Change detection in remote sensing images with graph cuts

Fernando Pérez Nava[a], Alejandro Pérez Nava[a], José M. Gálvez Lamolda[b],
Marcos Fuentes Redondo[c]

[a] Dept. de Estadística, I. O. y Computación. Universidad de La Laguna. Islas Canarias. 38271. Spain
[b] Dept. de Física Fundamental. Universidad de La Laguna. Islas Canarias. 38271. Spain
[c] Esc. Téc. Sup. de Ingeniería Informática. Universidad de La Laguna. Islas Canarias. 38271. Spain

ABSTRACT

Change detection is an important part of many remote sensing applications. This paper addresses the problem of unsupervised pixel classification into 'Change' and 'No Change' classes based on Hidden Markov Random Field (HMRF) models. HMRF models have long been recognized as a method to enforce spatially coherent class assignment. The optimal classification under these models is usually obtained under the Maximum a Posteriori (MAP) criterion. However, the MAP classification in HMRF models leads in general to problems with exponential complexity, so approximate techniques are needed. In this paper we show that the simple structure of the change detection problem makes that MAP classification can be exactly and efficiently calculated using graph cut techniques. Another problem related to HMRF modelling (and to any change detection technique) is the determination of the parameters or thresholds for classification. This learning problem is solved in our HMRF model by the Expectation Maximization (EM) algorithm. Experimental results obtained on four sets of multispectral remote sensing images confirm the validity of the proposed approach.

Keywords: Change detection, hidden Markov random models, expectation maximization, graph cuts, multitemporal images, remote sensing.

1. INTRODUCTION

Automatic change detection from images of the same scene acquired at different times is a problem of widespread interest due to a large number of applications in diverse disciplines [1], [2]. Important applications of change detection include video surveillance, medical diagnosis treatment and remote sensing [3].

Usually change detection in remote sensing involves the analysis of two registered, aerial or satellite multispectral images from the same geographical area obtained at two different times. Such an analysis aims at identifying changes that have occurred in the same geographical area between the two times considered.

Two main approaches have been proposed to solve the change detection problem [4]: the supervised approach and the unsupervised approach. The former is based on supervised classification methods, which require a learning set with multitemporal ground truth while the latter perform change detection without relying on any additional information. As the generation of a learning set is usually a difficult and expensive task, the use of unsupervised methods is of great interest in many applications in which a learning set is not available.

A variety of approaches have been proposed for unsupervised classification. Statistical approaches, both parametric and non-parametric have been employed. This type of methods labels pixels according to probability values, which are determined based on some distribution on the features for the pixels in the image. The simplest distributions are based on finite mixture models (FM), both parametric and non-parametric [5]. However, the FM model has an intrinsic limitation: spatial information is not taken into account because all the data points are considered to be independent samples drawn from a population. An approach to obtain a spatially coherent clustering is to use a hidden Markov random field (HMRF), which is an stochastic process generated by an Markov random field (MRF) [6], [7], and whose class labels cannot be observed directly but which can be observed through a field of observations.

Statistical approaches attempt to solve the problem of estimating the associated class label given only the values for each pixel. The most popular technique to estimate an HMRF is maximum a posteriori (MAP) estimation [6]. The MAP framework was popularised in the field of image analysis by Geman and Geman [8]. MAP estimation consists of maximizing the posterior probability of the labels given the pixel intensities. This is an optimisation problem that can be solved by a variety of techniques. The main problem of this approach is its high computational cost. For many

interesting problems finding the exact solution is NP-hard and so there is no hope to find a tractable method to find an exact maximum.

Recently, Min-Cut algorithms on graphs [9] have emerged as an increasingly useful tool for exact or approximate optimization in low-level vision. The basic technique is to construct a specialized graph for the function to be optimized, such that the minimum cut on the graph also provide the optimum of the function (either globally or locally). The minimum cut in turn can be computed very efficiently by max-flow algorithms. These methods have been successfully used for a wide variety of image analysis including image restoration [10] or stereo and motion [11]. The minimum cut solution comes with some interesting theoretical quality guarantee. In some cases it is the global minimum, in other cases a local minimum in a strong sense that is within a known factor of the global minimum [12].

2. AN HMRF FORMULATION OF THE CHANGE DETECTION PROBLEM

Problems where part of the data is missing or unobservable are common in image analysis. The observations may represent measurements in the form of multidimensional variables for each pixel in the image while the hidden data could consist of an unknown label assignment to be estimated from the observations for each pixel. Depending on the particular problem, labels may represent classes (segmentation problem), disparities (stereo problem) or displacement (flow problem). In this paper we focus on change detection as a bitemporal segmentation problem. However, unlike a simple segmentation problem, unknown lightning transforms between images may be present. In Section 2.1 we present the basic definitions concerning the Markov models for the unobservable data. In Section 2.2 we specify the complete parametric models for the observed and unobserved data. In Section 2.3 we restrict this general formulation for the change detection problem and in Section 2.4 we present the MAP solution to the change detection problem.

2.1. Markov Random Fields

A Markov Random Field (MRF) is composed of three sets. A finite set of sites S. A neighborhood system N defined as $N = \{ N_i \ i \in S \}$, where each N_i is a subset of S which form the neighborhood of site i. In this paper we will use a first order neighborhood system: for each site, the neighbors are the four sites surrounding it. The third set is a collection of random variables (also called the field) $X = \{ x_i, \ i \in S \}$. In order for X to be a MRF the joint probability distribution must satisfy:

$$P(x_i \mid \boldsymbol{x}_{S-\{i\}}) = P(x_i \mid \boldsymbol{x}_{N_i}) \tag{1}$$
$$\forall \boldsymbol{x} \ P(\boldsymbol{x}) > 0 \tag{2}$$

where $\boldsymbol{x}_M$ denotes a realization of the field restricted to M. Property (1) means that the interactions between site i and the other sites actually reduce to interactions with its neighbors. Property (2) is important for the Hammersley-Clifford theorem to hold. This theorem states that the joint probability distribution of a Markov field is a Gibbs distribution given by:

$$P(\boldsymbol{x}) = Z^{-1} \exp(-H(\boldsymbol{x})) \tag{3}$$

where H is the energy function:

$$H(\boldsymbol{x}) = \sum_C V_c(\boldsymbol{x}_c) \tag{4}$$

where C comprises the cliques (sets where all sites are neighbors) in N. The V_C are the clique potentials and typically depend on parameters. Finally,

$$Z = \sum_X \exp(-H(\boldsymbol{x})) \tag{5}$$

is the normalizing factor also called the partition function. The calculation of Z involves all possible realizations of the Markov field. Therefore its exact computation is exponentially complex. This is a problem when using these models in situations where an expression of the joint distribution is required.

