AN ADVANCED NON-GAUSSIAN FEATURE SPACE METHOD FOR POL-SAR IMAGE SEGMENTATION

A. P. Doulgeris and T. Eltoft

Department of Physics and Technology, University of Tromsø, Norway

ABSTRACT

This work extends upon our simple feature-based multi-channel SAR segmentation method to incorporate highly desirable statistical properties into a computationally simple approach. The desirable properties include Markov random field contextual smoothing and goodness-of-fit testing to automatically obtain the significant number of classes. To achieve this we need to find an explicit class model to fit these non-Gaussian, non-symmetric or skewed feature space clusters. We take the skewed scale mixture of Gaussian scheme to model our classes and approximate it by a number of constrained Gaussians, thereby retaining much of the speed and simplicity of the original feature space method. The algorithm will be demonstrated on a real data and compared to an automatic Gaussian model.

1. INTRODUCTION

Understanding and implementing the best knowledge about PolSAR signal processing and speckle variation has led to highly complex and computationally slow clustering algorithms such as the U-distribution for multi-look complex (MLC) data segmentation [3]. These models are very flexible, and hence, good fits to the non-Gaussian class properties (radar texture) found in real radar imagery, and the statistical rigour has allowed some extremely useful additional information to be determined. Of main importance is the statistically significant number of classes, that may be determined through goodness-of-fit hypothesis testing [1, 2].

A computationally much faster approach is to extract simple features across the whole image (in a sliding-window method) and then apply a fast and simple segmentation method to the new features. The chosen features are extracted within the non-Gaussian product model framework and therefore include the concept of radar texture [3, 4]. Since the features were found to display simple compact clusters, we have obtained quite satisfactory results by using a simple mixture of Gaussian clustering in this feature space. That is, the results gave smooth segmented regions and captured the main categories observed in the images. However, some limitations were readily observed with this approach and their solution is addressed in this study.

Firstly, the observed features are not exactly Gaussian in distribution, even after smart transformations, and therefore the clustering often converged to having several Gaussian classes centred on approximately the same non-Gaussian data cluster. This produced onion-ring type effects in the segmented images and rather complicated results with very many classes. Secondly, too many statistically significant classes are found, because the Gaussian based goodness-of-fit testing is not appropriate for these non-Gaussian classes. Thirdly, the Markov random field (MRF) contextual smoothing had complications due to the the split-and-merge operations changing the number of classes and causing the optimised MRF smoothing parameter to be inappropriate and too dominant for the new number of classes.

The poor fitting of the Gaussian model to the clusters could be partially avoided by using a high sub-sampling factor and only finding the coarse-level, or most significant, classes on the reduced data-set before expanding to the whole image. Satisfactory results were often only obtained after an extensive, and time consuming, study at many different number of classes and often still found too many clusters for the non-Gaussian classes, before separating certain desirable classes.

By introducing non-Gaussian mixture models we expect to obtain better fitting and fewer total clusters. We have chosen the family of scale mixture of Gaussian models, also known as *normal variance-mean mixture* models [5], because they offer a simple mechanism for both non-Gaussianity and skewness. In addition, they are readily approximated by a discrete mixture of Gaussians, constrained to the equivalent scale variable properties, and thus retain much of the speed and simplicity that were the key advantages of the original feature space method.

We observed that in addition to having non-Gaussian peak shapes, the feature space domain retained some curvature and the resulting compact clusters were slightly skewed. We have had to introduce a skewed model to achieve sufficiently good fits for the hypothesis testing. The scale mixture of Gaussian scheme easily allows this through the "beta-z" term and easily translates into the finite Gaussian approximation. The overall assumption is that the feature parameters vary smoothly and continuously in a constrained manner around their mean value and, therefore, any modes and significant lumps in the distribution are due to additional classes or sub-classes.

The complications due to the Markov random field parameter and the changing number of classes is also addressed with a new, simpler strategy. We now consider it best to incorporate the MRF smoothing only after a non-contextual, adaptive, finite mixture modelling stage is complete and the classes are set. The *local prior* MRF stage subsequently evolves smoothly from the *global prior* classification stage. In addition, we now propose a sequential converge and split strategy, instead of an adaptive split and merge, because the merging approach requires an excessive amount of computation. This new strategy is simpler, often faster, repeatable, and will avoid the dominating effect of the MRF as the number of classes adjusts.

This paper briefly recaps the feature space method in Sec. 2, the non-Gaussian approximation technique in Sec. 3, and the strategy of the new approach in Sec. 4. We show initial results for a real PolSAR image, in Sec. 5, to demonstrate the method and discuss the results, and finish with our conclusions in Sec. 6.

