Modeling of Inhomogeneous Markov Random Fields with Applications to Cloud Screening

Technical Report No. 98–21,
Information and Computer Science Department,
University of California, Irvine

Igor Cadez and Padhraic Smyth
Department of Information and Computer Science
University of California, Irvine
CA 92697-3425
[icadez,smyth@ics.uci.edu]

August 1998
Revised, March 1999

Abstract

Cloud screening is the process of classifying pixels in satellite images which contain clouds and is an important step in processing remotely-sensed images. This paper applies inhomogeneous statistical spatial models in the form of Markov random field models (MRF) to this problem and develops an efficient algorithm for the estimation of model parameters. The algorithm has a natural parallel decomposition. The model is tested on synthesized images for which ground truth is known and applied to segmentation of clouds in several Advanced Very High-Resolution Radiometer (AVHRR) images. This paper concentrates on the abstract spatial aspects of the models rather than the details of specific remote-sensing mechanisms. The main results are (1) the formulation (in terms of inference and estimation) of the inhomogeneous MRF model, (2) the exact solution of the "pseudo-likelihood" equations used for parameter estimation in this model, and (3) experimental results which indicate that (a) inhomogeneous models perform better than homogeneous models and (b) that spatial models perform better than non-spatial models for cloud-screening problems.

1 Introduction and Background

This paper introduces an algorithm for locally-adaptive parameter estimation of spatially inhomogeneous Markov random fields (MRFs) using an exact solution to the pseudo-likelihood equations of Besag (1986). The standard MRF model assumes that the coupling between pixel labels (the β parameter) is globally constant throughout the image. An *inhomogeneous* MRF model allows this coupling parameter to vary spatially. Our interest in this model stems from the practical problem of "cloud screening," namely the automatic segmentation of cloud pixels from ground pixels in remote-sensing applications (Simpson and Humphrey, 1990). Existing cloud-screening algorithms make decisions on a pixel-by-pixel basis and do not take advantage of spatial information. The inhomogeneous MRF appears better suited for such images than the homogeneous MRF, since clouds have locally varying spatial scales (e.g., see Figure 1). The method is tested on synthesized images for which ground truth is known and applied to segmentation of clouds in remote-sensing images.

The paper concentrates on the abstract spatial aspects of our models rather than the details of specific remote-sensing mechanisms. Our main results are the formulation (in terms of inference and estimation) of the inhomogeneous MRF model, the exact solution of the pseudo-likelihood equations used for parameter estimation in this model, and experimental results on synthetic images which validate the use of both the pseudo-likelihood estimation methodology and the inhomogeneous model.

1.1 Background on Cloud Screening

Continuous remote sensing of the Earth's surface is now routine in areas such as atmospheric science, ocean science, ecology, agricultural science, etc. In recent decades, the performance of onboard satellite sensors has dramatically improved while the cost of communication has significantly dropped. As a result, very large sets of remotely-sensed images are routinely collected, analyzed, and archived.

For optical, infrared, and submillimeter wavelengths, the main obstacle in obtaining clear satellite images of the Earth's surface are clouds (Elachi, 1987). Clouds can contaminate any analysis carried out on the images. A typical AVHRR image with cloud contamination is shown in Figure 1.

Typical "products" from satellite imagery are calculations of standard indices like NDVI (normalized difference vegetation index) or SST (sea surface temperature). These indices change over time and their temporal evolution is of fundamental interest in various scientific endeavors such as global climate change studies (Gutman et al., 1995). The indices are calculated from the intensity information contained in the images. These calculations can be very sensitive to the misrepresentation of surface reflectance caused by cloud-contaminated pixels (e.g., Gutman, Ignatov, and Olson, 1994). For example, calculations of sea-surface temperature (SST) can be considerably biased by pixels containing clouds (Henderson-Sellers, 1982; Stewart, 1985). Image archiving applications (e.g., for military and intelligence purposes and for monitoring of forest growth) also involve cloud screening in the sense that they are intolerant of cloudy images.

Hence, the classification of pixels in remotely-sensed images into "cloud" and "non-cloud" is an important practical problem, especially given the increasing volume with which these images are being collected. Once classified, unwanted pixels can be rejected from further consideration. Depending on the particular application, the whole image or part of it can be discarded, or it may suffice to quantify the "cloudiness" of the image (e.g., for

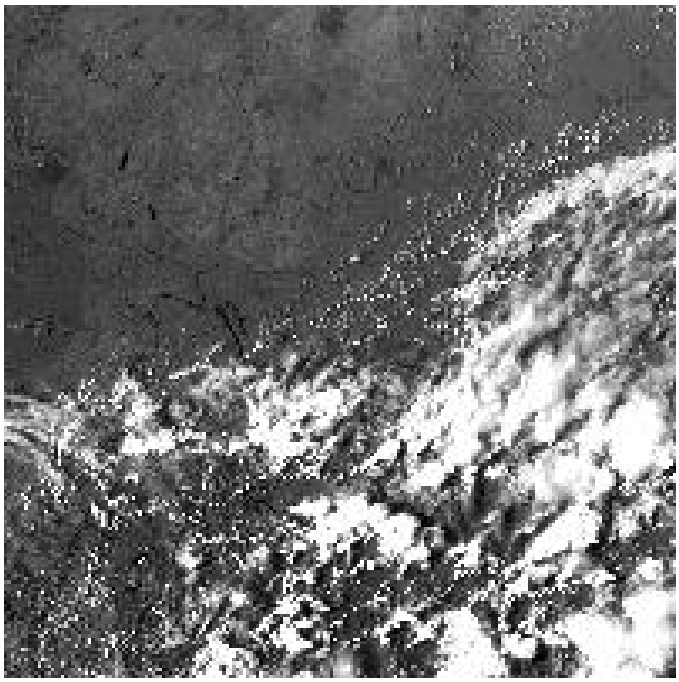

Figure 1: An example of an AVHRR image with cloud contamination.

characterization of optical Landsat imagery).

1.2 Current Cloud Screening Algorithms

Existing cloud screening algorithms are typically based on pixel-by-pixel classification into cloud and non-cloud classes. Early work on cloud screening was based on static thresholding of measured intensities using single (Bernstein, 1982) or multiple channels (McClain et al, 1983). The disadvantage of such thresholds is that they typically are reliable only in local regions under very specific circumstances. Simpson and Humphrey (1990) developed a more general approach based on dynamic thresholds which are computed using background knowledge of the physical radiation processes governing cloud appearance.

More recent work has focused on the application of both supervised and unsupervised learning algorithms to this problem. Yhann and Simpson (1995) used neural networks (combined with some post-processing) to classify pixels in a supervised fashion based on manually-labeled training images. Gallaudet and Simpson (1991) used unsupervised clustering algorithms to cluster pixels into groups which are then assigned cloud and non-cloud labels.

A disadvantage of these methods is that they do not model the spatial aspects of clouds, i.e., the classification decision is made locally at each pixel independent of other pixels. This ignores spatial context information which may be very useful for classifying ambiguous (noisy) pixels in particular. Coakley and Bretherton (1982) and Saunders and Kriebel (1988) use local windows to derive spatial indices for use in cloud screening, however, as pointed out in Gallaudet and Simpson (1990) their algorithms operate under significant practical restrictions.

