Fisher Vectors for PolSAR Image Classification

Javier Redolfi, Jorge Sánchez, and Ana Georgina Flesia

***Abstract*— In this letter, we study the application of the Fisher vector (FV) to the problem of pixelwise supervised classification of polarimetric synthetic aperture radar images. This is a challenging problem since information in those images is encoded as complex-valued covariance matrices. We observe that the real parts of these matrices preserve the positive semidefiniteness property of their complex counterpart. Based on this observation, we derive an FV from a mixture of real Wishart densities and integrate it with a Potts-like energy model in order to capture spatial dependencies between neighboring regions. Experimental results on two challenging data sets show the effectiveness of the approach.**

***Index Terms*— Fisher vectors (FVs), image classification, polarimetric synthetic aperture radar (PolSAR).**

I. Introduction

A POLARIMETRIC synthetic aperture radar (PolSAR) is an active sensing device capable of providing images that are robust against variations of weather and atmosphere conditions, irrespective of the time of the day they were acquired. These properties make PolSAR images a valuable resource in environmental monitoring applications and for the automated analysis of terrains and land covers [1]–[6].

PolSAR data are generated by transmitting orthogonally polarized electromagnetic pulses toward a target and recording the returned echo for each channel independently. Raw measurements are further processed in order to generate a multichannel image with complex-valued entries. As a consequence of the coherent illumination, the images are contaminated with a particular form of noise known as *speckle* [7]. To reduce the effect of this noise, PolSAR data are aggregated by averaging local information over small neighborhoods, resulting in the so-called *multilook* representation of the PolSAR data [8].

One of the characteristics that make PolSAR images unique—and perhaps one of the reasons why they are not widely used—is that the visual information (pixel values) is encoded in the form of complex-valued vectors or matrices. This makes the application of standard techniques, known from the statistical and machine learning literature, very difficult. Dealing with such data while taking into account its underlying structure is a challenging problem and the main purpose of this letter.

In the classification literature, one of the first methods proposed to address this problem was the complex Wishart classifier (CWC) of Lee *et al*. [1]. The CWC is based on the idea of fitting a complex Wishart pdf for each class and classifying pixels based on a maximum *a posteriori* criterion. Note that, by fitting a single pdf, the model implicitly assumes that samples within each class are homogeneous. Besides classification, the complex Wishart distribution has also been used successfully as a statistical model for other image analysis tasks like segmentation or boundary detection (see [7], [9] and the references therein). To move beyond the homogeneous case, Gao *et al*. [2] proposed to extend the CWC by considering finite mixtures of complex Wishart pdfs. Experimental results confirmed the advantages of this model in dealing with nonhomogeneous terrains. Mixture models have also been considered in [10] and [11] for PolSAR data modeling and classification. Doulgeris *et al.* [10] arrive to a generalization of the complex Wishart pdf by analyzing the distribution of the sample covariance for the scaled mixture of Gaussians (SMoG) model. This model originates from the product of a scalar (scale) and a multivariate complex Gaussian random variable, where the scalar and Gaussian terms account for the texture and speckle, respectively. In the SMoG, the distribution for the covariance of the composite variable can be estimated by integrating over the distribution of the scale term. As such, this model can be regarded as a *continuous* mixture model.

More recently, Wang *et al*. [3] proposed a model based on the polar decomposition of the Mueller matrix and showed good results. Jiao and Liu [4] proposed the Wishart deep stacking network (W-DSN) that, as the models proposed in [5] and [6], takes advantage of powerful techniques developed in the machine learning literature. However, training this kind of model requires large amounts of high-quality annotated data, which in some cases may be difficult or even impractical to gather.

