A Non-Parametric Texture Descriptor for Polarimetric SAR Data with Applications to Supervised Classification

Marc Jäger, German Aerospace Center (DLR), marc.jaeger@dlr.de
Andreas Reigber, German Aerospace Center (DLR), andreas.reigber@dlr.de

Abstract

The paper describes a novel representation of polarimetric SAR (PolSAR) data that is inherently non-parametric and therefore particularly suited for characterising data in which the commonly adopted hypothesis of Gaussian backscatter is not appropriate. The descriptor is also non-local and can capture image structure in terms of the arrangement of edge-, ridge- and point-like features, to yield a salient characerisation of semi-periodic spatial patterns. The basic approach is based closely on [1] and has been adapted for application to PolSAR data. As an example application, the descriptor is evaluated in the context of supervised classification. The performance is compared with conventional statistical approaches on both simulated and real PolSAR data [1].

1 Introduction

The term *texture* has, in recent years, acquired a very specific meaning in the context of SAR signal statistics. Recent advances [2] have been motivated by the fact that the well known hypothesis of Gaussian statistics of the backscattered field often fails to hold in practice: it does not apply to very high resolution SAR data, where speckle is not fully developed, and cannot describe inherently heterogeneous types of land cover such as urban areas. The definition of texture as a departure from Gaussian statistics, however, is quite restrictive in that it precludes spatial correlation and currently applies only to the backscattered intensity.

For the present purposes the term *texture* is understood to include spatial correlations, i.e. patterns in the spatial arrangement of image structures, including patterns involving more than one polarimetric scattering mechanism. This is closer to the meaning generally implied in image processing literature as well as several previous studies concerning texture in SAR images reviewed in [4].

The texture descriptor developed in section 2 is closely based on the approach described in [1], albeit extended from grayscale optical imagery to multi-channel, complex PolSAR data. Descriptor extraction is a two-stage process which begins with training a so-called dictionary containing entries for recurring image structures. A given region in a PolSAR dataset can then be characterised by the distribution of dictionary entries that match its contents.

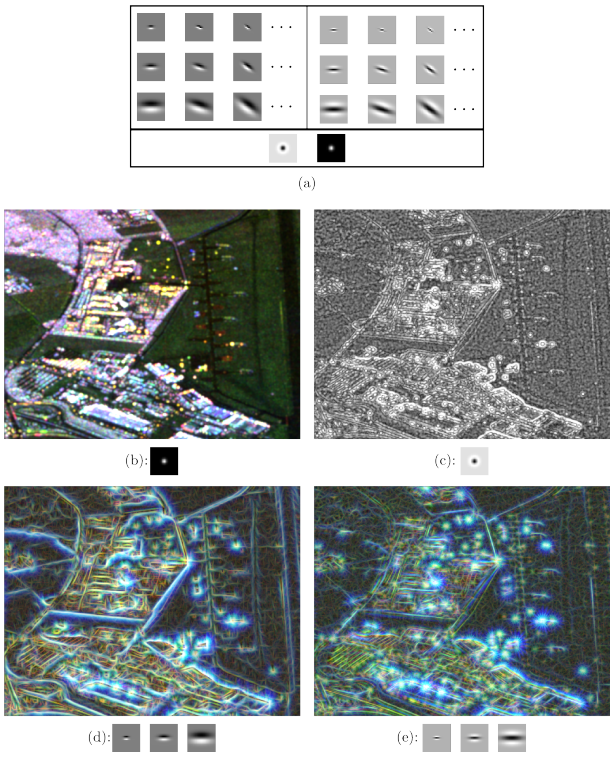

Figure 1: The filter bank applied to characterise local image structure. a) The eight families of filter applied. Clockwise from the top left: oriented 1^{st} order, oriented 2^{nd} order, Gaussian and Laplacian filters. b) Sample input data, convolved with the Gaussian filter. c) The Laplacian response. d) 1^{st} order responses for the three filter scales. e) 2^{nd} order responses for the three filter scales.

As an example application, section 3 describes a supervised classifier on the basis of the proposed texture descriptor. The classifier is based on mixture models trained using labled sample data and expectation maximisation.

[1] *Note to reviewers:* The current version of the paper merely includes preliminary results for simulated data. Additional results and evaluation will be included in the final version of the paper (or the presentation, due to space constraints).