In this work, we look at the statistical use of spatial information for cloud screening problems using an inhomogeneous MRF model. It imposes spatial continuity constraints

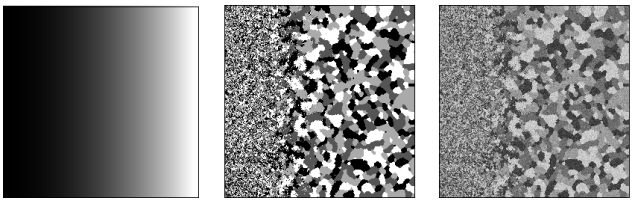

Figure 2: An example of a synthetic inhomogeneous MRF. From left to to right: β-map, label image, and intensity image.

on labeled regions but allows different scale patterns to exist within the same image.

2 Markov Image Models

2.1 MRF Priors on the Hidden Labels

The main object of interest is a rectangular $n \times m$ image S consisting of sites s_{ij} ordered in a matrix manner. The *neighborhood* of the site s_{ij} is any subset $\partial_{ij} \subset S$ such that $s_{ij} \notin \partial_{ij}$. The *neighborhood system* N is the set of all the neighborhoods: $N = \{\partial_{ij} | 1 \leq i \leq n, 1 \leq j \leq m\}$. At each site s_{ij} we define an intensity random variable X_{ij} (typically taking 256 gray-levels) and a hidden label random variable Y_{ij} (discrete-valued, k labels). The specific values the random variables take are denoted x_{ij} and y_{ij} respectively. This gives us two sets of variables defined on the image S: $X = \{X_{11}, \cdots, X_{nm}\}$, and $Y = \{Y_{11}, \cdots, Y_{nm}\}$.

MRFs have been widely used in image segmentation (see for example Besag (1986)) and have been proven equivalent to the Gibbs statistical model in physics by the Hammersley-Clifford (H-C) theorem (although they do not necessarily have to satisfy the rather strict symmetry constraints imposed on physical systems). From the H-C theorem, an MRF for $P(Y)$ is of the form

$$P(Y) = \frac{1}{z} e^{\beta V(Y)} = \frac{1}{z} e^{\beta \sum_{i,j} V_{ij}(\partial_{ij} \cup \{y_{ij}\})} \tag{1}$$

In this equation, z is the *partition function* or normalizing constant (yielding $\sum_Y P(Y) = 1$). β is the temperature parameter (frequently called the smoothness parameter in image segmentation). V is the so-called potential function and its extrema are tightly connected to the *optimal segmentation* within the MRF framework. The second term (on the right in Equation (1) above) shows how the potential function V can be factored into a sum of potentials defined at each site: $V = \sum_{ij} V_{ij}$.

2.2 Spatially Inhomogeneous MRF Priors

Traditionally the β parameter in an MRF model is a global constant across the whole image. This can be justified for small images, or images that are uniform in some way, but is less well motivated for many real-world images. For images that have high spatial variability in intensity, Pappas (1992) introduced the idea of inhomogeneous intensity parameters by slicing images into small windows and treating them as separate images. The argument is that the size of an image is a somewhat arbitrary concept. The variability of parameters within the image might be large, much like variability of parameters over different, separate images.

Here we allow for *scale variability* by allowing the parameter β to vary across the image. Large β values imply strong correlations and favors large, blob-like structures, smoothing out a lot of the noise. Small β values imply weak correlations and allows for 'broken' structures of 'small islands'. The variable β model is well suited for images that have underlying different scales. It prevents oversmoothing of small-scale regions, while still minimizing noise in large-scale regions. Clouds are a particularly good example of this. For example, notice the spatial scale variability in cloud regions (brighter pixels) in Figure 1. Figure 2 shows a synthetic image from an inhomogeneous model where β is increasing from left to right across the image. Ackroyd and Zimeras (1997) have also investigated the same type of model. However, their approach is somewhat different to that proposed here in that (1) they used Monte Carlo methods for estimation rather than the pseudo-likelihood approach described here, and (2) they investigated the model only on images for which ground truth is unknown (here we provide an objective evaluation of the method where truth is known).

More specifically, in this paper, $P(Y)$ is modeled using a second order neighborhood $\partial_{ij} = \{y_{kl} | i-1 \leq k \leq i+1, j-1 \leq l \leq j+1\}$. and a spatially varying β term, $\beta V_{ij}(\partial_{ij} \cup \{y_{ij}\}) = \beta_{ij}[\sum_{i'j' \in \partial_{ij}} \delta_{y_{ij}, y_{i'j'}}]$. This particular form of the potential function has one independent parameter β_{ij} per site. The potential is a count of neighbors with the same label, where δ is the Kronecker discrete delta function. Thus, β_{ij} is a per-site temperature parameter which defines an inhomogeneous MRF. We further use somewhat simpler notation $n_{ij}(y_{ij})$ for the number of pixels labeled y_{ij} in the neighborhood of the site s_{ij}. Thus, the potential function reduces to: $\beta V_{ij}(\partial_{ij} \cup \{y_{ij}\}) = \beta_{ij} n_{ij}(y_{ij})$.

In this paper we assume that observed intensities X_{ij} only depend on the local Y_{ij} labels and are conditionally Gaussian (normal) given the local label. We further set $\mu_{ij} = \mu_{y_{ij}}$, and $\sigma_{ij} = \sigma_{y_{ij}}$. This spatially homogeneous intensity model can be easily extended to the spatially inhomogeneous intensity model of Pappas (1992). Bayes' theorem yields a complete model coupling intensities and labels:

$$P(Y|X) \propto P(X|Y)P(Y) = \frac{1}{z} e^{\sum_{i,j} \beta_{ij} n_{ij}(y_{ij})} \prod_{ij} \frac{1}{\sqrt{2\pi\sigma^2_{y_{ij}}}} e^{-1/2\sigma^2_{y_{ij}}(x_{ij} - \mu_{y_{ij}})^2} \tag{2}$$

The goal of segmentation is to maximize $P(Y|X)$ with respect to labels Y. This corresponds to obtaining the maximum a-posteriori estimate of Y: $Y^* = \arg\max_Y \{P(Y|X)\}$. We use the method of Iterated Conditional Modes (ICM) by Besag (1986). ICM is a computationally efficient way to find a local maximum of $P(Y|X)$. Finding the global maximum is well-known to be intractable.

3 Likelihood-based Parameter Estimation

The intensity parameters (μ and σ for each class) can be easily estimated conditioned on a particular segmentation. This is an approximation to the full maximum likelihood estimate, (which would average over the hidden variable Y) which is intractable since the number of states of Y (possible labelings of the image) is k^{nm}. Thus, we use the tractable estimate where we simply condition on the most likely labeling Y^*; finding Y^* is discussed below.

The main difficulty in parameter estimation is the estimation of the MRF parameters β_{ij}. The estimation is intractable because β_{ij} appears both in the exponent of the probability density model and in the partition function z. As a first step, consider the estimation of the single parameter β as if it were global for the whole image. The direct approach is to

use the MLE:

$$\frac{\partial}{\partial\beta}P(Y|X,\beta) \propto P(X|Y)\frac{\partial}{\partial\beta}P(Y|\beta) = 0, \tag{3}$$

which reduces to:

$$\frac{\partial}{\partial\beta}\frac{1}{z(\beta)}e^{\beta V(Y,\beta)} = 0. \tag{4}$$

Solutions to this equation give the positions of likelihood minima and maxima. The optimal β is the one that corresponds to the largest maximum. The problem is that $z(\beta)$ represents the sum of k^{nm} terms in β (there is a term in the sum for every possible labeling Y).

