Smoothing parameter estimation for Markov random field classification of non-Gaussian distribution image

Hossein Aghighi, John Trinder, Ke Wang, Yuliya Tarabalka, Samsung Lim

HAL Id: hal-01068242
https://inria.hal.science/hal-01068242

Submitted on 25 Sep 2014

SMOOTHING PARAMETER ESTIMATION FOR MARKOV RANDOM FIELD CLASSIFICATION OF NON-GAUSSIAN DISTRIBUTION IMAGE

H. Aghighi[a,c] *, J. Trinder [a], K. Wang [a], Y. Tarabalka [b], S. Lim [a]

[a] School of Civil and Environmental Engineering, The University of New South Wales, UNSW SYDNEY NSW 2052, Australia
h.aghighi@student.unsw.edu.au, j.trinder@unsw.edu.au, ke.wang@student.unsw.edu.au, s.lim@unsw.edu.au
[b] Inria Sophia-Antipolis Méditerranée, AYIN team, 06902 Sophia Antipolis, France- yuliya.tarabalka@inria.fr
[c] Department of Remote Sensing & GIS, Faculty of Earth science, Shahid Beheshti University, Tehran, Iran

Commission VII, WG VII/4

KEY WORDS: Markov random field, smoothing parameter, *SVM*, non-Gaussian distribution

ABSTRACT:

In the context of remote sensing image classification, Markov random fields (*MRFs*) have been used to combine both spectral and contextual information. The *MRFs* use a smoothing parameter to balance the contribution of the spectral versus spatial energies, which is often defined empirically. This paper proposes a framework to estimate the smoothing parameter using the probability estimates from support vector machines and the spatial class co-occurrence distribution. Furthermore, we construct a spatially weighted parameter to preserve the edges by using seven different edge detectors. The performance of the proposed methods is evaluated on two hyperspectral datasets recorded by the AVIRIS and ROSIS and a simulated ALOS PALSAR image. The experimental results demonstrated that the estimated smoothing parameter is optimal and produces a classified map with high accuracy. Moreover, we found that the Canny-based edge probability map preserved the contours better than others.

1. INTRODUCTION

Hyperspectral imaging sensors provide a huge amount of data with rich spatial, spectral and temporal resolution information. These images have attracted much attention in the remote sensing community and have opened the doors to variety of new applications and challenges, which need to employ both spectral and spatial information for accurate data analysis and classification (Camps-Valls et al., 2014). Markov random fields (*MRFs*) as undirected graphical models are common methods for incorporating both spectral and contextual information (Chen et al., 2010a). They are formulated as the minimization of an energy function which consists of spectral and spatial energy terms and an important term which plays a key controlling role known as smoothing parameter (Aghighi et al., 2014). Larger values of the smoothing parameter result in over-smoothed classified maps and too small values do not fully utilize the available spatial information (Tolpekin and Stein, 2009). Several attempts have been made to estimate the smoothing parameter, which would maximize the image classification accuracy. Derin and Elliott (1987) employed the least squares method, but the performance of their method was limited by the choice of suitable data samples (Derin and Elliott, 1987, Tso and Mather, 1999). A number of studies have examined heuristic optimization algorithms to estimate the smoothing parameters, such as iterative conditional estimation (Salzenstein and Pieczynski, 1997), genetic algorithms (Tso and Mather, 1999), simulated annealing (Li et al., 2012), and Ho-Kashyap optimization method which was used for automatic weight parameters determination in the context of supervised classification by using training data (Serpico and Moser, 2006). In the recent years, Tolpekin and Stein (2009) demonstrated a new smoothing parameter estimation technique for super-resolution mapping based on class separability, the neighbourhood system size and the configuration of class labels. Then, Li et al. (2012) added local properties in estimating the optimal smoothing parameter and developed their spatial adaptive method. However, both meth-

ods introduced another parameter, which was set as a constant empirical value. Moreover, both methods suffered from similar covariance assumptions for the classes. Thus, we proposed a robust smoothing parameter estimation framework to overcome their limitations (Aghighi et al., 2014). One of the limitations of the described methods is that they were developed based on the assumption of the Gaussian class conditional distribution (Aghighi et al., 2014), (Tolpekin and Stein, 2009) and (Li et al., 2012). Although the Gaussian distribution is widely applied in many of image labelling applications, this assumption may not be tenable for remotely sensed mixed pixels (Xu et al., 2005). Moreover, some practical data such as polarimetric synthetic aperture radar (PolSAR) data are non-Gaussian (Doulgeris et al., 2008).

Another common drawback of contextual based classification approaches is that they generate unreliable classification results near edges between the land covers (Moser et al., 2013), or remove the edges of small features. Therefore, several attempts have been made to develop the edge-preserving methods, which have been used during the *MRF* optimization process. For instance, line processes (Moser et al., 2013) and (Solberg et al., 1996) and adaptive neighborhood systems (Hegarat-Mascle et al., 2007), are both based on the assumption of using an ideal edge map. However, each image band may provide different or even conflicting information based on its wavelength. Therefore, some other methods such as *fuzzy no-edge/edge* function using Sobel mask (Tarabalka et al., 2010), edge probability map using Canny operator (Aghighi et al., 2014) and graduated increase edge penalty (Yu and Clausi, 2008) were introduced in previous studies.