2.2. Hidden Markov models

Problems involving incomplete data, where part of the data is hidden or unobservable, are common in image analysis. Observations may represent measurements for each pixel of an image while the hidden data could consist of an unknown class assignment to be estimated from the observations for each pixel. In this paper, we focus on this case, usually referred to as image segmentation. The unobserved data is modeled as a discrete Markov random field X, as defined in (3) with energy function H depending on a parameter $\boldsymbol{b}$. In hidden Markov models, the observations Y are conditionally independent given X according to a likelihood $L(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{Q})$, where $\boldsymbol{Q}$ are the parameters that fully describe the likelihood. We will asume the likelihood to be:

$$L(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{Q}) = \prod_{i \in S} L_i(y_i \mid x_i, \boldsymbol{Q}) = \exp\left(\sum_{i \in S} \log(L_i(y_i \mid x_i, \boldsymbol{Q}))\right) = \exp\left(\sum_{i \in S} l_i(y_i \mid x_i, \boldsymbol{Q})\right) = \exp(-\sum_{i \in S} D_i(x_i)), \qquad (6)$$

$$D_i(x_i) = -l_i(y_i \mid x_i, \boldsymbol{Q})$$

assuming that all the l_i are positive. This makes the model similar to a finite mixture (FM) model. A FM model could be seen as a hidden Markov model where the hidden field X is one of independent variables. This is a particular case that makes FM models more tractable.

In the general case the complete likelihood is given by:

$$P(\boldsymbol{y}, \boldsymbol{x} \mid \boldsymbol{Q}, \boldsymbol{b}) = L(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{Q}) P(\boldsymbol{x} \mid \boldsymbol{b}) = Z^{-1} \prod_{i \in S} L_i(y_i \mid x_i, \boldsymbol{Q}) \prod_{c \in C} \exp(-V_c(\boldsymbol{x}_c)) = Z^{-1} \exp\left(-(H(\boldsymbol{x}) + \sum_{i \in S} D_i(x_i))\right) \qquad (7)$$

Note that both the normalization constant Z, the cluster potential V_c and the energy H depend on $\boldsymbol{b}$.

The conditional field X given $Y = y$ is a Markov field with energy function:

$$H(\boldsymbol{x} \mid \boldsymbol{y}) = \sum_{c \in C} V_c(\boldsymbol{x}_c) + \sum_{i \in S} D_i(x_i) \qquad (8)$$

In the image segmentation problem, we have to recover the unknown class labeling X given the observations Y. This classification problem usually requires values for the parameters $\boldsymbol{Y} = (\boldsymbol{Q}, \boldsymbol{b})$. If they are unknown they must be estimated.

2.3. HMRF formulation of the change detection problem

In this section we formulate the change detection problem as an image segmentation problem. In this case, observations $\boldsymbol{y}$ come from two multispectral images $\boldsymbol{u}$ and $\boldsymbol{v}$ acquired in the same geographical areas at two different times t_1 and t_2. A common assumption is to consider that images have been coregistered [13], and that the possible differences in light and atmospheric conditions have been corrected [14]. We will see in Section 4 that our formulation is invariant to global illumination changes. In the change detection problem we have two classes: "Change" (*CH*) and "No change" (*NC*). To formulate the change detection problem as a segmentation problem we will define an appropriate likelihood $L(Y \mid X)$ and Markov Random Field.

2.3.1. The likelihood model

The likelihood for both images given the classes will be based on a gaussian distribution $N(\boldsymbol{\mu}, S_{x_i})$, with common mean in each class:

$$D_i(x_i) = -l_i(\mathbf{y}_i \mid x_i, \mathbf{S}_{x_i}) = -\log(\mid 2\mathbf{pS}_{x_i} \mid^{-d/2} \exp(-1/2(\mathbf{y}_i - \boldsymbol{\mu})^\mathrm{T} \mathbf{S}_{x_i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}))), \tag{9}$$

$$\mathbf{y}_i^\mathrm{T} = (\mathbf{u}_i^\mathrm{T}, \mathbf{v}_i^\mathrm{T}) \quad x_i, x_j \in \{CH, NC\}$$

where d is the sum of the number bands for images $\mathbf{u}$ and $\mathbf{v}$. Superscript T denotes matrix transpose. We will restrict $\mathbf{S}_{CH}$ to represent independent variables so the correlation part between $\mathbf{u}_i$ and $\mathbf{v}_i$ of matrix $\mathbf{S}_{CH}$ is fixed to 0. Our formulation of the likelihood depend on the set of parameters $\mathbf{Q} = \{\mathbf{m}, \mathbf{S}_{CH}, \mathbf{S}_{NC}\}$.

Another possibility is to use a low-dimensional transform of the two images $T(\mathbf{y}) = T(\mathbf{u}, \mathbf{v})$. A common transform for change detection in monospectral images is the difference transform $T(\mathbf{y}) = \mathbf{u} - \mathbf{v}$ [4]. In this case we have:

$$D_i(x_i) = -l_i(T(\mathbf{y}_i) \mid x_i, \mathbf{s}_{x_i}^2) = -\log((2\mathbf{ps}_{x_i}^2)^{-1/2} \exp(-1/(2\mathbf{s}_{x_i}^2)(T(\mathbf{y}_i) - \boldsymbol{\mu})^2)), \tag{10}$$

$$\mathbf{y}_i^\mathrm{T} = (\mathbf{u}_i^\mathrm{T}, \mathbf{v}_i^\mathrm{T}), \quad x_i, x_j \in \{CH, NC\}$$

2.3.2. The Markov Random Field model

A widely used model for the hidden class labels is the isotropic Ising model. Here the set of sites S is a lattice and the neighbor system is composed of the four nearest neighbors of each pixel. Formally:

$$P(\mathbf{x}) = Z^{-1} \exp(-H(\mathbf{x})) = Z^{-1} \exp(\sum_{c \in C} -V_c(\mathbf{x}_c)) = Z^{-1} \exp(\mathbf{b} \sum_{(i,j) \in C} \mathbf{d}_{x_i = x_j}),$$

$$\mathbf{d}_{x_i = x_j} = \begin{cases} 1 & x_i = x_j \\ 0 & x_i \neq x_j \end{cases} \tag{11}$$

where (i,j) must be adjacent sites to be in the clique set C.