Instead of the maximum likelihood estimator above, we use the Pseudo Likelihood Estimator (PLE). The PLE was first introduced by Besag as an approximation to the maximum likelihood estimator (Besag, 1986). Geman and Graffigne (1986) showed that the pseudo-likelihood (PL) is a tractable approximation to the full likelihood and that gradient ascent is feasible for a particular potential function they define for texture modeling. In this work we look at a different class of potential functions (namely, those in the model described earlier for segmentation into a set of discrete hidden classes) and we present a way to reduce the maximum-pseudolikelihood parameter estimation problem to that of finding the roots of a polynomial.

The PL is defined as:

$$PL(Y|\beta) = \prod_{ij}\frac{1}{z_{ij}}e^{\beta V_{ij}(\partial_{ij}\cup\{y_{ij}\})}, \tag{5}$$

and its log:

$$l(Y|\beta) = \log PL(Y|\beta) = \sum_{ij}\log\frac{1}{z_{ij}}e^{\beta n_{ij}(y_{ij})}. \tag{6}$$

In both equations:

$$z_{ij} = \sum_{y=1}^{k} e^{\beta n_{ij}(y)}. \tag{7}$$

z_{ij} is a normalizing constant defined locally at each pixel. Intuitively, the PLE treats each pixel as a single image. However, each pixel is considered as an image with a frame (neighborhood) and hence the PL does not completely decouple neighbors. Instead, it captures "first order" dependence effects.

For further analysis, it is useful to define a new variable $\xi = e^{\beta}$ and use it instead of β itself. Estimation of β then reduces to estimation of ξ. Maximizing the log-pseudolikelihood l becomes:

$$\begin{aligned}
\frac{\partial}{\partial\beta}l(Y|\beta) &= \frac{\partial}{\partial\beta}\sum_{ij}[\beta n_{ij}(y_{ij}) - \log z_{ij}] && (8)\\
&= \sum_{ij} n_{ij}(y_{ij}) - \sum_{ij}\frac{1}{z_{ij}}\frac{\partial z_{ij}}{\partial\beta} && (9)\\
&= c - \sum_{ij}\frac{\sum_y n_{ij}(y)\xi^{n_{ij}(y)}}{\sum_y \xi^{n_{ij}(y)}} && (10)\\
&= c - \sum_{ij}\frac{P_{ij}(\xi)}{Q_{ij}(\xi)}. && (11)
\end{aligned}$$

In the above derivation c is a constant, and P_{ij} and Q_{ij} are polynomials with integer coefficients. The degree of P_{ij} and Q_{ij} is not more than $\max\{n_{ij}(y)\}$, which is bounded by the neighborhood size n.

Furthermore, P_{ij} and Q_{ij} are completely defined by nearby labels n_{ij}. Coefficients n_{ij} are constrained by $\sum_{y=1}^{k} n_{ij}(y) = n$, so there is a combinatorial number of possible different P_{ij}'s and Q_{ij}'s. For example, with $n = 8$ and $k = 4$, there are 15 possible P's and Q's. This is because there are 15 unordered partitions of length 4 of number 8. Although the number of possible neighborhoods is much larger than 15, each neighborhood will be equivalent to one of the 15 representatives. Each representative will have a particular P and Q associated with it. The two neighborhoods are equivalent if corresponding sets $\{n_{ij}(y)|\forall y\}$ are the same.

In general, let there be $\eta = \eta(n,k)$ possible unordered partitions of n that are of length k and let l denote an index that is in the range $[1,\eta]$. Let the set $\{n_l(y)|1 \le y \le k\}$ describe the lth neighborhood representative and let γ_l be the number of the neighborhoods that are equivalent to $\{n_l(y)\}$. Furthermore, let P_l and Q_l be corresponding polynomials P_{ij} and Q_{ij} defined on the representative $\{n_l(y)\}$, and let $Q(\xi) = \prod_{l=1}^{\eta} Q_l(\xi)$ be the product over all the possible (non equivalent) polynomials Q_l. Then, the PL maximizing condition becomes:

$$c - \sum_{l=1}^{\eta} \gamma_l \frac{P_l(\xi)}{Q_l(\xi)} = 0 \tag{12}$$

$$cQ(\xi) - \sum_{l=1}^{\eta} \gamma_l P_l(\xi) \frac{Q(\xi)}{Q_l(\xi)} = 0 \tag{13}$$

This reduces the problem to the one of finding roots of a polynomial of degree not more than the degree of Q (P_l and Q_l are of the same degree, and Q is divisible by any Q_l). The degree of Q can be no more than ηn as it is the product of the η terms of degree not more than n. Neither the coefficients nor the degree of Q depend on the image being analyzed. The same holds for all the P_l's and Q_l's. It is therefore possible (in principle) to find the analytical solution to the above optimization problem. In practice, since the involved polynomials are of degree ~ 100, in the results reported here we use a Newton-Raphson scheme for numerical root finding over the range of β of practical interest ($0 \le \beta \le 3$).

3.1 Uniqueness of the PL estimate for β

The polynomial equation for estimating β can in general have many roots, yielding many different ξ's and consequently many different β's. However, we can prove concavity of the log-pseudolikelihood, guaranteeing at most a single real root and justifying the use of numerical root bracketing.

The second derivative of the log-pseudolikelihood is:

$$\begin{aligned}
\frac{\partial^2}{\partial\beta^2} l(Y|\beta) &= \frac{\partial}{\partial\beta}\Big[\sum_{ij} n_{ij}(y_{ij}) - \sum_{ij} \frac{1}{z_{ij}} \frac{\partial z_{ij}}{\partial\beta}\Big] \\
&= -\sum_{ij}\Big[-\frac{1}{z_{ij}^2}\Big(\frac{\partial z_{ij}}{\partial\beta}\Big)^2 + \frac{1}{z_{ij}} \frac{\partial^2 z_{ij}}{\partial\beta^2}\Big] \\
&= \frac{1}{z_{ij}^2} \sum_{ij}\Big[\Big(\frac{\partial z_{ij}}{\partial\beta}\Big)^2 - z_{ij} \frac{\partial^2 z_{ij}}{\partial\beta^2}\Big]
\end{aligned} \tag{14}$$

The factor in front of the sum is always positive, so it remains to analyze each term in the sum:

$$
\begin{aligned}
&(\frac{\partial z_{ij}}{\partial \beta})^2 - z_{ij}\frac{\partial^2 z_{ij}}{\partial \beta^2} = \\
&= \sum_{y_1,y_2} \xi^{[n_{ij}(y_1)+n_{ij}(y_2)]} n_{ij}(y_1) n_{ij}(y_2) - \\
&\quad - \sum_{y_1,y_2} \xi^{[n_{ij}(y_1)+n_{ij}(y_2)]} n_{ij}^2(y_1) \\
&= \sum_{y_1,y_2} \xi^{[n_{ij}(y_1)+n_{ij}(y_2)]} n_{ij}(y_1)[n_{ij}(y_2) - n_{ij}(y_1)] \\
&= \sum_{a,b} \xi^{a+b} a[b-a]. \qquad (15)
\end{aligned}
$$

