Initialization of Markov Random Field clustering of large polarimetric SAR images

Thanh N. Tran, *member, IEEE*, Ron Wehrens, Lutgarde M.C. Buydens
Department of Analytical Chemistry
University of Nijmegen
Nijmegen, The Netherlands
lbuydens@sci.kun.nl

Dirk H. Hoekman, *member, IEEE*
Department of Environmental Sciences
Wageningen University
Wageningen, The Netherlands

***Abstract*— Markov Random Field clustering, utilizing both spectral and spatial inter-pixel dependency information, often provides higher accuracy for remote sensing images, such as polarimetric SAR images. However, it is heavily sensitive to initial conditions, i.e. the initialization of parameters and the choice of the number of clusters. In this paper, an initialization scheme for MRF clustering approaches for polarimetric SAR images is suggested. The method takes into account spatial relations between pixels and provides a guideline to the choice of the number of clusters using Pseudolikelihood Information Criterion (PLIC) criterion. A well-known polarimetric SAR image of Flevoland in the Netherlands is given as an example, showing that this approach gives very good performance.**

Keywords-Image clustering; Spatial information; Parameter estimation; ICM;

I. Introduction

Markov Random Field (MRF) clustering has been widely used for remote sensing [1][2][3][4]. The method provides a way to integrate spatial information in terms of inter-pixel dependency in the spatial domain with a mixture modeling approach [5][6]. The accuracy of MRF clustering methods, however, is much more dependent on the initial guess of cluster parameters than the classical mixture modeling algorithm. Classification results can be very poor when the estimation of the initial parameters fails [7]. This is because MRF clustering methods tend to converge very rapidly [8] and a locally optimal solution is often obtained instead of the global solution. The method thus requires the initial parameters to be not very far from the true parameters. The initialization scheme is often simply random, or sometimes it is obtained from other clustering techniques, such as k-means [7] or fuzzy c-means. An agglomerative hierarchical clustering (AHC) framework can also be used as an initialization scheme. This has been proposed for the mixture modeling in [9]. It provides a dendrogram, representing nested clusters. Then, the initial parameters for mixture modeling, as well as MRF clustering in this case, can easily extracted for different numbers of clusters. However, this AHC initialization method has difficulties dealing with large and noisy images.

In this paper, a new AHC initialization framework to MRF clustering, which is suitable for large multispectral remote sensing image data, e.g. SAR images, is proposed. Instead of building up the dendrogram from a huge number of singleton clusters, the method is applied to a limited number of spatially homogeneous regions, gathered from a so-called “multi-scale homogeneity test”. A deterministic Bhattacharyya or a probabilistic likelihood can be used as the merging criterion.

A practical example is evaluated on a polarimetric SAR image of an area in Flevoland in the Netherlands. The clustering of the SAR image in this case is based on statistical image processing techniques [10][11].

II. Basic Elements in Mixture Models and Markov Random Field Clustering

A. *Mixture models*

In mixture modeling or model-based clustering [5][6], every cluster c is described by a multivariate distribution f with parameters θ_c. Most commonly, f is the multivariate normal (Gaussian) distribution, and θ_c contains mean μ_c and covariance Σ_c. The total data set is described by a linear combination of individual clusters and the coefficients correspond to mixture proportions π_c. The probability density function of the pixel x_i under a g-component (cluster) mixture is given by:

$$f(x_i;\Psi)=\sum_{c=1}^{g}\pi_c f(x_i;\theta_c) \tag{1}$$

Now, the probabilistic likelihood function is given by the following expression:

$$L(\Psi)=\prod_{i=1}^{N} f(x_i;\Psi) \tag{2}$$

where N is the total number of pixels, and Ψ contains all cluster parameters and mixture proportions. The aim of the model-based clustering is to obtain a configuration Ψ in which it maximizes the likelihood $L(\Psi)$, This is equivalent to the optimizing the log-likelihood:

$$\log L(\Psi)=\sum_{c=1}^{g}\sum_{i}^{N}u_{ic}\log(\pi_c f(x_i;\theta_c)) \tag{3}$$

where u_{ic} corresponds to the conditional probability of object x_i belonging to cluster c. The maximization of the log-likelihood is usually performed by the EM (Expectation Maximum) algorithm.

