

POLSAR IMAGE CLASSIFICATION USING DIFFERENT CODIFICATIONS BASED ON FISHER VECTORS

Javier A. Redolfi^{1,2,*}, R. Gastón Araguás², A. Georgina Flesia³

¹ Grupo de Investigación Sobre Aplicaciones Inteligentes (GISAI),
Facultad Regional San Francisco, UTN, San Francisco, Córdoba, Argentina
jredolfi@sanfrancisco.utn.edu.ar

² Centro de Investigación en Informática para la Ingeniería (CIII),
Facultad Regional Córdoba, UTN, Córdoba, Argentina

³ Centro de Investigación y Estudios de Matemática (CIEM),
CONICET, FaMAF, Universidad Nacional de Córdoba, Córdoba, Argentina

KEY WORDS: PolSAR, Classification, Fisher Vector, Potts Model

ABSTRACT:

A 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. For an efficient use of these images it is necessary to have algorithms capable of classifying these images to generate maps with their content automatically. This paper presents the extension of a PolSAR image classification method based on exponential Fisher Vectors, a Potts smoothing model and different similarity measures. With the proposed extension, improvements in classification with respect to the base method are achieved. Future work consists in extending the codification so as not to have to discard the imaginary part of the data.

1. 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 [Lee et al., 1999, Gao et al., 2014, Wang et al., 2016, Jiao, Liu, 2016, Zhang et al., 2016, Hou et al., 2016]. However, the greatest difficulty in the use of PolSAR data is that the information (pixel values) is encoded as complex vectors or matrices, making it difficult to apply standard techniques of statistical analysis and machine learning literature. From the modeling perspective, dealing with this type of information in a well-founded way is a difficult problem that has attracted great attention in the past. PolSAR data is generated by transmitting orthogonally polarized electromagnetic pulses towards a target and recording the returned echo for each channel independently. Raw measurements are further processed in order to generate a multi-channel 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* [Frery et al., 2011]. To reduce the effect of this noise, PolSAR data is aggregated by averaging local information over small neighborhoods, resulting in the so called *multi-look* representation of the PolSAR data [Lee et al., 1994].

In this paper we present the application of models originated in the computer vision literature to the problem of land cover classification, this is the task of assigning labels to pixels based on the dispersion properties of the objective measured by a PolSAR sensor. Specifically, we propose a model that integrates the formalisms of exponential Fisher Vectors (eFV) [Sánchez, Redolfi, 2015] with a Potts-like energy model [Potts, 1952, Boykov

et al., 2001] that captures the spatial dependence between the variables. In the eFV scheme the content of the image (pixels, regions and/or the whole image) is characterized by the standardized gradient vector derived from different mixtures of convenient distributions. In this case, we consider the real part of the covariances measured by the PolSAR sensor and an eFV is derived from a mixture of real Wishart probability distributions functions (pdfs) and Gaussian pdfs. Then we define a Potts-like energy model where the unary terms are computed as different measures of similarity between the eFVs computed for each class and an eFV computed at pixel level. The minimization of this energy on the graph of four-connected pixel locations give us the desired classification.

The main contributions of this work are the extension of a previously presented method [Redolfi et al., 2017], adding the comparison of different probability distributions functions for the eFVs encoding and different similarity measures between the eFVs to generate a raw classification of pixels. The hypotheses of this work are that although the theoretical distribution of the data follows a Wishart distribution, with Gaussian distributions competitive results can be obtained and also that using other types of similarity measures the results in the classification can be improved.

2. PRELIMINARIES

In this section, we introduce the fundamental concepts on PolSAR image generation and the eFV image representation. For a deeper treatment of these topics, we refer the reader to [Lee, Pottier, 2009] and [Sánchez, Redolfi, 2015], respectively.

2.1 PolSAR imagery

A polarimetric SAR measures the backscattered signal from a medium in the four different combinations that result from

*Corresponding author

transmitting and receiving the radar signal with horizontal and vertical linear polarizations. The scattering information can be represented by the following complex-valued matrix:

$$\mathbf{S} = \begin{pmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{pmatrix}. \quad (1)$$