The last equation is just a simplified notation where a and b are specific to site (i,j). They are both integers and they take values that add up to n, where n is the neighborhood size thus, the last term can be analyzed as follows. Define

$$x = \sum_{a,b} \xi^{a+b} a[b-a].$$

By dummy exchange of variables a and b one gets the equivalent expression (note that a and b take values from the same set):

$$x = \sum_{b,a} \xi^{b+a} b[a-b] = \sum_{a,b} \xi^{a+b} b[a-b].$$

The two expressions differ only in order in which terms are summed. Upon adding them, we get:

$$
\begin{aligned}
2x &= \sum_{a,b} \xi^{a+b}(a[b-a] + b[a-b]) \\
&= -\sum_{a,b} \xi^{a+b}(a-b)^2 \leq 0 \qquad (16)
\end{aligned}
$$

The last inequality follows from the fact that $\xi > 0$ by definition. It also establishes that each term in the second derivative of the log-pseudolikelihood is negative, which in turn establishes the concavity of the log-pseudolikelihood. Concavity of the log-pseudolikelihood further establishes that there can be at most a single real solution for β, and that it represents a maximum rather than a minimum of the log-pseudolikelihood.

3.2 Local PL estimation of β

To extend the above methodology to the problem of finding *local* rather than *global* β's, we estimate β locally on small windows and then perform bilinear interpolation to get the intermediate values. For the results in this paper, the windows are non-overlapping and set to be $1/8 \times 1/8$ of the size of the original image. This choice represents a practical trade-off between having the window size be small enough to faithfully track local variability in spatial scale (i.e., reducing the bias of the model), while at the same time not making the window size so small that unreliable parameter estimates result from too few pixels (i.e.,

1. Given the observed intensities x, perform a non-spatial k-means segmentation to obtain the initial classification y_0.

2. Begin loop:

   (a) Set the *current* classification $y = y_0$.

   (b) Given the current classification y and the original intensities x, estimate the per class intensity parameters $(\hat{\mu}_c, \hat{\sigma}_c)$.

   (c) Given the current classification y, estimate the MRF parameters $\hat{\beta}$ on small windows.

   (d) Interpolate point estimates of $\hat{\beta}$ to obtain the full set of $\hat{\beta}_{ij}$.

   (e) Perform *one* ICM iteration. Evaluation of the potential function requires the current (estimated) $(\hat{\mu}_c, \hat{\sigma}_c, \hat{\beta}_{ij})$, as well as the current classification y and the original intensities x. Keep the updated classification in y_0.

   (f) While y and y_0 do not satisfy the convergence condition, go to step 2.

Figure 3: Pseudocode description of the estimation algorithm

increasing the variance). Alternatives to a fixed window size would be to (a) automatically estimate the window size from the pixel data, and/or (b) allow the windows to be of different sizes in different parts of the image. However, since β is a second order parameter in the MRF model, it is not clear if there is much to be gained by the extra effort of trying to determine the optimal window size.

The proposed algorithm uses the k-means algorithm for initial segmentation and then alternates between a parameter estimation step for a given segmentation, and an Iterative Conditional Mode (ICM) step (Besag, 1986) with fixed parameters. Relative to a standard EM approach, the ICM step is a deterministic approximation to the E-step, and the maximum PL estimation (MPLE) step is a local approximation to the M-step. Since both ICM and MPLE are guaranteed to locally increase the pseudolikelihood (the former by definition, the latter by the procedure described in this paper), the algorithm is guaranteed to converge to a local maximum of the PL. The time complexity of the algorithm scales linearly in the number of pixels in the image and the local nature of both the model and the estimation algorithms makes the method ideally suited to parallelization.

Figure 3 contains a pseudocode description of the complete algorithm for unsupervised MRF segmentation of an image.

3.3 Computational Complexity and Parallelization

As suggested in the pseudocode, the estimation procedure is iterative and the exact number of required iterations cannot be predicted. In the work described here, if the algorithm does not converge in the maximum allowed number of iterations, the last classification is returned as the result. The number 20 was used as the default value for the maximum number of iterations in our experiments. In the results below, the image size is $n \times m$, the neighborhood size is d, and there are k classes.

Estimation of the intensity parameters is linear in the number of pixels nm, as each pixel has to be visited exactly twice: once for calculating μ, and once for calculating σ.

To estimate the MRF parameters β, one needs to determine the *type* of the neighborhood of each pixel. This is done by counting differently classified pixels in the neighborhood, which takes time linear in the neighborhood size d (per pixel). Thus, for non-overlapping estimation windows, the total time required is $O(nmd)$. For overlapping windows, the time required is proportional to the sum of the window areas. There is also the slight overhead associated with solving the PLE equation, but it is a constant factor relative to the size of the image. The bilinear interpolation is linear in the number of pixels nm.

Each ICM iteration locally updates the classification labels and only requires looking at the neighborhood once. Hence, the time complexity is the same as for the estimation of the MRF parameters, $O(nmd)$.

The preceding analysis shows that the overall time complexity of a single iteration is linear in the number of pixels and that the inclusion of the parameter estimation step does not change the asymptotic time complexity. Furthermore, the choice of window size for the estimation of β does not effect the time complexity, as long as the windows are non-overlapping.

The locality of the MRF model and the local nature of the parameter estimation method makes the proposed algorithm ideally suited for parallelization. The natural way to break the problem into smaller problems is to assign each window to a single processor. Within each window, the problem reduces to a simpler, global MRF segmentation problem.

4 Experimental Results

We performed two types of experiments. One on images generated from synthetic MRF models, and the other on AVHRR images to investigate the application of the spatially-varying MRF model to the problem of cloud-screening. The main difference between the two sets of experiments is that for synthetic images ground truth is known, while for the real-life AVHRR images ground truth is not known. A general problem is that there are no well-calibrated ground truth annotations for evaluation of cloud-screening algorithms (J. Simpson, personal communication, 1997). The first set of experiments focuses on the performance of parameter estimation and the quantification of error rates. The second set of experiments focuses on the visual appeal of the segmentations and the adaptability of inhomogeneous model to scale variations within the images.

4.1 General Experimental Methodology

Images were generated as a realization of an inhomogeneous MRF with known parameters using a 1-dimensional (spatial) variation in the parameter β. The synthesized images consist of four labels and represent the ground truth for the experiment. The corresponding intensity images are obtained by choosing per-class means and adding Gaussian noise with per-class standard deviation. Depending on the separation of the means relative to the standard deviation, we generated intensity images with different noise levels. Although there are two parameters per class to control the noise (quality) of the image, there is actually only one relevant parameter, the ratio of separation between the means to the standard deviation.

Several different segmentation algorithms of interest were used to segment the images. Their performance was measured by comparison of the final segmentation to the true label

image. We used misclassification rate to quantify and compare the performance of the algorithms.

Two different types of experiments were performed on the synthetic images. The first is on images with variable β, the second is a control experiment on an image where β is constant. Thus, the visual difference between the images from the two experiments is in the spatial scale of label dependence.