2. EXTENDED POLARIMETRIC FEATURE SPACE METHOD

The approach of this paper is to extract a set of real valued features by the sliding window technique. We have previously found good results from a basic set of six features, but it is possible to use any other features if they are suitably transformed following the guidelines in the extended polarimetric feature space approach [4].

Given a fully polarimetric quad-pol data-set, in single-look complex vector form

$$\mathbf{s} = [S_{hh}, S_{hv}, S_{vh}, S_{vv}]^T,$$

then, for each sliding local neighbourhood, determine the multi-look covariance matrix

$$\mathbf{C} = \frac{1}{L} \sum_{i=1}^{L} \mathbf{s}_i \mathbf{s}_i^H,$$

and the basic set of six features

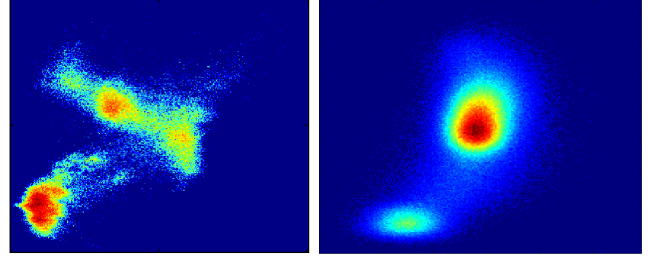1. A non-Gaussianity measure: relative kurtosis RK

$$\text{RK} = \frac{1}{Nd(d+1)} \sum_{i=1}^{N} [\mathbf{s}_i^H \mathbf{C}^{-1} \mathbf{s}_i]^2$$

2. An absolute backscatter value: multi-variate radar cross section

$$\text{MRCS} = \sqrt[d]{\det(\mathbf{C})}$$

3. A cross-polarisation fraction or ratio:

$$R_{\text{cr}} = \mathbf{C}_{hvhv}/\text{MRCS}$$

Co-pol ratio vs. MRCS   Correlation real part vs. MRCS

Fig. 1. Feature space examples from Radarsat-2 quad-pol scenes, San Francisco (left), Sea ice (right), showing distinct globular clusters, with potential skew.

4. A co-polarisation ratio: $R_{\text{co}} = \mathbf{C}_{vvvv}/\mathbf{C}_{hhhh}$

5. The co-polarisation correlation magnitude: $|\rho|$

$$\rho = \mathbf{C}_{hhvv}/\sqrt{(|\mathbf{C}_{hhhh}|\,|\mathbf{C}_{vvvv}|)}$$

6. The co-polarisation correlation angle:

$$\angle\rho = <\phi_{hh} - \phi_{vv}>$$

We have found several simple transforms, empirically, that reduce the non-linear spread, or curvature, in the feature domains. We take the inverse relative kurtosis, the logarithm for each of the multi-variate radar cross section, cross-pol fraction, and co-pol ratio, and take the real and imaginary parts of the co-polarisation correlation coefficient instead of magnitude and phase. This produces a reasonably independent set of features, that often show clear globular clusters in the feature space. These features are then (approximately) suitable for a simple distance-based clustering algorithm, such as the mixture of Gaussian model with the expectation maximisation (EM) algorithm [6]. Two examples in feature space are shown in Fig. 1.

3. SKEWED SCALE MIXTURE OF GAUSSIAN MODEL

The skewed scale mixture of Gaussian scheme, also called normal variance-mean mixtures [5], can be described by

$$\mathbf{y} = \boldsymbol{\mu} + z\boldsymbol{\beta} + \sqrt{z}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{x}$$

where $\boldsymbol{\mu}$ is a vector centre, $\boldsymbol{\beta}$ is a skew vector, $\boldsymbol{\Sigma}$ is the class covariance matrix, $\mathbf{x}$ is a normalised multivariate Gaussian variable, i.e., $\mathbf{x} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I})$, and z is a scalar scale random variable with some continuous, positive only distribution. To cover a very flexible range of scale distributions, and therefore very flexible range of non-Gaussian scale mixture distributions, we have chosen to use the F-distribution for the

scale parameter, because of its flexibility. Parameter estimation for the F-distribution is currently performed with a mix of moment expressions and a simple annealing type search algorithm for the vector beta and the two shape parameters of the F-distribution. This is a slow iterative process and would ideally be substituted with a faster fully moment or cumulant based method. However, we have not had time to solve this problem for the skewed-case and it currently runs fast enough to test the results.