Note that the clique potential function is:

$$V_c(\mathbf{x}_c) = -\mathbf{b}\,\mathbf{d}_{x_i = x_j}, \quad \mathbf{x}_c = (x_i, x_j) \tag{12}$$

This shows that the probability assigned to an image is a function only of the number of homogeneous cliques, or equivalently, the boundary length between the classes. We consider $\mathbf{b} > 0$ to enforce a spatially coherent classification.

2.4. MAP estimation of changes in bitemporal images

Statistical approaches attempt to solve the problem of estimating the class label, given only the data recorded for each pixel. The most popular way to estimate an HMRF is maximum a posteriori (MAP) estimation. MAP estimation consists of maximizing the posterior probability of the hidden class labels given the observed pixel values:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} P(\mathbf{x} \mid \mathbf{y}, \mathbf{Q}, \mathbf{b}) = \arg\max_{\mathbf{x}} Z^{-1} \exp(-H(\mathbf{x} \mid \mathbf{y}, \mathbf{Q}, \mathbf{b})) =$$

$$\arg\max_{\mathbf{x}} Z^{-1} \exp\left(-\sum_{c \in C} V_c(\mathbf{x}_c) - \sum_{i \in S} D_i(x_i)\right) = \arg\min_{\mathbf{x}} \sum_{c \in C} V_c(\mathbf{x}_c) + \sum_{i \in S} D_i(x_i) \tag{13}$$

Greig et al. [10] were first to discover that powerful min-cut/max-flow algorithms from combinatorial optimisation could be used to minimize the energies represented in (13). Greig et al. constructed a two terminal graph such that the minimum cost cut of the graph gives a globally optimal binary labelling in case of the Ising model of interaction (11). Previously, exact minimization of energies like (13) were not possible and such energies were approached mainly with iterative algorithms like simulated annealing. Unfortunately, the graph cut technique remained unnoticed for almost 10

years. In the late 90's a large number of new computer vision techniques appeared that figured how to use min-cut/max-flow algorithms on graphs for solving more complex problems [10], [11], [12].

To apply the min-cut approach it is necessary build a graph so that the global minimum of the energy function can be computed by solving a standard two terminal minimum cut problem over it. Consider a graph G defined as follows: there are two terminals: the source *CH* and the sink *NC*. For each pixel we create a vertex i. These vertex are connected to the terminals by t-links $\{CH, i\}$, $\{i,NC\}$ with weights $D_i(CH)+K$ and $D_i(NC)+K$. Constant K must be greater than $|N_i|\boldsymbol{b}$ where N_i is the number of neighbours (four in our case). For each pair of neighbouring pixels i and j we create an n-link $\{i,j\}$ with weight $\boldsymbol{b}$. An example for a 3x3 image is shown below:
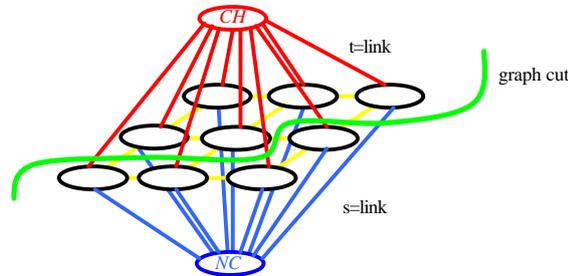

Figure 1 Example of graph construction for the image change problem (adapted from [12]).

The minimum cut partitions the pixels in two sets. Pixels in the same cut as vertex *CH* are classified as "Change" pixels and the rest are the "No Change" pixels. The minimum cut gives the global minimum of the energy function and can be computed in polynomial time. Notice that algorithms based in the image differences like [15] that were previously solved using approximate iterative techniques may in fact be solved exactly.

3. LEARNING THE PARAMETERS OF THE ENERGY FUNCTION

A potential problem with the presented approach is to set the correct value for the parameters $\boldsymbol{Y}=(\boldsymbol{\theta},\boldsymbol{b})$. Parameter estimation for Markov HMRF models has always been considered a difficult problem. This has lead to many ad-hoc schemes for estimating the parameters of the model [7], [16], [17], including choosing parameter values by hand.

The common approach to the parameter estimation problem is to obtain the maximum likelihood estimate (MLE) of the parameters given the observations $\boldsymbol{y}$. The log-likelihood of the model is:

$$\log(P(\boldsymbol{y}\,|\,\boldsymbol{Y})) = \log\left(\sum_{x} P(\boldsymbol{y},\boldsymbol{x}\,|\,\boldsymbol{Y})\right)$$

(14)

It is impossible to obtain an analytical solution of the MLE so numerical methods are needed. The Expectation Maximization (EM) [18] algorithm is an iterative algorithm that tries to maximize the log-likelihood by optimizing at iteration q:

$$Q(\boldsymbol{Y}\,|\,\boldsymbol{Y}^{(q)}) = \mathrm{E}(\log(P(\boldsymbol{y},X\,|\,\boldsymbol{Y})\,|\,Y=y,\boldsymbol{Y}^{(q)})$$

(15)

the expectation of the complete log-likelihood given the observation $\boldsymbol{y}$ and the current estimate of the parameters $\boldsymbol{Y}^{(q)}$

The EM algorithm proceeds as follows:

Initialization	Start from an initial guess $\boldsymbol{Y}^{(0)}$ of $\boldsymbol{Y}$
Expectation Step	Compute $Q(\boldsymbol{Y}\,|\,\boldsymbol{Y}^{(q)})$

Maximization Step   Update the current estimate $\boldsymbol{Y}^{(q)}$ to $\boldsymbol{Y}^{(q+1)} = \arg\max_{\boldsymbol{Y}} Q(\boldsymbol{Y} \mid \boldsymbol{Y}^{(q)})$

or just obtain $\boldsymbol{Y}^{(q+1)}$ increasing $Q(\boldsymbol{Y} \mid \boldsymbol{Y}^{(q)})$ (Generalized EM)

Repeat the Expectation and Maximization steps until convergence.