B. *Markov Random Field and Mixture Modeling*

Model-based clustering can be combined with the Markov Random Field (MRF) to take into account the spatial relation between pixels. In literature, the MRF model on mixture modeling first has been applied for the restoration method of 'dirty' images [8] and referred to a smoothing technique which gives more weight to the fuzzy class memberships of spatial neighbor clusters. It is assumed that the class probability of a pixel is only dependent on class memberships of its (spatial) neighbors. In MRF models, a w-th order neighborhood system for a particular pixel *i*, called ∂_i, is defined as a set of neighbor pixels belonging to a rectangular window of size w, centered at the pixel i. The estimation of the conditional probability of point x_i of belonging cluster c is:

$$P(x_i \mid c) \propto \exp\left[\beta \sum_{j \in \partial_i} u_{jc}\right] \tag{4}$$

where β is a spatial smoothness parameter. A higher positive β corresponds to higher spatial dependency of neighbor points. The EM algorithm is then adapted, leading to the log likelihood criterion:

$$\log L_{MRF}(\Psi) = \sum_{c=1}^{g} \sum_{i}^{N} u_{ic} \log(\pi_{ic} f(x_i;\theta_c)) \tag{5}$$

Where the mixture proportions π_c is now replaced by the transition probability π_{ic} [5]:

$$\pi_{ic} = \exp\left[\beta \sum_{j \in \partial_i} u_{jc}\right] \Big/ \sum_{h=1}^{g} \exp\left[\beta \sum_{j \in \partial_i} u_{jh}\right] \tag{6}$$

In remote sensing, this method frequently improves the separation between various ground cover classes [1][2][3][4]. However, the algorithm tends to converge too fast. The clustering accuracy is thus heavily dependent on the initial guess Ψ^o and the choice of number of clusters [7][8]. This requires the initial cluster parameters to be chosen very carefully. Hence, obtaining a good Ψ^o is a key element of Markov Random Field clustering.

III. Hierarchical Agglomeration Initialization Scheme

The AHC usually starts from *N* singleton clusters. Iteratively, the similarities between all cluster pairs, *i* and *j*, are calculated and two 'closest' clusters are merged. The algorithm ends if there is only one cluster. Variants differ mainly according to the criterion for optimality, the cluster similarities. Single-linkage, Complete-linkage, and Average-linkage are classical agglomerative methods with the merging criterion to be nearest, farthest, and average neighbor. In model-based hierarchical clustering, the probabilistic likelihood similarity is used, and a maximum-likelihood pair is merged at each stage according to a specific model [9].

The AHC algorithm yields a dendrogram, representing nested clusters and similarity levels where clusters are joined. The dendrogram can be cut at several levels in order to obtain any number of clusters. The equivalent parameters are extracted and they can be used for initialization of the mixture modeling [9]. The main disadvantages of this method lie in dealing with large and noisy images. The latter problem can be tackled by applying filtering/smoothing techniques. However, any incorrect use of these techniques can disturb structures in the image. Application of AHC to large images is difficult because the method demands high computation time and computational resources, proportional to the square of the number of pixels. In order to reduce the computation time, one solution is to apply the method on a small subset of pixels, which is a representative of the whole image data; usually, random samples are taken. An alternative method is to pre-process the image by grouping pixels into a small number of groups, so-call "representative partitions". The AHC clustering is then applied to these representative partitions rather than to all *N*. The minimum spanning tree [12] and k-means, for example, are used to create such partitions. The full clustering procedure is summarized in the flowchart below. The initialization scheme consists of first three steps.

<u>The algorithm steps:</u>

1. Determine a range M [Minclus, …, Maxclus] clusters.
2. Obtain representative partitions of the image.
3. Agglomerative Hierarchical Process (merging criterion: a deterministic similarity or a probabilistic likelihood).
4. Initial parameters for M models.
5. Apply MRF clustering of [Minclus, …, Maxclus] models.
6. Select the best model by using PLIC criterion.

Here, in **step 2**, the "representative partitions" are obtained from a "multi-scale homogeneity test".