For a reciprocal medium, $S_{HV} = S_{VH}$ holds and the scattering information can be alternatively encoded as a vector:

$$\mathbf{h} = (S_{HH} \quad \sqrt{2}S_{HV} \quad S_{VV})^T \quad (2)$$

where T denotes the transpose operator. The “multi-look” version of this vector is obtained by averaging the individual measurements in a local neighborhood. Alternatively, one can define the multi-look complex covariance as the matrix:

$$\mathbf{C} = [c_{ij}] = \frac{1}{m} \sum_{k=1}^m \mathbf{h}(k)\mathbf{h}(k)^*. \quad (3)$$

Here, $\mathbf{h}(k)$ is the scattering vector \mathbf{h} at location k , m is the number of looks and the superscript $*$ denotes conjugate transpose. The Hermitian matrix \mathbf{C} is positive semidefinite (PSD) and follows a complex Wishart distribution with m degrees of freedom (DoF) [Goodman, 1963, Lee et al., 1994].

In this work, instead of work with the complete matrix \mathbf{C} , we only consider its real part $\Re\{\mathbf{C}\} = [\Re\{c_{ij}\}]$.

2.2 Fisher vector codification

The eFV of a sample $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ given a probability distribution p_λ defined over a sample space Ω is defined as:

$$\mathcal{G}_\lambda(\mathbf{X}) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N \mathbf{L}_\lambda \nabla_\lambda \log p_\lambda(\mathbf{x}_i). \quad (4)$$

where L_λ is the Cholesky factor of the inverse of the Fisher information matrix, λ are the parameters of the model and ∇_λ denotes the gradient with respect to λ .

The selection of the parametric model p_λ depends on the particularities of the problem and for this case we define p_λ as a finite mixture distribution over the space of symmetric positive defined matrices of $q \times q$, $\mathcal{S}(q)$ of the form:

$$p_\lambda(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{k=1}^K w_k p_k(\mathbf{x}), \quad (5)$$

with $\sum_{k=1}^K w_k = 1$, $w_k > 0$, $\forall k$ and p_k an exponential family pdf.

For this work we propose the use of two different pdf for p_k . The first is the Gaussian distribution and the second is the Wishart distribution. Based on both distributions, two different types of eFV (equation (4)) are derived and we compare their accuracy on the problem of PolSAR image classification.

3. CLASSIFICATION METHOD

For classification we adopt an approach based on an energy model over a graph G , where each vertex or node represents a pixel of the image. This model contains two terms, the first

one that depends on the observed data and the second one that depends on the relation between neighboring labels; this second term gives smoothness to the solution.

The energy model can be written as:

$$E(Y) = E_{\text{data}}(Y) + E_{\text{smoothness}}(Y) \quad (6)$$

where Y is the labeling over the graph $G = (\mathcal{V}, \mathcal{E})$ with vertices $i \in \mathcal{V}$ and edges $(i, j) \in \mathcal{E}$. The energy term that depends on the data or observations analyzes the pixels individually and the other takes into account the relationship between neighboring pixels, penalizing adjacent pixels with different labels, achieving a smoother classification result.

For the smoothing function, we propose a simple model consisting of the following Potts-like energy over the graph of 4-connected pixel locations, which corresponds to a horizontal and vertical neighborhood of 3×3 :

$$E(Y) = \sum_{i \in \mathcal{V}} \phi_i(y_i) + \gamma \sum_{\{i,j\} \in \mathcal{E}} \mathbb{I}[i \neq j] \quad (7)$$

The first term of the energy, which is data dependent, is $\phi_i(y_i)$, also known as unary potential, which penalizes the incorrect assignment of the class label y_i to location i . The second term is composed of the indicator function $\mathbb{I}[z]$ which is equal to 1 if its predicate is true and 0 in another case, and for the penalty constant $\gamma \in \mathbb{R}$.

The unary potential $\phi_i(y_i)$ penalizes an incorrect assignment, therefore we can define it as the negative of the distance between the eFV computed over a local neighborhood to the location i and one eFV computed per class using all training samples for class $y_i \in \{1, \dots, c\}$, where c is the number of interest classes. We can also define this potential as the negative of an eFV classification score at the location i obtained using some type of classifier.

The figure 1 shows the connection between a central node and its neighbors, the edges with solid lines indicate the connection of the node with its neighbors and the arrow indicates the unary potential of the node in question. As we can see each node connects with four of its eight neighbors.

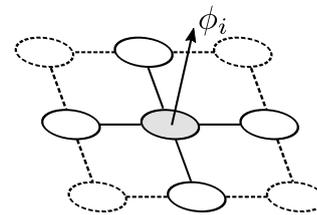


Figure 1. Connection between a central node and its four neighbors. The nodes with solid lines are those that connect to the central node and the edges in solid line indicate these connections. The arrow indicates the unary potential ϕ_i of the central node.

