

# Image Segmentation Using Binary Tree Structured Markov Random Fields

Kaan Ersahin

December 19, 2004

# 1 Introduction

Synthetic Aperture Radar (SAR) systems are important active sensors for Earth observation systems. They enable continuous monitoring of the targets (e.g., sea ice, agricultural areas, ships, etc.) through clouds and independent of solar illumination. Space-borne polarimetric SAR missions which are planned for the near future draw our attention to the crucial task of information extraction from such data. The radar imaging system is quite different than the optical counterpart, therefore it is important to understand the characteristics of such data to be able to make accurate interpretations. In particular, speckle (the noise-like phenomena) resulting from the coherent imaging system makes the interpretation of SAR images difficult and unreliable on a per pixel basis, so incorporating spatial information is necessary. This can be achieved through Markov Random Fields (MRFs) with a variety of approaches, where the additional information is expected to improve the classification accuracy.

There are several approaches suggested in the literature for image segmentation with the MRF modeling framework, mostly based on pairwise clique potentials and a selection of a suitable neighbourhood (e.g., 4-nearest neighbours, 8-nearest neighbours). An interesting approach, suggested by D'Elia *et al.* [1], uses recursive binary segmentation for solving the K-class problem, motivated by the relatively simple Ising model as opposed to the Potts model. It is an unsupervised technique, where the model parameters and the segmentation map are unknown and need to be estimated simultaneously. The solution is to alternate between parameter estimation and inference steps until convergence. However, an additional problem arises when the number of classes is unknown, and it needs to be estimated as well. D'Elia *et al.* provides a solution by defining a split gain, the ratio of the unnormalized posteriors obtained from two hypotheses, the tree structure before and after the split from a node. Such a test makes it possible to decide which representation fits better to the data and accordingly accepts or rejects the potential split. Although this is a good idea in an unsupervised framework, it should be noted that the class-based hierarchy may not have a physical meaning, thus may be inadequate for a specific data set.

The aim of this project was to form the foundations of applying such a framework on polarimetric radar data by reproducing similar results that are presented in [1]. Since some of the details were not clearly provided and the attempt to contact the author was not successful, due to time constraints the objective has been revised and limited to application of the algorithm to synthetic data. Some work on real data has also been accomplished, but is rather preliminary and is not included in this report. This work has provided enough insight to work with real data, but the statistical properties of such data and noise needs to be studied in more detail, to be able to provide satisfactory results. Therefore, I will not attempt to provide any results for SAR data in this report.

The organization of the report is as follows: Section 2 gives an overview of the MRF image modeling approach, TS-MRF algorithm is explained in Section 3, and data simulation and results are presented in Section 4. Finally, conclusions are drawn and future work is projected in Section 5.

## 2 Image Segmentation Based on MRF Models

The goal of segmentation is to estimate the correct label for each site. A quite common approach is to minimize the probability of error given the data, and hence resort to MAP estimation:

$$\hat{x} = \arg \max_x p(x|y) = \arg \max_x p(y|x)p(x) \quad (1)$$

The likelihood term depends only on the noise distribution, assumed to be known, and hence only two major problems remain: appropriately modeling the prior image distribution with its parameters, and finding the maximum given by (1).

The modeling problem is simplified by assuming that the label image is fully characterized by local dependencies. In particular, if the Markovian property holds for each site  $i$ , namely,  $X_i$  is conditionally independent of the rest of the image given the realization on a neighborhood  $N$ , then  $X$  is said to be an MRF with respect to the neighborhood system  $N$ .

Modeling using an MRF becomes possible with the Markov-Gibbs equivalence given by the Hammersley-Clifford theorem, where an MRF is characterized by its local property (the Markovianity), and Gibbs Random Field (GRF) is characterized by its global property, the Gibbs distribution in the following form

$$P(f) = Z^{-1} e^{-\frac{1}{T}U(f)} \quad (2)$$

where

$$Z = \sum_{f \in F} e^{-\frac{1}{T}U(f)} \quad (3)$$

is a normalizing constant called the partition function,  $T$  is the temperature and the energy function is

$$U(f) = \sum_{c \in C} V_c(f) \quad (4)$$

where  $V_c(f)$  stands for the clique potentials and the energy is the sum over all possible cliques. With the MRF-Gibbs equivalence, the model selection problem amounts to choose a suitable neighborhood system and potentials in order to express the desired spatial dependencies.