All experiments included running a non-spatial k-means algorithm and six different versions of MRF segmentation. The MRF segmentation algorithm of most interest (and the novel contribution of this work) is the one that is completely unsupervised and inhomogeneous (as described in section 2). The five other segmentations used were true β, fixed β (but chosen in an ad hoc manner), estimated global β (homogeneous model), maximum value of estimated β_{ij} and minimum value of estimated β_{ij}.

4.2 Experiments on Synthetic Inhomogeneous MRFs

The first experiment tests the performance of the parameter estimation algorithm for a variable-scale synthetic image. The image and the β-map are shown in the Figure 2. β was chosen to lie in the interval from 0.3 to 2.3, such that all points with the same horizontal pixel coordinate have the same value of β (1-dimensional variation in β).

The 1-dim variation in β was chosen to make it easy to plot and visualize the results. We were interested in performance on "tall narrow windows." These windows, consisting of pixels with the same horizontal pixel coordinate, were realizations of a single value of β. That way, the misclassification rate becomes a function of the single coordinate. However, instead of directly defining the "small width" of the windows and measuring the performance within each one of them, it was more convenient to look at the integral misclassification rate. In other words, we measured misclassification rate on windows that started at the left edge of the image and ended at the desired point x. The function obtained this way was not misclassification rate itself, but rather its integral. The "goodness" of the algorithm is represented by the slope of the resulting plot, rather than the value itself.

The main results for the inhomogeneous image from Figure 2 are contained in Figure 4. Several observations are apparent from the graph.

- At the left side of the image, spatial models do not perform that well (large slope), while at the right side they improve. The spatial information becomes increasingly relevant as the underlying patterns increase in size.

- The boundary value parameters (very large and very small β) are optimal within their respective regions (optimal slope), but their overall performance is generally poor. The measure of overall performance is the misclassification rate at the rightmost end of the graph.

- The ad-hoc value of β seems to be a reasonable tradeoff. Its performance is never optimal, but it still achieves better results than extremal values of β.

- The single estimated value of β is also never locally optimal, but it optimizes the overall accuracy over different global models. In principle, the fine-tuning of the ad-hoc β could produce results close to the estimated global β model. However, there is not much room for tuning if the ground truth is not known. Therefore, the estimated global β model on average performs much better than the ad-hoc model, and it never performs worse.

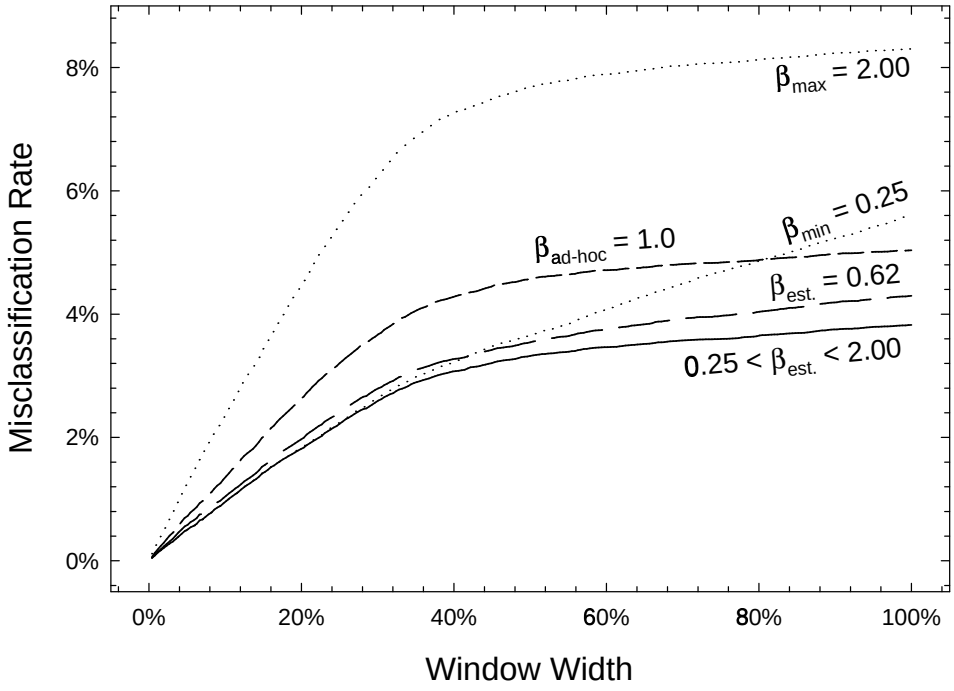

Figure 4: Performance of different MRF segmentation algorithms on a synthesized image from an inhomogeneous MRF model with increasing β from left to right across the image.

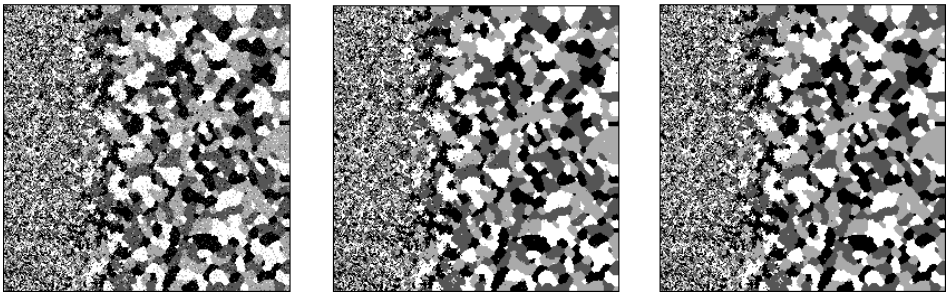

Figure 5: Segmentations returned by k-means, global β and variable β in an enlarged portion of the image.

- The variable (locally estimated) β model works optimally over the whole image. Its performance curve follows the best slope at all the times. It starts as the minimum β model at the beginning, but is able to adapt to the large scale and behaves as the maximum β model at the end.

- The true β model produced the same curve as the locally estimated β model and is therefore omitted from the graph.

The non-spatial k-means algorithm performed as expected with about a 10% error rate. Its performance is not shown in Figure 4.

It was also of interest to observe the "visual appeal" of the segmentation results. We were interested in comparison of homogeneous versus inhomogeneous models. For comparison purposes, the output of the non-spatial k-means segmentation is also included. The quality of the visual image from a human perceptual viewpoint becomes relevant when there is no objective way of quantifying the segmentation (most real-life images).

Figure 5 displays the MAP segmentation for the intensity image of Figure 2. The k-means segmentation looks very noisy, while the two MRF segmentations are able to use spatial information to their benefit. More details can be seen in the enlarged image of Figure 6. The portion of the image that is enlarged lies towards the right side of the original image (at the largest spatial scale). The noise in the non-spatial model is now obvious. Figure 6 also shows that the global β model cannot adapt to locally large scale blobs and that there is still some undersmoothing represented by small dots in the large blobs. The variable β model adapted to the large scale and produced results visually almost identical to the true image. It is worth emphasizing that Figure 4 does not capture fully the perceptual difference in the segmentations. In Figure 4, both MRF models seem to be close in the overall performance, whereas in Figure 6 there is a clear difference from a human perceptual viewpoint. This suggests that other objective functions could be used instead of direct misclassification rate.

To further check the quality of the MRF parameter estimation algorithm, we also compared the β map of the true image and the estimated β map. The two images are close. The estimated β map has slight variability (as one can expect given that it is completely data-driven). Some directional effects from interpolation can be seen. As β is a second order parameter in MRF models, we did not expect to gain much by fine-tuning the interpolation process. This is a possible direction for the future research.