The EM algorithm guarantees an increase in the log-likelihood until a local optima is found. To proceed with the Expectation step note that using (8) Q can be written as:

$$Q(\boldsymbol{Y} \mid \boldsymbol{Y}^{(q)}) = \sum_{i \in S} \sum_{x_i} P(x_i \mid \boldsymbol{y}, \boldsymbol{Y}^{(q)}) l_i(\boldsymbol{y}_i \mid x_i, \boldsymbol{\varphi}) - \log Z(\boldsymbol{b}) - \sum_{c \in C} \sum_{x_c} V_c(\boldsymbol{x}_c \mid \boldsymbol{b}) P(\boldsymbol{x}_c \mid \boldsymbol{y}, \boldsymbol{Y}^{(q)}) \tag{16}$$

The first term does not depend on $\boldsymbol{b}$ while the last two do not involve $\boldsymbol{\varphi}$. We rearrange those terms and write:

$$Q(\boldsymbol{\varphi} \mid \boldsymbol{Y}^{(q)}) = \sum_{i \in S} \sum_{x_i} P(x_i \mid \boldsymbol{y}, \boldsymbol{Y}^{(q)}) l_i(\boldsymbol{y}_i \mid x_i, \boldsymbol{\varphi}) \tag{17}$$

$$Q(\boldsymbol{b} \mid \boldsymbol{Y}^{(q)}) = -\log Z(\boldsymbol{b}) - \sum_{c \in C} \sum_{x_c} V_c(\boldsymbol{x}_c \mid \boldsymbol{b}) P(\boldsymbol{x}_c \mid \boldsymbol{y}, \boldsymbol{Y}^{(q)}) \tag{18}$$

There are several problems to compute both terms. In order to evaluate $Q(\boldsymbol{\varphi} \mid \boldsymbol{y}^{(q)})$ we need the conditional probability $P(x_i \mid \boldsymbol{y}, \boldsymbol{y}^{(q)})$ that cannot be computed exactly. To compute $Q(\boldsymbol{b} \mid \boldsymbol{y}^{(q)})$ we need both $P(\boldsymbol{x}_c \mid \boldsymbol{y}, \boldsymbol{y}^{(q)})$ and the partition function $Z(\boldsymbol{b})$ that also cannot be computed exactly. There are several approaches to approximately solve these problems: mean field-like approximations, Classification EM or Stochastic EM [16]. Our approach is based on Montecarlo EM (MCEM) [19] in which expectations are estimated from samples of the distributions.

3.1 Estimation of parameter $\boldsymbol{b}$

To estimate parameter $\boldsymbol{b}$ we differentiate in (18) and make the resulting expresion equal to zero obtaining the equation:

$$E\left(\sum_{c \in C} \sum_{x_c} \frac{\partial V_c(\boldsymbol{x}_c \mid \boldsymbol{b})}{\partial \boldsymbol{b}} \right) = E\left(\sum_{c \in C} \sum_{x_c} \frac{\partial V_c(\boldsymbol{x}_c \mid \boldsymbol{b})}{\partial \boldsymbol{b}} \mid \boldsymbol{y}, \boldsymbol{b}^{(q)} \right) \tag{19}$$

In our case (Ising model) we have:

$$E\left(\sum_{(i,j) \in C} \sum_{x_i, x_j} \boldsymbol{d}_{x_i = x_j} \right) = E\left(\sum_{(i,j) \in C} \sum_{x_i, x_j} \boldsymbol{d}_{x_i = x_j} \mid \boldsymbol{y}, \boldsymbol{b}^{(q)} \right) \tag{20}$$

The interpretation of the above equation is clear. The mean number of homogeneous cliques in both distributions must be the same. Notice that the first term is a function of $\boldsymbol{b}$ while the second term is a constant. Both terms are approximately computed by sampling from the correspondent distributions using the Swendsen-Wang algorithm [20], [21]. To solve this unidimensional equation we use the bisection method. This method has the advantage of reusing function evaluations done in previous iterations of the EM algorithm. Note also that the estimation of the function does not depend on the images to process (only of its size) and could be computed before learning.

3.2 Estimation of parameter $\boldsymbol{\varphi}$

In our case (9) the set of parameters $\boldsymbol{\varphi}$ comprises a common mean for the "Change" and "No Change" class and two separate covariance matrices $\boldsymbol{S}_{CH}, \boldsymbol{S}_{NC}$, that is $\boldsymbol{\varphi} = \{\boldsymbol{m}, \boldsymbol{S}_{CH}, \boldsymbol{S}_{NC}\}$. Instead of optimizing $Q(\boldsymbol{\varphi} \mid \boldsymbol{y}^{(q)})$ directly we will only

seek to increase it. This corresponds to the Generalised EM algorithm (GEM). We increase the log-likelihood this proceeding in two steps. First we maximize $\boldsymbol{m}$ with $\boldsymbol{S}_{CH}$, $\boldsymbol{S}_{NC}$ fixed to their previous values $\boldsymbol{S}_{CH}^{(q)}$, $\boldsymbol{S}_{NC}^{(q)}$ obtaining:

$$\boldsymbol{m}^{(q+1)} = \sum_{x_i} A_{x_i} \, \bar{y}_{x_i},$$

$$\bar{y}_{x_i} = \frac{\sum_{i \in S} P(x_i \mid y, \boldsymbol{Y}^{(q)}) y_i}{\boldsymbol{p}_{x_i} \mid S \mid}, \quad \boldsymbol{p}_{x_i} = \sum_{i \in S} P(x_i \mid y, \boldsymbol{Y}^{(q)}) / \mid S \mid, \quad A_{x_i} = \boldsymbol{p}_{x_i} \left(\boldsymbol{S}_{x_i}^{(q)}\right)^{-1} \left(\boldsymbol{p}_{x_i} \sum_{x_i} \left(\boldsymbol{S}_{x_i}^{(q)}\right)^{-1}\right)^{-1} \tag{21}$$

where $|S|$ is the number of sites (the number of pixels in the image).

Then we obtain the optimal values for $\boldsymbol{S}_{CH}$, $\boldsymbol{S}_{NC}$ with the value for $\boldsymbol{m}$ obtained before as:

$$\boldsymbol{S}_{x_i}^{(q+1)} = S_{x_i} + (\bar{y}_{x_i} - \boldsymbol{m}^{(q+1)})(\bar{y}_{x_i} - \boldsymbol{m}^{(q+1)})^{\mathrm{T}} \tag{22}$$

$$S_{x_i} = \frac{\sum_{i \in S} P(x_i \mid y, \boldsymbol{Y}^{(q)})(y_i - \bar{y}_{x_i})(y_i - \bar{y}_{x_i})^{\mathrm{T}}}{\boldsymbol{p}_{x_i} \mid S \mid}$$

Finally, covariance terms between the two images in $\boldsymbol{S}_{CH}$ are set to zero, to ensure modeling independence between the two images for the "Change" class. These update formulas in (21) and (22) guarantee an increase in the log-likelihood until a local maximum is found.