The labeling that minimizes energy $E(Y)$,

$$\hat{Y} = \arg \min_Y E(Y), \quad (8)$$

give us the desired classification.

3.1 Energy minimization

To minimize the energy of equation 7 there are several algorithms in the literature, among which we can highlight those based on maximum a posteriori estimates normally used for minimization of energy in Markov random field [Geman, Geman, 1987]. One of the disadvantages of these methods is that the solutions require exponential computation times. Because of this, the approximate solution presented in [Boykov et al., 2001] known as α -expansion or expansion-move was used for minimization, which is based on graph-cut.

3.1.1 Expansion-move algorithm This algorithm is a very powerful method and widely used in practice. The algorithm is based on finding the expansion of some of the labels that makes energy decrease. That energy minimization is solved using graph cuts.

Below are the steps of the algorithm:

1. Start with an arbitrary labeling.
2. Repeat for each label:
 - (a) Find the lowest energy labeling E with an expansion-move step using graph-cut.
 - (b) Go to the labeling if the energy is less than the energy of the current labeling.
3. If E does not decrease in the cycle, terminate the algorithm, otherwise return to 2.

The main advantages of this algorithm is that it has a convergence time proportional to the number of nodes, unlike other algorithms in which it is usually exponential. The fundamental disadvantage is that this algorithm does not find a global minimum of energy although it does find a local minimum and it is shown that this local minimum has an upper bound [Boykov et al., 2001], which is shown in the following equation:

$$E(\hat{Y}) \leq E(Y^*) \leq 2\beta E(\hat{Y}) \quad (9)$$

where \hat{Y} is the lowest global energy labeling, Y^* is the local minimum that finds the algorithm and β is a constant that depends on the smoothing function.

In figure 2 we show an example of the operation of the algorithm.

Obtaining the lower energy labeling. To find an expansion of any of the labels (step 2.a of the algorithm), the graph-cut algorithm is used. For this, the problem is transformed into a graph where the objective is to find the cut that divides the graph into 2 parts with the minimum energy known as the minimum cut [Boykov et al., 2001]. The graph cut finds the division of minimum energy between 2 labels, therefore to solve a multi-label problem as in this case, first a label is selected and the rest of the labels are grouped into a new temporary label. This can be seen in figure 3.

After this, if the new energy obtained for step $i + 1$ is less than the energy of step i we move to this new labeling, as indicated in step 2.b, otherwise we end the algorithm.

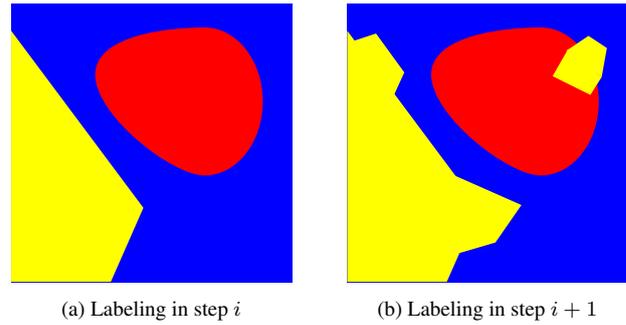


Figure 2. Example of a step of the expansion-move algorithm.

In this example we have 3 labels; on the left, the labeling obtained in step i is shown and on the right, the labeling for step $i + 1$ is shown after performing an expansion-move step of the label represented by the color yellow.

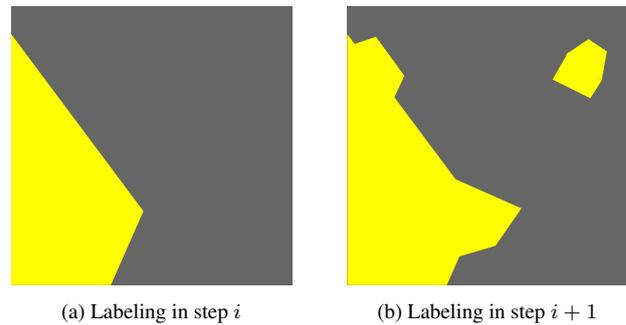


Figure 3. Example of a step of the graph-cut algorithm. In this example, the blue and red labels of figure 2 became a single label; on the left the label obtained in step i is shown and on the right the label for step $i + 1$ is shown after performing an expansion-move step of the label represented by the color yellow.