In this letter, we present a first study on the application of a family of models, originated in the computer vision literature, and apply it to the problem of terrain-type classification, i.e., the task of assigning labels to pixels based on the scattering properties of the target as measured by the PolSAR sensor. We propose a model that integrates the Fisher vector (FV) formalism [12], [13] with a Potts-like energy model [14] that captures the spatial dependency between the variables. In the FV framework, the image content (pixels, regions, and/or the whole image) is characterized by a normalized gradient vector derived from a suitable mixture distribution. In our case, we consider the real part of the PoLSAR covariances and derive an FV from a mixture of real Wishart pdfs

Manuscript received April 5, 2017; revised July 20, 2017; accepted August 26, 2017. This work was supported in part by CONICET under Grant PIP-11220150100656-CO, in part by ANPCyT under Grant PICT-2016-0118, and in part by SeCyT-UNC under Grant RES.313/16. *(Corresponding author: Jorge Sánchez.)*

J. Redolfi is with CONICET and CIII, Universidad Tecnológica Nacional, Facultad Regional Córdoba, Córdoba X5016ZAA, Argentina (e-mail: jredolfi@frc.utn.edu.ar).

J. Sánchez and A. G. Flesia are with the CIEM-CONICET and FaMAF, Universidad Nacional de Córdoba, Córdoba X5000HUA, Argentina (e-mail: jsanchez@famaf.unc.edu.ar; flesia@famaf.unc.edu.ar).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2017.2750800

by first showing that these matrices preserve the symmetric positive semidefiniteness property of their complex counterpart. We, then, define a Potts-like energy where the unary terms are computed as the negative inner product between per-class and pixel-level FVs. Minimization of this energy over the graph of four-connected pixel locations gives us the desired classification.

Our main contribution is a novel method for solving the problem of PolSAR image classification based on FVs (Sections II-B and III). To the best of our knowledge, this is the first time the FV is applied to the analysis of PolSAR imagery. As a second contribution, we made available a set of ground truth annotations and a well-defined training/testing procedure based on two popular data sets (Section IV-A) found in the literature.

II. Fundamentals

In this section, we introduce the fundamental concepts on PolSAR image generation and the FV image representation. For a more detailed treatment of these topics, we refer the reader to [15] and [16], respectively.

A. PolSAR Data

A PolSAR measures the backscattered signal from a medium in the four different combinations that result from transmitting and receiving the radar signal with horizontal and vertical linear polarizations. The scattering information can be represented by the following complex-valued matrix:

$$S = \begin{pmatrix} S_{\mathrm{HH}} & S_{\mathrm{HV}} \\ S_{\mathrm{VH}} & S_{\mathrm{VV}} \end{pmatrix}. \tag{1}$$

For a reciprocal medium, $S_{\mathrm{HV}} = S_{\mathrm{VH}}$ holds and the scattering information can be alternatively encoded as a vector $h = (S_{\mathrm{HH}}\ \sqrt{2}S_{\mathrm{HV}}\ S_{\mathrm{VV}})^T$ where T denotes the transpose operator. The "multilook" version of this vector is obtained by averaging the individual measurements in a local neighborhood. Alternatively, one can define the multilook complex covariance as the matrix

$$C = [c_{ij}] = \frac{1}{m}\sum_{k=1}^{m} h(k)h(k)^{\dagger}. \tag{2}$$

Here, $h(k)$ is the scattering vector h at location k, m is the number of looks, and the superscript $\dagger$ denotes conjugate transpose. The Hermitian matrix mC is positive semidefinite (PSD) and follows a complex Wishart distribution with m degrees of freedom (DoF) [8], [17]. It can be shown that the real part of the matrix C, i.e., $\Re\{C\} = [\Re\{c_{ij}\}]$, is a symmetric PSD (SPSD) real matrix (see the Appendix). It can, thus, be considered as being drawn from a real Wishart pdf with n DoF. In this case, n is a free parameter that has to be set empirically.

B. Fisher Vector Principle

Let $X = \{x_i\}_{i=1}^N$ be an independent identically distributed sample drawn from a parametric distribution p_λ defined over a sample space Ω, with λ the parameters of the model. Let $T(X)$ denote the sufficient statistics of X with respect to the distribution p_λ. The FV of X under p_λ is a vectorial representation of the sample that encodes the difference between $T(X)$ and its expected value under p_λ [16]. Formally, it is defined as the normalized gradient of the log-likelihood of X with respect to the parameters of the model p_λ

$$g(X) = \frac{1}{N}\sum_{i=1}^{N} L_\lambda \nabla_\lambda \log p_\lambda(x_i) \tag{3}$$