Since texture is intrinsically linked to scale (i.e. texture only becomes perceptibly semi-periodic at a certain, coarse scale of observation), the underlying model is designed to operate at several scales simultaneously and can even be said to incorporate a mechanism for automatic texture scale selection.
Section 4 presents preliminary classification results obtained for simulated PolSAR data. The problems considered are challenging, as the classes considered have virtually identical intensity distributions and involve the same spectrum of polarimetric scattering mechanisms.
Finally, section 5 summarises the contributions and presents conclusions drawn from the results obtained.

2 The Texton Descriptor for PolSAR Data

The process of texton extraction is perhaps easiest to convey by considering the much simplified case of discrete valued, single channel grayscale imagery. In this case, the dictionary could be an identity mapping: gray value V is mapped to dictionary entry V. This type of dictionary requires no training and the resulting texton descriptor is merely the local histogram of grayscale values. Importantly, the local histogram thus obtained is a non-parametric density estimate, such that no prior knowledge concerning its shape is required. The remainder of this section outlines the steps taken to extend this simple approach to accommodate both geometric image structure as well as fully polarimetric SAR data.
PolSAR data are initially represented as a collection of rank-deficient $q \times q$ covariance matrices $[\mathbf{X}] = [\mathbf{X}_1, \mathbf{X}_2, \ldots]$ obtained from q dimensional scattering vectors $\vec{\mathbf{k}}_i$ via the outer product

$$\mathbf{X}_i = \vec{\mathbf{k}}_i \vec{\mathbf{k}}_i^{\dagger}, \tag{1}$$

where $\dagger$ denotes the conjugate transpose.
As in [1], both dictionary learning and descriptor extraction begin with the application of a bank of filters, illustrated in 1a), to the input $[\mathbf{X}]$. The Gaussian filter is applied to $[\mathbf{X}]$ to obtain full-rank, smoothed covariance matrices

$$\left[\mathbf{X}^G\right] = [\mathbf{X}] \otimes G, \tag{2}$$

where G denotes the Gaussian kernel and the spatial convolution is applied to each element of the covariance matrix independently.
For the remaining derivative filters, two separate convolutions are carried out for positive and negative filter coefficients. Given the signed filter kernel F, this initially results in

$$\left[\mathbf{X}^{F\pm}\right] = [\mathbf{X}] \otimes \max\left(\pm F, 0\right) \tag{3}$$

A scalar filter response, indicating the presence of edge-, ridge- or point-like image structure (depending on the filter) in the neighbourhood of sample $\mathbf{X}_i$, is then derived from the generalised likelihood ratio for edge detection introduced in [5]:

$$r_i^F = \ln\left(2^{-2q} \frac{\left|\mathbf{X}_i^{F+} + \mathbf{X}_i^{F-}\right|^2}{\left|\mathbf{X}_i^{F+}\right|\left|\mathbf{X}_i^{F-}\right|}\right), \tag{4}$$

where $|\cdot|$ denotes the matrix determinant. After ensuring rotation invariance by maximising the response over all orientations for the oriented filters, this leaves a vector of seven scalar responses $\vec{\mathbf{r}}$ (three each for the 1^{st} and 2^{nd} order oriented filters and one Laplacian response). The outcome of filtering can therefore be denoted

$$[T] = [t_1, t_2, \ldots] \text{ with } t_i = \left(\mathbf{X}_i^G, \vec{\mathbf{r}}_i\right). \tag{5}$$

The contents of T are visualised on the basis of a sample PolSAR data set in figure 1b)-e).

2.1 Dictionary Learning

The so-called texton dictionary is obtained by analysing a large amount of representative PolSAR data to identify recurring filter responses t. In this approach, the dictionary is the result of clustering using expectation maximisation (the *K-means* approach of [1] is not applicable due to the complex internal structure of response tuples t). Clustering fits a mixture model with a large but fixed number N_D of mixture components to the response vectors in the training data base. Denoting the set of mixture model parameters $D = \{d_1, d_2, \ldots, d_{N_D}\}$, with d_j representing the parametrisation of the j^{th} component density, the likelihood of a response t can be written