It is also of interest to examine the performance of different models as a function of the noise in the image as measured by $\Delta\mu/\sigma$. Figure 8 shows the misclassification rate for

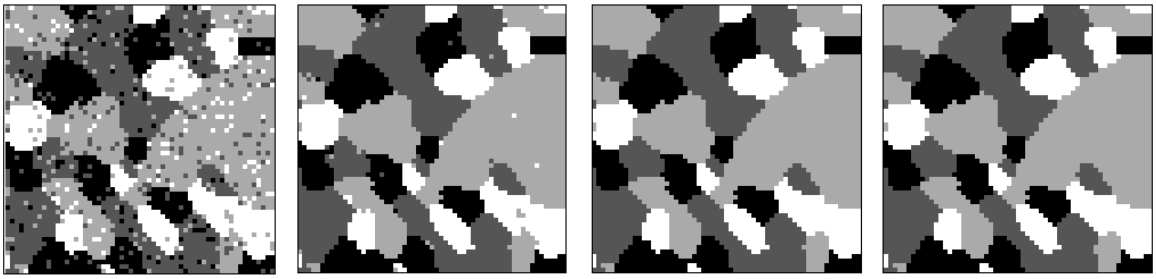

Figure 6: k-means, global β, variable β and Ground Truth on Enlarged Image Portion

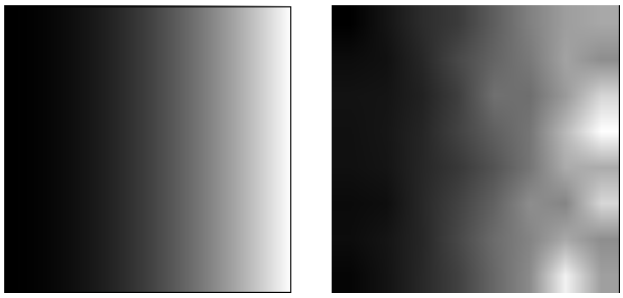

Figure 7: True and locally estimated β maps.

several models as a function of the noise, where the underlying label model is the same inhomogeneous MRF as used earlier in this section.

Non-spatial k-means was included as a reference. Other than showing the usual relative performance of MRFs versus k-means, they also give us a rough idea about what performance to expect based on measurable image properties, namely, that the spatial models have a greater advantage as the intensity noise decreases, and that the inhomogeneous MRF shows a consistent (about 10 to 15%) improvement in classification accuracy over the homogeneous MRF. It is very important to keep in mind that the results above are obtained on the images that *are* realizations of an MRF. It would be incorrect to try to generalize these results to real-life images. Nonetheless, the above results provide an indication of the likely relative performance of the different algorithms for different qualities of images.

4.3 Experiments on Synthetic Homogeneous MRF's

The second experiment tested the performance of the inhomogeneous model on an image that was generated from an homogeneous MRF. The single β value was chosen to be 0.8. An example of such an image is shown in figure 9. This experiment tests the inhomogeneous model for possible overfitting.

Figure 10 shows the resulting segmentation for the models of interest. As before, the non-spatial k-means model performed distinguishably worse than the two MRF models. The enlarged view on figure 11 gives more information.

Both the homogeneous and inhomogeneous MRF models perform almost exactly the same and are very close to the ground truth. The variable β model adapted locally to the underlying scale without overfitting. The generated β map for the locally estimated β is given in Figure 12.

The map has a mean value of $\beta = 0.92$ and a standard deviation of 0.05; it is quite

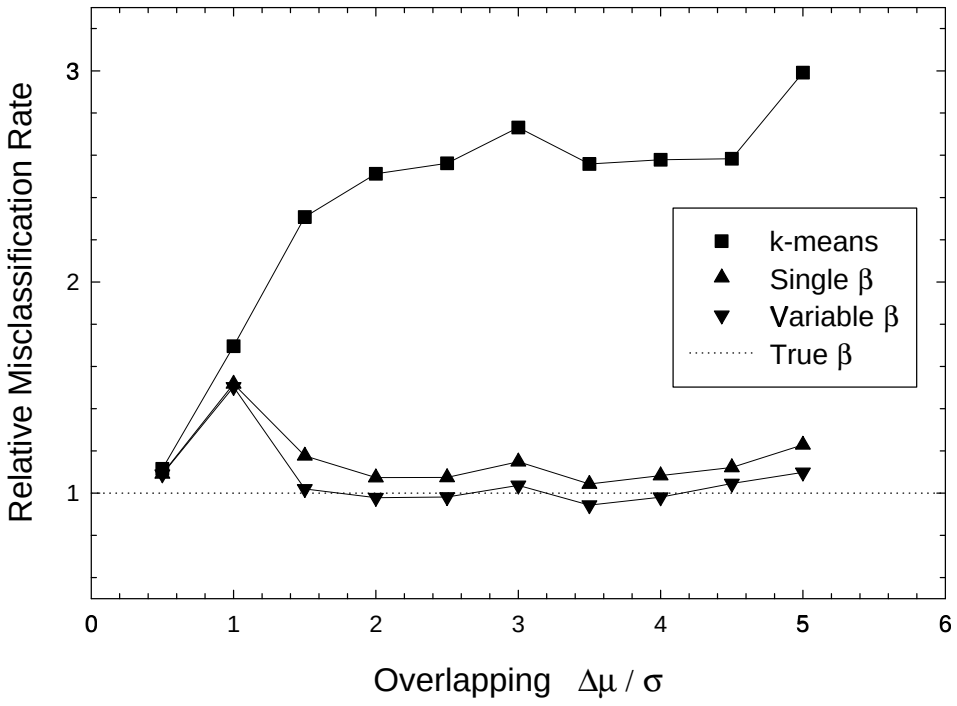

Figure 8: The relative performance of different segmentation algorithms on images of varying noise levels.

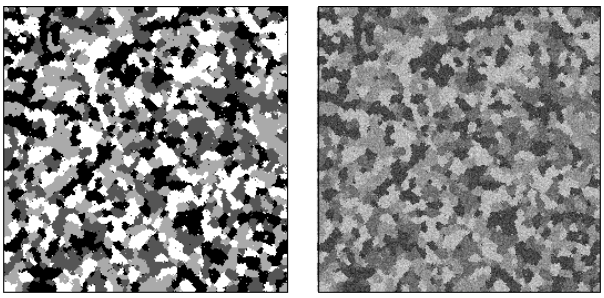

Figure 9: A synthesized label image from a homogeneous MRF and a corresponding intensity image.

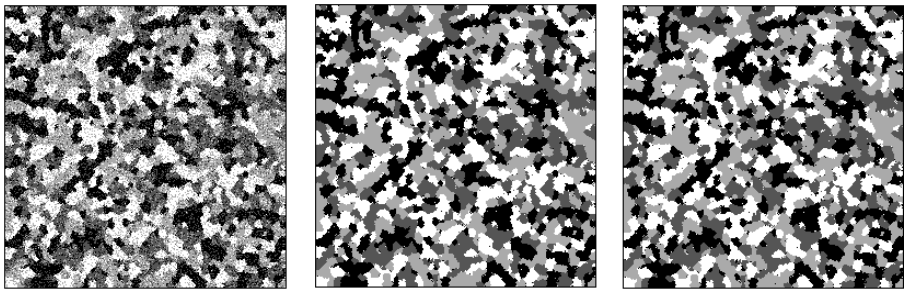

Figure 10: From left to right: k-means, global β and variable β segmentations.