where L_λ is the Cholesky factor of the inverse of the Fisher information matrix and ∇_λ denotes the gradient with respect to λ. Here, p_λ can be considered as a *universal* (class-independent) model for the generation of samples on Ω. The choice of its particular form is constrained by the nature of Ω and the particularities of the problem. In our case, we are constrained to $\Omega = \mathcal{S}(q)$, i.e., the space of SPSD $q \times q$ matrices (Section II-A). We must, therefore, seek for a sensible model on this space. Following [16], we define p_λ as a finite mixture distribution with K components as follows:

$$p_\lambda(x) := \sum_{k=1}^{K} w_k p_k(x) \tag{4}$$

$\sum_{k=1}^K w_k = 1$, $w_k > 0$, $\forall k$, and p_k a Wishart pdf with n DoF, $k = 1, \ldots, K$. Being a member of the exponential family, p_k can be parameterized as follows:

$$p_k(x) := H(x)\exp[\mathrm{vec}(\eta_k)^T \mathrm{vec}(x) - \Psi(\eta_k)]. \tag{5}$$

$\mathrm{vec}(\cdot)$ denotes the vectorization operator, $\Psi(\eta_k) = \log\Gamma_q(n/2) - (n/2)\log|-\eta_k|$ is known as log-partition function, $\eta_k \in \mathbb{R}^{q\times q}$ the parameters of the distribution in natural form, $H(x) = |x|^{(n-q-1)/2}$ is a normalizer, and $\Gamma_q(\cdot)$ denotes the multivariate gamma function.

With the above-mentioned definitions, the FV for a sample X results in a vector of dimensionality $D = K(q^2+1)$. This vector is subsequently power- and L2-normalized since this was shown to improve the overall classification performance [13].

C. Model Parameter Estimation

The distribution given by (4) is a loose model for the generation of samples in any image, regardless of the classes they might belong to. Its parameters must, thus, be estimated based on a diverse set of elements sampled at random from the images under consideration. Let $X = \{x_i\}_{i=1}^N$ be such a sample. To estimate the parameters of the model, we use the expectation–maximization (EM) algorithm [18]. EM iterates the following two steps.

1) *E-Step:* Compute the soft assignment of samples to mixture components for each element of the training set

$$\gamma_k(x_i) = \frac{w_k p_k(x_i)}{\sum_{j=1}^K w_j p_j(x_i)} \tag{6}$$

$k = 1, \ldots, K$ and $i = 1, \ldots, N$, using the current estimates of the parameters.

2) *M-Step:* Update the parameters of the model using the posteriors computed in the E-Step. For the mixture of real Wishart pdfs with fixed n, and parameterized as

in (5), the update equations adopt a particularly simple form

$$w_k = \frac{1}{N}\sum_{i=1}^{N}\gamma_k(x_i), \quad \eta_k = -\frac{n}{2}\left(\frac{\sum_{i=1}^{N}\gamma_k(x_i)x_i}{\sum_{i=1}^{N}\gamma_k(x_i)}\right)^{-1} \tag{7}$$

for $k = 1, \ldots, K$.

These two steps are repeated until the expected complete data log-likelihood ceases to improve or until a maximum number of iterations are reached. Initialization of the algorithm can be performed by randomly sampling K prototypes from the training set or by running a clustering algorithm on a subset of the data [19].

III. Pixelwise Classification With the FV

We adopt a simple model consisting of the following Potts-like energy over the graph of four-connected pixel locations:

$$E(\mathrm{Y}) = \sum_{i\in|\mathcal{V}|}\phi_i(y_i) + \lambda\sum_{\{i,j\}\in\mathcal{E}}\mathbb{I}[y_i \neq y_j] \tag{8}$$

where Y is labeling over the graph $(\mathcal{V}, \mathcal{E})$ with vertices $i \in \mathcal{V}$ and edges $(i, j) \in \mathcal{E}$, $\mathbb{I}[z]$ denotes the indicator function that equals 1 if its predicate is true and 0 otherwise, $\lambda \in \mathbb{R}$, and $\phi_i(y_i)$ is a unary term that penalizes the (incorrect) assignment of location i to class label y_i. In our case, $\phi_i(y_i)$ is defined as the negative inner product between the FV computed on a local neighborhood at location i and a per-class FV computed using all training samples for class $y_i \in \{1, \ldots, C\}$