4. RADIOMETRIC INVARIANCE

Before remote sensing images can be used for a change detection study, the images must first be standardized for conditions outside of real surface change. Differences in the sensor, solar illumination or atmospheric conditions make it difficult, if not impossible, to accurately compare satellite or aerial images acquired on different dates and/or different platforms. Several methods [14] have been proposed for the relative radiometric normalization of multispectral images taken under different conditions at different times. All proceed under the assumption that the relationship between the radiances recorded at two different times from regions of constant reflectance is spatially homogeneous and can be approximated by linear functions. In this section we show that our approach is invariant to global affine transforms of both images

Let us write again our original observations for pixel i as $y_i = (\boldsymbol{u}_i^{\mathrm{T}}, \boldsymbol{v}_i^{\mathrm{T}})^{\mathrm{T}}$. If each image is affinely transformed we obtain for each pixel a new value:

$$Ty_i = \begin{pmatrix} T\boldsymbol{u}_i \\ T\boldsymbol{v}_i \end{pmatrix} = \boldsymbol{M} \begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{pmatrix} + \boldsymbol{n}, \quad \boldsymbol{M} = \begin{pmatrix} \boldsymbol{M}_u & 0 \\ 0 & \boldsymbol{M}_v \end{pmatrix} \boldsymbol{n} = \begin{pmatrix} \boldsymbol{n}_u \\ \boldsymbol{n}_v \end{pmatrix} \tag{23}$$

The update equations in each iteration of the EM are now (assuming that $\boldsymbol{M}$ is invertible):

$$T\boldsymbol{\mu}^{(q+1)} = \boldsymbol{M}\boldsymbol{\mu}^{(q)} + \boldsymbol{n}, \tag{24}$$

$$T\boldsymbol{S}_{x_i}^{(q+1)} = \boldsymbol{M}\boldsymbol{S}_{x_i}^{(q+1)}\boldsymbol{M}^T$$

Then we obtain as optimal parameters $T\boldsymbol{\varTheta} = \{ T\boldsymbol{m}, T\boldsymbol{S}_{CH}, T\boldsymbol{S}_{NC} \}$ and the log-likelihood for the transformed images $l_i(T\boldsymbol{y}_i \mid x_i, T\boldsymbol{\varTheta})$ is equal to the original log-likelihood $l_i(y_i \mid x_i, T\boldsymbol{\varTheta})$ making the whole change detection process invariant.

5. EXPERIMENTAL RESULTS

In order to asses the effectiveness of the proposed techniques for change detection, we consider four different data sets: the three first ones are aerial images and the last one is a satellite image. The first dataset is a multispectral (RGB) pair of images corresponding to the island of Gran Canaria (Spain). The three last data sets are monospectral (black and white images), two corresponding to the island of Tenerife (Spain) and the last one corresponding to Madrid (Spain). In the following both the data sets and the experiments are detailed.

5.1. Data sets

The first of the four data sets consisted on two multispectral images acquired from an airborne platform in the island of Gran Canaria (Spain) in 1996 and 1998. They are color images and have 3 channels for the Red, Green, and Blue components. The area selected for the experiments was a section (400x400 pixels) of the full images. The second data set consisted on two aerial monospectral images in an area of La Laguna in the island of Tenerife (Spain) acquired in 1996 and 1998. The area selected for the experiments was a section (640x480 pixels) of the full images. The third data set consisted on two aerial monospectral images in an area of the city of Santa Cruz de Tenerife in the island of Tenerife (Spain) acquired in 1996 and 1998. The area selected for the experiments was a section (492x492 pixels) of the full images. The last data set consisted on two satellite monospectral images acquired by the IRS-1C satellite in an area of the city of Madrid (Spain) in 1999 and 2000. The area selected for the experiments was a section (500x500 pixels) of the full images.

All images are orthophotos so no geometric registration was employed. Histograms for each band in the first image were transformed to match the corresponding band in the second image to partially remove any non linear transform in the light conditions at the time of the two acquisitions. The noise affecting the data values was reduced by applying a smoothing filter (3x3 window size) in all images.

5.2. Description of the experiments

For all the datasets we assumed that the model parameters were unknown so we employed the EM algorithm. The initialization of parameter in the EM algorithm was done as follows: Parameter b was set to 1.5. An initial estimation of the posterior field was done computing the difference image. We set a probability of change of 1 to all pixels whose Euclidean distance between the multispectral data in the two images was above a threshold (40% of the maximum distance). The initial value for m was taken from the mean of both images. From the initial posterior field and mean we calculated the starting values for S_{CH}, S_{NC} using equation (22). The max-flow/min-cut code used for MAP computation is a freely available C++ implementation from [12]. The evaluation of the MAP solution in all the experiments took no more than two seconds. The EM algorithm converged typically in less that ten iterations but the learning time is high due to the complexity of the sampling process.

5.2.1. Experimental results for the first data set

In this example we show the results of the change detection process for the two colour orthophotos with 1:10000 scale (1 meter per pixel) of the first data set. In Figure 2 we can see the two colour images (printed in black and white) that were taken in 1996 and 1998 over an urban development area in Gran Canaria (Canary Islands). In this case we have a truly multispectral image with 3 components corresponding to the Red, Green and Blue (RGB) components. The main changes are new buildings on the right of the images and the completion of part of the roads.

The parameters found by the EM algorithm were:

b=1.07

$$
\boldsymbol{m} = \begin{pmatrix} 138.39 \\ 141.95 \\ 133.70 \\ 139.03 \\ 142.92 \\ 134.43 \end{pmatrix},
\boldsymbol{S}_{CH} = 10^3 \begin{pmatrix} 5.14 & 5.05 & 4.52 & 0 & 0 & 0 \\ 5.05 & 5.11 & 4.64 & 0 & 0 & 0 \\ 4.52 & 4.64 & 4.34 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4.82 & 4.53 & 3.66 \\ 0 & 0 & 0 & 4.53 & 5.03 & 4.44 \\ 0 & 0 & 0 & 3.66 & 4.44 & 4.37 \end{pmatrix},
\boldsymbol{S}_{NC} = 10^3 \begin{pmatrix} 3.66 & 3.59 & 3.19 & 2.90 & 2.77 & 2.31 \\ 3.59 & 3.64 & 3.29 & 2.83 & 2.82 & 2.43 \\ 3.19 & 3.29 & 3.07 & 2.49 & 2.56 & 2.31 \\ 2.90 & 2.83 & 2.49 & 3.20 & 2.98 & 2.37 \\ 2.77 & 2.82 & 2.56 & 2.98 & 3.07 & 2.62 \\ 2.31 & 2.43 & 2.31 & 2.37 & 2.62 & 2.53 \end{pmatrix}
$$
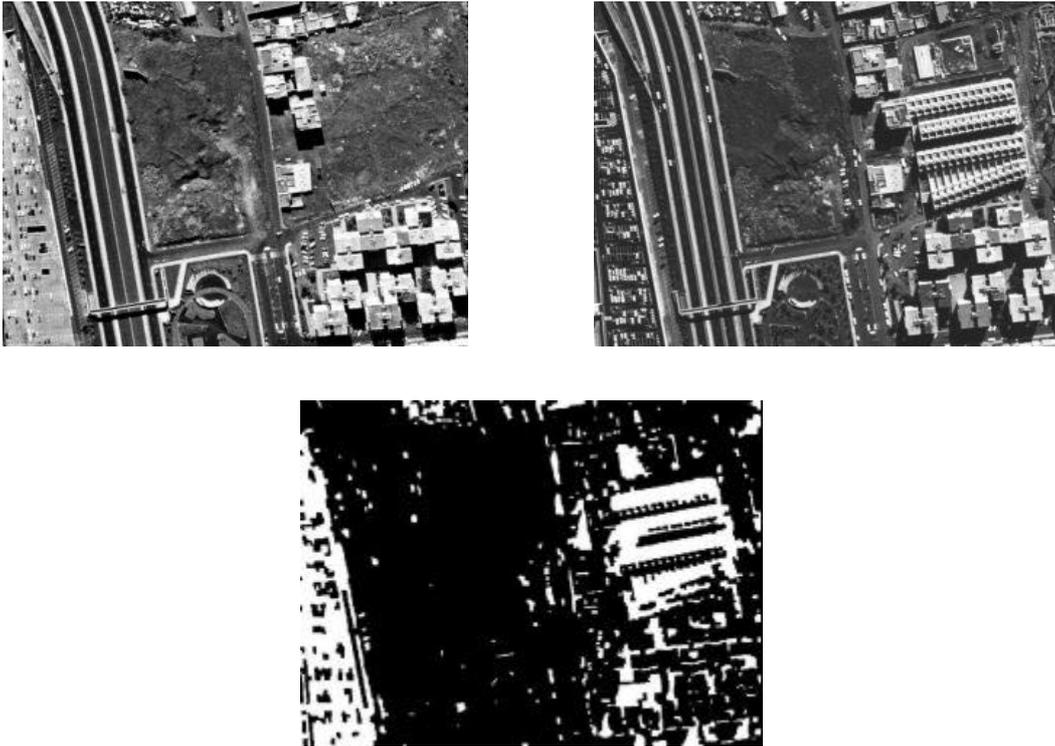
and the change detection results are:
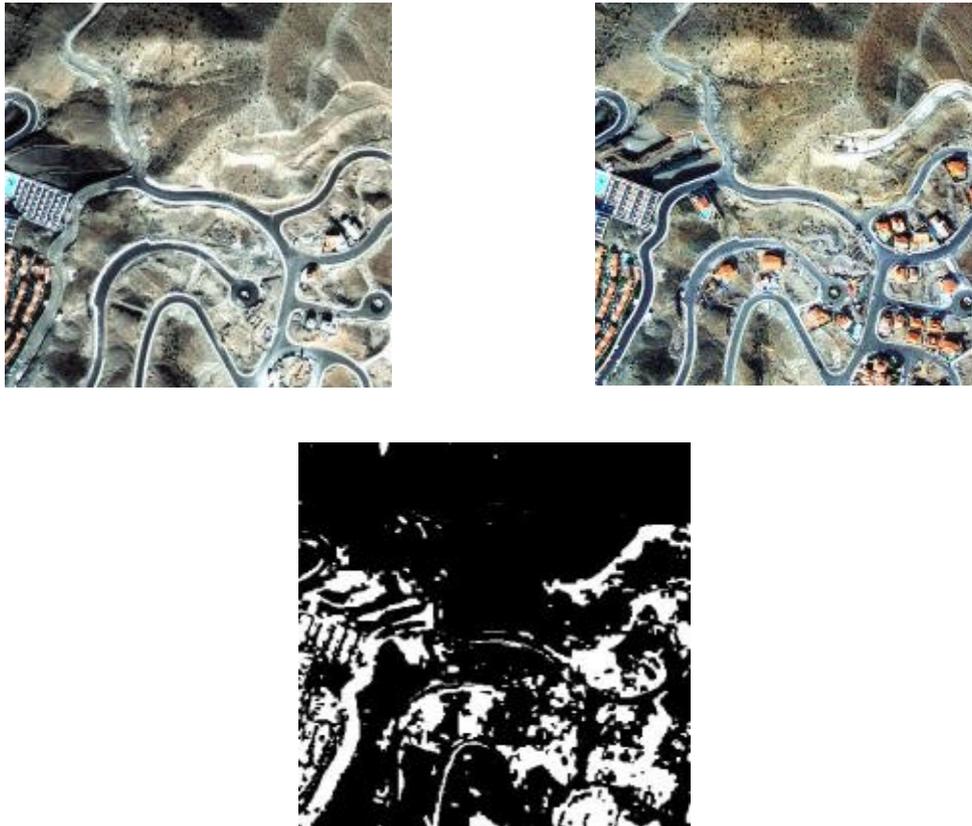
Figure 2 Up: Aerial images from Gran Canaria used in the experiments. Down: MAP solution, black pixels are classified as 'No Change' and white pixels as 'Change'.

The MAP classification image shows that main changes are detected. There are two main sources of errors. The first source is due to geometric registration errors in the ortophotos and can be detected in the borders of the roads and buildings. The second source is due to non uniform lightning changes in the images. We also compared the change detection results for a monochrome version of the two images. As expected the use of 3 components decreases the change detection errors.