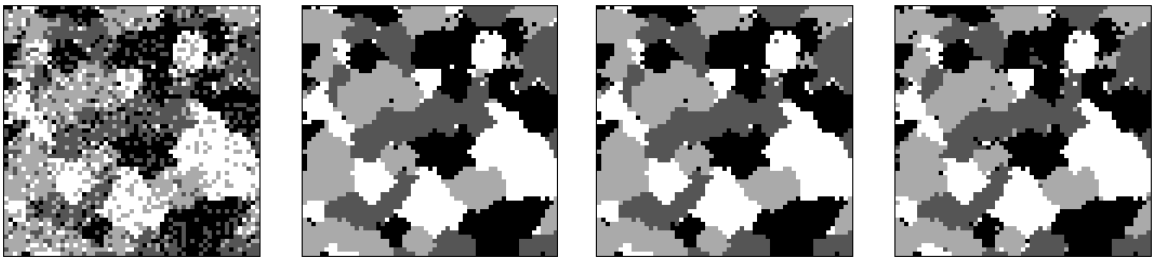

Figure 11: k-means, global β, variable β and Ground Truth on Enlarged Image Portion

Figure 12: Locally estimated β map for the homogeneous image

	Image 1 (variable β)	Image 2 (fixed β)
k-means	10.0	24.0
max β	8.0	5.0
min β	6.0	4.4
ad-hoc β	5.0	4.4
single β	4.3	4.3
true β	3.8	4.3
variable β	3.8	4.3

Table 1: Summary of percentage misclassification rates on both images.

uniform. The minimum value across the image is $\beta = 0.79$. Despite the slight discrepancy from the true value, the variable β segmentation performs the same as the global model.

Table 1 summarizes the overall results of both experiments in terms of misclassification rate for the two synthesized images.

4.4 Experiments on AVHRR Images with Clouds

AVHRR is a sensing device onboard many satellites orbiting the Earth. The primary physical characteristics of the AVHRR instrument are the recording of images in four or five spectral bands, a roughly 1.1km ground resolution, a roughly 2700km swath width, and 6 scans per second. AVHRR images are used for calculating standard indices like SST (Berstein, 1982) or NDVI (Gutman et al., 1995).

The images we use here are preprocessed versions which are composites (averages) from several spectral bands. The underlying scenes were cloudy, but there was no ground truth available. This means that the quality of any particular segmentation could not be obtained in the objective way it was done for the synthesized images. In the following figures we compare several classifications and look at various enlarged details. The observed pixel classifications agree with our intuition about large scale vs. small scale β parameters.

Figure 13 represents a scene that has clouds ranging from thick cover to small broken clouds. Both MRF models appear to perform much better than simple k-means. On the enlarged view in figure 14, some differences become more visible. The enlarged portion of the image is the image of the ground. It looks uniform on the raw image and the inhomogeneous model was able to reflect this fact.

Figure 15 shows a second AVHRR image which is mostly cloudy. There is not much scale variability; both MRF models perform very similarly and they both appear to perform much better than k-means method.

5 Conclusion

In this paper we developed an efficient algorithm for parameter estimation in inhomogeneous MRF models. The algorithm was based on the concept of a pseudo-likelihood that approximates the regular likelihood. We presented an inhomogeneous MRF model and applied the parameter estimation algorithm to it. Results on the test images favored the inhomogeneous model over the homogeneous model whenever there was a significant scale variation across the image. On the images with no scale variations, the inhomogeneous model performed as well as the homogeneous model and both models performed better

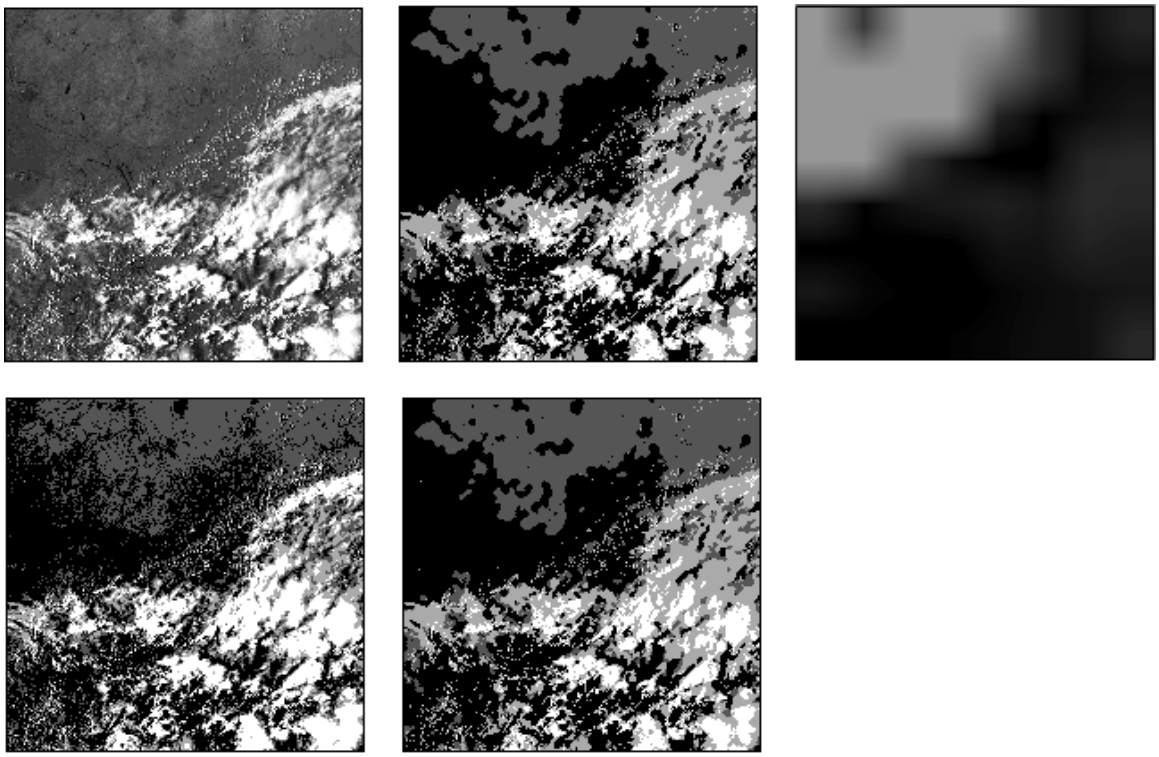

Figure 13: First row: First AVHRR raw image, variable β segmentation and corresponding β map. Second row: k-means and single β segmentation.