A number of models have been proposed in the literature for various applications. For example, a GRF is said to be homogeneous if  $V_c(f)$  is independent of the relative position of the clique  $c$  in  $S$ . This is assumed in most MRF vision models for mathematical and computational convenience. It is said to be isotropic if  $V_c$  is independent of the orientation of  $c$ , which results in blob-like regions. An important special case of Gibbs distribution is when only cliques of size up to two are considered. In this case, the energy can be written as

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{i' \in N_i} V_2(f_i, f_{i'}) \quad (5)$$

This implies the lowest order constraints on two labels to convey contextual information, and they are widely used because of their simple form and relatively low computational cost. A further categorization of this pair-wise MRF model is called auto-models [2] when  $V_1(f_i) = f_i G_i(f_i)$  and  $V_2(f_i, f_{i'}) = \beta_{i,i'} f_i f_{i'}$  where  $G_i(\cdot)$  are arbitrary functions and  $\beta_{i,i'}$  are constants reflecting the pair-wise interaction between  $i$

and  $i'$ . In addition if  $f_i$ 's take on values from a discrete label set  $\{0, 1\}$  or  $\{-1, 1\}$ , this is said to be an auto-logistic model and the corresponding energy is of the following form

$$U(f) = \sum_{\{i\} \in C_1} \alpha_i f_i + \sum_{\{i, i'\} \in C_2} \beta_{i, i'} f_i f_{i'} \quad (6)$$

When  $N$  is the nearest neighborhood system on a 2D lattice (4 nearest neighbors), the auto-logistic model is reduced to the Ising model. The generalization of auto-logistic model to more than 2 states is called multi-level logistic (MLL) also known as Potts model. For a second-order MLL model, where only single-site and pair-site clique parameters ( $\alpha$  and  $\beta$ ) are non-zero, the potential function for a pair-wise clique is

$$V_2(f_i, f_{i'}) = \begin{cases} \beta_c & \text{if states on clique } \{i, i'\} \in C_2 \text{ have the same label} \\ -\beta_c & \text{otherwise} \end{cases}$$

where  $\beta_c$  is the parameter for type- $c$  cliques and reduces to a single  $\beta$  value when the model is isotropic.

Assuming that one has been able to select a satisfactory model for the prior distribution, the problem remains of maximizing  $p(x)p(y|x)$  over  $x$ , which is when computational complexity becomes a dominant concern due to the necessity to evaluate the partition function  $Z$ . This is prohibitive even for moderate size problems, so approximate methods need to be employed (e.g., ICM - iterated conditional modes). Even resorting to the ICM, renouncing global optimality and computational complexity remains the major concern. In that case, the problem gets worse for the unsupervised segmentation, where a number of important parameters such as the number of classes, the parameters of the likelihood term, and the parameters of the Gibbs prior, are not known, and must be estimated from the data together with the segmentation. These parameters are collectively represented by a random vector  $\theta$  and now the problem is to solve the following equation

$$(\hat{x}, \hat{\theta}) = \arg \max_{x, \theta} p(x, y | \theta) \quad (7)$$

Since exact joint optimization is computationally intractable, a two-step Expectation Maximization (EM) procedure is often used, where first the model parameters are estimated from the observed data, and then the MAP segmentation is carried out in a second step using these estimated parameter values.

### 3 Tree Structured Markov Random Fields (TS-MRFs)

The algorithm proposed by [1] is based on the recursive binary segmentation of image regions. The entire image is modeled originally as a binary MRF and segmented accordingly. The generated segments, if any, are modeled as binary MRFs as well. This process goes on recursively until some stopping condition is met. The choice of binary partitions was motivated by the fact that the computational complexity reduces dramatically when compared to the  $K$ -state MRF.