This article presents a novel robust framework for the smoothing parameter estimation which is not dependent on assumptions of a specific statistical distribution of the image data (Section 2.1). This contextually adaptive smoothing parameter estimation method is proposed on the basis of the balance of spatial and spectral energies and the global spatial frequency distribution of a co-occurrence class label. For this purpose, we have introduced a new spectral energy change function and two new concepts called the class label co-occurrence matrix of the categories

(*CLCMC*) and global class label co-occurrence matrix of the categories (*GCLCMC*), which can all be computed using pairwise coupling of probability estimates from support vector machines (*SVM*) one-versus-one classification outputs. Furthermore, we compared seven different edge filters to incorporate the edge information into the *MRF* framework: Canny, Sobel, Roberts, Prewitt, Laplacian of Gaussian, curvature edge indicator, graduated edge penalty. The performance of the proposed method is evaluated using two hyperspectral images collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and the Reflective Optics System Imaging Spectrometer (ROSIS) (Section 3). Finally, conclusions are drawn in Section 4.

2. PROPOSED METHOD

In the development of our new smoothing parameter estimation framework, we denote an image by $Y = \{\mathscr{Y}_i \in \mathbb{R}^B, i = 1, 2, \cdots, m\}$, where B is a number of spectral channels, and m is a number of pixels. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$ be a set of M thematic classes of interest. The classification task consists in assigning for each pixel $\mathscr{Y}_i$ a class label l_j, yielding the classification map $L = \{\ell_j, j = 1, 2, \cdots, m\}$.

2.1 The Potts MRF model

The optimal classification map L^* given the image Y can be generated by solving the maximization problem for the a posteriori probability (*MAP*) decision rule (1).

$$L^* = \underset{L}{\arg\max}\{p(L|Y)\} = \underset{L}{\arg\max}\{p(Y|L)p(L)\} \quad (1)$$

where, $p(Y \mid L)$ is the class-conditional distribution and $p(L)$ is the prior probability distribution. Based on the complexity of (1) which involves the optimization of a global distribution model of the image and due to the equivalence of *MRF* and Gibbs random field (Duggin and Robinove, 1990), (Geman and Geman, 1984), this optimization problem can be simplified and resolved by minimizing the sum of local posterior energies (2) (Geman and Geman, 1984):

$$U(L|Y) = U(Y|L) + U(L). \quad (2)$$

In this research, we applied expectation-maximization (*EM*) algorithm (Levitan and Herman, 1987) to compute the *MAP*. In Equation (2), $U(Y \mid L)$ and $U(L)$ denote spectral and spatial energy terms, respectively. The spatial term $U(L)$ is defined by using the Potts model, which penalizes different class labels for neighboring pixels (3):

$$U(L) = \sum_{\mathscr{Y}_j \in N_i} \mathscr{W}(\ell_j)(1 - \delta(\ell_i, \ell_j)) \quad (3)$$

where, ℓ_j is the class label of a pixel $\mathscr{Y}_j$ which is a member of the symmetric neighborhood for pixel $\mathscr{Y}_i$ denoted by N_i. In this equation, $\delta(\ell_i, \ell_j)$ is the Kronecker delta function ($\delta(\ell_i, \ell_j) = 1$ *if* $\ell_i = \ell_j$ and $\delta(\ell_i, \ell_j) = 0$ *if* $\ell_i \neq \ell_j$. In Equation (3), $\mathscr{W}(\ell_j)$ denotes the weight of contribution from pixel $\ell_j \in N_i$ to the spatial energy term and can be modeled as:

$$\mathscr{W}(\ell_j) = q \times \phi(\ell_j) \quad (4)$$

where, $0 \leq q < \infty$ and $\sum_{\ell_j \in N_i} \phi(\ell_j) = 1$. In these equations, q controls the overall magnitude of weights and consequently the spatial energies; thus, larger values of q leads to smoother solutions (Tolpekin and Stein, 2009). Here, $\phi(\ell_j)$ inversely depends to $d(\ell_i, \ell_j)$ which is a geometric distance between the pixel ℓ_i and

its spatial neighbors ℓ_j (5) (Li et al., 2012).

$$\phi(\ell_j) = \frac{1}{\eta}\left(\frac{d(\ell_i, \ell_j)}{r}\right)^{-g} \quad (5)$$

where, r is the spatial resolution, η is a normalization constant to $\sum_{\ell_j \in N_i} \phi(\ell_j) = 1$ and g is a power-law index which is assumed to be one. By substitution (3- 5) in (2) we can write (2) as (6).

$$U(L|Y) = U(Y|L) + q\sum_{\mathscr{Y}_j \in N_i} \phi(\ell_j)(1 - \delta(\ell_i, \ell_j)) \quad (6)$$

In order to normalize $U(L \mid Y)$, we multiply (6) by $1/(1+q)$:

$$(L|Y) \propto \frac{1}{1+q}U(Y|L) + \frac{q}{1+q}\sum_{\mathscr{Y}_j \in N_i} \phi(\ell_j)(1 - \delta(\ell_i, \ell_j)) \quad (7)$$

We call $q/(1+q)$ smoothing parameter and denote it by λ; therefore, $1/(1+q)$ can be written as $1 - \lambda$. As mentioned, the optimal classified map L^* depends on the maximization of the posterior probability (1) or the minimization of local posterior energies (2 or 7), thus the absolute value of $U(L \mid Y)$ in (7) is not important. From Equation (8) and on, $U(L \mid Y)$ is referred to as the normalised energy.

$$U(L|Y) = (1 - \lambda)U(Y|L) + \lambda \sum_{\mathscr{Y}_j \in N_i} \phi(\ell_j)(1 - \delta(\ell_i, \ell_j)) \quad (8)$$