$$\mathrm{p}(t|\, D) = \sum_{j \in [1, N_D]} \mathrm{p}(t|\, d_j). \tag{6}$$

Once the likelihood of the training data base has been maximised with respect to D, individual component parameters d_j correspond to the dictionary entries. In the component density chosen, the elements of $t = \left(\mathbf{X}^G, \vec{\mathbf{r}}\right)$ are distributed independently according to the complex Wishart and the Γ distributions, respectively. Thus,

$$\begin{aligned}\mathrm{p}\left(t = \left(\mathbf{X}^G, \vec{\mathbf{r}}\right)\middle|\, d_j\right) =& \mathrm{p}\left(\mathbf{X}^G\middle|\, (n, \mathbf{\Sigma}) \in d_j\right) \\ & \prod_{k \in [1,7]} \mathrm{p}(\vec{\mathbf{r}}_k|\, (m_k, s_k) \in d_j),\end{aligned} \tag{7}$$

where the Γ distribution can be parametrised as

$$\mathrm{p}(r|\, m, s) = \frac{m^m}{\Gamma(m) s^m} r^{m-1} \exp\left(-m\frac{r}{s}\right) \tag{8}$$

and $\mathrm{p}\left(\mathbf{X}^G\middle|\, (n, \mathbf{\Sigma}) \in d_j\right)$ denotes the Wishart distribution [6] with n degrees of freedom and mean $\mathbf{\Sigma}$. For the Γ distribution, (m, s) denote the shape parameter and the mean response, respectively.
The EM iteration begins with the *E-Step*, which merely computes the posteriors

$$\mathrm{p}(d_j|\, t_i) = \frac{\mathrm{p}(t_i|\, d_j)}{\sum_{j \in [1, N_D]} \mathrm{p}(t_i|\, d_j)}. \tag{9}$$

The subsequent *M-Step* updates the component density parameters d_j. The update equations for the Wishart parameters, including the degrees of freedom n, have been derived in [3]. The update equations for the parameters (m, s) of Γ-distributions in (7) follow trivially, since the Γ distribution is simply the complex Wishart distribution for 1×1 covariance matrices.
In terms of implementation, the EM iteration begins with initialising $\mathrm{p}(t_i|\, d_j)$ with uniform random values and iterating the *E-* and *M-Steps* until convergence.

2.2 Descriptor Extraction

Given a fully trained dictionary D, texture descriptors for a given PolSAR dataset $[\mathbf{X}] = [\mathbf{X_1}, \mathbf{X_2}, \ldots]$ are obtained by applying the filter bank to obtain responses $T = [t_1, t_2, \ldots]$ as described in section 2 above. Instead of associating each sample i with a single lable corresponding to the most likely dictionary component according to (9), as in [1], a vector valued signal $B = \left[\vec{\mathbf{b}}_1, \vec{\mathbf{b}}_2, \ldots\right]$ is obtained directly from the posteriors to make the descriptor more robust when associations are uncertain:

$$\vec{\mathbf{b}}_i = \begin{pmatrix} \mathrm{p}(\, d_1|\, t_i) \\ \ldots \\ \mathrm{p}(\, d_{N_D}|\, t_i) \end{pmatrix} \tag{10}$$

Each vector can be thought of a normalised histogram. An entire region R of a given dataset can now be characterised by averaging the descriptors it contains.

$$\vec{\mathbf{b}}^R = \mathrm{E}\Big\{ \vec{\mathbf{b}}_i \Big|\, i \in R \Big\} \tag{11}$$

This vector now summarises the polarimetric scattering mechanisms as well as the type, scale and abundance of elementary geometric image structures within R. Note that, to obtain meaningful results, the size of the region R must cover one or more periods of any underlying semi-periodic texture pattern.

3 An Application to Supervised Classification

The vector signal thus obtained can be used as a basis for supervised classification, just as the underlying observations $[\mathbf{X}]$ often are. The goal of the statistical classifier developed in this section is to associate a class lable L_i with each observation $\vec{\mathbf{b}}_i$ by evaluating

$$L_i = \arg\max_c \left(\mathrm{p}\Big(\theta_c |\, \vec{\mathbf{b}}_i \Big) \right), \tag{12}$$

where θ_c denotes the set of parameters governing the probability density of $\vec{\mathbf{b}}$ in class c. These parameters are estimated from sample training data.