### 3.1 TS-MRF Model

A binary tree  $T$ , is identified by its nodes and their mutual relationships. Except for the root, each node  $t$  has one parent  $u(t)$ , and each internal node has two children  $l(t)$  and  $r(t)$ . As shown in Figure 1, the set of terminal nodes or leaves are the ones who has no children (black nodes), and the rest of the nodes in the tree are defined as the internal nodes (empty nodes).

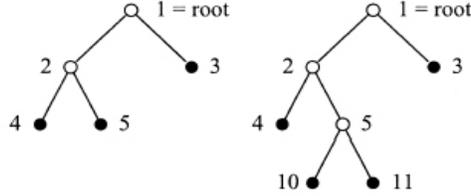


Figure 1: Binary Tree Structure (a)after 2 splits (b)result of an additional split

### 3.2 Unsupervised Segmentation

The unsupervised segmentation has a recursive nature, starts with a single-node tree, which grows according to the following procedure until all the splits from the leaf nodes are rejected.

1. Initialize segmentation  $\hat{x}$  using k-means
2. Find  $\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}_1, \hat{\Sigma}_2$
3. Find  $\hat{\beta}$
4. Run inference to find the new segmentation  $\hat{x}$
5. If not converged, go to step 2
6. Calculate split gain  $G$ , and decide to accept or reject the split

### 3.3 Parameter Estimation

D'Elia *et al.* used the homogeneous, isotropic Ising model and set  $\alpha = 0$ , where single-site cliques are not used and only pair-site cliques are involved. In this case energy takes the following form

$$U(f|\theta) = \sum_{i,i' \in C_2} \beta f_i f_{i'} \quad (8)$$

Parameter estimation can be carried out using ML Estimation. However, for an MRF in general, this requires to evaluate the partition function in the corresponding Gibbs distribution. The partition function is also a function of  $\theta$  and the computation of  $Z(\theta)$  is intractable, which is the main difficulty with parameter estimation for MRFs. Therefore, approximate techniques are used such as maximizing the pseudo-likelihood [3-5], the coding method [2-3] or the mean field approximation [3].

In this study, maximum pseudo-likelihood is used as suggested in [1]. The pseudo-likelihood is defined as the simple product of the conditional likelihood as follows

$$PL(f) = \prod_{i \in S - \partial S} P(f_i | f_{N_i}) = \prod_{i \in S - \partial S} \frac{e^{-U_i(f_i, f_{N_i})}}{\sum_{f_i} e^{-U_i(f_i, f_{N_i})}} \quad (9)$$

Since the model used by D'Elia et al. is homogeneous isotropic Ising model, where  $\alpha = 0$ , the energy function, is given by (8), the objective function to maximize can be written as follows

$$\ln(PL(f|\theta)) = \sum_{i \in S - \partial S} \log \frac{e^{\sum_{i' \in N_i} \beta f_i f_{i'}}}{e^{\beta \sum_{i' \in N_i} f_{i'}} + e^{-\beta \sum_{i' \in N_i} f_{i'}}} \quad (10)$$

and taking the derivative w.r.t.  $\beta$  results in

$$\frac{\partial(\ln(PL(f|\theta)))}{\partial \beta} = \frac{f_i \sum_{i' \in N_i} f_{i'} (e^{\beta \sum_{i' \in N_i} f_{i'}} + e^{-\beta \sum_{i' \in N_i} f_{i'}}) - \sum_{i' \in N_i} f_{i'} (e^{\beta \sum_{i' \in N_i} f_{i'}} - e^{-\beta \sum_{i' \in N_i} f_{i'}})}{e^{\beta \sum_{i' \in N_i} f_{i'}} + e^{-\beta \sum_{i' \in N_i} f_{i'}}} \quad (11)$$

This procedure is carried out using the minimization routine by Rasmussen [6].