The spectral energy $U(Y \mid L)$ can be computed as (9) (Tarabalka et al., 2010):

$$U(Y|L) = -\ln\{P(\mathscr{Y}_i|\ell_i)\} \quad (9)$$

In order to estimate the smoothing parameter, consider that a given pixel i with the true label $\ell_i = \alpha$ is assigned to an incorrect class label $\ell_i = \beta$. Therefore, based on (1) we can infer that:

$$p(\ell_i = \beta|\mathscr{Y}_i) \leq p(\ell_i = \alpha|\mathscr{Y}_i) \quad (10)$$

Which is same as (11)

$$U(\ell_i = \alpha|\mathscr{Y}_i) \geq U(\ell_i = \beta|\mathscr{Y}_i) \quad (11)$$

By substituting the corresponding terms in (11) and solving this inequality equation, we will have the local likelihood energy change $\Delta U_{\alpha\beta}^l$ and the change of a local prior energy which is simplified as $\Delta U_{\alpha\beta}^P = q\psi_{\alpha\beta}$. Furthermore, λ for each pair of classes (α and β) can be estimated by (12) (Aghighi et al., 2014):

$$\lambda_{\alpha\beta} = \frac{1}{1 + \frac{\psi_{\alpha\beta}}{\Delta U_{\alpha\beta}^l}} \quad (12)$$

Due to the assumption of Gaussian class conditional densities, the value of $\Delta U_{\alpha\beta}^l$ was defined as a Mahalanobis distance using the equal covariance matrix for all the classes in the case of (Tolpekin and Stein, 2009, Li et al., 2012); or using the mean of the covariance of each pair of classes in the case of (Aghighi et al., 2014). However, this assumption may not be tenable for remotely sensed mixed pixels (Xu et al., 2005). Thus, in order to avoid the Gaussian distribution assumption we propose a new equation to compute the change in the likelihood energy as:

$$\Delta U_{\alpha\beta}^l = -\ln|\{P(\mathscr{Y}_i|\ell_i = \beta)\} - \{P(\mathscr{Y}_i|\ell_i = \alpha)\}| \quad (13)$$

where, $P(\mathscr{Y}_i \mid \ell_i)$ can be estimated by pairwise coupling of probability estimates from one-versus-one *SVM* outputs (Wu et al., 2004). In order to compute $\Delta U_{\alpha\beta}^l$, the image pixels are catego-

Table 1: Global matrix of $\Delta U^l_{\alpha\beta}$ for the image's selected pixels

$\Delta U^l_{1,1} = 0$	$\Delta U^l_{1,2}$	...	$\Delta U^l_{1,M}$
$\Delta U^l_{2,1}$	$\Delta U^l_{2,2} = 0$	...	$\Delta U^l_{2,M}$
...	...	0	...
$\Delta U^l_{M,1}$	$\Delta U^l_{M,2}$	...	$\Delta U^l_{M,M} = 0$

rized based on the computed class labels and the pixels of each category are sorted in descending order based on the maximum probability in each pixel. Then, by employing the *SVM* accuracy assessment results, the minimum value between user and producer accuracies of each class are used to select the proportion of pixels with the higher probability in each category. These selected pixels in each category are assumed to be reliably classified.

For a given pixel amongst the selected pixels of each category, the maximum probability should belong to the $P(\mathscr{Y}_i \mid \ell_i = \alpha)$ and other probabilities of that pixel belong to $P(\mathscr{Y}_i \mid \ell_i = \beta)$. Then the means of pixels in each category were computed using $\Delta U^l_{\alpha\beta}$ (13). In the next step, each mean vector of $\Delta U^l_{\alpha\beta}$ is normalized and a square matrix of $\Delta U^l_{\alpha\beta}$ with size M was generated (see Table 1). This matrix is a zero diagonal square matrix, because each element on its diagonal indicates $\Delta U^l_{\alpha\alpha}$ is zero.

Then, the class label of the neighbours of each pixel in each category are extracted to compute the *CLCMC* and *GCLCMC*, which are square matrices of size M. Let N_P be the number of pixels for a class ω_i; since the second order *MRF* neighboring system is chosen, each pixel $\mathscr{Y}_i$ is surrounded by $N_S = 8$ pixels $\mathscr{Y}_i$. The *CLCMC* index is calculated for pixels of each class ω_i as:

$$CLCMCC_{\omega_i,\omega_j} = \frac{1}{N_P \times N_s} \sum_{i=1}^{N_P} \sum_{s=1}^{N_s} \delta(\ell_c, \omega_i) \delta(\ell_s, \omega_j) \quad (14)$$

$CLCMC_{\omega_i,\omega_j}$ shows the probability of co-occurrence of class ω_i with other classes ω_j. In this equation, ℓc is the class label of a given pixel (i) of the category and ℓs is the class label of surrounding pixel s. Due to the *MRF* neighboring concept which says that 1) a site cannot be a neighbor with itself $i \notin N'_i$; and 2) the neighborhood relationship is mutual ($i \in N_j \Longleftrightarrow j \in N_i$) (Levitan and Herman, 1987), *CLCMC* is converted to *GCLCMC* to show the global spatial frequency distribution of each pair of classes in the image:

$$GCLCMC_{\omega_i,\omega_j} = CLCMC_{\omega_i,\omega_j} + CLCMC_{\omega_j,\omega_i} \quad (15)$$

The index of *GCLCMC* indicates the probability that a given pixel with true label (α) is misclassified as a false label (β) due to the spatial energy. Therefore, $\psi_{\alpha\beta}$ for each pair of classes ($\psi_{\alpha\beta}$) is the corresponding element of *GCLCMC* matrix for classes α and β. Although the general concepts of the *CLCMC* and *GCLCMC* indexes are similar to the concepts of class label co-occurrence matrix of the blocks (*CLCMB*) and global class label co-occurrence matrix of the blocks (*GCLCMB*) which we proposed earlier (Aghighi et al., 2014), for this paper both *CLCMC* and *GCLCMC* are normalized and indicate the probability of co-occurrence of class ω_i with another classes ω_j in each category over the whole image.