<u>Test of complete homogeneity</u>: Given two groups of pixels, A and B, with mean vectors and covariance matrices, $\{\mu_A \hat{\Sigma}_A\}$ and $\{\mu_B, \hat{\Sigma}_B\}$, respectively. The test of complete homogeneity of two groups under the hypothesis H_c; $\mu_A = \mu_B$ and $\hat{\Sigma}_A = \hat{\Sigma}_B$ is the likelihood ratio test $\lambda = L_c / L$, where L_c and L are the maximized likelihoods under the hypothesis H_c and the unconstrained maximum likelihoods, respectively. The statistic $-2\log\lambda = (n_A + n_B)\log|\Sigma_{<AB>}| - (n_A \log|\Sigma_A| + n_B \log|\Sigma_B|)$ has an asymptotic chi-squared distribution with $0.5d(d+3)$. degrees of freedom, where *d* is the number of input bands of the image data.

Hence, two groups of pixels, *A* and *B*, are said to be "completely homogeneous" at significance level α if and only if $-2\log\lambda$ is not significantly than (larger than) the critical value provided by the chi-squared distribution.

A "spatial region" (or "representative partition"), *r*, in a multispectral image is defined to be a group of pixels forming a continuous region in spatial domain, e.g. a rectangle or an ellipse.

Homogeneity test for "spatial region": Given a "spatial region" r, e.g. a square window, and a set of sub spatial-regions r_i, the "spatial region" *r* is said to be "totally homogeneous" at significance level α, for instance 0.05, if for all pairs of spatial sub-regions $<r_i, r_j>$ the test of complete homogeneity is not rejected at the significance level α.

The homogeneity test for "spatial region" can also be repeatedly applied to all sub-spatial regions ri, as in Fig.1. This is called the "multi-scaled" test and it is stronger than the single test, and less representative partitions are obtained. The two levels multi-scaled test (double-scaled) is used in this study.

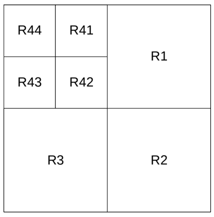

Figure 1: Multi-scale homogeneity test for "spatial region".

In **step 3**, the merging criterion in AHC clustering can be a deterministic or a probabilistic likelihood similarity. For Gaussian distribution data, the probabilistic likelihood similarity is used in the model-based clustering [6][9]. In this study, Bhattacharyya distance is used for the deterministic similarity. The Bhattacharyya distance gives the distance between two Gaussian regions r1 and r2:

$$B(r1,r2)=\frac{1}{8}(\mu_{r1}-\mu_{r2})^T\left(\frac{\Sigma_{r1}+\Sigma_{r2}}{2}\right)^{-1}(\mu_{r1}-\mu_{r2})+\frac{1}{2}\ln\left(\frac{|\Sigma_{r1}+\Sigma_{r2}|}{2\sqrt{|\Sigma_{r1}||\Sigma_{r2}|}}\right) \quad (7)$$

The Bhattacharyya distance consists of two terms, which are dominated by the differences in means, and covariance, respectively. It is very close to the Bayes error of two clusters.

In many cases, outliers are also present in the "representative partitions". By the hierarchical mechanism, they are trapped into isolated singleton clusters. The real number of clusters, thus, can be defined after these singleton clusters (outliers) are eliminated.

In **step 4**, a range of interesting mixture models [*Minclus*, …, *Maxclus*] can be gathered from the dendrogram, after removing the outliers. Then, the followed MRF clustering uses the corresponding initial parameters extracted for each model in **step 5**.

The computational complexity of the initialization scheme is dependent on **step 2** and **step 3**. In **step 2, t**he complexity is small of O(N.*w*), in which *w* is the size of the spatial regions. The number of representative partitions, say *M*, is relatively small in the hierarchical merging process, **step 3**. The total complexity of this step is O(M^M), which can be managed to be reasonably small depending on the value of *M*. If *M* is still too high to get results within a reasonable time, a sampling technique can be used.

The Bayesian Information Criterion (BIC) criterion is often used for the traditional mixture modeling [6]. The Pseudolikelihood Information Criterion (PLIC) criterion is adapted from the BIC criterion for the MRF modeling [13]: $PLIC = 2\log L_{MRF}(\Psi_k) - d_k \log(N)$, where the MRF log-likelihood function, $\log L_{MRF}(\Psi_k)$, is used instead of the log-likelihood function. PLIC criterion is used in **step 6**.