$$\phi_i(y) = -\langle g(X_i), g_y\rangle \tag{9}$$

with X_i the sample extracted from a 3×3 neighborhood centered on location i, and g_y the FV computed from training samples of class y. The formulation above can be seen as a nearest mean classifier [20] with an additional spatial regularity constraint given by the second term in (8). The labeling $\hat{\mathrm{Y}} = \arg\min_{\mathrm{Y}} E(\mathrm{Y})$ gives us the desired classification.

IV. Experiments

We begin this section by providing a detailed description of the data and evaluation procedure we have followed in our experiments. Next, we evaluate different aspects of our model and compare its performance with other methods found in the literature. To facilitate reproducibility, data and scripts will be available at the project Web site.[1]

A. Data Set

For evaluation, we consider a subset of the images that were available trough PolSARpro[2] by the European Space Agency. This subset consists of two fully polarimetric images in the L-band acquired by the NASA/JPL Airborne Synthetic Aperture Radar (AIRSAR) sensor over the San Francisco Bay (SFB) area, USA, and over an agricultural region in the Flevoland (FL) province in the Netherlands. Figs. 1 and 2 show the pseudocolor representation of the polarimetric data (left) and the ground truth labels (right) for the SFB and FL regions, respectively. Segmentation masks for these two sets are based on [2] and [21]. For SFB, we cropped the original 900×1024 image and considered a region of 500×500 pixels since we have not ground truth annotations for the rest of the image. For FL, we consider the full 750×1024 image.

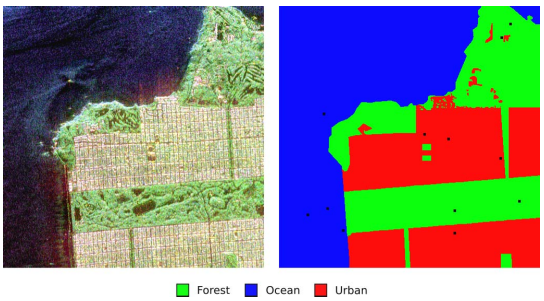

Fig. 1. (Left) SFB image. (Right) Ground truth labels. Training samples are marked as black squares (best viewed with magnification).

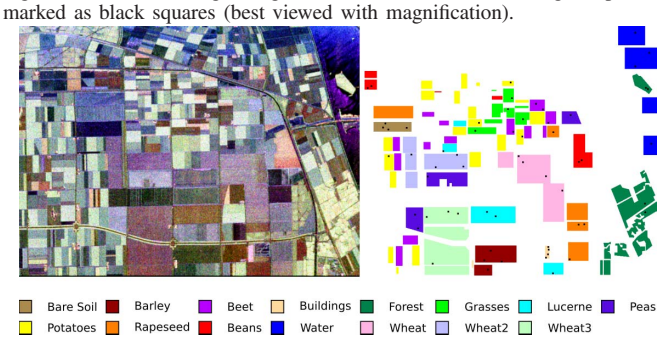

Fig. 2. (Left) FL image. (Right) Ground truth labels. Training samples are marked as black squares (best viewed with magnification).

For evaluation, we generate 10 different train/test splits and report the mean accuracy (and standard deviation) over the 10 runs. Train/test splits are generated following a process that lies between two common strategies found in the literature, namely, random selection [3]–[5] and manual annotation of training and testing samples [10], [22], [23]. The process is as follows. For each class, we sample r annotated pixels, and at each pixel location, we crop a small window of size $s \times s$ pixels. To enforce data variability, we only consider nonoverlapping windows. Following this procedure, we end up with rs^2 samples per class that we use for training, while the rest is used for testing. From now on, we set $r = 4$ and $s = 5$.