4. PROPOSED EXTENSION

In this paper we propose two modifications to the work presented in [Redolfi et al., 2017]. The first one is with respect to the selection of the pdf of equation (5). While it is known that the data measured by a PolSAR sensor is modeled as a Wishart distribution [Goodman, 1963] we also know that with a sum of Gaussian distributions we can model any type of distribution [Papoulis, Pillai, 2002], therefore it is feasible to use a Gaussian pdf in equation (5).

The $p_k(\mathbf{x})$ of equation (5) looks like:

$$p_k(\mathbf{x}) := H(\mathbf{x}) \exp \left[\text{vec}(\eta_k)^T \text{vec}(\mathbf{x}) - \Psi(\eta_k) \right]. \quad (10)$$

where $H(\mathbf{x})$ is a normalizer, $\text{vec}(\cdot)$ denotes the vectorization operator, $\Psi(\eta_k)$ is known as log-partition function and $\eta_k \in \mathbb{R}^{q \times q}$ are the parameters of the distribution in natural form. For the definition of these terms for each of the distributions see [Sánchez, Redolfi, 2015]. Then using this definition for the probability distributions we calculate the gradient of equation (4) on a set of pixels obtaining the eFV encoding for those pixels.

The second modification proposed is in the calculation of the $\phi_i(y_i)$ unary potential of equation (7). In this work we propose to use three similarity measures, the first one is the dot product (DP) between eFVs, the second is the Euclidean distance to the

nearest neighbor (NN) and the third is the distance to a hyper-plane obtained with Vector Support Machines (SVM) [Cortes, Vapnik, 1995].

5. EXPERIMENTS

This section describes the data used to perform the experiments, then what are the proposed experiments and the evaluation procedure for the performed experiments. Finally, the results obtained are shown.

5.1 Dataset

For evaluation, we consider a subset of the images that were available through PolSARpro¹ by the European Space Agency (ESA). This subset consists of two fully polarimetric images in the L-band acquired by the NASA/JPL AIRSAR sensor over the San Francisco Bay (SFB) area, USA, and over an agricultural region in the Flevoland (FL) province in The Netherlands. Figures 4 and 5 show the Pauli decomposition 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 [Gao et al., 2014] and [Anfinsen et al., 2007]. 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. The cropped image is shown in 4. For FL, we consider the full 750×1024 image.

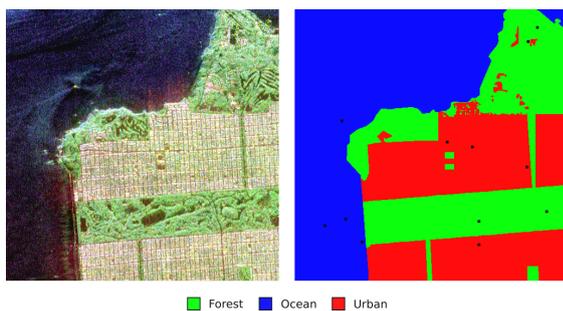


Figure 4. San Francisco Bay image (left) and ground truth labels (right). Training samples are marked as black squares (best viewed with magnification).

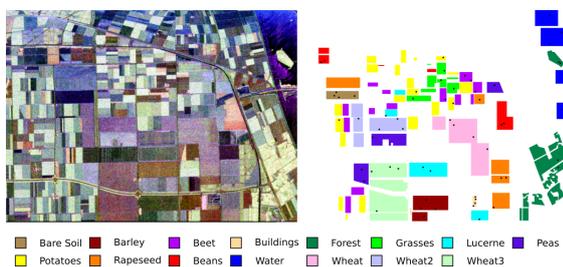


Figure 5. Flevoland image (left) and ground truth labels (right). Training samples are marked as black squares (best viewed with magnification).

The SFB image will be used for parameter setting and the FL image will be used for evaluation. For each image we generate 5 different train/test splits and report the mean accuracy (and standard deviation) over the 5 runs. Train/test splits are generated following a process that lies between two common strategies found in the literature, namely random selection [Wang et

al., 2016, Jiao, Liu, 2016, Zhang et al., 2016] and manual annotation of training and testing samples [Zhang et al., 2009, Dougeris et al., 2008, Zhang et al., 2015]. 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 non-overlapping 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 $s = 5$, which gives a total of 100 samples per class.