### 3.4 Inference

In the case of MRFs, computational complexity becomes a dominant concern due to the necessity to evaluate the partition function  $Z$ . This is prohibitive even for moderate size problems, so exact inference is intractable and approximate methods are required. There are a few techniques for approximate inference (e.g., ICM - Iterated Conditional Modes, LBP - Loopy Belief Propagation), which will only find a local minimum. ICM suggested by D'Elia *et al.* [1] is used for inference, since it is known to be very fast. This is required due to the fact that inference is used as a subroutine in the two-step alternating optimization procedure (EM), which needs to be performed more than once (at multiple nodes) because of the binary tree structure. LBP has also been considered, but the slow nature of this algorithm compared to ICM was the main drawback.

### 3.5 Stopping Criteria

The first hypothesis corresponds to the case in which the whole image, associated with the root node ( $S^1 = S, y^1 = y$ ), is represented as a single region, in this case all sites have the same label attached. The second hypothesis corresponds to the case in which the image is represented by two regions. The MAP estimate divides the image into two new regions,  $S^2$  and  $S^3$  with their associated data  $y^2$  and  $y^3$ .

The two statistical descriptions of the image, based on a single-class model (tree  $T$ ) or a two-class model (tree  $T'$ ), are compared by checking the condition

$$G^1 = \frac{p(\hat{x}^{T'})p(y|\hat{x}^{T'})}{p(\hat{x}^T)p(y|\hat{x}^T)} > 1 \quad (12)$$

If the test succeeds, namely the split gain  $G^1$  is greater than 1, the two-region description better fits the data, so the split is accepted and procedure continues, otherwise stops. From the implementation point of view, The node is first split into two, segmentation is performed, according to the result of the split gain test, those leaves are either kept or removed from the tree structure. Being at node  $\tau$ , the test depends only on region  $S^\tau$ .

## 4 Results

To be able to discuss the results of parameter estimation, inference, and to comment both on the effectiveness of the recursive binary segmentation in an unsupervised fashion, synthetic data is suitable. In this section, the methodology which is used to simulate data will be explained and the obtained results will be presented.

### 4.1 Synthetic Data Generation

To generate a synthetic dataset, the Ising model with a first-order neighborhood system is used. The region  $S$  is visited site-wise and if  $i$  is the current site and  $f_i$  is the configuration, it is updated according to the sampling of associated local conditional distribution. So it is set to  $-1$  with probability  $\propto \exp\{-\beta \sum_{f_j \in N_i} f_j\}$ , and to  $1$  with probability  $\propto \exp\{\beta \sum_{f_j \in N_i} f_j\}$ . Increasing  $\beta$  value will result in increased penalty for non-homogeneity.

This procedure is adapted to follow a given binary tree structure to form the data, which will be used as a reference. After the synthetic data is generated, gaussian noise is added and the noisy data is used as input to the unsupervised algorithm.