We make the following observations. First, a sample drawn as described above is—by construction—composed of r subsets of s^2 neighboring pixels. It has, therefore, the ability to retain some of the local dependencies that may exist between pixels in a window of size $s \times s$. These dependencies are less likely to be reflected on a sample drawn at random. Second, the process is not biased due to the subjectivity of manual selection, favoring reproducibility and fairness when comparing results.

B. Implementation Details

Classifying an image with our model involves the following steps: 1) fitting the parameters of the mixture distribution (4); 2) computing FVs' signatures at each pixel location; and 3) solving the classification problem posed by (8). Details regarding each step are next provided. First, images are

[1] http://ciii.frc.utn.edu.ar/JavierAndresRedolfi/sartb

[2] https://earth.esa.int/web/polsarpro

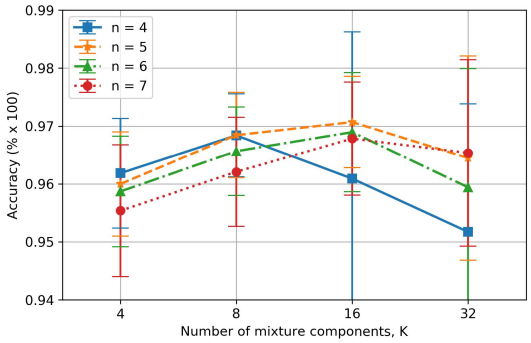

Fig. 3. Mean accuracy and standard deviation measured on the SFB subset for different choices of n and K.

converted to the multilook complex covariance format (2) using PolSARPro v4.2. The parameters of the mixture model (4) are estimated under a maximum-likelihood criterion with the EM algorithm using around 1 K points chosen at random, as described in Section II-C. In practice, we consider only those whose determinant is within the 95th percentile of the sample population. Empirically, we observed this has the effect of reducing the noise during estimation by removing samples that are badly conditioned. Once the model has been fitted, FVs are computed pixelwise as in [16]. Finally, the inference problem associated to the minimization of (8) is solved via graph cuts using the approximate solver of [24] as implemented in the GCO library.[3]

C. Selection of Parameters n and K

In this section, we evaluate the impact on the classification accuracy of the two main free parameters in our model, i.e., the effective number of DoF, n [see (5)], and the number of mixture components, K [see (4)]. Evaluation is performed on SFB. Fig. 3 shows the results obtained on the SFB subset for different values of n and K. From Fig. 3, we see that K has the highest influence on the overall accuracy. The best performance is obtained with $n \in \{5, 6\}$ and $K = 16$, although it is quite stable within this range. In what follows, we set $n = 5$ and $K = 16$.

D. Comparison With Other Methods

We first compare our approach to two common models found in the literature, namely, the CWC[4] [1] and a system similar to that of [2] but restricted to mixtures of real Wishart pdfs (RWM). In both cases, we fit a model for each class and use the negative log-likelihood as unary potentials in (8). Evaluation is performed on both the SFB and FL data sets and results are shown in Table I. Parameters for the RWM method were chosen following the procedure described in Section IV-C. From Table I, we observe that CWC and RWM perform similarly, with our approach outperforming both methods by a large margin. The larger variability observed with CWC and RWM can be attributed to the low number of samples available for some of the classes in order to properly fit the models. On the contrary, in the FV case, the underlying pdf is class independent and it was fitted from a more diverse and bigger set of samples.

[3]http://vision.csd.uwo.ca/code/

[4]CWC can be regarded as a model based on fitting a per-class mixture of complex Wishart pdfs with a single component ($K = 1$). Results are shown following this convention.

TABLE I

MEAN ACCURACY ON SFB AND FL FOR THE CWC, RWM, AND THE FV-BASED APPROACHES

Dataset	CWC	RWM (K=32, n=3)	Ours (K=16, n=5)
SFB	0.8848 (0.0602)	0.9081 (0.0238)	0.9707 (0.0079)
FL	0.8762 (0.0223)	0.8608 (0.0312)	0.9144 (0.0191)