5.2 Implementation details

Classifying an image with our model involves the following steps:

1. fitting the parameters of the mixture distribution (5) based on the selected distribution,
2. computing eFVs signatures at each pixel location,
3. compute unary potentials of equation (7) for each pixel depending on the similarity measure selected, and
4. solving the classification problem posed by equation (7).

Details regarding each step are next provided. First, since the images are the result of a single pass (single look), they are converted to a multi-look image using PolSARPro v4.2. The parameters of the mixture model (5) are estimated under a maximum-likelihood criteria with the EM algorithm using around 1000 points chosen at random. 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 which are badly conditioned. Once the model has been fitted, FVs are computed pixel-wise as in [Sánchez, Redolfi, 2015]. For similarity measures we use the implementations of the scikit-learn library [Pedregosa et al., 2011]. Finally, the inference problem associated to the minimization of (7) is solved via Graph Cuts using the approximate solver of [DeLong et al., 2012] as implemented in the GCO library².

5.3 Parameter selection

In this subsection, we evaluate the impact on the classification accuracy of the number of mixture components K , equation (5). In figure 6 we show the results obtained on the image of SFB for different values of K . As can be seen in figure 6, the accuracy using Wishart is more stable and remains similar for the range of K analyzed. In the case of Gaussian there is more variation with the number of components and for more than 16 the accuracy begins to decrease. Due to the latter, in what follows we set $K = 16$.

5.4 Results

In table 1 we show the experiments results on the FL image. As can be seen in the results table, the Wishart distribution gives better accuracy than the Gaussian distribution for any of the similarity measures used. This is consistent with the assertion that the selection of the parametric model p_λ depends of the particularities of the problem as we claim in sub-section 2.2.

¹<https://earth.esa.int/web/polсарpro>

²<http://vision.csd.uwo.ca/code/>

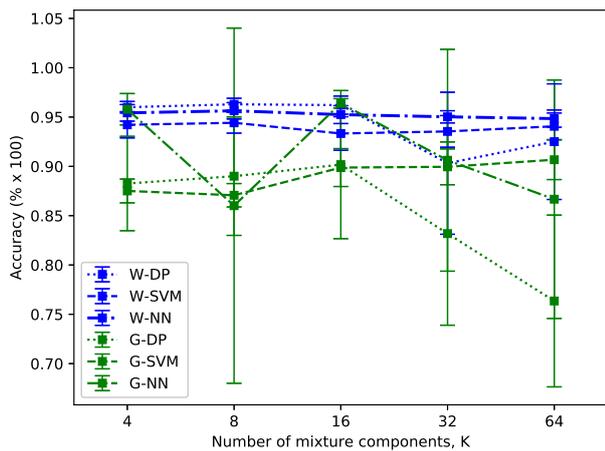


Figure 6. Mean accuracy and standard deviation measured on the SFB subset for different choices of K . In blue for eFV based on Wishart distribution and in green for eFV based on Gaussian distribution. Similarity measures are dot product (DP), support vector machines (SVM) and nearest neighbors (NN).

	Wishart	Gaussian
Dot product	0.9019 (0.0205)	0.5273 (0.0340)
Nearest neighbor	0.9222 (0.0270)	0.7247 (0.0514)
SVM	0.8296 (0.0112)	0.6339 (0.0965)

Table 1. Comparison of results on the FL image.

In addition, it can be seen that for both probability distributions, the use of a similarity measure based on the nearest neighbor produces better results than the other proposed measures.

In figure 7 we show the segmentation results on the FL image using Wishart distribution and nearest neighbor distance.

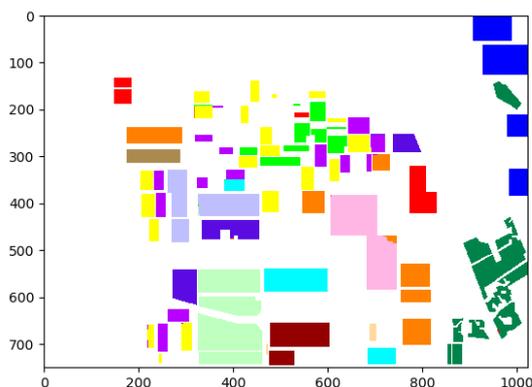


Figure 7. Segmentation results on the FL image for $K = 16$, and eFV based on Wishart distribution and nearest neighbors as similarity measure.