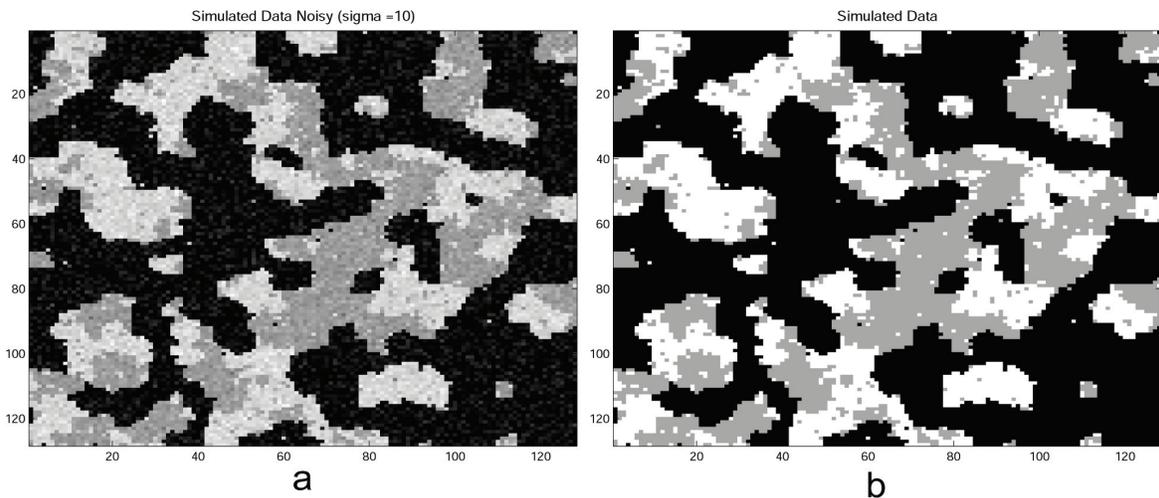


Figure 2: Synthetic Data (a) noisy (b) original segmentation

## 4.2 Results Obtained From Synthetic Data

Data presented in Figure 2 was generated according to the tree structure in Figure 1.a,  $\beta = 0.8$  was used for the first split at node 1, and  $\beta = 0.5$  was used for the second split at node 2. The parameter estimation results after EM converged, were used to make the plots in Figure 3, both results show agreement with the actual values.

In Figure 4.a, the segmentation results obtained after the first split is given, while 4.b shows the result after the second split (from node 2). Both results do make sense in the context of tree structured binary segmentation and the final result agrees well with the original data. The split gain,  $G$  is used to accept or reject a potential split, and the correct tree structure with 3 leaves (classes) was found for the synthetic data used.

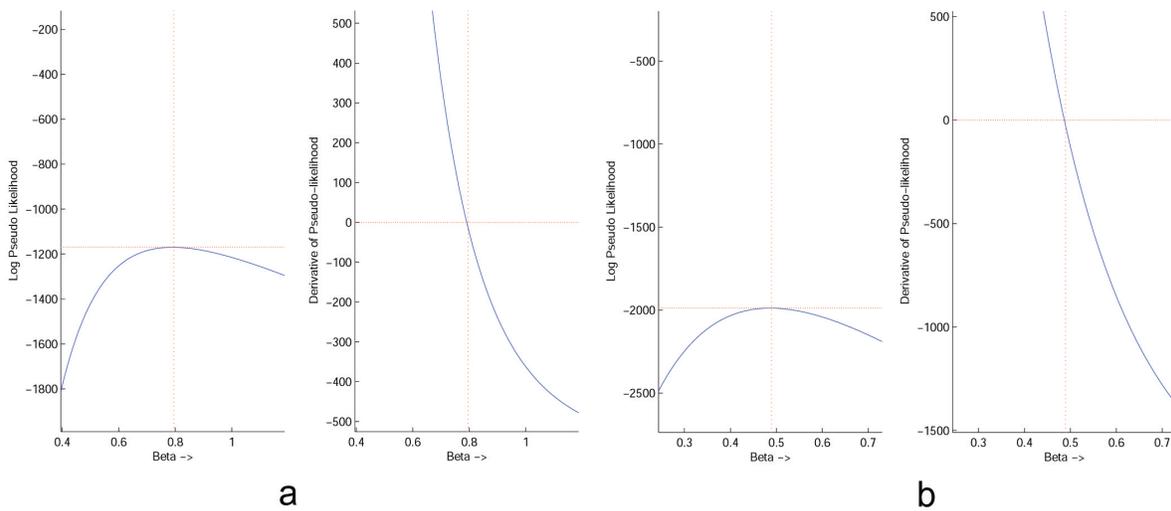


Figure 3: Parameter Estimation (a) first split  $\hat{\beta} = 0.79$  (b) second split  $\hat{\beta} = 0.49$