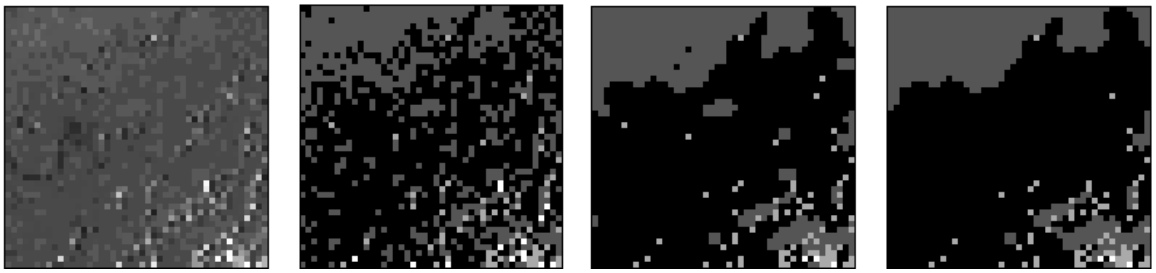

Figure 14: Details of segmentation: raw image, k-means, single β and variable β models.

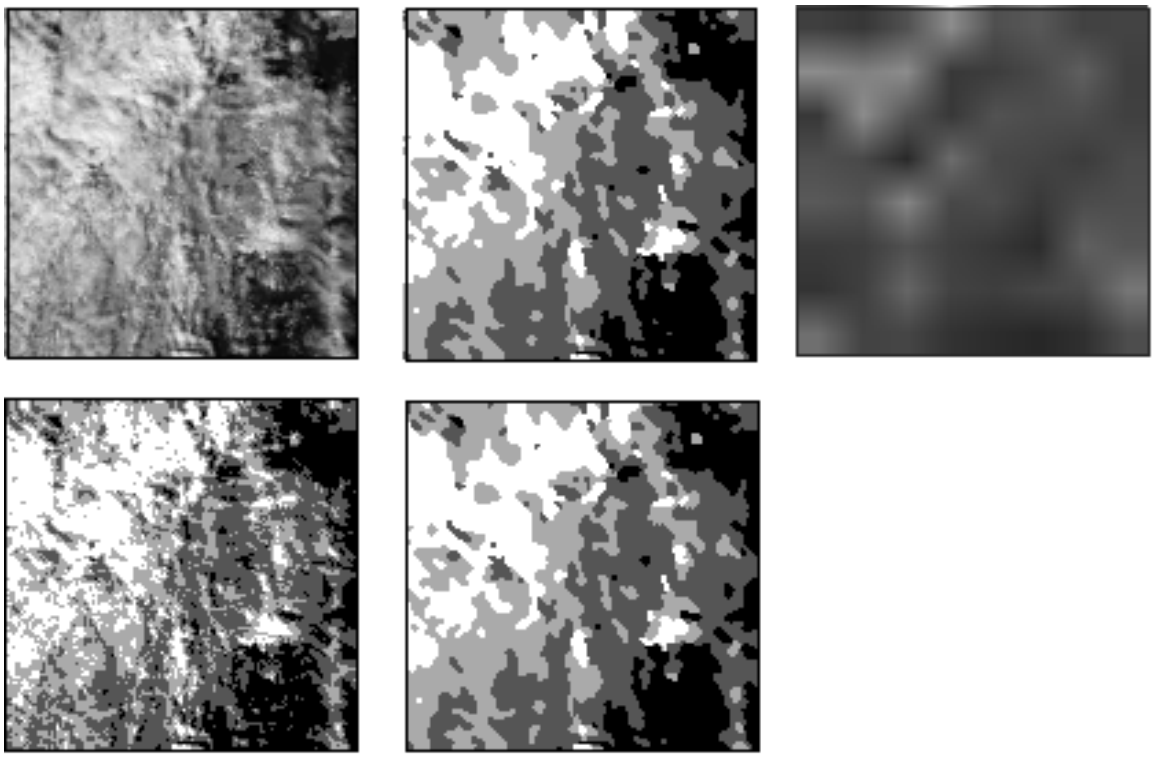

Figure 15: First row: Second AVHRR raw image, variable β segmentation and corresponding β map. Second row: k-means and single β segmentation.

than the ad-hoc parameter model. On the AVHRR images, the segmentation obtained with the inhomogeneous MRF model was visually appealing and followed our intuition about scale variability. The experimental evidence suggests that inhomogeneous MRF models are able to adapt to local scale in images with large scale variations while still performing as well as homogeneous models on images with no scale variability.

6 Acknowledgements

This work was supported in part by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration and in part by the National Science Foundation under NSF Grant IRI-9703120.

7 References

Aykroyd, R. G., and S. Zimeras, 'Automatic reconstruction with inhomogeneous nodels,' *Proceedings of The Art And Science of Bayesian Image Analysis*, Leeds, UK, pp. 161-168, 1997.

Berstein, R. L., 'Sea-surface temperature estimation using the NOAA-6 satellite advanced high resolution radiometer,' *J. Geophys.*, 87, 9445–9465, 1982.

Besag, J., 'On the statistical analysis of dirty pictures,' *J. R. Statist. Soc. B*, 48(3), pp.259-302, 1986.

Coakley, J. A. and Bretherton, F. P., 'Cloud cover from high resolution scanner data: detecting and allowing for partially filled fields of view,' *J. Geophys. Res.*, 87, 4917–4932.

Elachi, C., *An Introduction to the Physics of Remote Sensing*, New York: John Wiley and Sons, 1987.

Gallaudet, T. C. and Simpson, J. J., 'Automated cloud screening of AVHRR imagery using split-and-merge clustering,' *Remote Sens. Environ.*, 38, 77–121, 1991.

Geman, S. and C. Graffigne, 'Markov random field image models and their applications to computer vision,' *Proceedings of the International Congress of Mathematicians*, pp. 1496-1517, Berkeley, 1986.

Gray, A. J., J. W. Kay, and D. M. Titterington, 'An empirical study of the simulation of various models used for images,' *IEEE Trans. Pattern Anal. Machine Intell.*, 16(5), pp. 507-512, May 1994.

Gutman, G., Ignatov, D. and Olson, S., 'Towards better quality of AVHRR composite images over land: reduction of cloud contamination,' *Rem. Sens. Environ*, 50, 134–148, 1994.

Gutman, G., Tarpley, D., Ignatov, D. and Olson, S., 'The enhanced NOAA global land dataset from the Advanced Very High Resolution Radiometer,' *Bull. Amer. Met. Soc.*, 76(7), 1141–1156, 1995.

Henderson-Sellers, A., 'Defogging cloud determination algorithms,' *Nature*, 298, 419–420, 1982.

McClain, E. P., Pichel, W. P., Walton, C. C., Amhad, Z., and Sutton. J., 'Multi-channel improvements to satellite-derived global sea-surface temperatures,' *Adv. Space Res.*, 2, 43–47, 1983.

Pappas, T. N., 'An adaptive algorithm for image segmentation,' *IEEE Trans. Signal Process.*, 40(4), pp. 901-913, April 1992.

Saunders, R. W. and Kriebel, K. T., 'An improved algorithm for detecting clear sky and cloudy radiances from AVHRR data,' *Int. J. Remote Sens.*, 9, 123–150.

Simpson, J. J. and C. Humphrey, 'An automated cloud screening algorithm for daytime Advanced Very High Resolution Radiometer imagery,' *Journal of Geophysical Research*, 95(8), pp. 13459-13481, August 1990.

Stewart, R. H., *Methods of Satellite Oceanography*, Berkeley, CA: University of California Press, 1985.

Yhann, S. R. and J. J. Simpson, 'Application of neural networks to AVHRR cloud segmentation,' *IEEE Trans. Geo. Rem. Sens.*, 33(3), 590–604, 1995.