3.1 Multinomial Model

Since the observations are histograms, the use of a mixture-of-multinomials model for each class suggests itself. The corresponding density for the j^{th} component is given by

$$\mathrm{p}\Big(\vec{\mathbf{b}}_i \Big|\, (p_j, \vec{\mathbf{p}}_j) \in \theta_c \Big) = \frac{p_j^{N_D - 1} \Gamma(p_j + 1)}{\prod_{d=1}^{N_D} \Gamma\Big(p_j \vec{\mathbf{b}}_{i,d} + 1 \Big)} \prod_{d=1}^{N_D} (\vec{\mathbf{p}}_{j,d})^{p_j \vec{\mathbf{b}}_{i,d}}, \tag{13}$$

where p_j and $\vec{\mathbf{p}}_j$ denote the precision and the 1^{st} moment of the mixture component, respectively, and $\vec{\mathbf{b}}_{i,d}$ and $\vec{\mathbf{p}}_{j,d}$ denote the d^{th} elements of the N_D dimensional vectors $\vec{\mathbf{b}}_i$ and $\vec{\mathbf{p}}_j$.
Mixture models are trained independently for each class using expectation maximisation. The *E-Step* computes the posteriors $\mathrm{p}\Big((p_j, \vec{\mathbf{p}}_j) |\, \vec{\mathbf{b}}_i \Big)$ in the manner of (9), while the *M-Step* derives updated parameters $\big(p'_j, \vec{\mathbf{p}}'_j\big)$. The update equation for the 1^{st} moment is obtained as

$$\vec{\mathbf{p}}'_j = \frac{1}{\sum_i w_{ij}} \sum_i w_{ij} \vec{\mathbf{b}}_i \tag{14}$$

with $w_{ij} = \mathrm{p}\Big((p_j, \vec{\mathbf{p}}_j) |\, \vec{\mathbf{b}}_i \Big)$. Precision parameters can be updated using the techniques developed in [7].
Once a mixture model has been trained for each class, evaluating the *E-Step* followed by (12) results in a class label L_i for any given observation $\vec{\mathbf{b}}_i$.

3.2 Multi-Resolution Model

The multinomial model above does not account for the fact that descriptors $\vec{\mathbf{b}}$ must be spatially averaged to provide a meaningful texture characterisation. In practice, different textures with different characteristic scales will require different amounts of spatial integration. This section introduces a simple modification to the multinomial mixture model that accounts for the notion of texture scale.
Instead of introducing scale as an free parameter to be estimated, the classifier is modified to simultaneously act on several versions of the observation set $\left[\vec{\mathbf{b}}^{\sigma_1}\right], \ldots, \left[\vec{\mathbf{b}}^{\sigma_{N_S}}\right]$, each of which is obtained by spatially smoothing using a Gaussian kernel of standard deviation σ_k. The number of scales N_S and the scales σ_k are fixed *a priori*. The multi-scale mixture model proposed allocates several components to each scale and results in a likelihood function of the form

$$\mathrm{p}\Big(\vec{\mathbf{b}}_i \Big|\, \theta_c \Big) = \sum_{j=1}^{N_M} \sum_{k=1}^{N_S} \mathrm{p}\Big(\vec{\mathbf{b}}_i^{\sigma_k} \Big|\, (p_{jk}, \vec{\mathbf{p}}_{jk}) \in \theta_c \Big). \tag{15}$$

The EM update equations remain essentially unchanged. In this approach, scale selection happens implicitly through precision parameters p_{jk}: for scales smaller than the intrinsic texture scale, observations appear noisy and the precision is low, such that component densities are flat and have little impact on the classification outcome. Similarly, for very large scales, non-stationarities near the boarders of textured regions begin to degrade precision.

4 Preliminary Results

The following results, based on simulated PolSAR data, are included as a proof of concept for the proposed approach. The problems considered below are challenging, since the classes involved have virtually identical intensity distributions and mean covariance matrices. Furthermore, the textures (at least in figure 3) involve variations in the underlying polarimetric scattering mechanism.

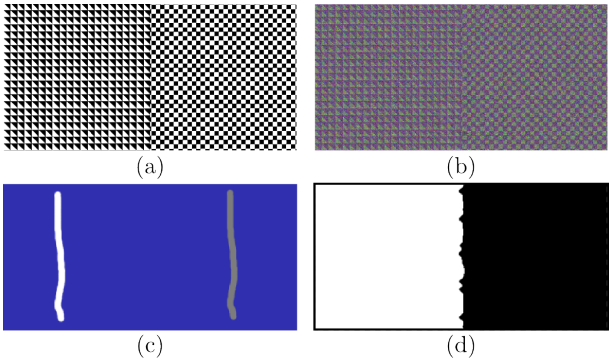

Figure 2: A simulated two-class classification problem. a) Grayscale image. b) Derived PolSAR image. c) Training samples marked in shades of gray. d) Classification result.