In the next step, the global matrix $\Delta U^l_{\alpha\beta}$ and $\psi_{\alpha\beta}$ are used to compute $\lambda_{\alpha\beta}$ for each pair of classes using Eq. (12). Finally, the optimized smoothing parameter λ^* is computed by averaging the estimated smoothing parameter $\lambda_{\alpha\beta}$ for each pair of classes (16) (Aghighi et al., 2014):

$$\lambda^* = \frac{\sum_{\alpha=1}^{M} \sum_{\beta=1}^{M} \lambda_{\alpha\beta}}{M(M-1)} \quad (16)$$

2.2 Edge preserving

The first step for preserving the edges is to determine the edge locations. Hence, we employed five of the best known edge detection methods including to Canny (Canny, 1986), Sobel (Roushdy, 2006), Roberts (Jain et al., 1995), Prewitt (Prewitt, 1970), Laplacian of Gaussian (Marr and Hildreth, 1980) to extract the binary edge map I_i, as well as utilize the different curvature indicators to generate the curvature edge indicator D_i in each case (21) (Chen et al., 2010b). Since conflicting information is derived from the different wavelengths in hyperspectral images, I_i and D_i were computed for each image band (B-band). Then a one-band edge probability map from the B-band image (n_b) is computed by using (17) for the Canny, Sobel, Roberts, Prewitt and Laplacian of Gaussian as well as (18) for different curvature indicators.

$$w_{(i)} = 1 - \frac{1}{n_b} \sum_{i=1}^{n_b} I_{(i)} \quad , \quad 0 < w_{(i)} \leq 1 \quad (17)$$

$$w_{(i)} = 1 - \frac{1}{n_b} \sum_{i=1}^{n_b} \frac{D_{(i)}}{max\left(D_{(i)}\right)} \quad (18)$$

In order to preserve the small structures and edges in the classified map, the spatial energy component (3) can be formulated as (19 or 20)(Aghighi et al., 2014).

$$U^E_{spatial}(\mathscr{Y}_i) = \sum_{\mathscr{Y}_j \in N_i} w_{(i)} \left(1 - \delta(\ell_i, \ell_j)\right) \quad (19)$$

$$U^E_{spatial}(\mathscr{Y}_i) = \sum_{\mathscr{Y}_j \in N_i} g(\nabla_s) \left(1 - \delta(\ell_i, \ell_j)\right) \quad (20)$$

where the superscript E refers to the edge probability map and $w_{(i)}$ can be computed using (17 or 18) based on edge detection method and $g(\nabla_s)$ can be calculated using (24).

2.3 Different curvature indicator

This edge indicator was developed by Chen et al (2010) and called different curvature indicator. It can effectively distinguish edges from areas with flat and ramp intensity distributions in the image data (Chen et al., 2010b). The difference curvature D_i for a given pixel i of the image is defined as:

$$D_{(i)} = \left| \left| u_{\eta\eta(i)} \right| - \left| u_{\varepsilon\varepsilon(i)} \right| \right|, \quad (21)$$

where $u_{\eta\eta}$ and $u_{\varepsilon\varepsilon}$ represent the second derivation of the gradient ∇u and perpendicular to ∇u, respectively, and $|.|$ denotes the absolute value

$$u_{\varepsilon\varepsilon} = \frac{u_{xy}^2 u_{xx} + 2u_x u_y u_{xy} + u_x^2 u_{yy}}{u_x^2 + u_y^2} \quad (22)$$

$$u_{\eta\eta} = \frac{u_x^2 u_{xx} + 2u_x u_y u_{xy} + u_y^2 u_{yy}}{u_x^2 + u_y^2}. \quad (23)$$

In these equations, u_x and u_{xx} denote the first and second derivation in x, respectively, u_y and u_{yy} are the first and second derivation in y, respectively, and u_{xy} indicates the first derivation in y of the first derivation in x. Table 2 summarizes the behaviour analysis of the different curvature edge indicator.

2.4 Graduated edge penalty

In order to use the graduated edge penalty, we defined the edge penalty term $g(\nabla_s)$ which can be any monotonically decreasing function of edge strength. In this function, the penalty decreases

Table 2: Behaviour analysis of the different curvature edge indicator (Chen et al., 2010)

	$u_{\eta\eta}$	$u_{\varepsilon\varepsilon}$	D_i
Edge pixel	large	small	large
Flat and ramp pixel	small	small	small
Noise pixel	large	large	small

as the edge strength increases between two groups of pixels assigned to different classes; thus all the edges pixels are not penalized equally. The penalty function is formulated as (Yu and Clausi, 2008):

$$g(\nabla_s) = e^{-(\nabla_s/K)^2}, \tag{24}$$

where ∇_s represents the normalized gradient magnitude ($\nabla_s \in [0,1]$) as the edge strength measurement on site s. K is a positive value that controls the strength of the edge penalty term. By increasing K to infinity, all the edges penalties are equal to 1; and by gradually increasing K, the edges are penalized differently based on the strength or weakness of the edge gradient magnitude (Yu and Clausi, 2008). In order to calculate ∇_s, Equation (25) which is based on the gradient magnitude $\sqrt{\vartheta}$ on site s was used (Yu et al., 2012):

$$\nabla_s = \sqrt{\vartheta}/max_{s \in S}\sqrt{\vartheta}, \tag{25}$$

where ϑ can be computed using (Lee and Cok, 1991):

$$\vartheta = \frac{1}{2}\left(p_s + q_s + \sqrt{(p_s + q_s)^2 - 4(p_s q_s - t_s^2)}\right) \tag{26}$$