6. CONCLUSIONS AND FUTURE WORK

In this work we present an extension of a PolSAR image classification model presented in a previous work. The proposed extension consists in the use of different probability distribu-

tions to capture the underlying structure of the data and the use of different similarity measures for the comparison of vectors.

With respect to the first hypothesis we can say as a conclusion that the underlying structure is very important for the choice of the base probability distribution for eFV encoding. For this particular case, the Wishart distribution is the one that best adapts to the data as can be seen in the results.

With respect to the similarity measure, for both probability distributions analyzed, the measure of similarity that showed better accuracy was nearest neighbor.

We are currently working on the extension of the eFV model to adapt it to complex probability distributions. With this type of distributions, we avoid discarding the imaginary part of the data and work with all the available information and we expect to obtain an improvement in the classification.

REFERENCES

- Anfinsen, S. N., Jenssen, R., Eltoft, T., 2007. Spectral clustering of polarimetric sar data with wishart-derived distance measures. *Proc. POLinSAR*, 7.
- Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11).
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine learning*, 20(3), 273–297.
- DeLong, A., Osokin, A., Isack, H. N., Boykov, Y., 2012. Fast approximate energy minimization with label costs. *Int. J. Computer Vision*, 96(1).
- Doulgeris, A. P., Anfinsen, S. N., Eltoft, T., 2008. Classification with a non-Gaussian model for PolSAR data. *IEEE Trans. Geosci. Remote Sens.*, 46(10), 2999–3009.
- Frery, A. C., Nascimento, A. D. C., Cintra, R. J., 2011. Information theory and image understanding: An application to polarimetric SAR imagery. *Chilean J. of Stat.*, 2(2), 81–100.
- Gao, W., Yang, J., Ma, W., 2014. Land cover classification for polarimetric SAR images based on mixture models. *Remote Sensing*, 6(5), 3770–3790.
- Geman, S., Geman, D., 1987. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Readings in Computer Vision*, Elsevier, 564–584.
- Goodman, N., 1963. Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann. Math. Statist.*, 34(1).
- Hou, B., Kou, H., Jiao, L., 2016. Classification of Polarimetric SAR Images Using Multilayer Autoencoders and Superpixels. *IEEE J. of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(7), 3072–3081.
- Jiao, L., Liu, F., 2016. Wishart Deep Stacking Network for Fast POLSAR Image Classification. *IEEE Trans. on Image Processing*, 25(7), 3273–3286.
- Lee, J.-S., Grunes, M. R., Ainsworth, T. L., Du, L.-J., Schuler, D. L., Cloude, S. R., 1999. Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. *IEEE Trans. Geosci. Remote Sens.*, 37(5).

- Lee, J.-S., Grunes, M. R., Kwok, R., 1994. Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution. *Int. J. Remote Sens.*, 15(11), 2299–2311.
- Lee, J.-S., Pottier, E., 2009. *Polarimetric radar imaging: from basics to applications*. CRC press.
- Papoulis, A., Pillai, S. U., 2002. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Potts, R. B., 1952. Some generalized order-disorder transformations. *Mathematical Proc. of the Cambridge Philosophical Society*, 48number 01, Cambridge Univ Press.
- Redolfi, J., Sánchez, J., Flesia, A. G., 2017. Fisher Vectors for PolSAR Image Classification. *IEEE Geoscience and Remote Sensing Letters*, 14(11), 2057–2061.
- Sánchez, J., Redolfi, J., 2015. Exponential family Fisher vector for image classification. *Pattern Recognition Letters*, 59, 26–32.
- Wang, H., Zhou, Z., Turnbull, J., Song, Q., Qi, F., 2016. PolSAR Classification Based on Generalized Polar Decomposition of Mueller Matrix. *IEEE Geosc. Remote Sens. Lett.*, 13(4), 565–569.
- Zhang, L., Ma, W., Zhang, D., 2016. Stacked Sparse Autoencoder in PolSAR Data Classification Using Local Spatial Information. *IEEE Geosc. Remote Sens. Lett.*, 13(9), 1359-1363.
- Zhang, L., Sun, L., Zou, B., Moon, W. M., 2015. Fully polarimetric SAR image classification via sparse representation and polarimetric features. *IEEE J. Sel. Topics Appl. Earth Observ.*, 8(8), 3923–3932.
- Zhang, Y., Wu, L., Neggaz, N., Wang, S., Wei, G., 2009. Remote-sensing image classification based on an improved probabilistic neural network. *Sensors*, 9(9), 7516–7539.