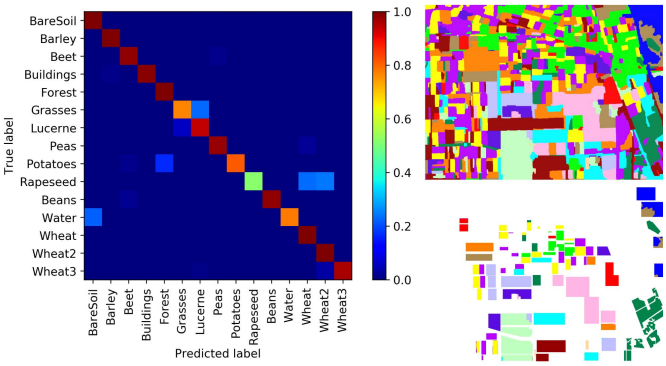

Fig. 4. (Left) CM for the FV-based model on the FL image. (Right) Classification results on the full image and test regions only.

In Fig. 4, we show the confusion matrix (CM) as well as classification results obtained with our model on the FL image. From the CM, we observe that for most of the classes, the method behaves accurately. The worst case is for the *Rapeseed* class that is often confused with the classes *Wheat* and *Wheat2*. There is also some confusion between the classes *Watter* and *BareSoil*, as well as between *Grasses* and *Lucerne*.

Next, we compare our model against two other methods recently proposed in the literature. The first is a method based on a polar decomposition of the Mueller matrix [3] while the second corresponds to the W-DSN of [4]. For a fair comparison, we followed the same evaluation procedures as those chosen by the authors in the reference works. Evaluation is performed on the FL subset only.

In the first case, results are reported on 11 of the 15 classes, as in [3]. Models are trained using 2000 samples drawn at random. Our results along with those reported by the authors are shown in Table II. From Table II, it can be observed that our model performs better on all cases except for the *Water* class. The mean accuracy of our system (0.9728) is 4% above the other methods.

We also compare our model with the deep-learning-based approach of [4]. In this case, following the configuration of the authors, classification performance is reported as the mean accuracy over the 15 classes, where 5% of the (labeled) samples are used for training and the remaining 95% for testing. The W-DSN system achieved an average accuracy (over the 15 classes) of 0.9268 while ours achieved a score of 0.9688, which represents a gain of 4% compared with W-DSN.

Finally, we report computation times for the different parts of our algorithm. With our implementation, fitting

TABLE II
COMPARISON WITH THE BEST PERFORMING METHOD OF [3] IN FL

Method	BareSoil	Beet	Forest	Grasses	Lucerne	Peas	Potatoes	Rapeseed	Beans	Water	Wheat
Wang *et al.* [3]	0.9878	0.9371	0.9496	0.8489	0.9231	0.9555	0.8896	0.9486	0.9653	0.9642	0.8864
Ours ($K = 16, n = 5$)	1.0	0.9926	0.9770	0.9920	0.9261	0.9993	0.9959	0.9983	0.9757	0.8472	0.9971

the mixture takes approximately 1 min on an Intel Core i7-2600K CPU machine. FV computation takes an average of 100 μs per pixel. It is by far the most costly operation. However, since FVs are computed on each pixel location independently, this operation could be trivially parallelized. Our implementation, however, does not exploit parallelism at this level. Inference depends on the size of the image as well as on the number of classes. Solving the inference problem for our largest image took approximately 12 s.

V. CONCLUSION

In this letter, we applied the FV framework to the problem of classification of PolSAR images and obtained encouraging results on two data sets commonly used in the literature. A natural extension of our model is the use of mixtures of complex Wishart pdfs or more complex models specifically tailored to PolSAR data [25]. These lines of research will be pursued in future works.

APPENDIX

A matrix $C \in \mathbb{C}^{q\times q}$ is said to be PSD if $z^*Cz \geq 0$ and $\forall z \in \mathbb{C}^q \setminus \{0\}$. Since $\mathbb{R}^q \subset \mathbb{C}^q$, it follows that $x^TCx \geq 0$ and $\forall x \in \mathbb{R}^q \setminus \{0\}$, i.e., the complex matrix C is also PSD on $\mathbb{R}^q$. Without loss of generality, let us consider the case $m = 1$ in (2) and let $h = (h_1, \ldots, h_q)^T$. Since C is Hermitian, the bilinear form on $\mathbb{R}^q$ above can be written as

$$\begin{aligned} x^TCx &= \sum_i x_i^2|h_i|^2 + \sum_i \sum_{j<i} x_i x_j (c_{ij} + \overline{c_{ij}}) \\ &= \sum_i x_i^2 \Re\{c_{ii}\} + 2\sum_i \sum_{j<i} x_i x_j \Re\{c_{ij}\} \\ &= x^T \Re\{C\} x \geq 0. \end{aligned}$$