Each of the variables of (26) can be estimated as:

$$p_s = \sum_{b=1}^{N_b}\left(\frac{\partial \mathscr{Y}_i^{(b)}}{\partial x}\right)^2 \tag{27}$$

$$t_s = \sum_{b=1}^{N_b}\left(\frac{\partial \mathscr{Y}_i^{(b)}}{\partial x}\right)\left(\frac{\partial \mathscr{Y}_i^{(b)}}{\partial y}\right) \tag{28}$$

$$q_s = \sum_{b=1}^{N_b}\left(\frac{\partial \mathscr{Y}_i^{(b)}}{\partial y}\right)^2 \tag{29}$$

In order to compute ϑ, its items can be calculated using:

$$p_s q_s - t_s^2 = \sum_{b_1=1}^{N_b}\sum_{b_2=1}^{N_b}\left(\frac{\partial \mathscr{Y}_i^{(b_1)}}{\partial x}\frac{\partial \mathscr{Y}_i^{(b_2)}}{\partial y} - \frac{\partial \mathscr{Y}_i^{(b_2)}}{\partial x}\frac{\partial \mathscr{Y}_i^{(b_1)}}{\partial y}\right)^2 \tag{30}$$

Therefore, $p_s q_s - t_s^2 \geq 0$ and $p_s + q_s \geq \lambda$. Since

$$p_s + q_s = \sum_{b=1}^{N_b}\left(\frac{\partial \mathscr{Y}_i^{(b)}}{\partial x}\right)^2 + \left(\frac{\partial \mathscr{Y}_i^{(b)}}{\partial y}\right)^2 = \sum_{b=1}^{N_b}\parallel \nabla \mathscr{Y}_i^{(b)} \parallel^2, \tag{31}$$

where $\frac{\partial \mathscr{Y}_i^{(b)}}{\partial x}$ and $\frac{\partial \mathscr{Y}_i^{(b)}}{\partial y}$ denote the first partial derivatives of the bth univariate band of image $\mathscr{Y}$ on site i with respect to vertical and horizontal directions.

3. EXPERIMENTAL RESULTS AND DISCUSSION

In order to compare the performance of the proposed method with our previous method (Aghighi et al., 2014), the same sets of training and test pixels for each image were selected by the stratified random method (Foody, 2004), which can be found in (Aghighi et al., 2014). In this research, two benchmark hyperspectral datasets and all classes were collected to evaluate the proposed smoothing parameter estimation method.

1) The Indian Pines image: This hyperspectral image was recorded by the AVIRIS sensor from Indian Pines test site which was an agricultural area in North-western Indiana, USA. The image comprises 145 by 145 pixels with 20 m/pixel spatial resolution and 200 spectral bands within a wavelength range of 0.4 to 2.5 m. The reference map contains sixteen classes, namely Alfalfa, Corn-notill, Corn-min, Corn, Grass-pasture, Grass-trees, Grass-pasture-mowed, Hay-windrowed, Oats, Soybean-notill, Soybean-mintill, Soybean clean, Wheat, Woods, Buildings-Grass-Trees-Drives and Stone-Steel-Towers.

2) The Pavia University image: The Pavia University hyperspectral image was recorded by ROSIS sensor from Pavia district in north Italy (Figure 2(a)). This dataset comprising of 610 by 340 pixels with spatial resolution of 1.3m and 103 spectral bands. The reference map (Figure 2(b)) contains sixteen classes, namely asphalt, meadows, gravel, trees, painted metal sheets, bare soil, bitumen, self-blocking bricks and shadows.

In order to evaluate the procedure on non-Gaussian data, we also simulated a 210 by 250 pixel test image using fully polarized L-band data from Phased Array type L-band Synthetic Aperture Radar (PALSAR) sensor at the Japanese Advanced Land Observing Satellite (ALOS)(Pantze et al., 2014). The quad polarizations data (HH, HV, VH and VV) was selected to reach the best classification results (Doulgeris et al., 2011). After pre-processing of the data using software package (NEST version 5.0.12) provided by ESA (European Space Agency), some samples of five class, namely water, forest, urban, farm and open field were extracted. Then, they utilized to produce a simulated PALSAR data (Fig. 1(b)), such that each class has at least on boundary to every other classes (Fig. 1(a)) (Doulgeris et al., 2011). The produced data is non-Gaussian distributed due to using high resolution image and highly textured regions (Doulgeris et al., 2011).
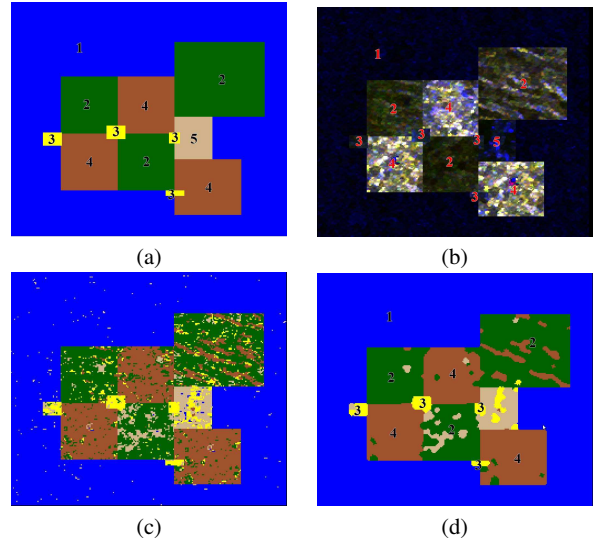

