging*, vol. 3, no. 4, p. 43, 2017.
- [39] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-Euclidean metrics for fast and simple calculus on diffusion tensors," *Magnetic Resonance in Medicine*, vol. 56, no. 2, pp. 411–421, 2006.
- [40] R. Rosu, M. Donias, L. Bombrun, S. Said, O. Regniers, and J.-P. Da Costa, "Structure tensor Riemannian statistical models for CBIR and classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 248–260, 2017.
- [41] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, p. 27, 2011.
- [42] R. M. Haralick et al., "Textural features for image classification," *IEEE Trans. Sys. Man Cybern.*, vol. 3, no. 6, pp. 610–621, 1973.
- [43] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Patt. Recog.*, vol. 24, no. 12, pp. 1167–1186, 1991.
- [44] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, "WLD: A robust local image descriptor," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1705–1720, 2010.
- [45] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, 2010.
- [46] M.-T. Pham, S. Lefèvre, E. Aptoula, and L. Bruzzone, "Recent developments from attribute profiles for remote sensing image classification," in *ICPRAI*, 2018.
- [47] P. R. Marpu, M. Pedernana, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Automatic generation of standard deviation attribute profiles for spectral-spatial classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 293–297, 2013.
- [48] I. Ilea, L. Bombrun, S. Said, and Y. Berthoumieu, "Covariance matrices encoding based on the log-Euclidean and affine invariant Riemannian metrics," in *CVPR Workshops*, 2018, pp. 393–402.



Minh-Tan Pham (S'13-M'17) received the M.Eng. and M.Res. degrees in Electronics and Telecommunications from the Institute Mines-Telecom, Telecom Bretagne, Brest, France in 2013. He obtained his Ph.D. in Information and Image processing from Telecom Bretagne in collaboration with the French Space Agency (CNES) in 2016. He was intern at the Department of Geography, Laval University, Quebec, Canada in 2013 and at CNES, Toulouse, France in 2015. He is now a post-doctoral researcher at the OBELIX team, IRISA laboratory, Vannes,

France. His research interests include image processing, computer vision and machine learning applied to remote sensing imagery with the current focus on mathematical morphology, hierarchical representation, graph signal processing and deep networks for feature extraction, object detection and classification of remote sensing data. He serves as reviewer of IEEE TIP, IEEE TGRS, IEEE JSTARS, IEEE GRSL and several MDPI journals including Remote Sensing, Sensors and Journal of Imaging. Dr. Pham has been visiting researcher at Technische Universität Berlin (Germany), University of Trento (Italy) and Imperial College London (UK) in June, July and September 2018.