IV. Application on SAR Data

The data is a well-known image of Flevoland, an agricultural area in The Netherlands, acquired by the NASA/JPL AirSAR system (C-, L- and P-band polarimetric) on 3 July 1991. For a clustering procedure statistical descriptions are needed for pixels belonging to a certain cluster, rather than belonging to a certain homogeneous area. In polarimetric SAR image, such distributions do not follow directly from theoretical considerations. Assumptions should be made, which have to be carefully verified with experimental data. In [14][15], field averaged backscatter intensities of a certain class are well described by the log-normal distribution. In [15] the validity of this assumption is demonstrated for the Flevoland data set which used in this paper. The classification can be performed on these logarithmically scaled (dB values) intensity images and a multivariate distribution can be applied, which simply is the multivariate normal. Note that for an "individual homogeneous field" the complex Wishart distribution and its marginal distributions are appropriate. Consequently, using the logarithmic dB-scale to express radar backscatter intensities, 3 multivariate Gaussian models, termed 9I (nine intensities), have been proposed. Readers are referred to reference [15] for a detailed discussion of this transformation. In this study, 18 intensities from the C- and L-band of full polarimetry model I9 are used for clustering .

The study area has a size of 400 x 400 pixels and is taken from the original data without any aggregation process. Fig. 2a shows the false-color image of the first 3 intensities of the C band. The crop type map which is the ground truth for the clustering is shown in Fig. 2b. The Yellow color is a mask where the ground truth is uncertain, or not recorded. 227 homogeneous regions, showing in Fig. 2c, are identified by double-scaled homogeneity test for spatial region on the grid-image of 3136 windows of 15x15 pixels size.

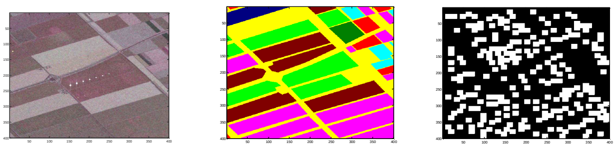
a) b) c)

Figure 2. a) shows the false-color image by 3 first intensities on C-band, b) the ground-truth information of the site to 7 clusters (Barley, Sugar Beet, Winter Wheat, Lucerne, Rapeseed, Peas, and Potato), and c) 227 homogeneous regions.

Model parameters are estimated by using the Bhattacharyya model. Six cluster-models corresponding to [5, ..,10] clusters are extracted from the dendrogram after removing outliers (singleton regions), shown in Fig. 3.

Then, MRF clustering is applied for each cluster-model, using the initial parameters obtained from the Bhattacharyya model, with a rectangular window of size 5 x 5 as the neighborhood and β = 1. The results of four selected models are shown in Fig. 4 and the corresponding PLIC values are in Table 1.

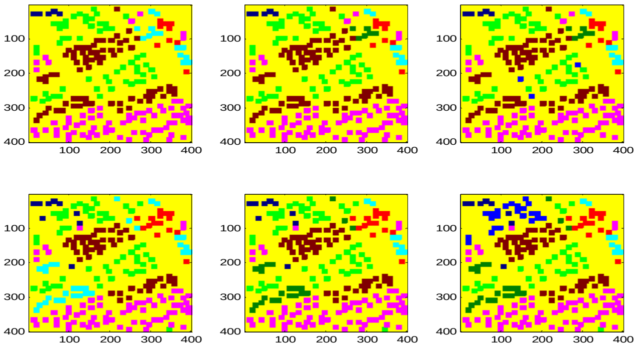

Figure 3. : Initial parameters images of deterministic Bhattacharyya [5, 6, …, 10] cluster-models.

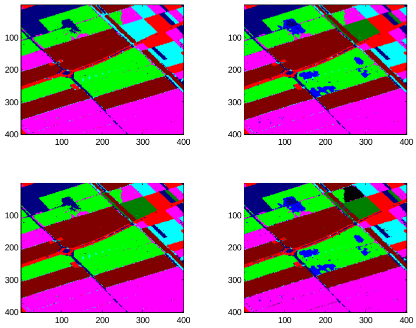

Figure 4. CMM clustering result for four selected models for: a) six, b) seven, c) eight, and d) ten clusters.

TABLE I. PLIC VALUES FOR SEVEN MODELS

Models	PLIC	Models	PLIC
5	-7.359557	8	-7.097460
6	-7.154923	9	-7.088239
7	-7.131478	10	-7.081267

The best model, according to the PLIC values, is the ten-cluster model shown in Fig. 4d. It obtains extra three small clusters, capturing the variations of the Barley, Winter Wheat and Rapeseed clusters in the spatial domain. In this case, the true number of clusters is seven. This model obtains more than 97 % accuracy on the area for which reference information is available (not the yellow area in Fig 2b). The 7-cluster model has a significantly higher PLIC value than the 5- or 6-cluster models.