(a) (b)

(c) (d)

Figure 1: Simulated PALSAR data . (a) Reference data ((1) water, (2) Forest, (3) Farm, (4) Urban and (5) Open field). (b) Simulated image (R (VH), G (HV), and B (HH)). (c) *SVM* pixelwise classification map. (d) *SVMMRF* classification map.

In this research, the probabilistic one-versus-one *SVM* classification method was adopted as a nonlinear classifier through the use of Gaussian radial basis function (*RBF*) kernel for hyperspectral data (Tarabalka et al., 2010), as well as for simulated

ALOS PALSAR data (Sambodo and Indriasari, 2014). The optimal *SVM* parameter C and γ were chosen by fivefold cross validation (Aghighi et al., 2014). The *SVMLIB* library (Chang and Lin, 2011) was used to estimate the probability of individual classes for each pixel and produce the classification map (see Fig. 1(c) and Fig. 1(c)). Then, the results of *SVM* classification were used to estimate the smoothing parameter λ^* for both hyperspectral dataset and classify the image using MRF (Table 3). The smoothing parameters for simulated PALASAR data were estimated based on the current method and (Aghighi et al., 2014) (Table 4).

Table 3: The estimated smoothing parameter (λ^*) for *MRF*

Studying area	Indian pines	Pavia University
λ^*	0.976	0.943

Table 4: The estimated smoothing parameter (λ^*) for simulated ALOS PALASAR data

Methodology	(Aghighi et al., 2014)	Current method
λ^*	0.785	0.895

In order to compare the efficiency of the proposed parameter estimation method with (Aghighi et al., 2014) and (Tarabalka et al., 2010), both hyperspectral datasets were used. Then, the λ^* value derived in this paper was employed to manage the contribution of spectral and spatial energies for the both non-edge based *MRF* method (Fig. 1(d)) and the edge-based MRF (Fig. 1(e)) by using our previous edge probability map (Aghighi et al., 2014) (see Table 5).

The difference between results obtained by applying the proposed method and (Aghighi et al., 2014) are about -0.48 precent and with (Tarabalka et al., 2010) are +2.8 precents. Furthermore, comparisons of the accuracy of maps produced using λ^* and those produced by (Tarabalka et al., 2010) show that the overall accuracy of the classification by *SVMMRF_E* for the University of Pavia data have increased by 6.27 percent and 0.1 percent for the Indian Pines data (Table 5). Then we evaluated the statistical significance for each pair of corresponding classification maps in terms of accuracy by using *McNemars* test with the 5% significance level (Zhang et al., 2011). Due to the calculated χ^2 and Z values, the null hypothesis (H_0) of no significant difference between two corresponding maps accuracy values of this paper and (Aghighi et al., 2014) is not rejected for both Indian Pines and University of Pavia images. However, the null hypothesis (H_0) of no significant difference between two corresponding maps accuracy values of this paper and (Tarabalka et al., 2010) for the University of Pavia is rejected.

In another experiment for non-Gaussian dataset, we utilized *SVMMRF* to classify PALSAR simulated image using both estimated smoothing parameters (Table 4). The difference between results derived by λ^* and (Aghighi et al., 2014) is about 1.9 percent (Table 6).

We evaluated the statistical significance for both Pines and University of Pavia classified maps in terms of accuracy by using *McNemars* test with the 5% significance level (Zhang et al., 2011). According to the calculated χ^2 and Z values, the null hypothesis (H_0) of no significant difference between the classified map accuracy values using λ^* and (Aghighi et al., 2014) is not rejected. Moreover, the classification accuracy in percentage for each class of for PALSAR images is presented in Table 7).

As mentioned in Section 2.2, different edge preserving maps are produced using Canny, Sobel, Roberts, Prewitt, Laplacian of Gaussian operators, different curvature indicator and the graduated

Table 5: Overall accuracy assessment of the classified maps for SVMMRF method with different smoothing parameters and SVM for AVIRIS and ROSIS images

Smoothing Parameter	Indian pines		Pavia University	
	SVMMRF		*SVMMRF*	
	NE	E	NE	E
(Aghighi et al., 2014)	91.8	92.3	91.5	94.1
(Tarabalka et al., 2010)	92.1	91.8	86.9	87.6
λ^*	91.2	91.9	90.8	93.9
SVM	82.2	82.2	82.2	82.2

edge penalty (17, 18, 24). Then, the new edge preserving information was incorporated in the *MRF* using the spatial energy term (19, 20) to produce the *SVMMRF_E* maps (Table 8).

Table 8 reports the global classification accuracies for all the datasets using seven edge preserving methods. The computed overall accuracies and *Kappa* coefficients in this research show that *SVMMRF_E* using the edge probability map, in particular the Canny edge method (Aghighi et al., 2014) results in the best overall classification accuracies in these experiments. Then, *McNemars* test with the 5% significance level (Zhang et al., 2011) is used to evaluate the statistical significance between *SVMMRF_E* using Canny based edge probability map and other edge preserving method. Due to the calculated χ^2 and Z values, the null hypothesis (H_0) of no significant difference between Canny based edge preserving method and others is not rejected for the University of Pavia image. However, the performance of the edge preserving method depends on the datasets, the land cover classes and the size of misclassified regions in the initial pixelwise classification map.