We propose to approximate the continuous scale mixture of Gaussian scheme with a finite number of scales to retain the speed and simplicity of the clustering stage, and to avoid the need for numerically difficult and slow hypergeometric functions. The required level of the approximation depends on the degree of non-Gaussianity of the clusters and the sample size used in the estimation. The more samples involved, the more accurate the approximation needs to be to avoid unnecessarily failing the goodness-of-fit tests. A coarse level approximation might use 5 scale mixture Gaussians per class, and a more detailed on could use 9 or more. The speed of the finite mixture of Gaussian clustering is approximately linear with the number of clusters and is not unreasonable when using several Gaussians per class, since it is already quite fast. The sequential splitting of one cluster at a time means that most cluster centres already start each stage very close to their final values, which also benefits the overall speed.

We choose the set number of scales at the equal probable partition of the F-distribution domain, and thus can constrain their individual prior probabilities to be equal. per sub-cluster. This makes the mixture modelling simple, with fewer degrees of freedom, and leads to a simple visualisation of the non-Gaussianity.

A simulated test sample with skew and the fitted mixture is plotted in Fig. 2. The left plots show the data histograms for each dimension with the individual sub-cluster fitted curves in red, with the overall mixture density in dashed black, and demonstrates how the sum of the sub-cluster Gaussian models can sum up to the skew-non-Gaussian peak-shape. The other plot shows the 2-D data scatter in grey with the 3-sigma confidence ellipses corresponding to each scaled sub-cluster Gaussian. The skew and non-Gaussian scaling can be clearly interpreted by the scaling and off-set of each ellipse. These indicate that the model fitting appears to be working correctly. This flexible non-Gaussian and potentially skewed model should hopefully capture the main cluster density centres and domain curvature and lead to robust clustering results.

4. AUTOMATIC SEGMENTATION

We now propose a new simpler strategy for the automatic adaptive number of clusters. The previous strategy would both split poor fitting clusters and merge pair-wise similar clusters. The pair-wise merge testing very quickly involves a large number of combinations, each requiring estimation of
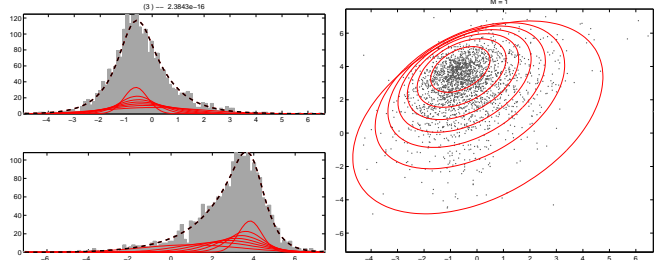
Fig. 2. Simulated data. (Left) Fitted sub-cluster models (red) and total density (dashed black) over the data histogram (grey) for each dimension. The Gaussian sub-clusters sum up to the non-Gaussian peak shape. (Right) 2-D scatter plot of data (grey) with 3-sigma ellipses (red) for each scaled Gaussian sub-cluster models. The skewness and non-Gaussianity can be interpreted from the scaling and off-set of each ellipse.

the merged parameters, and was simply impractical. In addition, the adaptive number of clusters interfered with the MRF model parameter estimation and often caused the contextual information to dominate the pixel-wise information and reduced the effectiveness of the merge testing.

The new strategy is to completely separate the adaptive number of clusters from the MRF contextual smoothing. The first adaptive number of clusters stage finds the number and parameters that satisfactorily fit the entire data-set, constrained to the given model. This includes global prior probabilities for each class.. The second stage with the MRF smoothing will take into account the local information to enhance the local prior probabilities and produce smooth results. It is easy to transition smoothly between the two stages by starting the MRF with the global prior probabilities for the contextual weights.

The new adaptive strategy is to run the EM-algorithm for each number of clusters to full convergence before testing the goodness-of-fit of each cluster, and then only splitting the single worst fitting cluster if over the statistical confidence level. We still propose to start with a single cluster for the entire image, as this solves the initialisation influence problem and seems to repeatedly produce the same end result. This new sequential strategy seems to adapt robustly and avoids the need to re-merge any overly split classes.