V. CONCLUSION

We proposed in this work a fairly simple, straightforward and robust initialization method for MRF clustering. The clustering method is applied to a large polarimetric SAR image, utilizing the full polarimetric information content through a transformation described in [14] The initialization scheme uses an AHC framework on spatially continuous regions, which are gathered from the "multi-scale homogeneity test", in order to be able to work with large images. The method works best for an image consisting of many large homogeneous regions, such as agricultural crops areas. Small and spatially isolated clusters may not be recognized by the method and incremental model-based clustering [16] is suggested as a post processing step. The method is quite stable and by the use of the PLIC criterion automatically identifies an optimal number of clusters. A range of number of clusters around the suggested cluster can also be studied. Furthermore, the proposed method does not need pre-processing or a prior segmentation. The example in this work shows excellent results.

REFERENCES

[1] A.H.S. Solberg, T. Taxt, A.K. Jain, "A Markov random field model for classification of multisource satellite imagery", IEEE Trans. on Geosci. Remote Sensing, vol. 34, pp. 100 -113, Jan. 1996.

[2] P.C. Smits, S.G. Dellepiane, "Synthetic aperture radar image segmentation by a detail preserving Markov random field approach", IEEE Trans. on Geosci. Remote Sensing, vol. 35, pp. 844 -857, Jul. 1997.

[3] Q. Jackson, D. A. Landgrebe, "Adaptive Bayesian Contextual Classification Based on Markov Random Fields", IEEE Trans. on Geosci. Remote Sensing, vol. 40, pp. 2454-2463, Nov. 2002.

[4] A. Sarkar, M. Kumar Biswas, B Kartikeyan, V. Kumar, K. L. Majumder, D. K. Pal, "A MRF Model-based Segmentation Approach to Classification for Multispectral Imagery", IEEE Trans. on Geosci. Remote Sensing, vol. 40, pp. 1102-1113, May. 2002.

[5] G. McLachlan and D. Peel, "Finite Mixture Models", Willey series in probability and statistic, Canada, 2000.

[6] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation", J. the Amer. Statist. Asso., vol. 97, pp. 611-631, 2002.

[7] J. Besag, "On the statistical analysis of dirty pictures", Journal R. Statistic, Soc., B, 1986.

[8] R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised Classification of Radar Images using Hidden Markov Chains and Hidden Markov Random Fields", IEEE Trans. on Geosci. Remote Sensing., vol. 41, pp. 675 – 686, Mar. 2003.

[9] C. Fraley, "Algorithms for Model-Based Gaussian Hierarchical Clustering", SIAM J. Sci. Comput., vol. 20, pp. 270-281, 1998.

[10] H. A. Yueh, A. A. Swartz, J. A. Kong, R. T. Shin, and L. M. Novak, "Bayes classification of terrain cover using normalized polarimetric data," J. Geophys. Res. B-12, vol. 93, pp. 15.261–15.267, 1998.

[11] 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

[12] C. Posse, "Hierarchical Model-Based Clustering for Large Datasets", Journal of Computational and Graphical Statistics, vol. 10, pp. 464-486, 2001.

[13] D.C. Stanford, A. E. Raftery, "Approximate Bayes Factors for Image Segmentation: The Pseudolikelihood Information Criterion (PLIC)", IEEE Trans. on Pattern Anal. Mach. Intell., vol. 24 ,pp. 1517-1520, 2002.

[14] D.H. Hoekman and M.J. Quiñones, "Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Colombian Amazon", IEEE Trans. on Geosci. Remote Sensing, vol.38, pp.685-696, 2000.

[15] D.H. Hoekman, and M.A.M. Vissers, "A new polarimetric classification approach evaluated for agricultural crops", IEEE Trans. on Geosci. Remote Sensing, vol. 41, pp. 2881-2889, Dec. 2003.

[16] C. Fraley, A. E. Raftery and R. Wehrens, "Incremental Model-Based Clustering for Large Datasets with Small Clusters", Techn. Rep. no. 439, Dept. of Statistics, University of Washington, Dec. 2003.