Table 6: Overall accuracy assessment of the classified maps for PALSAR images

SmoothingParameter	PALSAR Simulated Image
	Accuracy (%)
(Aghighi et al., 2014)	94.2
λ^*	96.1
SVM	91.4

Table 7: Overall accuracy assessment of the classified maps for PALSAR images

Class label	SVM	(Aghighi et al., 2014)	λ^*
Water	99.9	100	100
Forest	82.1	86	86.4
Farm	66.2	86.9	87.8
Urban	84.1	95.7	96.3
Open field	81.2	79.3	78.4

4. CONCLUSION

In this article, we adress the issue of the automatic smoothing parameter estimation to manage the contribution of the spectral and contextual information in the context of *MRF* classification. An innovative *MRF* smoothing parameter estimation method has been proposed, which is not dependent on assumptions of a specific statistical distribution of the image data. This method consists of employing *SVM* classification to produce the probability of individual classes for each pixel, followed by two new indexes named *CLCMC* and *GCLCMC* to estimate the smoothing parameter. In addition, the performance of seven edge preserving techniques using Canny, Sobel, Roberts, Prewitt, Laplacian of Gaussian operators, different curvature indicator and the graduated edge penalty were evaluated. Experimental results have demonstrated that the proposed method can estimate the optimal
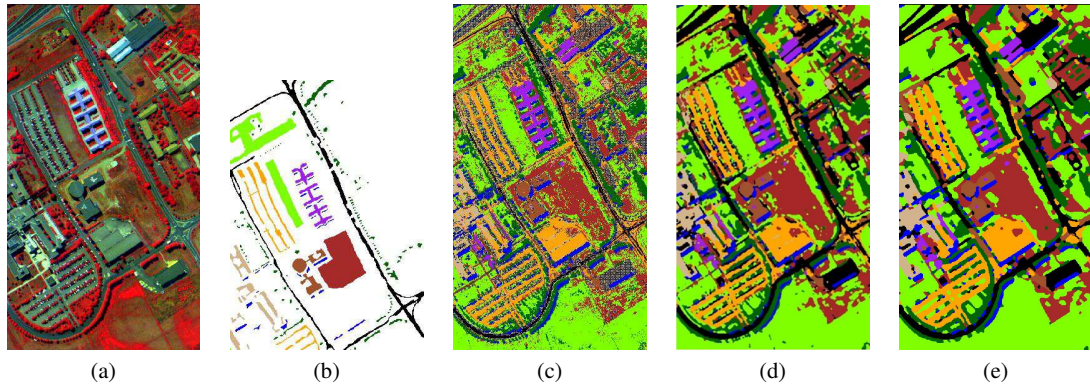
Figure 2: The University of Pavia. (a) Three-band colour composite. (b) Reference data. (c) *SVM* pixelwise classification map. (d) *SVMMRF-NE* classification map . (e)*SVMMRF-E* classification map.

Table 8: Overall accuracy assessment of the classified maps for *SVMMRF-NE* and *SVMMRF-E* method using different edge preserving methods; *OA* (Overall accuracy), *AA* (Average accuracy), *NE* (*SVMMRF-NE*), *C* (Canny), *S* (Sobel), *R* (Roberts), *P* (Prewitt), *L* (Laplacian of Gaussian), *D.C.* (Different curvature indicator) and *GIED* (Graduated edge penalty)

The Indian Pines image								
	SVM	*NE*	*C*	*S*	*R*	*P*	*L*	*D.C.*	*GIED*
OA	82.2	91.2	91.9	89.9	89.5	89.9	88.7	89.6	88.9
AA	85.1	95.1	95.5	91.7	91.3	91.7	91.7	91.7	92.1
K	0.797	0.901	0.911	0.885	0.880	0.852	0.871	0.882	0.873
The Pavia University								
	SVM	*NE*	*C*	*S*	*R*	*P*	*L*	*D.C.*	*GIED*
OA	84.8	90.8	93.9	91.7	89.6	89.5	93.7	91.7	92.4
AA	88.0	92.7	94.1	92.6	92.3	92.3	90.6	90.2	93.5
K	0.803	0.885	0.912	0.889	0.858	0.857	0.916	0.889	0.899

smoothing parameter for both non-Gaussian and Gaussian distributed data. Furthermore, the edge probability map using Canny filter yields accurate classification maps for both hyperspectral datasets.

REFERENCES

Aghighi, H., Trinder, J., Tarabalka, Y. and Lim, S., 2014. Dynamic block-based parameter estimation for mrf classification of high-resolution images. Geoscience and Remote Sensing Letters, IEEE 11(10), pp. 1687–1691.

Camps-Valls, G., Tuia, D., Bruzzone, L. and Benediktsson, J., 2014. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. Signal Processing Magazine, IEEE 31(1), pp. 45–54.

Canny, J., 1986. A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on (6), pp. 679–698.

Chang, C.-C. and Lin, C.-J., 2011. Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2(3), pp. 27.

Chen, Q., Montesinos, P., Sun, Q. S., Heng, P. A. and Xia, D. S., 2010a. Adaptive total variation denoising based on difference curvature. Image and vision computing 28(3), pp. 298–306.