5. RESULTS

Preliminary results for the San Francisco, Radarsat-2 sample scene from 2008 are shown in Fig. 3. The three images are the Pauli RGB image, on the left, the result from an automatic number of classes clustering with the standard Gaussian model, in the middle, and the result from the new skew-non-Gaussian model on the right. The Gaussian model found 6 classes and the skew-non-Gaussian model found 7 classes.
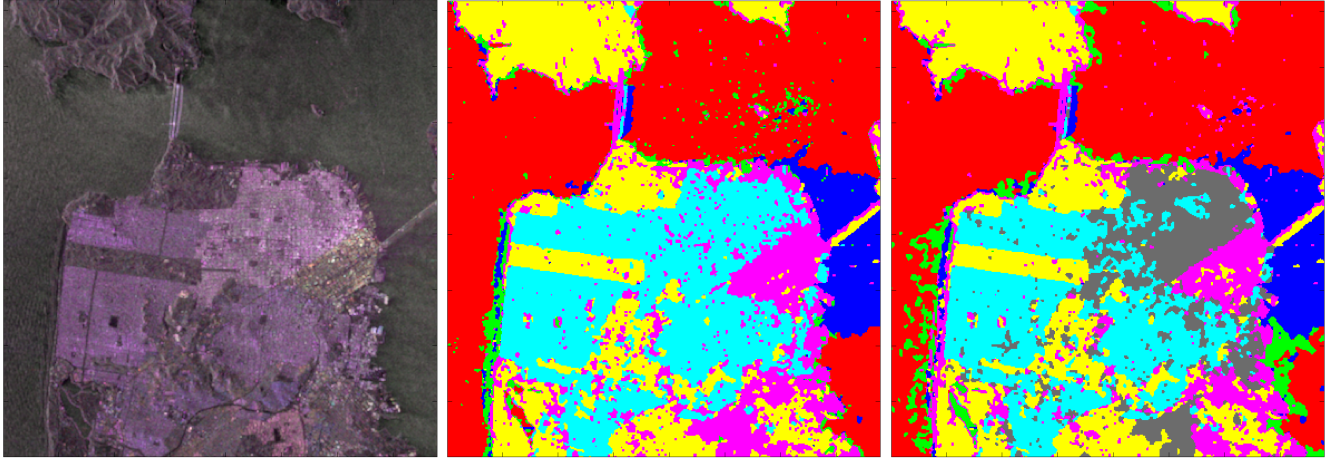
Fig. 3. Results for San Francisco, Radarsat-2. (Left) Pauli RGB; (middle) auto Gaussian, 6 classes; and (right) auto skew-non-Gaussian model, 7 classes.

We argue that the latter has a better visual separation of the main classes and smoother solid areas than the former. So although the non-Gaussian model can better absorb the variation of the non-Gaussian peak shapes into one cluster, it then displays more polarimetric distinguishing power for the remaining classes and resulted in one extra class. A more detailed version (less sub-sampling, not shown) resulted in 20 classes for the Gaussian model and 22 classes for the non-Gaussian model, which again seemed to have a better visual distinction.

6. CONCLUSIONS

We have developed a simple approximation model that includes non-Gaussianity and skewness and is suitable for the automatic number of classes technique. The skewed scale mixture of Gaussians model is chosen for each cluster, because it easily incorporates both non-Gaussianity and skewness, and can be approximated by a discrete number of scaled Gaussians in order to retain some clustering simplicity and speed. Initial tests still use a slow parameter estimation routine that would need to be improved for practical use.

The full effect of this non-Gaussian skewed modelling, and whether it achieves any significant benefits has not yet been fully explored. There are several factors that complicate the interpretation. The number of sub-clusters for each class limits the level of non-Gaussianity that can be achieved and the goodness-of-fit false alarms due to the approximation. We needed 7, or more, sub-classes to clearly obtain smoother results for some real data-sets. The sub-sampling factor, and hence the sample size used in the testing, also affects the final outcome - fewer samples can support fewer distinct clusters. The non-Gaussian clustering does appear to have better regional smoothing in some classes, has sharper boundaries

in others, and also obtained more overall classes due to improved polarimetric distinction. A detailed investigation is still on-going.

7. REFERENCES

[1] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Automated non-Gaussian clustering of polarimetric synthetic aperture radar images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3665–3676, 2011.

[2] A. P. Doulgeris, V. Akbari, and T. Eltoft, "Automatic PolSAR segmentation with the U-distribution and Markov random fields.," in *9th European Conference on Synthetic Aperture Radar (EUSAR2012)*, Nuremberg, Germany, April 23-26 2012.

[3] A. P. Doulgeris and T. Eltoft, "Scale Mixture of Gaussian Modelling of Polarimetric SAR Data," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 874592, pp. 1–12, 2010.

[4] A.P. Doulgeris, "A simple and extendable segmentation method for multi-polarisation sar images," in *POLinSAR 2013*, Frascati, Italy, January 28 - February 1 2013.

[5] O. Barndorff-Nielsen, J. Kent, and M. Sorensen, "Normal variance-mean mixtures and z distributions," *International Statistical Review / Revue Internationale de Statistique*, vol. 50, no. 2, pp. 145–159, 1982.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1–38, 1977.