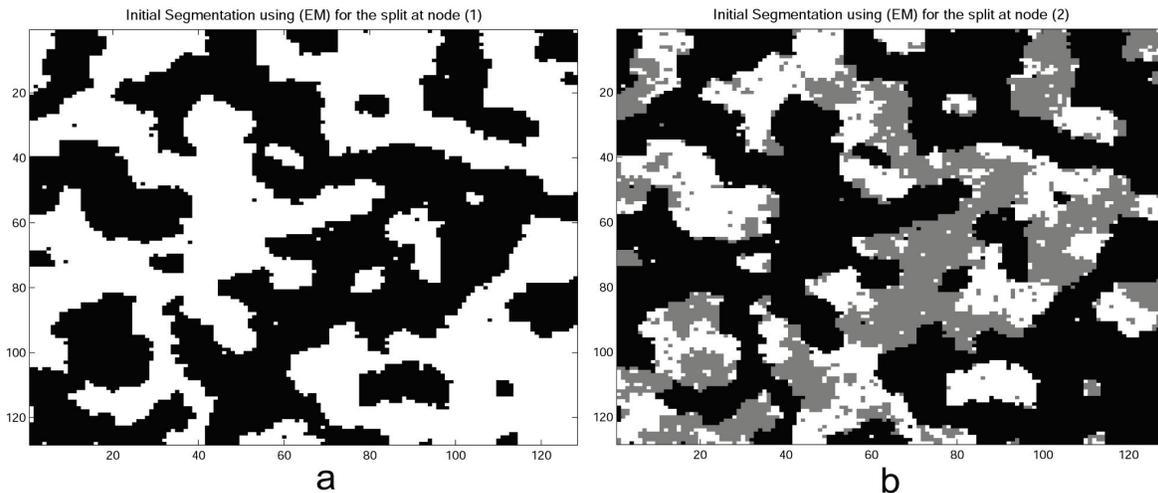


Figure 4: Segmentation Results using TS-MRF (a) after first split (b) final result

## 5 Conclusion and Future Work

Using the synthetic data was helpful for understanding the procedure and implementing the algorithm, and good results for parameter estimation, number of classes and segmentation has been obtained, but at this stage these are far from conclusive. This was also the case in D'Elia *et al.* [1] where the presented results were mostly qualitative except one case, which was a comparison of TS-MRF to the K-state MRF, but not to the ground truth.

This project was originally planned to be extended to the case of real data from a polarimetric radar, but this turned out to be rather complex, due to the fact that, the model used here is not suitable in the case of SAR data and in order to get satisfactory results, one has to study the characteristics of such data set in more detail and choose a model accordingly. The second part of the project, using real data has been partially accomplished (i.e. feature extraction), but suitable model selection is the important task that needs to be investigated further.

Future work includes, using anisotropic models for texture generation and segmentation, comparing approximate inference algorithms on synthetic data, studying the properties of data and the speckle in Polarimetric SAR images to be able to provide a suitable model.

## 6 References

- [1] C.D'Elia, G.Poggi, G.Scarpa, "A Tree-Structured Markov random field model for Bayesian image segmentation," IEEE Transactions on Image Processing, vol.12, no.10, pp.1259-1273, Oct'03.
- [2] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," J. R. Statist. Soc., ser. B 36, pp. 192-236, 1974
- [3] S. Z. Li, Markov random field modeling in image analysis: Springer-Verlag, 2001
- [4] J. Besag, "Statistical analysis of non-lattice data" The Statistician, vol. 24, no. 3, pp. 179-195, 1975
- [5] J. Besag, "Efficiency of pseudolikelihood estimation for simple gaussian fields" Biometrika, vol. 64, no. 3, pp. 616-618, 1977
- [6] C. E. Rasmussen, "Conjugate Gradient Minimizer - 'minimize.m' function in MATLAB", <http://www.kyb.tuebingen.mpg.de/bs/people/carl/code/minimize/>
- [7] C. D'Elia, G. Poggi, and G. Scarpa, "Advances in the segmentation and compression of multispectral images", Proc.IEEE Int.Geoscience Remote Sensing Symp., vol.6, July '01, pp.2671-73.
- [8] C. D'Elia, G. Poggi, and G. Scarpa, "An adaptive MRF model for boundary-preserving segmentation of multispectral images", Proc. XI Eur. Signal Proc. Conf., Toulouse, France, Sept.'02.
- [9] P. Perez, "Markov random fields and images" CWI Quarterly, vol. 11, no. 4, pp. 413-437, 1998