Chen, Q., Montesinos, P., Sun, Q. S., Heng, P. A. and Xia, D. S., 2010b. Adaptive total variation denoising based on difference curvature. Image and vision computing 28(3), pp. 298–306.

Derin, H. and Elliott, H., 1987. Modeling and segmentation of noisy and textured images using gibbs random fields. Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-9(1), pp. 39–55.

Doulgeris, A. P., Anfinsen, S. N. and Eltoft, T., 2008. Classification with a non-gaussian model for polsar data. Geoscience and Remote Sensing, IEEE Transactions on 46(10), pp. 2999–3009.

Doulgeris, A. P., Anfinsen, S. N. and Eltoft, T., 2011. Automated non-gaussian clustering of polarimetric synthetic aperture radar images. Geoscience and Remote Sensing, IEEE Transactions on 49(10), pp. 3665–3676.

Duggin, M. J. and Robinove, C. J., 1990. Assumptions implicit in remote sensing data acquisition and analysis. International Journal of Remote Sensing 11(10), pp. 1669–1694.

Foody, G. M., 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. Photogrammetric Engineering and remote sensing 70(5), pp. 627–633.

Geman, S. and Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions on (6), pp. 721–741.

Hegarat-Mascle, L., Kallel, A. and Descombes, X., 2007. Ant colony optimization for image regularization based on a nonstationary markov modeling. Image Processing, IEEE Transactions on 16(3), pp. 865–878.

Jain, R., Kasturi, R. and Schunck, B. G., 1995. Machine vision, vol. 5.

Lee, H.-C. and Cok, D. R., 1991. Detecting boundaries in a vector field. Signal Processing, IEEE Transactions on 39(5), pp. 1181–1194.

Levitan, E. and Herman, G. T., 1987. A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography. Medical Imaging, IEEE Transactions on 6(3), pp. 185–192.

Li, X., Du, Y. and Ling, F., 2012. Spatially adaptive smoothing parameter selection for markov random field based sub-pixel mapping of remotely sensed images. International Journal of Remote Sensing 33(24), pp. 7886–7901.

Marr, D. and Hildreth, E., 1980. Theory of edge detection. Proceedings of the Royal Society of London. Series B. Biological Sciences 207(1167), pp. 187–217.

Moser, G., Serpico, S. B. and Benediktsson, J. A., 2013. Land-cover mapping by markov modeling of spatialcontextual information in very-high-resolution remote sensing images. Proceedings of the IEEE 101(3), pp. 631–651.

Pantze, A., Santoro, M. and Fransson, J. E., 2014. Change detection of boreal forest using bi-temporal alos palsar backscatter data. Remote Sensing of Environment.

Prewitt, J. M., 1970. Object enhancement and extraction. Picture processing and Psychopictorics 10(1), pp. 15–19.

Roushdy, M., 2006. Comparative study of edge detection algorithms applying on the grayscale noisy image using morphological filter. GVIP journal 6(4), pp. 17–23.

Salzenstein, F. and Pieczynski, W., 1997. Parameter estimation in hidden fuzzy markov random fields and image segmentation. Graphical Models and Image Processing 59(4), pp. 205–220.

Sambodo, K. A. and Indriasari, N., 2014. Land cover classification of alos palsar data using support vector machine. International Journal of Remote Sensing and Earth Sciences (IJReSES).

Serpico, S. B. and Moser, G., 2006. Weight parameter optimization by the hokashyap algorithm in mrf models for supervised image classification. Geoscience and Remote Sensing, IEEE Transactions on 44(12), pp. 3695–3705.

Solberg, A. H., Taxt, T. and Jain, A. K., 1996. A markov random field model for classification of multisource satellite imagery. Geoscience and Remote Sensing, IEEE Transactions on 34(1), pp. 100–113.

Tarabalka, Y., Fauvel, M., Chanussot, J. and Benediktsson, J. A., 2010. Svm-and mrf-based method for accurate classification of hyperspectral images. Geoscience and Remote Sensing Letters, IEEE 7(4), pp. 736–740.

Tolpekin, V. A. and Stein, A., 2009. Quantification of the effects of land-cover-class spectral separability on the accuracy of markov-random-field-based superresolution mapping. Geoscience and Remote Sensing, IEEE Transactions on 47(9), pp. 3283–3297.

Tso, B. C. and Mather, P. M., 1999. Classification of multisource remote sensing imagery using a genetic algorithm and markov random fields. Geoscience and Remote Sensing, IEEE Transactions on 37(3), pp. 1255–1260.

Wu, T.-F., Lin, C.-J. and Weng, R. C., 2004. Probability estimates for multi-class classification by pairwise coupling. Journal of Machine Learning Research 5(975-1005), pp. 4.

Xu, M., Watanachaturaporn, P., Varshney, P. K. and Arora, M. K., 2005. Decision tree regression for soft classification of remote sensing data. Remote Sensing of Environment 97(3), pp. 322–336.

Yu, P., Qin, A. and Clausi, D. A., 2012. Unsupervised polarimetric sar image segmentation and classification using region growing with edge penalty. Geoscience and Remote Sensing, IEEE Transactions on 50(4), pp. 1302–1317.

Yu, Q. and Clausi, D. A., 2008. Irgs: image segmentation using edge penalties and region growing. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30(12), pp. 2126–2139.

Zhang, B., Li, S., Jia, X., Gao, L. and Peng, M., 2011. Adaptive markov random field approach for classification of hyperspectral imagery. Geoscience and Remote Sensing Letters, IEEE 8(5), pp. 973–977.