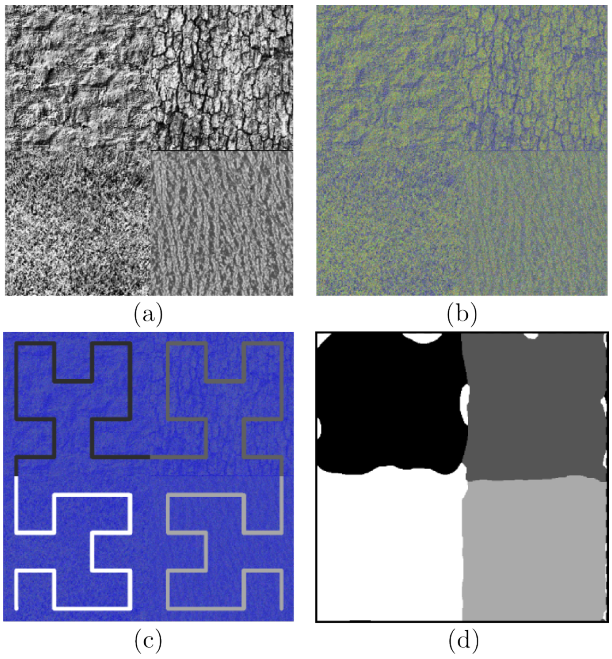

Figure 3: A simulated four-class classification problem. a)-d) as in figure 2.

The PolSAR data are simulated on the basis of arbitrary grayscale images with values in the range $[0, 1]$. These values are used to obtain the *true* covariance matrix for each sample by a weighted average of two fixed matrices given *a priori*. Scattering vectors with multiplicative Gaussian speckle are obtained on the basis of these matrices and result in the polarimetric input data shown above.

5 Summary

The paper presents a novel descriptor for polarimetric SAR data that characterises image regions in terms of polarimetric scattering mechanisms as well as constellations of elementary geometric image features such as edges, ridges and points. The multi-scale, supervised classifier developed on the basis of this descriptor has been shown to effectively solve challenging classification problems involving complex textures including variations in the polarimetric scattering mechanism. Further investigations will focus on the analysis of real SAR data and an evaluation of the scale selection mechanism in the classifier.

References

[1] Leung, T.; Malik, J.: *Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons*, International Journal of Computer Vision, vol. 43, no.1, pp.29-44, June 2001

[2] Anfinsen, S.N.; Eltoft, T.: *Application of the Matrix-Variate Mellin Transform to Analysis of Polarimetric Radar Images,* Geoscience and Remote Sensing, IEEE Transactions on , vol.49, no.6, pp.2281-2295, June 2011

[3] Doulgeris, A.P.; Anfinsen, S.N.; Eltoft, T.: *Automated Non-Gaussian Clustering of Polarimetric Synthetic Aperture Radar Images,* Geoscience and Remote Sensing, IEEE Transactions on , vol.49, no.10, pp.3665,3676, Oct. 2011

[4] De Grandi, G.D.; Jong-Sen Lee; Schuler, D.L.: *Target Detection and Texture Segmentation in Polarimetric SAR Images Using a Wavelet Frame: Theoretical Aspects,* Geoscience and Remote Sensing, IEEE Transactions on , vol.45, no.11, pp.3437-3453, Nov. 2007

[5] Schou, J.; Skriver, H.; Nielsen, A.A.; Conradsen, K.: *CFAR edge detector for polarimetric SAR images,* Geoscience and Remote Sensing, IEEE Transactions on , vol.41, no.1, pp.20-32, Jan 2003

[6] Conradsen, K.; Nielsen, A.A.; Schou, J.; Skriver, H.: *A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data,* Geoscience and Remote Sensing, IEEE Transactions on , vol.41, no.1, pp.4-19, Jan 2003

[7] T. Minka: *Estimating a Dirichlet distribution.* Technical report, M.I.T., 2000.