However, note that in general $x^T\Re\{C\}x = x^T\Re\{hh^*\}x \neq x^T\Re\{h\}\Re\{h\}^Tx$, and for the multilook case [$m > 1$ in (2)], the matrix $\Re\{mC\}$ cannot be considered as generated by a Wishart distribution *with m DoF*, as we did before. In our case, the number of DoF is a parameter that has to be set empirically.

ACKNOWLEDGMENT

The authors would like to thank S. N. Anfinsen for providing the segmentation mask for the Flevoland image.

REFERENCES

[1] 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, Sep. 1999.

[2] W. Gao, J. Yang, and W. Ma, "Land cover classification for polarimetric SAR images based on mixture models," *Remote Sens.*, vol. 6, no. 5, pp. 3770–3790, 2014.

[3] 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, Apr. 2016.

[4] L. Jiao and F. Liu, "Wishart deep stacking network for fast POLSAR image classification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3273–3286, Jul. 2016.

[5] L. Zhang, W. Ma, and D. Zhang, "Stacked sparse autoencoder in PolSAR data classification using local spatial information," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1359–1363, Sep. 2016.

[6] B. Hou, H. Kou, and L. Jiao, "Classification of polarimetric SAR images using multilayer autoencoders and superpixels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 3072–3081, Jul. 2016.

[7] A. C. Frery, A. D. C. Nascimento, and R. J. Cintra, "Information theory and image understanding: An application to polarimetric SAR imagery," *Chilean J. Stat.*, vol. 2, no. 2, pp. 81–100, 2011.

[8] 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.

[9] A. C. Frery, A. D. C. Nascimento, and R. J. Cintra, "Analytic expressions for stochastic distances between relaxed complex Wishart distributions," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1213–1226, Feb. 2014.

[10] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Classification with a non-Gaussian model for PolSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2999–3009, Oct. 2008.

[11] A. P. Doulgeris and T. Eltoft, "Scale mixture of Gaussian modelling of polarimetric SAR data," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, p. 874592, Dec. 2009.

[12] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[13] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 222–245, 2013.

[14] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[15] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*. Boca Raton, FL, USA: CRC Press, 2009.

[16] J. Sánchez and J. Redolfi, "Exponential family Fisher vector for image classification," *Pattern Recognit. Lett.*, vol. 59, pp. 26–32, Jul. 2015.

[17] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, 1963.

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B (Methodol.)*, vol. 39, no. 1, pp. 1–38, 1977.

[19] C. Saint-Jean and F. Nielsen, "A new implementation of k-MLE for mixture modeling of Wishart distributions," in *Geometric Science of Information*. Berlin, Germany: Springer, 2013, pp. 249–256.

[20] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. San Diego, CA, USA: Academic, 2013.

[21] S. N. Anfinsen, R. Jenssen, and T. Eltoft, "Spectral clustering of polarimetric SAR data with Wishart-derived distance measures," in *Proc. POLinSAR*, vol. 7. 2007, pp. 1–9.

[22] Y. Zhang, L. Wu, N. Neggaz, S. Wang, and G. Wei, "Remote-sensing image classification based on an improved probabilistic neural network," *Sensors*, vol. 9, no. 9, pp. 7516–7539, 2009.

[23] L. Zhang, L. Sun, B. Zou, and W. M. Moon, "Fully polarimetric SAR image classification via sparse representation and polarimetric features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3923–3932, Aug. 2015.

[24] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov, "Fast approximate energy minimization with label costs," *Int. J. Comput. Vis.*, vol. 96, no. 1, pp. 1–27, 2012.

[25] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric $\mathcal{G}$ distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, 2005.