5.2.2. Experimental results for the second data set

In this example we show the results of the change detection process for two black and white orthophotos with 1:5000 scale (0.5 meters per pixel). In Figure 3 we can see two images that were taken in 1996 and 1998 over a region in La Laguna, (Tenerife). The main sources of change are a new parking floor on the left and new buildings on the right. The parameters found by the EM algorithm were:

$$b = 1.08,$$

$$m = \begin{pmatrix} 108.03 \\ 107.61 \end{pmatrix} \quad S_{CH} = 10^3 \begin{pmatrix} 5.80 & 0 \\ 0 & 8.31 \end{pmatrix} \quad S_{NC} = 10^3 \begin{pmatrix} 4.01 & 2.78 \\ 2.78 & 3.54 \end{pmatrix}$$

and the change detection results are:

Figure 3 Up: Aerial images from La Laguna used in the experiments. Down: MAP solution, black pixels are classified as 'No Change' and white pixels as 'Change'.

The MAP classification image shows that main changes are detected. We can see again two main sources of errors. The first source is due to geometric registration errors in the ortophotos and can be detected in the borders of the streets and buildings. The second source is due to non uniform lightning changes in the images mainly in the roofs of the buildings present in the two images.

5.2.3. Experimental results for the third data set

The third example we show the results of the change detection process for two black and white orthophotos with 1:5000 scale (0.5 meters per pixel). In Figure 4 we can see two images that were taken in 1996 and 1998 over a region in the city of Santa Cruz de Tenerife (Tenerife). The main changes are the disappearing of the petrol tanks on the right and the construction of a new congress center in the bottom of the image.

The parameters found by the EM algorithm were:

$$b = 1.09,$$

$$m = \begin{pmatrix} 118.77 \\ 118.35 \end{pmatrix} \quad S_{CH} = 10^4 \begin{pmatrix} 1.05 & 0 \\ 0 & 0.59 \end{pmatrix} \quad S_{NC} = 10^3 \begin{pmatrix} 3.59 & 2.58 \\ 2.58 & 4.10 \end{pmatrix}$$
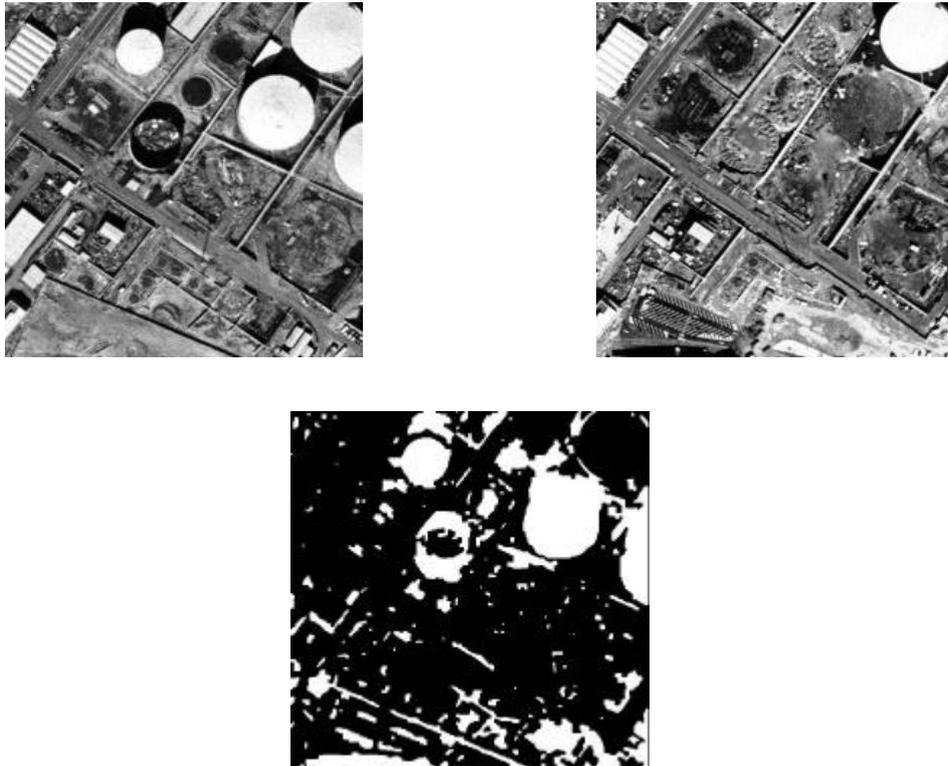
and the change detection results are:

Figure 4 Up: Aerial images from Santa Cruz de Tenerife used in the experiments. Down: MAP solution, black pixels are classified as 'No Change' and white pixels as 'Change'.

The MAP classification image shows that main changes are detected. There are again two main sources of errors. The first source is due to geometric registration errors in the ortophotos and can be detected in the borders of the buildings. The second source is due to the detection of shadows as changes.

5.2.4. Experimental results for the fourth data set

The fourth example we show the results of the change detection process for two black and white satellite images with 5 meters per pixel. In Figure 5 we can see two images over a region in the city of Madrid (Spain) that were taken in 1999 and 2000. The main changes are land movements in several zones of the image.

The parameters found by the EM algorithm were:

$$\boldsymbol{b} = 1.43,$$

$$\boldsymbol{m} = \begin{pmatrix} 170.77 \\ 170.57 \end{pmatrix}, \quad \boldsymbol{S}_{CH} = 10^3 \begin{pmatrix} 0.41 & 0 \\ 0 & 7.11 \end{pmatrix}, \quad \boldsymbol{S}_{NC} = 10^3 \begin{pmatrix} 0.77 & 0.67 \\ 0.67 & 0.88 \end{pmatrix}$$
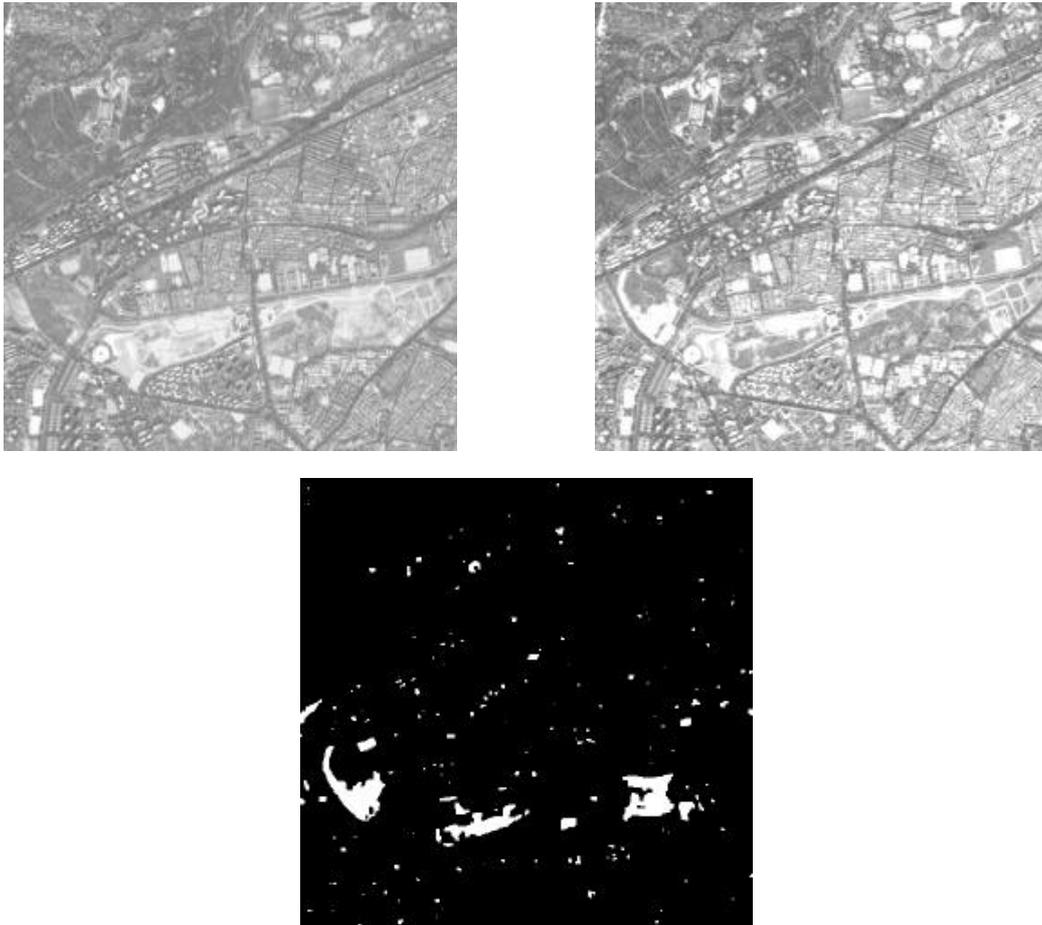
and the change detection results are:

Figure 5 Up: Satellite images from Madrid used in the experiments. Down: MAP solution, black pixels are classified as 'No Change' and white pixels as 'Change'.

Note the increase the value of parameter **b** with respect to the previous images. This means that gross changes are detected. The learned value of 1.43 for parameter beta implies that in mean each pixel shares his class assignment with 3.96 of his 4 neighbours. Again main changes are detected in this case. Geometric registration errors and lightning variations have less importance.

6. CONCLUSIONS

In this paper, we have presented an unsupervised approach to the change detection problem on remote sensing images. From a methodological viewpoint, the main innovation of this paper lies in the formulation of the change detection problem in the HMRF framework and its exact and efficient resolution using graph cuts. This makes the use of Markov Random Fields for change detection a viable alternative to existing approaches. Another problem addressed in this paper is how to determine the parameters in these models. The EM algorithm is used to solve the learning problem. This approach shares two errors sources with other change detection techniques: geometric registration uncertainty and non uniform lightning variations. Future extensions of this work will try to solve these problems integrating them into the HMRF model.

REFERENCES

1. Deer, P.J., Digital Change Detection Techniques: Civilian And Military Applications, *International Symposium on Spectral Sensing Research (ISSSR)*, (1995).
2. Singh A., Digital Change Detection Techniques Using Remotely-Sensed Data, *International Journal Of Remote Sensing,* 10 (6): 989-1003, (1989).
3. Coppin P., Jonckheere I., Nackaerts K., Muys I. and Lambin E., Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing* 25:1565-1596, (2004).
4. Radke R.J., Andra S., Al-Kofani O. and Roysan B., Image change detection algorithms: A systematic survey, *IEEE Trans. Image Processing*, 14 (3):291-307, (2005).
5. Black M. J., Fleet D. J., and Yaccob Y., Robustly estimating changes in image appearance, *Computer Vision and Image Understanding*, Special Issue on Robust Statistical Techniques in Image Understanding, 78 (1): 8-31, (2000).
6. Li, S. Z., *Markov Random Field Modeling in Image Analysis* Springer 2nd ed., (2001).
7. Pérez, P., Markov Random Fields and Images, *CWI Quaterly* 11 (4): 413-437 (1998).
8. Geman S and Geman D, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 :721-741 (1984).
9. Boykov Y., Vexler O. and Zabih R., Markov random fields with efficient approximations. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 648-655, (2001).
10. Greig D., Porteous B. and Seheult A., Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, 51(2):271–279, (1989).
11. Kolmogorov V. and Zabih R, Visual correspondence with occlusions using graph cuts. *International Conference on Computer Vision*, 508–515 (2001).
12. Boykov, Y., Vexler, O. and Zabih R., Fast approximate energy minimization via graph cuts *, IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222-1239, (2001).
13. Townshend J., Justice C., Gurney and McManus, J., The impact of misregistration on change detection, *IEEE Trans. Geosci. Remote Sensing*, 30: 1054-1060, (1992).
14. Chavez P. S., Radiometric calibration of Landsat Thematic Mapper multispectral images.*Photogrammetric Engineering and Remote Sensing* 55(9): 1285-1294 (1989).

15. Bruzzone L. and Prieto D. F., Automatic analysis of the difference image for unsupervised change detection, *IEEE Trans. on Geosci. Remote Sensing*, 38 (3): 1171-1182, (2000).

16. Celeux G. N., Forbes F. and Peyrard N., EM Procedures Using Mean Field-Like Approximations for Markov Model-Based Image Segmentation, *Pattern Recognition*, 36: 131-144, (2003).

17. Wang L., Liu J. and. Li S.Z., MRF Parameter Estimation by MCMC Method, *Pattern Recognition*, 33 (11), 1919-1925, (2000).

18. Dempster A.P., Laird N.M. and Rubin D.B., Maximum likelihood from incomplete data via the EM algorithm, *J. Royal Statistical Soc.*, Ser. B 39 (1), 1-38. (1977).

19. Wei, G. and Tanner, M, A Monte Carlo implementation of the EM algorithm and the poor mans data augmentation algorithms. *Journal of the American Statistical Association* 85 (1990).

20. Swendsen R.H. and Wang J.S., "Nonuniversal critical dynamics in Monte Carlo simulations", Physical Reviews Letters 58 (2): 86-88 (1987).

21. Barbu A. and Zhu S. C., Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (27) (2005).